8

"Would you tell me, please, which way I ought to go from here?"

"That depends a good deal on where you want to get to," said the Cat.

"I don't much care where—" said Alice.

"Then it doesn't matter which way you go," said the Cat.

—LEWIS CARROLL
ALICE'S ADVENTURES IN WONDERLAND

Sam Abell/National Geographic/Getty Images

Mary Evans Picture Library/Arthur Rackham/ The Image Works

Charles Lutwidge Dodgson (1832–1898) was an English mathematician who loved to write children's stories in his free time. The dialogue between Alice and the Cheshire Cat occurs in the masterpiece *Alice's Adventures in Wonderland*, written by Dodgson under the pen name Lewis Carroll. These lines relate to our study of hypothesis testing. Statistical tests cannot answer all of life's questions. They cannot always tell us "where to go," but after this decision is made on other grounds, they can help us find the best way to get there.

For online student resources, visit the Brase/Brase, *Understandable Statistics,* 10th edition web site at **http://www.cengage.com/statistics/brase**.

HYPOTHESIS TESTING

PREVIEW QUESTIONS

Many of life's questions require a yes or no answer. When you must act on incomplete (sample) information, how do you decide whether to accept or reject a proposal? (SECTION 8.1)

What is the P-value of a statistical test? What does this measurement have to do with performance reliability? (SECTION 8.1)

How do you construct statistical tests for μ? Does it make a difference whether σ is known or unknown? (SECTION 8.2)

How do you construct statistical tests for the proportion p of successes in a binomial experiment? (SECTION 8.3)

What are the advantages of pairing data values? How do you construct statistical tests for paired differences? (SECTION 8.4)

How do you construct statistical tests for differences of independent random variables? (SECTION 8.5)

Comstock Images/Jupiter Images

FOCUS PROBLEM

Benford's Law: The Importance of Being Number 1

Benford's Law states that in a wide variety of circumstances, numbers have "1" as their first nonzero digit disproportionately often. Benford's Law applies to such diverse topics as the drainage areas of rivers; properties of chemicals; populations of towns; figures in newspapers, magazines, and government reports; and the half-lives of radioactive atoms!

Specifically, such diverse measurements begin with "1" about 30% of the time, with "2" about 18% of time, and with "3" about 12.5% of the time. Larger digits occur less often. For example, less than 5% of the numbers in circumstances such as these begin with the digit 9. This is in dramatic contrast to a random sampling situation, in which each of the digits 1 through 9 has an equal chance of appearing.

The first nonzero digits of numbers taken from large bodies of numerical records such as tax returns, population studies, government records, and so forth, show the probabilities of occurrence as displayed in the table on the next page.

Corbis

409

First nonzero digit	1	2	3	4	5	6	7	8	9
Probability	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046

More than 100 years ago, the astronomer Simon Newcomb noticed that books of logarithm tables were much dirtier near the fronts of the tables. It seemed that people were more frequently looking up numbers with a low first digit. This was regarded as an odd phenomenon and a strange curiosity. The phenomenon was rediscovered in 1938 by physicist Frank Benford (hence the name *Benford's Law*).

More recently, Ted Hill, a mathematician at the Georgia Institute of Technology, studied situations that might demonstrate Benford's Law. Professor Hill showed that such probability distributions are likely to occur when we have a "distribution of distributions." Put another way, large random collections of random samples tend to follow Benford's Law. This seems to be especially true for samples taken from large government data banks, accounting reports for large corporations, large collections of astronomical observations, and so forth. For more information, see *American Scientist,* Vol. 86, pp. 358–363, and *Chance,* American Statistical Association, Vol. 12, No. 3, pp. 27–31.

Can Benford's Law be applied to help solve a real-world problem? Well, one application might be accounting fraud! Suppose the first nonzero digits of the entries in the accounting records of a large corporation (such as Enron or WorldCom) do not follow Benford's Law. Should this set off an accounting alarm for the FBI or the stockholders? How "significant" would this be? Such questions are the subject of statistics.

In Section 8.3, you will see how to use sample data to test whether the proportion of first nonzero digits of the entries in a large accounting report follows Benford's Law. Problems 7 and 8 of Section 8.3 relate to Benford's Law and accounting discrepancies. In one problem, you are asked to use sample data to determine if accounting books have been "cooked" by "pumping numbers up" to make the company look more attractive or perhaps to provide a cover for money laundering. In the other problem, you are asked to determine if accounting books have been "cooked" by artificially lowered numbers, perhaps to hide profits from the Internal Revenue Service or to divert company profits to unscrupulous employees. (See Problems 7 and 8 of Section 8.3.)

SECTION 8.1   Introduction to Statistical Tests

FOCUS POINTS
- Understand the rationale for statistical tests.
- Identify the null and alternate hypotheses in a statistical test.
- Identify right-tailed, left-tailed, and two-tailed tests.
- Use a test statistic to compute a *P*-value.
- Recognize types of errors, level of significance, and power of a test.
- Understand the meaning and risks of rejecting or not rejecting the null hypothesis.

In Chapter 1, we emphasized the fact that one of a statistician's most important jobs is to draw inferences about populations based on samples taken from the populations. Most statistical inference centers around the parameters of a population (often the mean or probability of success in a binomial trial). Methods for drawing inferences about parameters are of two types: Either we make decisions concerning the value of the parameter, or we actually estimate the value of the

Hypothesis testing

Hypothesis

parameter. When we estimate the value (or location) of a parameter, we are using methods of estimation such as those studied in Chapter 7. Decisions concerning the value of a parameter are obtained by *hypothesis testing,* the topic we shall study in this chapter.

Students often ask which method should be used on a particular problem—that is, should the parameter be estimated, or should we test a *hypothesis* involving the parameter? The answer lies in the practical nature of the problem and the questions posed about it. Some people prefer to test theories concerning the parameters. Others prefer to express their inferences as estimates. Both estimation and hypothesis testing are found extensively in the literature of statistical applications.

Stating Hypotheses

Null hypothesis H_0

Our first step is to establish a working hypothesis about the population parameter in question. This hypothesis is called the *null hypothesis,* denoted by the symbol H_0. The value specified in the null hypothesis is often a historical value, a claim, or a production specification. For instance, if the average height of a professional male basketball player was 6.5 feet 10 years ago, we might use a null hypothesis $H_0: \mu = 6.5$ feet for a study involving the average height of this year's professional male basketball players. If television networks claim that the average length of time devoted to commercials in a 60-minute program is 12 minutes, we would use $H_0: \mu = 12$ minutes as our null hypothesis in a study regarding the average length of time devoted to commercials. Finally, if a repair shop claims that it should take an average of 25 minutes to install a new muffler on a passenger automobile, we would use $H_0: \mu = 25$ minutes as the null hypothesis for a study of how well the repair shop is conforming to specified average times for a muffler installation.

Alternate hypothesis H_1

Any hypothesis that differs from the null hypothesis is called an *alternate hypothesis.* An alternate hypothesis is constructed in such a way that it is the hypothesis to be accepted when the null hypothesis must be rejected. The alternate hypothesis is denoted by the symbol H_1. For instance, if we believe the average height of professional male basketball players is taller than it was 10 years ago, we would use an alternate hypothesis $H_1: \mu > 6.5$ feet with the null hypothesis $H_0: \mu = 6.5$ feet.

> **Null hypothesis H_0:** This is the statement that is under investigation or being tested. Usually the null hypothesis represents a statement of "no effect," "no difference," or, put another way, "things haven't changed."
>
> **Alternate hypothesis H_1:** This is the statement you will adopt in the situation in which the evidence (data) is so strong that you reject H_0. A statistical test is designed to assess the strength of the evidence (data) against the null hypothesis.

EXAMPLE 1   NULL AND ALTERNATE HYPOTHESES

A car manufacturer advertises that its new subcompact models get 47 miles per gallon (mpg). Let μ be the mean of the mileage distribution for these cars. You assume that the manufacturer will not underrate the car, but you suspect that the mileage might be overrated.

(a) What shall we use for H_0?

   SOLUTION: We want to see if the manufacturer's claim that $\mu = 47$ mpg can be rejected. Therefore, our null hypothesis is simply that $\mu = 47$ mpg. We denote the null hypothesis as

   $$H_0: \mu = 47 \text{ mpg}$$

(b) What shall we use for H_1?

SOLUTION: From experience with this manufacturer, we have every reason to believe that the advertised mileage is too high. If μ is not 47 mpg, we are sure it is less than 47 mpg. Therefore, the alternate hypothesis is

$$H_1: \mu < 47 \text{ mpg}$$

GUIDED EXERCISE 1 | **Null and alternate hypotheses**

A company manufactures ball bearings for precision machines. The average diameter of a certain type of ball bearing should be 6.0 mm. To check that the average diameter is correct, the company formulates a statistical test.

(a) What should be used for H_0? (*Hint:* What is the company trying to test?)

➡ If μ is the mean diameter of the ball bearings, the company wants to test whether $\mu = 6.0$ mm. Therefore, $H_0: \mu = 6.0$ mm.

(b) What should be used for H_1? (*Hint:* An error either way, too small or too large, would be serious.)

➡ An error either way could occur, and it would be serious. Therefore, $H_1: \mu \neq 6.0$ mm (μ is either smaller than or larger than 6.0 mm).

COMMENT: NOTATION REGARDING THE NULL HYPOTHESIS   In statistical testing, the null hypothesis H_0 always contains the equals symbol. However, in the null hypothesis, some statistical software packages and texts also include the inequality symbol that is opposite that shown in the alternate hypothesis. For instance, if the alternate hypothesis is "μ is less than 3" ($\mu < 3$), then the corresponding null hypothesis is sometimes written as "μ is greater than or equal to 3" ($\mu \geq 3$). The mathematical construction of a statistical test uses the null hypothesis to assign a specific number (rather than a range of numbers) to the parameter μ in question. The null hypothesis establishes a single fixed value for μ, so we are working with a single distribution having a specific mean. In this case, H_0 assigns $\mu = 3$. So, when $H_1: \mu < 3$ is the alternate hypothesis, we follow the commonly used convention of writing the null hypothesis simply as $H_0: \mu = 3$.

Types of Tests

The null hypothesis H_0 always states that the parameter of interest *equals* a specified value. The alternate hypothesis H_1 states that the parameter is *less than,* *greater than,* or simply *not equal to* the same value. We categorize a statistical test as *left-tailed,* *right-tailed,* or *two-tailed* according to the alternate hypothesis.

Types of statistical tests

A statistical test is:

Left-tailed test
  left-tailed if H_1 states that the parameter is less than the value claimed in H_0

Right-tailed test
  right-tailed if H_1 states that the parameter is greater than the value claimed in H_0

Two-tailed test
  two-tailed if H_1 states that the parameter is different from (or not equal to) the value claimed in H_0

TABLE 8-1	The Null and Alternate Hypotheses for Tests of the Mean μ		
Null Hypothesis	**Alternate Hypotheses and Type of Test**		
Claim about μ or historical value of μ $H_0: \mu = k$	You believe that μ is less than value stated in H_0. $H_1: \mu < k$ Left-tailed test	You believe that μ is more than value stated in H_0. $H_1: \mu > k$ Right-tailed test	You believe that μ is different from value stated in H_0. $H_1: \mu \neq k$ Two-tailed test

In this introduction to statistical tests, we discuss tests involving a population mean μ. However, you should keep an open mind and be aware that the methods outlined apply to testing other parameters as well (e.g., p, σ, $\mu_1 - \mu_2$, $p_1 - p_2$, and so on). Table 8-1 shows how tests of the mean μ are categorized.

Hypothesis Tests of μ, Given x Is Normal and σ Is Known

Once you have selected the null and alternate hypotheses, how do you decide which hypothesis is likely to be valid? Data from a simple random sample and the sample test statistic, together with the corresponding sampling distribution of the test statistic, will help you decide. Example 2 leads you through the decision process.

First, a quick review of Section 6.4 is in order. Recall that a population *parameter* is a numerical descriptive measurement of the entire population. Examples of population parameters are μ, p, and σ. It is important to remember that for a given population, the parameters are *fixed* values. They do not vary! The null hypothesis H_0 makes a statement about a population parameter.

A *statistic* is a numerical descriptive measurement of a sample. Examples of statistics are $\bar{x}$, $\hat{p}$, and s. Statistics usually *vary* from one sample to the next. The probability distribution of the statistic we are using is called a *sampling distribution*.

For hypothesis testing, we take a *simple* random sample and compute a *sample test statistic* corresponding to the parameter in H_0. Based on the sampling distribution of the statistic, we can assess how compatible the sample test statistic is with H_0.

Sample test statistic for μ, given x normal and σ known

In this section, we use hypothesis tests about the mean to introduce the concepts and vocabulary of hypothesis testing. In particular, let's suppose that x has *a normal distribution* with mean μ and standard deviation σ. Then, Theorem 6.1 tells us that $\bar{x}$ has a *normal distribution* with mean μ and standard deviation $\sigma/\sqrt{n}$.

PROCEDURE

Requirements The x distribution is *normal* with known standard deviation σ.

Then $\bar{x}$ has a normal distribution. The standardized test statistic is

$$\textbf{test statistic} = z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

where $\bar{x}$ = mean of a simple random sample

μ = value stated in H_0.

n = sample size

EXAMPLE 2   STATISTICAL TESTING PREVIEW

Rosie is an aging sheep dog in Montana who gets regular checkups from her owner, the local veterinarian. Let x be a random variable that represents Rosie's resting heart rate (in beats per minute). From past experience, the vet knows that

x has a normal distribution with $\sigma = 12$. The vet checked the *Merck Veterinary Manual* and found that for dogs of this breed, $\mu = 115$ beats per minute.

Over the past six weeks, Rosie's heart rate (beats/min) measured

93	109	110	89	112	117

The sample mean is $\bar{x} = 105.0$. The vet is concerned that Rosie's heart rate may be slowing. Do the data indicate that this is the case?

SOLUTION:

(a) Establish the null and alternate hypotheses.

If "nothing has changed" from Rosie's earlier life, then her heart rate should be nearly average. This point of view is represented by the null hypothesis

$$H_0\text{: } \mu = 115$$

However, the vet is concerned about Rosie's heart rate slowing. This point of view is represented by the alternate hypothesis

$$H_1\text{: } \mu < 115$$

(b) Are the observed sample data compatible with the null hypothesis?

Are the six observations of Rosie's heart rate compatible with the null hypothesis H_0: $\mu = 115$? To answer this question, we need to know the *probability* of obtaining a sample mean of 105.0 or less from a population with true mean $\mu = 115$. If this probability is small, we conclude that H_0: $\mu = 115$ is not the case. Rather, H_1: $\mu < 115$ and Rosie's heart rate is slowing.

(c) How do we compute the probability in part (b)?

Well, you probably guessed it! We use the sampling distribution for $\bar{x}$ and compute $P(\bar{x} < 105.0)$. Figure 8-1 shows the $\bar{x}$ distribution and the corresponding standard normal distribution with the desired probability shaded.

Check Requirements Since x has a normal distribution, $\bar{x}$ will also have a normal distribution for any sample size n and given σ (see Theorem 6.1).

Note that using $\mu = 115$ from H_0, $\sigma = 12$, and $n = 6$ the sample $\bar{x} = 105.0$ converts to

$$\text{test statistic} = z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{105.0 - 115}{12/\sqrt{6}} \approx -2.04$$
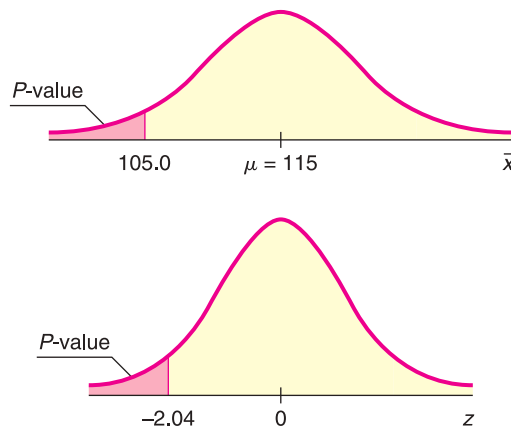
Using the standard normal distribution table, we find that

$$P(\bar{x} < 105.0) = P(z < -2.04) = 0.0207$$

The area in the left tail that is more extreme than $\bar{x} = 105.0$ is called the *P-value* of the test. In this example, P-value = 0.0207. We will learn more about P-values later.

P-value

FIGURE 8-1

Sampling Distribution for $\bar{x}$ and Corresponding z Distribution

(d) *Interpretation* What conclusion can be drawn about Rosie's average heart rate?

If H_0: $\mu = 115$ is in fact true, the probability of getting a sample mean of $\bar{x} \leq 105.0$ is only about 2%. Because this probability is small, we reject H_0: $\mu = 115$ and conclude that H_1: $\mu < 115$ Rosie's average heart rate seems to be slowing.

(e) Have we proved H_0: $\mu = 115$ to be false and H_1: $\mu < 115$ to be true?

No! The sample data do not *prove* H_0 to be false and H_1 to be true! We do say that H_0 has been "discredited" by a small *P*-value of 0.0207. Therefore, we abandon the claim H_0: $\mu = 115$ and adopt the claim H_1: $\mu < 115$.

The *P*-value of a Statistical Test

Rosie the sheep dog has helped us to "sniff out" an important statistical concept.
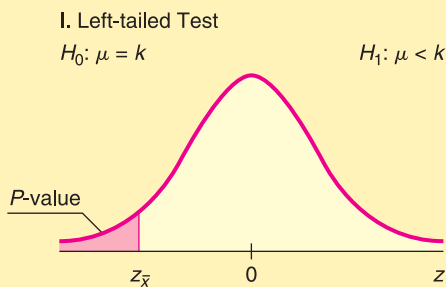
> **P-value**
>
> Assuming H_0 is true, the *probability* that the test statistic will take on values as extreme as or more extreme than the observed test statistic (computed from sample data) is called the **P-value** of the test. The smaller the *P*-value computed from sample data, the stronger the evidence against H_0.

The *P*-value, sometimes called the *probability of chance*, can be thought of as the probability that the results of a statistical experiment are due only to chance. The lower the *P*-value, the greater the likelihood of obtaining the same (or very similar) results in a repetition of the statistical experiment. Thus, a low *P*-value is a good indication that your results are not due to random chance alone.

The *P*-value associated with the observed test statistic takes on different values depending on the alternate hypothesis and the type of test. Let's look at *P*-values and types of tests when the test involves the mean and standard normal distribution. Notice that in Example 2, part (c), we computed a *P*-value for a left-tailed test. Guided Exercise 3 asks you to compute a *P*-value for a two-tailed test.

P-values and types of tests

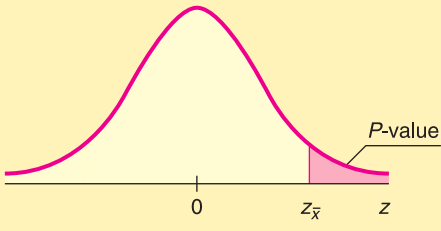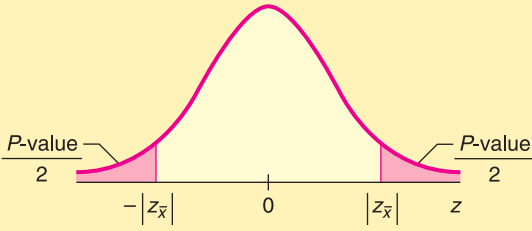Let $z_{\bar{x}}$ represent the standardized sample test statistic for testing a mean μ using the standard normal distribution. That is, $z_{\bar{x}} = (\bar{x} - \mu)/(\sigma/\sqrt{n})$.

I. Left-tailed Test

H_0: $\mu = k$ H_1: $\mu < k$

P-value = $P(z < z_{\bar{x}})$

This is the probability of getting a test statistic as low as or lower than $z_{\bar{x}}$.

Continued

II. Right-tailed Test

$H_0: \mu = k$ $H_1: \mu > k$

$P\text{-value} = P(z > z_{\bar{x}})$

This is the probability of getting a test statistic as high as or higher than $z_{\bar{x}}$

III. Two-tailed Test

$H_0: \mu = k$ $H_1: \mu \neq k$

$$\frac{P\text{-value}}{2} = P(z > |z_{\bar{x}}|); \text{ therefore,}$$

$$P\text{-value} = 2P(z > |z_{\bar{x}}|)$$

This is the probability of getting a test statistic either lower than $-|z_{\bar{x}}|$ or higher than $|z_{\bar{x}}|$.

Types of Errors

If we *reject the null hypothesis when it is,* in fact, *true,* we have made an error that is called a *type I error.* On the other hand, if we *accept the null hypothesis when it is,* in fact, *false,* we have made an error that is called a *type II error.* Table 8-2 indicates how these errors occur.

Type I error
Type II error

For tests of hypotheses to be well constructed, they must be designed to minimize possible errors of decision. (Usually, we do not know if an error has been made, and therefore, we can talk only about the probability of making an error.) Usually, for a given sample size, an attempt to reduce the probability of one type of error results in an increase in the probability of the other type of error. In practical applications, one type of error may be more serious than another. In such a case, careful attention is given to the more serious error. If we increase the sample size, it is possible to reduce both types of errors, but increasing the sample size may not be possible.

Good statistical practice requires that we announce in advance how much evidence against H_0 will be required to reject H_0. The probability with which we are willing to risk a type I error is called the *level of significance* of a test. The level of significance is denoted by the Greek letter α (pronounced "alpha").

Level of significance α

The **level of significance** α is the probability of rejecting H_0 when it is true. This is the probability of a type I error.

TABLE 8-2   **Type I and Type II Errors**

Truth of H_0	Our Decision	
	And if we do not reject H_0	And if we reject H_0
If H_0 is true	Correct decision; no error	Type I error
If H_0 is false	Type II error	Correct decision; no error

TABLE 8-3	**Probabilities Associated with a Statistical Test**	
	Our Decision	
Truth of H_0	**And if we accept H_0 as true**	**And if we reject H_0 as false**
If H_0 is true	Correct decision, with corresponding probability $1 - \alpha$	Type I error, with corresponding probability α, called the *level of significance of the test*
If H_0 is false	Type II error, with corresponding probability β	Correct decision; with corresponding probability $1 - \beta$, called the *power of the test*

Probability of a type II error β

> The *probability of making a type II error* is denoted by the Greek letter β (pronounced "beta").

Methods of hypothesis testing require us to choose α and β values to be as small as possible. In elementary statistical applications, we usually choose α first.

Power of a test $(1 - \beta)$

> The quantity $1 - \beta$ is called the *power of a test* and represents the probability of rejecting H_0 when it is, in fact, false.

For a given level of significance, how much power can we expect from a test? The actual value of the power is usually difficult (and sometimes impossible) to obtain, since it requires us to know the H_1 distribution. However, we can make the following general comments:

1. The power of a statistical test increases as the level of significance α increases. A test performed at the $\alpha = 0.05$ level has more power than one performed at $\alpha = 0.01$. This means that the less stringent we make our significance level α, the more likely we will be to reject the null hypothesis when it is false.

2. Using a larger value of α will increase the power, but it also will increase the probability of a type I error. Despite this fact, most business executives, administrators, social scientists, and scientists use *small* α values. This choice reflects the conservative nature of administrators and scientists, who are usually more willing to make an error by failing to reject a claim (i.e., H_0) than to make an error by accepting another claim (i.e., H_1) that is false. Table 8-3 summarizes the probabilities of errors associated with a statistical test.

COMMENT   Since the calculation of the probability of a type II error is treated in advanced statistics courses, we will restrict our attention to the probability of a type I error.

GUIDED EXERCISE 2 | **Types of errors**

Let's reconsider Guided Exercise 1, in which we were considering the manufacturing specifications for the diameter of ball bearings. The hypotheses were

$H_0: \mu = 6.0$ mm (manufacturer's specification) $H_1: \mu \neq 6.0$ mm (cause for adjusting process)

(a) Suppose the manufacturer requires a 1% level of significance. Describe a type I error, its consequence, and its probability.

A type I error is caused when sample evidence indicates that we should reject H_0 when, in fact, the average diameter of the ball bearings being produced is 6.0 mm. A type I error will cause a needless

Continued

	adjustment and delay of the manufacturing process. The probability of such an error is 1% because $\alpha = 0.01$.
(b) Discuss a type II error and its consequences.	A type II error occurs if the sample evidence tells us not to reject the null hypothesis H_0: $\mu = 6.0$ mm when, in fact, the average diameter of the ball bearing is either too large or too small to meet specifications. Such an error would mean that the production process would not be adjusted even though it really needed to be adjusted. This could possibly result in a large production of ball bearings that do not meet specifications.

Concluding a Statistical Test

Usually, α is specified in advance before any samples are drawn so that results will not influence the choice for the level of significance. To conclude a statistical test, we compare our α value with the *P*-value computed using sample data and the sampling distribution.

PROCEDURE

HOW TO CONCLUDE A TEST USING THE *P*-VALUE AND LEVEL OF SIGNIFICANCE α

Statistical significance

If *P*-value $\leq \alpha$, we **reject** the null hypothesis and say the data are **statistically significant** at the level α.

If P-value $> \alpha$, we **do not reject** the null hypothesis.

In what sense are we using the word *significant*? *Webster's Dictionary* gives two interpretations of *significance*: (1) having or signifying *meaning*: or (2) being important or momentous.

In statistical work, significance does not necessarily imply momentous importance. For us, "significant" at the α level has a special meaning. It says that at the α level of risk, the evidence (sample data) against the null hypothesis H_0 is sufficient to discredit H_0, so we adopt the alternate hypothesis H_1.

In any case, we do not claim that we have "proved" or "disproved" the null hypothesis H_0. We can say that the probability of a type I error (rejecting H_0 when it is, in fact, true) is α.

Basic components of a statistical test

A statistical test can be thought of as a package of five basic ingredients.

1. **Null hypothesis H_0, alternate hypothesis H_1,** and **preset level of significance α**

   If the evidence (sample data) against H_0 is strong enough, we reject H_0 and adopt H_1. The level of significance α is the probability of rejecting H_0 when it is, in fact, true.

2. **Test statistic and sampling distribution**

   These are mathematical tools used to measure compatibility of sample data and the null hypothesis.

3. **P-value**

   This is the probability of obtaining a test statistic from the sampling distribution that is as extreme as, or more extreme (as specified by H_1) than, the sample test statistic computed from the data under the assumption that H_0 is true.

4. **Test conclusion**

   If P-value $\leq \alpha$, we reject H_0 and say that the data are significant at level α. If P-value $> \alpha$, we do not reject H_0.

5. **Interpretation of the test results**

   Give a simple explanation of your conclusions in the context of the application.

GUIDED EXERCISE 3 | **Constructing a statistical test for μ (normal distribution)**

The Environmental Protection Agency has been studying Miller Creek regarding ammonia nitrogen concentration. For many years, the concentration has been 2.3 mg/l. However, a new golf course and new housing developments are raising concern that the concentration may have changed because of lawn fertilizer. Any change (either an increase or a decrease) in the ammonia nitrogen concentration can affect plant and animal life in and around the creek (Reference: *EPA Report* 832-R-93-005). Let x be a random variable representing ammonia nitrogen concentration (in mg/l). Based on recent studies of Miller Creek, we may assume that x has a normal distribution with $\sigma = 0.30$ Recently, a random sample of eight water tests from the creek gave the following x values.

       2.1   2.5   2.2   2.8   3.0   2.2   2.4   2.9

The sample mean is $\bar{x} \approx 2.51$.

Let us construct a statistical test to examine the claim that the concentration of ammonia nitrogen has changed from 2.3 mg/l. Use level of significance $\alpha = 0.01$.

(a) What is the null hypothesis? What is the alternate hypothesis? What is the level of significance α?

> H_0: $\mu = 2.3$
> H_1: $\mu \neq 2.3$
> $\alpha = 0.01$

(b) Is this a right-tailed, left-tailed, or two-tailed test?

> Since H_1: $\mu \neq 2.3$, this is a two-tailed test.

(c) *Check Requirements* What sampling distribution shall we use? Note that the value of μ is given in the null hypothesis, H_0.

> Since the x distribution is normal and σ is known, we use the standard normal distribution with
>
> $$z = \frac{\bar{x} - \mu}{\dfrac{\sigma}{\sqrt{n}}} = \frac{\bar{x} - 2.3}{\dfrac{0.3}{\sqrt{8}}}$$

(d) What is the sample test statistic? Convert the sample mean $\bar{x}$ to a standard z value.

> The sample of eight measurements has mean $\bar{x} = 2.51$. Converting this measurement to z, we have
>
> $$\text{test statistic} = z = \frac{2.51 - 2.3}{\dfrac{0.3}{\sqrt{8}}} \approx 1.98$$
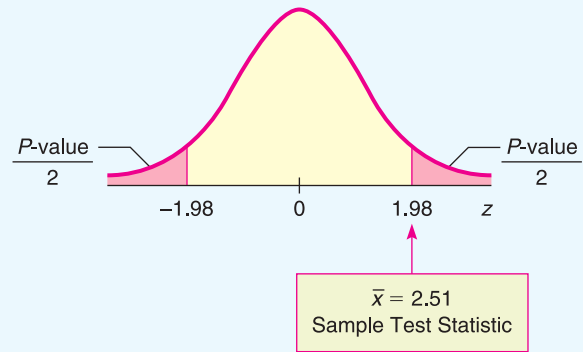
Continued

GUIDED EXERCISE 3   *continued*

(e) Draw a sketch showing the *P*-value area on the standard normal distribution. Find the *P*-value.

➡ $P\text{-value} = 2P(z > 1.98) = 2(0.0239) = 0.0478$

FIGURE 8-2 *P*-value

(f) Compare the level of significance α and the *P*-value. What is your conclusion?

➡ Since *P*-value $0.0478 \geq 0.01$, we see that *P*-value $> \alpha$. We fail to reject H_0.

(g) *Interpret* your results in the context of this problem.

➡ The sample data are not significant at the $\alpha = 1\%$ level. At this point in time, there is not enough evidence to conclude that the ammonia nitrogen concentration has changed in Miller Creek.

Interpretation of level of significance

In most statistical applications, the level of significance is specified to be $\alpha = 0.05$ or $\alpha = 0.01$, although other values can be used. If $\alpha = 0.05$, then we say we are using a 5% level of significance. This means that in 100 similar situations, H_0 will be rejected 5 times, on average, when it should not have been rejected.

Meaning of accepting H_0

Using Technology at the end of this chapter shows a simulation of this phenomenon.

When we accept (or fail to reject) the null hypothesis, we should understand that we are *not proving the null hypothesis*. We are saying only that the sample evidence (data) is not strong enough to justify rejection of the null hypothesis. The word *accept* sometimes has a stronger meaning in common English usage than we are willing to give it in our application of statistics. Therefore, we often use the expression *fail to reject H_0* instead of *accept H_0*. "*Fail to reject* the null hypothesis" simply means that the evidence in favor of rejection was not strong enough (see Table 8-4). Often, in the case that H_0 cannot be rejected, a confidence interval is used to estimate the parameter in question. The confidence interval gives the statistician a range of possible values for the parameter.

TABLE 8-4   **Meaning of the Terms *Fail to Reject H_0* and *Reject H_0***

Term	Meaning
Fail to reject H_0	Fail to reject H_0	There is not enough evidence in the data (and the test being used) to justify a rejection of H_0. This means that we retain H_0 with the understanding that we have not proved it to be true beyond all doubt.
Reject H_0	Reject H_0	There is enough evidence in the data (and the test employed) to justify rejection of H_0. This means that we choose the alternate hypothesis H_1 with the understanding that we have not proved H_1 to be true beyond all doubt.

COMMENT   Some comments about *P*-values and level of significance α should be made. The level of significance α should be a fixed, prespecified value. Usually, α is chosen before any samples are drawn. The level of significance α is the probability of a type I error. So, α is the probability of rejecting H_0 when, in fact, H_0 is true.

Interpreting the P-value of a test statistic

The *P*-value should *not* be interpreted as the probability of a type I error. The level of significance (in theory) is set in advance before any samples are drawn. The *P*-value cannot be set in advance, since it is determined from the random sample. The *P*-value, together with α, should be regarded as tools used to conclude the test. If *P*-value $\leq \alpha$, then reject H_0, and if *P*-value $> \alpha$, then do not reject H_0.

In most computer applications and journal articles, only the *P*-value is given. It is understood that the person using this information will supply an appropriate level of significance α. From an historical point of view, the English statistician F. Y. Edgeworth (1845–1926) was one of the first to use the term *significant* to imply that the sample data indicate a "meaningful" difference from a previously held view.

In this book, we are using the most popular method of testing, which is called the *P-value method*. At the end of the next section, you will learn about another (equivalent) method of testing called the *critical region method*. An extensive discussion regarding the *P*-value method of testing versus the critical region method can be found in *The American Statistician,* Vol. 57, No. 3, pp. 171–178, American Statistical Association.

VIEWPOINT | Lovers, Take Heed!!!

If you are going to whisper sweet nothings to your sweetheart, be sure to whisper them in the left ear. Professor Sim of Sam Houston State University (Huntsville, Texas) found that emotionally loaded words have a higher recall rate when spoken into a person's left ear, not the right. Professor Sim presented his findings at the British Psychology Society European Congress. He told the Congress that his findings are consistent with the hypothesis that the brain's right hemisphere has more influence in the processing of emotional stimuli. (The left ear is controlled by the right side of the brain.) Sim's research involved statistical tests like the ones you will study in this chapter.

SECTION 8.1 PROBLEMS

1. *Statistical Literacy* Discuss each of the following topics in class or review the topics on your own. Then write a brief but complete essay in which you answer the following questions.
   (a) What is a null hypothesis H_0?
   (b) What is an alternate hypothesis H_1?
   (c) What is a type I error? a type II error?
   (d) What is the level of significance of a test? What is the probability of a type II error?

2. *Statistical Literacy* In a statistical test, we have a choice of a left-tailed test, a right-tailed test, or a two-tailed test. Is it the null hypothesis or the alternate hypothesis that determines which type of test is used? Explain your answer.

3. *Statistical Literacy* If we fail to reject (i.e., "accept") the null hypothesis, does this mean that we have *proved* it to be true beyond *all* doubt? Explain your answer.

4. *Statistical Literacy* If we reject the null hypothesis, does this mean that we have *proved* it to be false beyond *all* doubt? Explain your answer.

5. *Statistical Literacy* What terminology do we use for the probability of rejecting the null hypothesis when it is true? What symbol do we use for this probability? Is this the probability of a type I or a type II error?

6. *Statistical Literacy* What terminology do we use for the probability of rejecting the null hypothesis when it is, in fact, false?

7. *Statistical Literacy* If the P-value in a statistical test is greater than the level of significance for the test, do we reject or fail to reject H_0?

8. *Statistical Literacy* If the P-value in a statistical test is less than or equal to the level of significance for the test, do we reject or fail to reject H_0?

9. *Statistical Literacy* Suppose the P-value in a right-tailed test is 0.0092. Based on the same population, sample, and null hypothesis, what is the P-value for a corresponding two-tailed test?

10. *Statistical Literacy* Suppose the P-value in a two-tailed test is 0.0134. Based on the same population, sample, and null hypothesis, and assuming the test statistic z is negative, what is the P-value for a corresponding left-tailed test?

11. *Basic Computation: Setting Hypotheses* Suppose you want to test the claim that a population mean equals 40.
    (a) State the null hypothesis.
    (b) State the alternate hypothesis if you have no information regarding how the population mean might differ from 40.
    (c) State the alternate hypothesis if you believe (based on experience or past studies) that the population mean may exceed 40.
    (d) State the alternate hypothesis if you believe (based on experience or past studies) that the population mean may be less than 40.

12. *Basic Computation: Setting Hypotheses* Suppose you want to test the claim that a population mean equals 30.
    (a) State the null hypothesis.
    (b) State the alternate hypothesis if you have no information regarding how the population mean might differ from 30.
    (c) State the alternate hypothesis if you believe (based on experience or past studies) that the population mean may be greater than 30.
    (d) State the alternate hypothesis if you believe (based on experience or past studies) that the population mean may not be as large as 30.

13. *Basic Computation: Find Test Statistic, Corresponding P-value, and Conclude Test* A random sample of size 20 from a normal distribution with $\sigma = 4$ produced a sample mean of 8.
    (a) *Check Requirements* Is the $\bar{x}$ distribution normal? Explain.
    (b) Compute the sample test statistic z under the null hypothesis $H_0: \mu = 7$.
    (c) For $H_1: \mu \neq 7$, estimate the P-value of the test statistic.
    (d) For a level of significance of 0.05 and the hypotheses of parts (b) and (c), do you reject or fail to reject the null hypothesis? Explain.

14. *Basic Computation: Find the Test Statistic and Corresponding P-value* A random sample of size 16 from a normal distribution with $\sigma = 3$ produced a sample mean of 4.5.
    (a) *Check Requirements* Is the $\bar{x}$ distribution normal? Explain.
    (b) Compute the sample test statistic z under the null hypothesis $H_0: \mu = 6.3$.
    (c) For $H_1: \mu < 6.3$, estimate the P-value of the test statistic.
    (d) For a level of significance of 0.01 and the hypotheses of parts (b) and (c), do you reject or fail to reject the null hypothesis? Explain.

15. *Veterinary Science: Colts* The body weight of a healthy 3-month-old colt should be about $\mu = 60$ kg (Source: *The Merck Veterinary Manual,* a standard reference manual used in most veterinary colleges).
    (a) If you want to set up a statistical test to challenge the claim that $\mu = 60$ kg, what would you use for the null hypothesis H_0?

(b) In Nevada, there are many herds of wild horses. Suppose you want to test the claim that the average weight of a wild Nevada colt (3 months old) is less than 60 kg. What would you use for the alternate hypothesis H_1?

(c) Suppose you want to test the claim that the average weight of such a wild colt is greater than 60 kg. What would you use for the alternate hypothesis?

(d) Suppose you want to test the claim that the average weight of such a wild colt is *different* from 60 kg. What would you use for the alternate hypothesis?

(e) For each of the tests in parts (b), (c), and (d), would the area corresponding to the *P*-value be on the left, on the right, or on both sides of the mean? Explain your answer in each case.

16. *Marketing: Shopping Time* How much customers buy is a direct result of how much time they spend in a store. A study of average shopping times in a large national housewares store gave the following information (Source: *Why We Buy: The Science of Shopping* by P. Underhill):

> Women with female companion: 8.3 min.
> Women with male companion: 4.5 min.

Suppose you want to set up a statistical test to challenge the claim that a woman with a female friend spends an average of 8.3 minutes shopping in such a store.

(a) What would you use for the null and alternate hypotheses if you believe the average shopping time is less than 8.3 minutes? Is this a right-tailed, left-tailed, or two-tailed test?

(b) What would you use for the null and alternate hypotheses if you believe the average shopping time is different from 8.3 minutes? Is this a right-tailed, left-tailed, or two-tailed test?

Stores that sell mainly to women should figure out a way to engage the interest of men—perhaps comfortable seats and a big TV with sports programs! Suppose such an entertainment center was installed and you now wish to challenge the claim that a woman with a male friend spends only 4.5 minutes shopping in a housewares store.

(c) What would you use for the null and alternate hypotheses if you believe the average shopping time is more than 4.5 minutes? Is this a right-tailed, left-tailed, or two-tailed test?

(d) What would you use for the null and alternate hypotheses if you believe the average shopping time is different from 4.5 minutes? Is this a right-tailed, left-tailed, or two-tailed test?

17. *Meteorology: Storms* *Weatherwise* magazine is published in association with the American Meteorological Society. Volume 46, Number 6 has a rating system to classify Nor'easter storms that frequently hit New England states and can cause much damage near the ocean coast. A *severe* storm has an average peak wave height of 16.4 feet for waves hitting the shore. Suppose that a Nor'easter is in progress at the severe storm class rating.

(a) Let us say that we want to set up a statistical test to see if the wave action (i.e., height) is dying down or getting worse. What would be the null hypothesis regarding average wave height?

(b) If you wanted to test the hypothesis that the storm is getting worse, what would you use for the alternate hypothesis?

(c) If you wanted to test the hypothesis that the waves are dying down, what would you use for the alternate hypothesis?

(d) Suppose you do not know whether the storm is getting worse or dying out. You just want to test the hypothesis that the average wave height is *different* (either higher or lower) from the severe storm class rating. What would you use for the alternate hypothesis?

(e) For each of the tests in parts (b), (c), and (d), would the area corresponding to the *P*-value be on the left, on the right, or on both sides of the mean? Explain your answer in each case.

18. *Chrysler Concorde: Acceleration* *Consumer Reports* stated that the mean time for a Chrysler Concorde to go from 0 to 60 miles per hour is 8.7 seconds.
   (a) If you want to set up a statistical test to challenge the claim of 8.7 seconds, what would you use for the null hypothesis?
   (b) The town of Leadville, Colorado, has an elevation over 10,000 feet. Suppose you wanted to test the claim that the average time to accelerate from 0 to 60 miles per hour is longer in Leadville (because of less oxygen). What would you use for the alternate hypothesis?
   (c) Suppose you made an engine modification and you think the average time to accelerate from 0 to 60 miles per hour is reduced. What would you use for the alternate hypothesis?
   (d) For each of the tests in parts (b) and (c), would the *P*-value area be on the left, on the right, or on both sides of the mean? Explain your answer in each case.

For Problems 19–24, please provide the following information.
   (a) What is the level of significance? State the null and alternate hypotheses. Will you use a left-tailed, right-tailed, or two-tailed test?
   (b) *Check Requirements* What sampling distribution will you use? Explain the rationale for your choice of sampling distribution. Compute the value of the sample test statistic.
   (c) Find (or estimate) the *P*-value. Sketch the sampling distribution and show the area corresponding to the *P*-value.
   (d) Based on your answers in parts (a) to (c), will you reject or fail to reject the null hypothesis? Are the data statistically significant at level α?
   (e) *Interpret* your conclusion in the context of the application.

19. *Dividend Yield: Australian Bank Stocks* Let x be a random variable representing dividend yield of Australian bank stocks. We may assume that x has a normal distribution with $\sigma = 2.4\%$. A random sample of 10 Australian bank stocks gave the following yields.

| 5.7 | 4.8 | 6.0 | 4.9 | 4.0 | 3.4 | 6.5 | 7.1 | 5.3 | 6.1 |

The sample mean is $\bar{x} = 5.38\%$. For the entire Australian stock market, the mean dividend yield is $\mu = 4.7\%$ (Reference: *Forbes*). Do these data indicate that the dividend yield of all Australian bank stocks is higher than 4.7%? Use $\alpha = 0.01$.

20. *Glucose Level: Horses* Gentle Ben is a Morgan horse at a Colorado dude ranch. Over the past 8 weeks, a veterinarian took the following glucose readings from this horse (in mg/100 ml).

| 93 | 88 | 82 | 105 | 99 | 110 | 84 | 89 |

The sample mean is $\bar{x} \approx 93.8$. Let x be a random variable representing glucose readings taken from Gentle Ben. We may assume that x has a normal distribution, and we know from past experience that $\sigma = 12.5$. The mean glucose level for horses should be $\mu = 85$ mg/100 ml (Reference: *Merck Veterinary Manual*). Do these data indicate that Gentle Ben has an overall average glucose level higher than 85? Use $\alpha = 0.05$.

21. *Ecology: Hummingbirds* Bill Alther is a zoologist who studies Anna's hummingbird (*Calypte anna*) (Reference: *Hummingbirds* by K. Long and W. Alther). Suppose that in a remote part of the Grand Canyon, a random sample of six of these birds was caught, weighed, and released. The weights (in grams) were

| 3.7 | 2.9 | 3.8 | 4.2 | 4.8 | 3.1 |

The sample mean is $\bar{x} = 3.75$ grams. Let x be a random variable representing weights of Anna's hummingbirds in this part of the Grand Canyon. We assume that x has a normal distribution and $\sigma = 0.70$ gram. It is known that for the population of all Anna's hummingbirds, the mean weight is $\mu = 4.55$ grams. Do the data indicate that the mean weight of these birds in this part of the Grand Canyon is less than 4.55 grams? Use $\alpha = 0.01$.

22. *Finance: P/E of Stocks* The price-to-earnings (P/E) ratio is an important tool in financial work. A random sample of 14 large U.S. banks (J.P. Morgan, Bank of America, and others) gave the following P/E ratios (Reference: *Forbes*).

24	16	22	14	12	13	17
22	15	19	23	13	11	18

The sample mean is $\bar{x} \approx 17.1$. Generally speaking, a low P/E ratio indicates a "value" or bargain stock. A recent copy of the *Wall Street Journal* indicated that the P/E ratio of the entire S&P 500 stock index is $\mu = 19$. Let x be a random variable representing the P/E ratio of all large U.S. bank stocks. We assume that x has a normal distribution and $\sigma = 4.5$. Do these data indicate that the P/E ratio of all U.S. bank stocks is less than 19? Use $\alpha = 0.05$.

23. *Insurance: Hail Damage* Nationally, about 11% of the total U.S. wheat crop is destroyed each year by hail (Reference: *Agricultural Statistics*, U.S. Department of Agriculture). An insurance company is studying wheat hail damage claims in Weld County, Colorado. A random sample of 16 claims in Weld County gave the following data (% wheat crop lost to hail).

15	8	9	11	12	20	14	11
7	10	24	20	13	9	12	5

The sample mean is $\bar{x} = 12.5\%$. Let x be a random variable that represents the percentage of wheat crop in Weld County lost to hail. Assume that x has a normal distribution and $\sigma = 5.0\%$. Do these data indicate that the percentage of wheat crop lost to hail in Weld County is different (either way) from the national mean of 11%? Use $\alpha = 0.01$.

24. *Medical: Red Blood Cell Volume* Total blood volume (in ml) per body weight (in kg) is important in medical research. For healthy adults, the red blood cell volume mean is about $\mu = 28$ ml/kg (Reference: *Laboratory and Diagnostic Tests* by F. Fischbach). Red blood cell volume that is too low or too high can indicate a medical problem (see reference). Suppose that Roger has had seven blood tests, and the red blood cell volumes were

32	25	41	35	30	37	29

The sample mean is $\bar{x} \approx 32.7$ ml/kg. Let x be a random variable that represents Roger's red blood cell volume. Assume that x has a normal distribution and $\sigma = 4.75$. Do the data indicate that Roger's red blood cell volume is different (either way) from $\mu = 28$ ml/kg? Use a 0.01 level of significance.

SECTION 8.2

Testing the Mean μ

FOCUS POINTS

- Review the general procedure for testing using *P*-values.
- Test μ when σ is known using the normal distribution.
- Test μ when σ is unknown using a Student's *t* distribution.
- Understand the "traditional" method of testing that uses critical regions and critical values instead of *P*-values.

In this section, we continue our study of testing the mean μ. The method we are using is called the *P*-value method. It was used extensively by the famous statistician R. A. Fisher and is the most popular method of testing in use today. At the end of this section, we present another method of testing called the *critical region method* (or *traditional method*). The critical region method was used extensively

by the statisticians J. Neyman and E. Pearson. In recent years, the use of this method has been declining. It is important to realize that for a fixed, preset level of significance α, both methods are logically equivalent.

In Section 8.1, we discussed the vocabulary and method of hypothesis testing using P-values. Let's quickly review the basic process.

1. We first state a proposed value for a population parameter in the null hypothesis H_0. The alternate hypothesis H_1 states alternative values of the parameter, either $<$, $>$, or $\neq$ the value proposed in H_0. We also set the level of significance α. This is the risk we are willing to take of committing a type I error. That is, α is the probability of rejecting H_0 when it is, in fact, true.

2. We use a corresponding sample statistic from a simple random sample to challenge the statement made in H_0. We convert the sample statistic to a test statistic, which is the corresponding value of the appropriate sampling distribution.

3. We use the sampling distribution of the test statistic and the type of test to compute the P-value of this statistic. Under the assumption that the null hypothesis is true, the P-value is the probability of getting a sample statistic as extreme as or more extreme than the observed statistic from our random sample.

4. Next, we conclude the test. If the P-value is very small, we have evidence to reject H_0 and adopt H_1. What do we mean by "very small"? We compare the P-value to the preset level of significance α. If the P-value $\leq \alpha$, then we say that we have evidence to reject H_0 and adopt H_1. Otherwise, we say that the sample evidence is insufficient to reject H_0.

5. Finally, we *interpret* the results in the context of the application.

Knowing the sampling distribution of the sample test statistic is an essential part of the hypothesis testing process. For tests of μ, we use one of two sampling distributions for $\overline{x}$: the standard normal distribution or a Student's t distribution. As discussed in Chapters 6 and 7, the appropriate distribution depends upon our knowledge of the population standard deviation σ, the nature of the x distribution, and the sample size.

Part I: Testing μ When σ Is Known

In most real-world situations, σ is simply not known. However, in some cases a preliminary study or other information can be used to get a realistic and accurate value for σ.

PROCEDURE

HOW TO TEST μ WHEN σ IS KNOWN

Requirements

Let x be a random variable appropriate to your application. Obtain a simple random sample (of size n) of x values from which you compute the sample mean $\overline{x}$. The value of σ is already known (perhaps from a previous study). If you can assume that x has a normal distribution, then any sample size n will work. If you cannot assume this, then use a sample size $n \geq 30$.

Procedure

1. In the context of the application, state the *null and alternate hypotheses* and set the *level of significance* α.

2. Use the known σ, the sample size n, the value of x from the sample, and μ from the null hypothesis to compute the standardized *sample test statistic*.

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

3. Use the standard normal distribution and the type of test, one-tailed or two-tailed, to find the *P-value* corresponding to the test statistic.

4. *Conclude* the test. If $P\text{-value} \leq \alpha$, then reject H_0. If $P\text{-value} > \alpha$, then do not reject H_0.

5. *Interpret your conclusion* in the context of the application.

In Section 8.1, we examined *P*-value tests for normal distributions with relatively small sample sizes ($n < 30$). The next example does not assume a normal distribution, but has a large sample size ($n \geq 30$).

EXAMPLE 3   TESTING μ, σ KNOWN

Sunspots have been observed for many centuries. Records of sunspots from ancient Persian and Chinese astronomers go back thousands of years. Some archaeologists think sunspot activity may somehow be related to prolonged periods of drought in the southwestern United States. Let x be a random variable representing the average number of sunspots observed in a four-week period. A random sample of 40 such periods from Spanish colonial times gave the following data (Reference: M. Waldmeir, *Sun Spot Activity,* International Astronomical Union Bulletin).

12.5	14.1	37.6	48.3	67.3	70.0	43.8	56.5	59.7	24.0
12.0	27.4	53.5	73.9	104.0	54.6	4.4	177.3	70.1	54.0
28.0	13.0	6.5	134.7	114.0	72.7	81.2	24.1	20.4	13.3
9.4	25.7	47.8	50.0	45.3	61.0	39.0	12.0	7.2	11.3

The sample mean is $\bar{x} \approx 47.0$. Previous studies of sunspot activity during this period indicate that $\sigma = 35$. It is thought that for thousands of years, the mean number of sunspots per four-week period was about $\mu = 41$. Sunspot activity above this level may (or may not) be linked to gradual climate change. Do the data indicate that the mean sunspot activity during the Spanish colonial period was higher than 41? Use $\alpha = 0.05$.

SOLUTION:

(a) Establish the null and alternate hypotheses.

Since we want to know whether the average sunspot activity during the Spanish colonial period was higher than the long-term average of $\mu = 41$,

$$H_0\colon \mu = 41 \quad \text{and} \quad H_1\colon \mu > 41$$

(b) *Check Requirements* What distribution do we use for the sample test statistic? Compute the test statistic from the sample data. Since $n \geq 30$ and we know σ, we use the standard normal distribution. Using $\bar{x} = 47$ from the sample, $\sigma = 35$, $\mu = 41$ from H_0, and $n = 40$,

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \approx \frac{47 - 41}{35/\sqrt{40}} \approx 1.08$$

(c) Find the *P*-value of the test statistic.

Figure 8-3 shows the *P*-value. Since we have a right-tailed test, the *P*-value is the area to the right of $z = 1.08$ shown in Figure 8-3. Using Table 5 of Appendix II, we find that
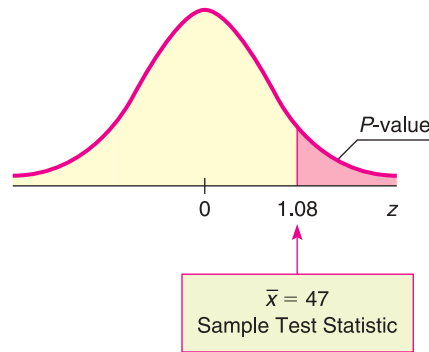
$$P\text{-value} = P(z > 1.08) \approx 0.1401.$$
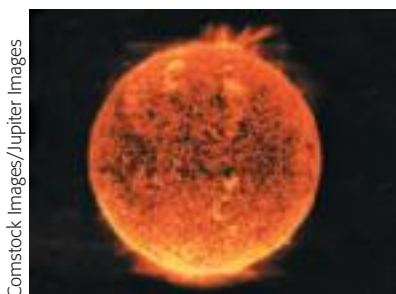
P-value Area

P-value

0   1.08   *z*

$\bar{x} = 47$
Sample Test Statistic

(d) Conclude the test.
    Since the *P*-value of $0.1401 > 0.05$ for α we do not reject H_0.

(e) *Interpretation* Interpret the results in the context of the problem.
    At the 5% level of significance, the evidence is not sufficient to reject H_0. Based on the sample data, we do not think the average sunspot activity during the Spanish colonial period was higher than the long-term mean.

Part II: Testing μ When σ Is Unknown

In many real-world situations, you have only a random sample of data values. In addition, you may have some limited information about the probability distribution of your data values. Can you still test μ under these circumstances? In most cases, the answer is yes!

PROCEDURE

HOW TO TEST μ WHEN σ IS UNKNOWN

Requirements

Let *x* be a random variable appropriate to your application. Obtain a simple random sample (of size *n*) of *x* values from which you compute the sample mean $\bar{x}$ and the sample standard deviation *s*. If you can assume that *x* has a normal distribution or simply a mound-shaped and symmetric distribution, then any sample size *n* will work. If you cannot assume this, use a sample size $n \geq 30$.

Procedure

1. In the context of the application, state the *null and alternate hypotheses* and set the *level of significance* α.

2. Use $\bar{x}$, *s*, and *n* from the sample, with μ from H_0, to compute the *sample test statistic*.

d.f. for testing μ when σ unknown

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \quad \text{with degrees of freedom } d.f. = n - 1$$

3. Use the Student's *t* distribution and the type of test, one-tailed or two-tailed, to find (or estimate) the *P-value* corresponding to the test statistic.

4. *Conclude* the test. If P-value $\leq \alpha$, then reject H_0. If P-value $> \alpha$, then do not reject H_0.

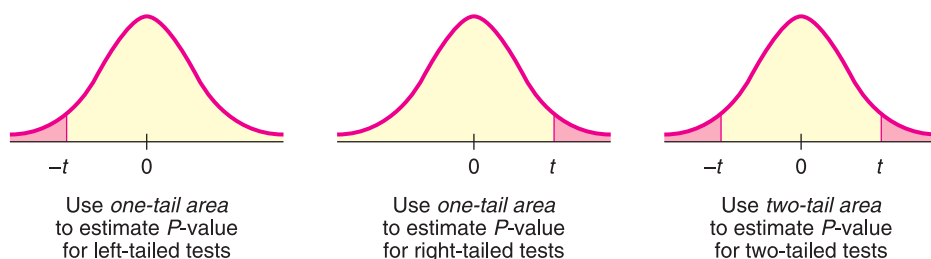5. *Interpret your conclusion* in the context of the application.

In Sections 7.2 and 7.4, we used Table 6 of Appendix II, Student's
t Distribution, to find critical values t_c for confidence intervals. The critical values are in the body of the table. We find *P*-values in the *rows* headed by "one-tail area" and "two-tail area," depending on whether we have a one-tailed or two-tailed test. If the test statistic *t* for the sample statistic $\bar{x}$ is negative, look up the *P*-value for the corresponding *positive* value of *t* (i.e., look up the *P*-value for $|t|$).

Note: In Table 6, areas are given in *one tail* beyond positive *t* on the right or negative *t* on the left, and in *two tails* beyond $\pm t$. Notice that in each column, two-tail area = 2(one-tail area). Consequently, we use *one-tail areas* as endpoints of the interval containing the *P*-value for *one-tailed tests*. We use *two-tail areas* as endpoints of the interval containing the *P*-value for *two-tailed tests*. (See Figure 8-4.)

Example 4 and Guided Exercise 4 show how to use Table 6 of Appendix II to find an interval containing the *P*-value corresponding to a test statistic *t*.

FIGURE 8-4

P-value for One-Tailed Tests and for
Two-Tailed Tests

| Use *one-tail area* to estimate *P*-value for left-tailed tests | Use *one-tail area* to estimate *P*-value for right-tailed tests | Use *two-tail area* to estimate *P*-value for two-tailed tests |

EXAMPLE 4   TESTING μ, σ UNKNOWN

The drug 6-mP (6-mercaptopurine) is used to treat leukemia. The following data represent the remission times (in weeks) for a random sample of 21 patients using 6-mP (Reference: E. A. Gehan, University of Texas Cancer Center).

10	7	32	23	22	6	16	34	32	25	11
20	19	6	17	35	6	13	9	6	10	

The sample mean is $\bar{x} \approx 17.1$ weeks, with sample standard deviation $s \approx 10.0$. Let *x* be a random variable representing the remission time (in weeks) for all patients using 6-mP. Assume the *x* distribution is mound-shaped and symmetric. A previously used drug treatment had a mean remission time of $\mu = 12.5$ weeks. Do the data indicate that the mean remission time using the drug 6-mP is different (either way) from 12.5 weeks? Use $\alpha = 0.01$.

SOLUTION:

(a) Establish the null and alternate hypotheses.
Since we want to determine if the drug 6-mP provides a mean remission time that is different from that provided by a previously used drug having $\mu = 12.5$ weeks,

$H_0: \mu = 12.5$ weeks   and $H_1: \mu \neq 12.5$ weeks

(b) *Check Requirements* What distribution do we use for the sample test statistic *t*? Compute the sample test statistic from the sample data.
The *x* distribution is assumed to be mound-shaped and symmetric. Because we don't know σ, we use a Student's *t* distribution with *d.f.* = 20. Using $\bar{x} \approx 17.1$ and $s \approx 10.0$ from the sample data, $\mu = 12.5$ from H_0, and $n = 21$,

$$t \approx \frac{\bar{x} - \mu}{s/\sqrt{n}} \approx \frac{17.1 - 12.5}{10.0/\sqrt{21}} \approx 2.108$$

(c) Find the *P*-value or the interval containing the *P*-value.
Figure 8-5 shows the *P*-value. Using Table 6 of Appendix II, we find an interval containing the *P*-value. Since this is a two-tailed test, we use entries from the row headed by *two-tail area*. Look up the *t* value in the row headed by $d.f. = n - 1 = 21 - 1 = 20$. The sample statistic $t = 2.108$ falls between 2.086 and 2.528. The *P*-value for the sample *t* falls between the corresponding two-tail areas 0.050 and 0.020. (See Table 8-5.)

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

$\bar{x} = 17.1$
Sample Test Statistic
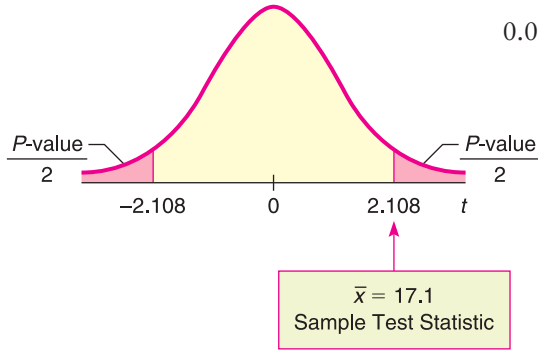
TABLE 8-5	**Excerpt from Student's *t* Distribution (Table 6, Appendix II)**		
one-tail area		...	...
✓ two-tail area		0.050	0.020
d.f. = 20		2.086	2.528
		Sample *t* = 2.108	

(d) Conclude the test.
The following diagram shows the interval that contains the single *P*-value corresponding to the test statistic. Note that there is just one *P*-value corresponding to the test statistic. Table 6 of Appendix II does not give that specific value, but it does give a range that contains that specific *P*-value. As the diagram shows, the entire range is greater than α. This means the specific *P*-value is greater than α, so we cannot reject H_0.

α
0.01   0.020   0.050
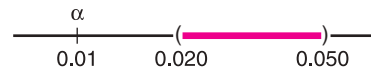
Note: Using the raw data, computer software gives *P*-value ≈ 0.048. This value is in the interval we estimated. It is larger than the α value of 0.01, so we do not reject H_0.

(e) *Interpretation* Interpret the results in the context of the problem.
At the 1% level of significance, the evidence is not sufficient to reject H_0. Based on the sample data, we cannot say that the drug 6-mP provides a different average remission time than the previous drug.

GUIDED EXERCISE 4 | **Testing μ, σ unknown**

Archaeologists become excited when they find an anomaly in discovered artifacts. The anomaly may (or may not) indicate a new trading region or a new method of craftsmanship. Suppose the lengths of projectile points (arrowheads) at a certain archaeological site have mean length $\mu = 2.6$ cm. A random sample of 61 recently discovered projectile points in an adjacent cliff dwelling gave the following lengths (in cm) (Reference: A. Woosley and A. McIntyre, *Mimbres Mogollon Archaeology*, University of New Mexico Press).

3.1	4.1	1.8	2.1	2.2	1.3	1.7	3.0	3.7	2.3	2.6	2.2	2.8	3.0
3.2	3.3	2.4	2.8	2.8	2.9	2.9	2.2	2.4	2.1	3.4	3.1	1.6	3.1
3.5	2.3	3.1	2.7	2.1	2.0	4.8	1.9	3.9	2.0	5.2	2.2	2.6	1.9
4.0	3.0	3.4	4.2	2.4	3.5	3.1	3.7	3.7	2.9	2.6	3.6	3.9	3.5
1.9	4.0	4.0	4.6	1.9									

Continued

GUIDED EXERCISE 4 *continued*

The sample mean is $\bar{x} \approx 2.92$ cm and the sample standard deviation is $s \approx 0.85$, where x is a random variable that represents the lengths (in cm) of all projectile points found at the adjacent cliff dwelling site. Do these data indicate that the mean length of projectile points in the adjacent cliff dwelling is longer than 2.6 cm? Use a 1% level of significance.

(a) State H_0, H_1, and α.

H_0: $\mu = 2.6$ cm; H_1: $\mu > 2.6$ cm; $\alpha = 0.01$

(b) *Check Requirements* What sampling distribution should you use? What is the value of the sample test statistic t?

Because $n \geq 30$ and σ is unknown, use the Student's t distribution with $d.f. = n - 1 = 61 - 1 = 60$. Using $\bar{x} \approx 2.92$, $s \approx 0.85$, $\mu = 2.6$ from H_0 and $n = 61$,

$$t = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \approx \frac{2.92 - 2.6}{0.85/\sqrt{61}} \approx 2.940$$

(c) When you use Table 6, Appendix II, to find an interval containing the P-value, do you use one-tail or two-tail areas? Why? Sketch a figure showing the P-value. Find an interval containing the P-value.

This is a right-tailed test, so use a one-tail area.
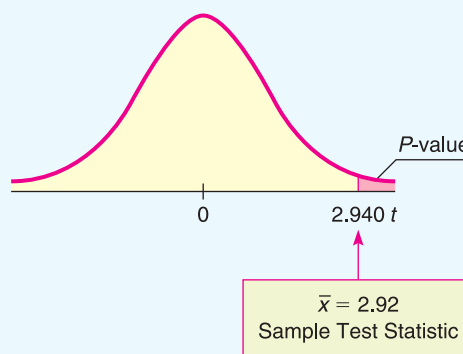
FIGURE 8-6 *P-value*

TABLE 8-6   **Excerpt from Student's *t* Table**

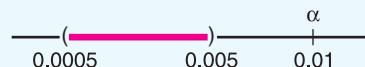|		
✓ one-tail area	... 0.005	0.0005
two-tail area	... 0.010	0.0010
d.f. = 60	... 2.660	3.460
	↑	
	Sample *t* = 2.940	

Using $d.f. = 60$, we find that the sample $t = 2.940$ is between the critical values 2.660 and 3.460. The sample P-value is then between the one-tail areas 0.005 and 0.0005.

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

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

Since the interval containing the P-value lies to the left of $\alpha = 0.01$, we reject H_0.

Note: Using the raw data, computer software gives P-value ≈ 0.0022. This value is in our estimated range and is less than $\alpha = 0.01$ so we reject H_0.

(e) *Interpretation* Interpret your results in the context of the application.

At the 1% level of significance, sample evidence is sufficiently strong to reject H_0 and conclude that the average projectile point length at the adjacent cliff dwelling site is longer than 2.6 cm.

TECH NOTES   The TI-84Plus/TI-83Plus/TI-*n*spire calculators, Excel 2007, and Minitab all support testing of μ using the standard normal distribution. The TI-84Plus/TI-83Plus/TI-*n*spire and Minitab support testing of μ using a Student's t distribution. All the technologies return a P-value for the test.

TI-84Plus/TI-83Plus/TI-_n_spire (with TI-84Plus keypad) You can select to enter raw data (**Data**) or summary statistics (**Stats**). Enter the value of μ_0 used in the null hypothesis $H_0\colon \mu = \mu_0$. Select the symbol used in the alternate hypothesis ($\neq\mu_0$, $<\mu_0$, $>\mu_0$). To test μ using the standard normal distribution, press **Stat,** select **Tests,** and use option **1:Z-Test.** The value for σ is required. To test μ using a Student's t distribution, use option **2:T-Test.** Using data from Example 4 regarding remission times, we have the following displays. The P-value is given as p.

```
T-Test
  Inpt:Data Stats
  μ₀ :12.5
  List:L₁
  Freq:1
  μ:≠μ₀ <μ₀ >μ₀
  Calculate Draw
```

```
T-Tests
  μ≠12.5
  t=2.105902924
  p=.0480466063
  x̄=17.0952381
  Sx=9.999523798
  n=21
```

Excel 2007 In Excel, the **ZTEST** function finds the P-values for a right-tailed test. Click the ribbon choice **Insert Function** (f_x). In the dialogue box, select **Statistical** for the category and **ZTEST** for the function. In the next dialogue box, give the cell range containing your data for the array. Use the value of μ stated in H_0 for x. Provide σ. Otherwise, Excel uses the sample standard deviation computed from the data.

Minitab Enter the raw data from a sample. Use the menu selections **Stat ➤ Basic Stat ➤ 1-Sample z** for tests using the standard normal distribution. For tests of μ using a Student's t distribution, select **1-Sample t.**

Part III: Testing μ Using Critical Regions (Traditional Method)

The most popular method of statistical testing is the P-value method. For that reason, the P-value method is emphasized in this book. Another method of testing is called the *critical region method* or *traditional method*.

Critical region method

For a fixed, preset value of the level of significance α, both methods are logically equivalent. Because of this, we treat the traditional method as an "optional" topic and consider only the case of testing μ when σ is known.

Consider the null hypothesis $H_0\colon \mu = k$. We use information from a random sample, together with the sampling distribution for $\overline{x}$ and the level of significance α, to determine whether or not we should reject the null hypothesis. The essential question is, "How much can $\overline{x}$ vary from $\mu = k$ before we suspect that $H_0\colon \mu = k$ is false and reject it?"

The answer to the question regarding the relative sizes of $\overline{x}$ and μ, as stated in the null hypothesis, depends on the sampling distribution of $\overline{x}$, the alternate hypothesis H_1, and the level of significance α. If the sample test statistic $\overline{x}$ is sufficiently different from the claim about μ made in the null hypothesis, we reject the null hypothesis.

The values of $\overline{x}$ for which we reject H_0 are called the *critical region* of the $\overline{x}$ distribution. Depending on the alternate hypothesis, the critical region is located on the left side, the right side, or both sides of the $\overline{x}$, distribution. Figure 8-7 shows the relationship of the critical region to the alternate hypothesis and the level of significance α.

Notice that the total area in the critical region is preset to be the level of significance α. This is *not* the P-value discussed earlier! In fact, you cannot set the P-value in advance because it is determined from a random sample. Recall that the level of significance α should (in theory) be a fixed, preset number assigned before drawing any samples.

Another method for concluding two-tailed tests involves the use of confidence intervals. Problems 25 and 26 at the end of this section discuss the confidence interval method.

FIGURE 8-7

Critical Regions for H_0: $\mu = k$

(a) H_1: $\mu < k$
     Left-tailed

(b) H_1: $\mu > k$
     Right-tailed

(c) H_1: $\mu \neq k$
     Two-tailed

Area = α

Area = α

Area = $\alpha/2$ Area = $\alpha/2$

k k k

Critical region   Critical region   Critical regions
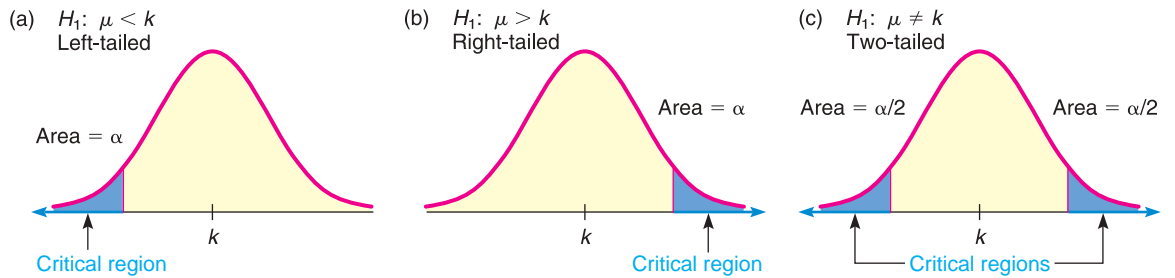
Critical values

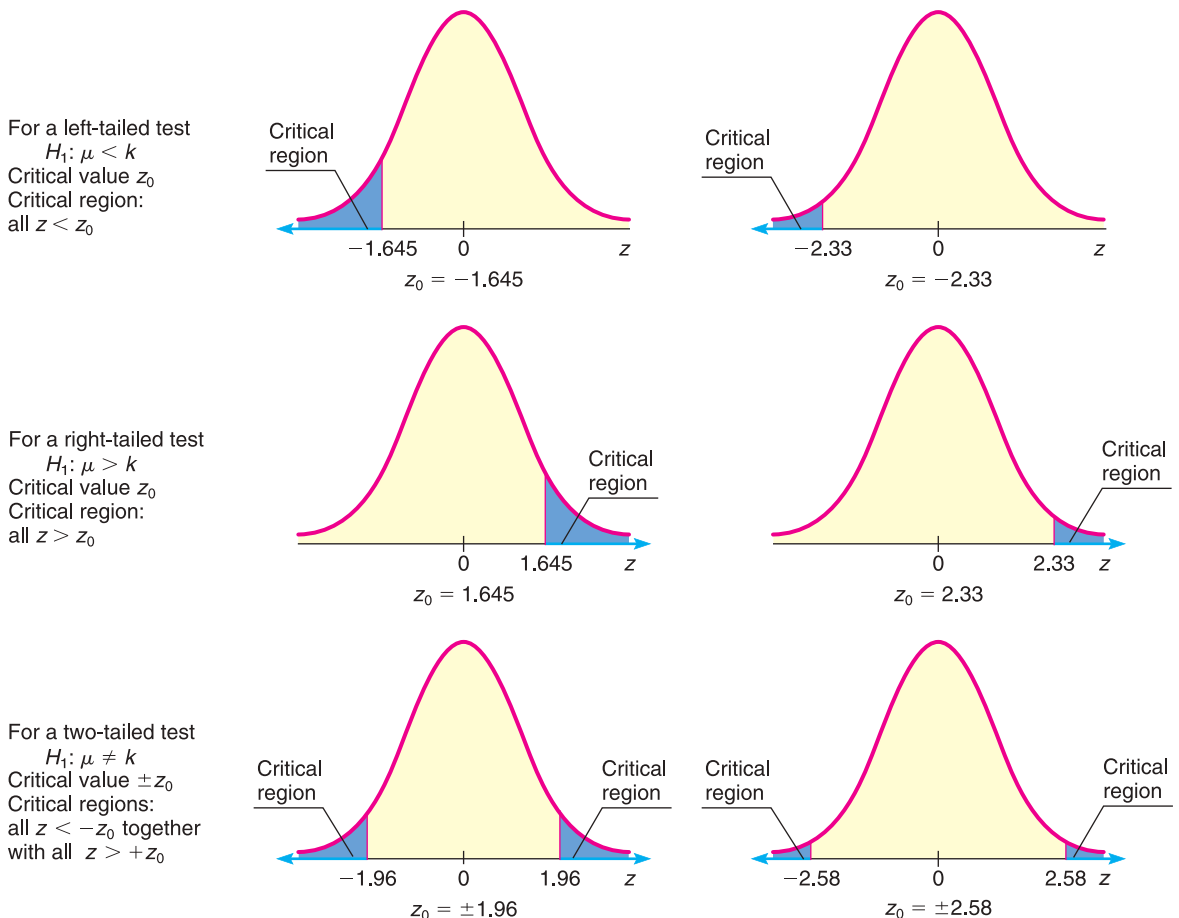The most commonly used levels of significance are $\alpha = 0.05$ and $\alpha = 0.01$. Critical regions of a standard normal distribution are shown for these levels of significance in Figure 8-8. *Critical values* are the boundaries of the critical region. Critical values designated as z_0 for the standard normal distribution are shown in Figure 8-8. For easy reference, they are also included in Table 5 of Appendix II, Areas of a Standard Normal Distribution.

The procedure for hypothesis testing using critical regions follows the same first two steps as the procedure using *P*-values. However, instead of finding a *P*-value for the sample test statistic, we check if the sample test statistic falls in the critical region. If it does, we reject H_0. Otherwise, we do not reject H_0.

FIGURE 8-8

Critical Values z_0 for Tests Involving a Mean (Large Samples)

Level of significance $\alpha = 0.05$ $\alpha = 0.01$

For a left-tailed test
H_1: $\mu < k$
Critical value z_0
Critical region:
all $z < z_0$

Critical region -1.645 0 z
$z_0 = -1.645$

Critical region -2.33 0 z
$z_0 = -2.33$

For a right-tailed test
H_1: $\mu > k$
Critical value z_0
Critical region:
all $z > z_0$

Critical region 0 1.645 z
$z_0 = 1.645$

Critical region 0 2.33 z
$z_0 = 2.33$

For a two-tailed test
H_1: $\mu \neq k$
Critical value $\pm z_0$
Critical regions:
all $z < -z_0$ together
with all $z > +z_0$

Critical region   Critical region -1.96 0 1.96 z
$z_0 = \pm 1.96$

Critical region   Critical region -2.58 0 2.58 z
$z_0 = \pm 2.58$

PROCEDURE	HOW TO TEST μ WHEN σ IS KNOWN (CRITICAL REGION METHOD)

Let x be a random variable appropriate to your application. Obtain a simple random sample (of size n) of x values from which you compute the sample mean $\bar{x}$. The value of σ is already known (perhaps from a previous study). If you can assume that x has a normal distribution, then any sample size n will work. If you cannot assume this, use a sample size $n \geq 30$. Then $\bar{x}$ follows a distribution that is normal or approximately normal.

1. In the context of the application, state the *null and alternate hypotheses* and set the *level of significance* α. We use the most popular choices, $\alpha = 0.05$ or $\alpha = 0.01$.

2. Use the known σ, the sample size n, the value of $\bar{x}$ from the sample, and μ from the null hypothesis to compute the standardized sample *test statistic*.

$$z = \frac{\bar{x} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$$

3. Show the *critical region* and *critical value(s)* on a graph of the sampling distribution. The level of significance α and the alternate hypothesis determine the locations of critical regions and critical values.

4. *Conclude* the test. If the test statistic z computed in Step 2 is in the critical region, then reject H_0. If the test statistic z is not in the critical region, then do not reject H_0.

5. *Interpret your conclusion* in the context of the application.

EXAMPLE 5   CRITICAL REGION METHOD OF TESTING μ

Consider Example 3 regarding sunspots. Let x be a random variable representing the number of sunspots observed in a four-week period. A random sample of 40 such periods from Spanish colonial times gave the number of sunspots per period. The raw data are given in Example 3. The sample mean is $\bar{x} \approx 47.0$. Previous studies indicate that for this period, $\sigma = 35$. It is thought that for thousands of years, the mean number of sunspots per four-week period was about $\mu = 41$. Do the data indicate that the mean sunspot activity during the Spanish colonial period was higher than 41? Use $\alpha = 0.05$.

SOLUTION:

(a) Set the null and alternate hypotheses.
   As in Example 3, we use H_0: $\mu = 41$ and H_1: $\mu > 41$.

(b) Compute the sample test statistic.
   As in Example 3, we use the standard normal distribution, with $\bar{x} = 47$, $\sigma = 35$, $\mu = 41$ from H_0, and $n = 40$.

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \approx \frac{47 - 41}{35/\sqrt{40}} \approx 1.08$$

(c) Determine the critical region and critical value based on H_1 and $\alpha = 0.05$.
   Since we have a right-tailed test, the critical region is the rightmost 5% of the standard normal distribution. According to Figure 8-8, the critical value is $z_0 = 1.645$.

(d) Conclude the test.

We conclude the test by showing the critical region, critical value, and sample test statistic $z = 1.08$ on the standard normal curve. For a right-tailed test with $\alpha = 0.05$ the critical value is $z_0 = 1.645$. Figure 8-9 shows the critical region. As we can see, the sample test statistic does not fall in the critical region. Therefore, we fail to reject H_0.

FIGURE 8-9

Critical Region, $\alpha = 0.05$
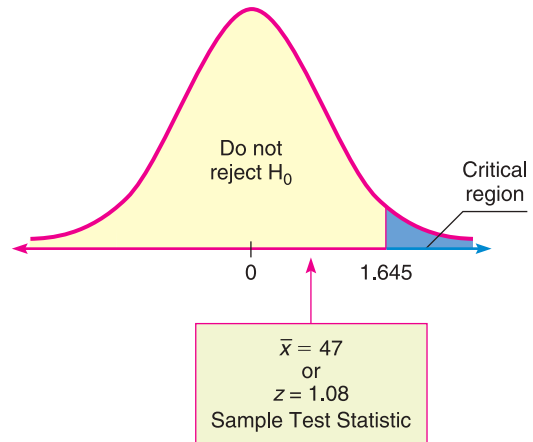
Do not reject H_0

Critical region

0   1.645

$\bar{x} = 47$
or
$z = 1.08$
Sample Test Statistic

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

At the 5% level of significance, the sample evidence is insufficient to justify rejecting H_0. It seems that the average sunspot activity during the Spanish colonial period was the same as the historical average.

(f) How do results of the critical region method compare to the results of the *P*-value method for a 5% level of significance?

The results, as expected, are the same. In both cases, we fail to reject H_0.

The critical region method of testing as outlined applies to tests of other parameters. As with the *P*-value method, you need to know the sampling distribution of the sample test statistic. Critical values for distributions are usually found in tables rather than in computer software outputs. For example, Table 6 of Appendix II provides critical values for Student's *t* distributions.

The critical region method of hypothesis testing is very general. The following procedure box outlines the process of concluding a hypothesis test using the critical region method.

PROCEDURE

HOW TO CONCLUDE TESTS USING THE CRITICAL REGION METHOD

1. Compute the sample test statistic using an appropriate sampling distribution.

2. Using the same sampling distribution, find the critical value(s) as determined by the level of significance α and the nature of the test: right-tailed, left-tailed, or two-tailed.

3. Compare the sample test statistic to the critical value(s).

   (a) For a right-tailed test,

       i. if sample test statistic $\geq$ critical value, reject H_0.

       ii. if sample test statistic $<$ critical value, fail to reject H_0.

Continued

(b) For a left-tailed test,

    i. if sample test statistic $\leq$ critical value, reject H_0.

    ii. if sample test statistic $>$ critical value, fail to reject H_0.

(c) For a two-tailed test,

    i. if sample test statistic lies at or beyond critical values, reject H_0.

    ii. if sample test statistic lies between critical values, fail to reject H_0.

VIEWPOINT | Predator or Prey?

Consider animals such as the arctic fox, gray wolf, desert lion, and South American jaguar. Each animal is a predator. What are the total sleep time (hours per day), maximum life span (years), and overall danger index from other animals? Now consider prey such as rabbits, deer, wild horses, and the Brazilian tapir (a wild pig). Is there a statistically significant difference in average sleep time, life span, and danger index? What about other variables such as the ratio of brain weight to body weight or the sleep exposure index (sleeping in a well-protected den or out in the open)? How did prehistoric humans fit into this picture? Scientists have collected a lot of data, and a great deal of statistical work has been done regarding such questions. For more information, see the web site **http://lib.stat.cmu.edu/** *and follow the links to Datasets and then Sleep.*

SECTION 8.2 PROBLEMS

1. *Statistical Literacy* For the same sample data and null hypothesis, how does the *P*-value for a two-tailed test of μ compare to that for a one-tailed test?

2. *Statistical Literacy* To test μ for an x distribution that is mound-shaped using sample size $n \geq 30$, how do you decide whether to use the normal or the Student's t distribution?

3. *Statistical Literacy* When using the Student's t distribution to test μ, what value do you use for the degrees of freedom?

4. *Critical Thinking* Consider a test for μ. If the *P*-value is such that you can reject H_0 at the 5% level of significance, can you always reject H_0 at the 1% level of significance? Explain.

5. *Critical Thinking* Consider a test for μ. If the *P*-value is such that you can reject H_0 for $\alpha = 0.01$, can you always reject H_0 for $\alpha = 0.05$? Explain.

6. *Critical Thinking* If sample data is such that for a one-tailed test of μ you can reject H_0 at the 1% level of significance, can you always reject H_0 for a two-tailed test at the same level of significance? Explain.

7. *Basic Computation: P-value Corresponding to t Value* For a Student's t distribution with $d.f. = 10$ and $t = 2.930$,
(a) find an interval containing the corresponding *P*-value for a two-tailed test.
(b) find an interval containing the corresponding *P*-value for a right-tailed test.

8. *Basic Computation: P-value Corresponding to t Value* For a Student's t distribution with $d.f. = 16$ and $t = -1.830$,
(a) find an interval containing the corresponding *P*-value for a two-tailed test.
(b) find an interval containing the corresponding *P*-vaiue for a left-tailed test.

9. *Basic Computation: Testing μ, σ Unknown* A random sample of 25 values is drawn from a mound-shaped and symmetric distribution. The sample mean is 10 and the sample standard deviation is 2. Use a level of significance of 0.05 to conduct a two-tailed test of the claim that the population mean is 9.5.
   (a) *Check Requirements* Is it appropriate to use a Student's t distribution? Explain. How many degrees of freedom do we use?
   (b) What are the hypotheses?
   (c) Compute the sample test statistic t.
   (d) Estimate the P-value for the test.
   (e) Do we reject or fail to reject H_0?
   (f) *Interpret* the results.

10. *Basic Computation: Testing μ, σ Unknown* A random sample has 49 values. The sample mean is 8.5 and the sample standard deviation is 1.5. Use a level of significance of 0.01 to conduct a left-tailed test of the claim that the population mean is 9.2.
   (a) *Check Requirements* Is it appropriate to use a Student's t distribution? Explain. How many degrees of freedom do we use?
   (b) What are the hypotheses?
   (c) Compute the sample test statistic t.
   (d) Estimate the P-value for the test.
   (e) Do we reject or fail to reject H_0?
   (f) *Interpret* the results.

Please provide the following information for Problems 11–22.
   (a) What is the level of significance? State the null and alternate hypotheses.
   (b) *Check Requirements* What sampling distribution will you use? Explain the rationale for your choice of sampling distribution. Compute the value of the sample test statistic.
   (c) Find (or estimate) the P-value. Sketch the sampling distribution and show the area corresponding to the P-value.
   (d) Based on your answers in parts (a) to (c), will you reject or fail to reject the null hypothesis? Are the data statistically significant at level α?
   (e) *Interpret* your conclusion in the context of the application.
   Note: For degrees of freedom *d.f.* not given in the Student's t table, use the closest *d.f.* that is *smaller*. In some situations, this choice of *d.f.* may increase the P-value by a small amount and therefore produce a slightly more "conservative" answer.

11. *Meteorology: Storms* *Weatherwise* is a magazine published by the American Meteorological Society. One issue gives a rating system used to classify Nor'easter storms that frequently hit New England and can cause much damage near the ocean. A severe storm has an average peak wave height of $\mu = 16.4$ feet for waves hitting the shore. Suppose that a Nor'easter is in progress at the severe storm class rating. Peak wave heights are usually measured from land (using binoculars) off fixed cement piers. Suppose that a reading of 36 waves showed an average wave height of $\bar{x} = 17.3$ feet. Previous studies of severe storms indicate that $\sigma = 3.5$ feet. Does this information suggest that the storm is (perhaps temporarily) increasing above the severe rating? Use $\alpha = 0.01$.

12. *Medical: Blood Plasma* Let x be a random variable that represents the pH of arterial plasma (i.e., acidity of the blood). For healthy adults, the mean of the x distribution is $\mu = 7.4$ (Reference: *Merck Manual,* a commonly used reference in medical schools and nursing programs). A new drug for arthritis has been developed. However, it is thought that this drug may change blood pH. A random sample of 31 patients with arthritis took the drug for 3 months. Blood tests showed that $\bar{x} = 8.1$ with sample standard deviation $s = 1.9$. Use a 5% level of significance to test the claim that the drug has changed (either way) the mean pH level of the blood.

13. *Wildlife: Coyotes* A random sample of 46 adult coyotes in a region of northern Minnesota showed the average age to be $\bar{x} = 2.05$ years, with sample standard deviation $s = 0.82$ years (based on information from the book *Coyotes: Biology, Behavior and Management* by M. Bekoff, Academic Press). However, it is thought that the overall population mean age of coyotes is $\mu = 1.75$. Do the sample data indicate that coyotes in this region of northern Minnesota tend to live longer than the average of 1.75 years? Use $\alpha = 0.01$.

14. *Fishing: Trout* Pyramid Lake is on the Paiute Indian Reservation in Nevada. The lake is famous for cutthroat trout. Suppose a friend tells you that the average length of trout caught in Pyramid Lake is $\mu = 19$ inches. However, the *Creel Survey* (published by the Pyramid Lake Paiute Tribe Fisheries Association) reported that of a random sample of 51 fish caught, the mean length was $\bar{x} = 18.5$ inches, with estimated standard deviation $s = 3.2$ inches. Do these data indicate that the average length of a trout caught in Pyramid Lake is less than $\mu = 19$ inches? Use $\alpha = 0.05$.

15. *Investing: Stocks* Socially conscious investors screen out stocks of alcohol and tobacco makers, firms with poor environmental records, and companies with poor labor practices. Some examples of "good," socially conscious companies are Johnson and Johnson, Dell Computers, Bank of America, and Home Depot. The question is, are such stocks overpriced? One measure of value is the P/E, or price-to-earnings, ratio. High P/E ratios may indicate a stock is overpriced. For the S&P stock index of all major stocks, the mean P/E ratio is $\mu = 19.4$. A random sample of 36 "socially conscious" stocks gave a P/E ratio sample mean of $\bar{x} = 17.9$, with sample standard deviation $s = 5.2$ (Reference: *Morningstar*, a financial analysis company in Chicago). Does this indicate that the mean P/E ratio of all socially conscious stocks is different (either way) from the mean P/E ratio of the S&P stock index? Use $\alpha = 0.05$.

16. *Agriculture: Ground Water* Unfortunately, arsenic occurs naturally in some ground water (Reference: *Union Carbide Technical Report K/UR-1*). A mean arsenic level of $\mu = 8.0$ parts per billion (ppb) is considered safe for agricultural use. A well in Texas is used to water cotton crops. This well is tested on a regular basis for arsenic. A random sample of 37 tests gave a sample mean of $\bar{x} = 7.2$ ppb arsenic, with $s = 1.9$ ppb. Does this information indicate that the mean level of arsenic in this well is less than 8 ppb? Use $\alpha = 0.01$.

17. *Medical: Red Blood Cell Count* Let x be a random variable that represents red blood cell (RBC) count in millions of cells per cubic millimeter of whole blood. Then x has a distribution that is approximately normal. For the population of healthy female adults, the mean of the x distribution is about 4.8 (based on information from *Diagnostic Tests with Nursing Implications*, Springhouse Corporation). Suppose that a female patient has taken six laboratory blood tests over the past several months and that the RBC count data sent to the patient's doctor are

4.9   4.2   4.5   4.1   4.4   4.3

i. Use a calculator with sample mean and sample standard deviation keys to verify that $\bar{x} = 4.40$ and $s \approx 0.28$.
ii. Do the given data indicate that the population mean RBC count for this patient is lower than 4.8? Use $\alpha = 0.05$.

18. *Medical: Hemoglobin Count* Let x be a random variable that represents hemoglobin count (HC) in grams per 100 milliliters of whole blood. Then x has a distribution that is approximately normal, with population mean of about 14 for healthy adult women (see reference in Problem 17). Suppose that a female patient has taken 10 laboratory blood tests during the past year. The HC data sent to the patient's doctor are

15   18   16   19   14   12   14   17   15   11

    i. Use a calculator with sample mean and sample standard deviation keys to verify that $\bar{x} = 15.1$ and $s \approx 2.51$.

   ii. Does this information indicate that the population average HC for this patient is higher than 14? Use $\alpha = 0.01$.

19. *Ski Patrol: Avalanches* Snow avalanches can be a real problem for travelers in the western United States and Canada. A very common type of avalanche is called the slab avalanche. These have been studied extensively by David McClung, a professor of civil engineering at the University of British Columbia. Slab avalanches studied in Canada have an average thickness of $\mu = 67$ (Source: *Avalanche Handbook* by D. McClung and P. Schaerer). The ski patrol at Vail, Colorado, is studying slab avalanches in its region. A random sample of avalanches in spring gave the following thicknesses (in cm):

59	51	76	38	65	54	49	62
68	55	64	67	63	74	65	79

    i. Use a calculator with mean and standard deviation keys to verify that $\bar{x} \approx 61.8$ and $s \approx 10.6$ cm.

   ii. Assume the slab thickness has an approximately normal distribution. Use a 1% level of significance to test the claim that the mean slab thickness in the Vail region is different from that in Canada.

20. *Longevity: Honolulu* *USA Today* reported that the state with the longest mean life span is Hawaii, where the population mean life span is 77 years. A random sample of 20 obituary notices in the *Honolulu Advertizer* gave the following information about life span (in years) of Honolulu residents:

72	68	81	93	56	19	78	94	83	84
77	69	85	97	75	71	86	47	66	27

    i. Use a calculator with mean and standard deviation keys to verify that $\bar{x} = 71.4$ years and $s \approx 20.65$ years.

   ii. Assuming that life span in Honolulu is approximately normally distributed, does this information indicate that the population mean life span for Honolulu residents is less than 77 years? Use a 5% level of significance.

21. *Fishing: Atlantic Salmon* Homser Lake, Oregon, has an Atlantic salmon catch and release program that has been very successful. The average fisherman's catch has been $\mu = 8.8$ Atlantic salmon per day (Source: *National Symposium on Catch and Release Fishing*, Humboldt State University). Suppose that a new quota system restricting the number of fishermen has been put into effect this season. A random sample of fishermen gave the following catches per day:

12	6	11	12	5	0	2
7	8	7	6	3	12	12

    i. Use a calculator with mean and sample standard deviation keys to verify that $\bar{x} = 7.36$ and $s \approx 4.03$.

   ii. Assuming the catch per day has an approximately normal distribution, use a 5% level of significance to test the claim that the population average catch per day is now different from 8.8.

22. *Archaeology: Tree Rings* Tree-ring dating from archaeological excavation sites is used in conjunction with other chronologic evidence to estimate occupation dates of prehistoric Indian ruins in the southwestern United States. It is thought that Burnt Mesa Pueblo was occupied around 1300 A.D. (based on evidence from potsherds and stone tools). The following data give tree-ring dates (A.D.) from adjacent archaeological sites (*Bandelier Archaeological Excavation Project: Summer 1990 Excavations at Burnt Mesa Pueblo,* edited by T. Kohler, Washington State University Department of Anthropology, 1992):

1189	1267	1268	1275	1275
1271	1272	1316	1317	1230

    i. Use a calculator with mean and standard deviation keys to verify that $\bar{x} = 1268$ and $s \approx 37.29$ years.

    ii. Assuming the tree-ring dates in this excavation area follow a distribution that is approximately normal, does this information indicate that the population mean of tree-ring dates in the area is different from (either higher or lower than) that in 1300 A.D.? Use a 1% level of significance.

23. *Critical Thinking: One-Tailed versus Two-Tailed Tests*

  (a) For the same data and null hypothesis, is the *P*-value of a one-tailed test (right or left) larger or smaller than that of a two-tailed test? Explain.

  (b) For the same data, null hypothesis, and level of significance, is it possible that a one-tailed test results in the conclusion to reject H_0 while a two-tailed test results in the conclusion to fail to reject H_0? Explain.

  (c) For the same data, null hypothesis, and level of significance, if the conclusion is to reject H_0 based on a two-tailed test, do you also reject H_0 based on a one-tailed test? Explain.

  (d) If a report states that certain data were used to reject a given hypothesis, would it be a good idea to know what type of test (one-tailed or two-tailed) was used? Explain.

24. *Critical Thinking: Comparing Hypothesis Tests with U.S. Courtroom System*
Compare statistical testing with legal methods used in a U.S. court setting. Then discuss the following topics in class or consider the topics on your own. Please write a brief but complete essay in which you answer the following questions.

  (a) In a court setting, the person charged with a crime is initially considered to be innocent. The claim of innocence is maintained until the jury returns with a decision. Explain how the claim of innocence could be taken to be the null hypothesis. Do we assume that the null hypothesis is true throughout the testing procedure? What would the alternate hypothesis be in a court setting?

  (b) The court claims that a person is innocent if the evidence against the person is not adequate to find him or her guilty. This does not mean, however, that the court has necessarily *proved* the person to be innocent. It simply means that the evidence against the person was not adequate for the jury to find him or her guilty. How does this situation compare with a statistical test for which the conclusion is "do not reject" (i.e., accept) the null hypothesis? What would be a type II error in this context?

  (c) If the evidence against a person is adequate for the jury to find him or her guilty, then the court claims that the person is guilty. Remember, this does not mean that the court has necessarily *proved* the person to be guilty. It simply means that the evidence against the person was strong enough to find him or her guilty. How does this situation compare with a statistical test for which the conclusion is to "reject" the null hypothesis? What would be a type I error in this context?

  (d) In a court setting, the final decision as to whether the person charged is innocent or guilty is made at the end of the trial, usually by a jury of impartial people. In hypothesis testing, the final decision to reject or not reject the null hypothesis is made at the end of the test by using information or data from an (impartial) random sample. Discuss these similarities between statistical hypothesis testing and a court setting.

  (e) We hope that you are able to use this discussion to increase your understanding of statistical testing by comparing it with something that is a well-known part of our American way of life. However, all analogies have weak points, and it is important not to take the analogy between statistical hypothesis testing and legal court methods too far. For instance, the judge does not set a level of significance and tell the jury to determine a verdict that is wrong only 5% or 1% of the time. Discuss some of these weak points in the analogy between the court setting and hypothesis testing.

25. ***Expand Your Knowledge: Confidence Intervals and Two-Tailed Hypothesis Tests*** Is there a relationship between confidence intervals and two-tailed hypothesis tests? Let c be the level of confidence used to construct a confidence interval from sample data. Let α be the level of significance for a two-tailed hypothesis test. The following statement applies to hypothesis tests of the mean.

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

(A corresponding relationship between confidence intervals and two-tailed hypothesis tests also is valid for other parameters, such as p, $\mu_1 - \mu_2$, and $p_1 - p_2$, which we will study in Sections 8.3 and 8.5.) Whenever the value of k given in the null hypothesis falls *outside* the $c = 1 - \alpha$ confidence interval for the parameter, we *reject* H_0. For example, consider a two-tailed hypothesis test with $\alpha = 0.01$ and

$$H_0: \mu = 20 \qquad H_1: \mu \neq 20$$

A random sample of size 36 has a sample mean $\overline{x} = 22$ from a population with standard deviation $\sigma = 4$.

(a) What is the value of $c = 1 - \alpha$? Using the methods of Chapter 7, construct a $1 - \alpha$ confidence interval for μ from the sample data. What is the value of μ given in the null hypothesis (i.e., what is k)? Is this value in the confidence interval? Do we reject or fail to reject H_0 based on this information?

(b) Using methods of this chapter, find the *P*-value for the hypothesis test. Do we reject or fail to reject H_0? Compare your result to that of part (a).

26. ***Confidence Intervals and Two-Tailed Hypothesis Tests*** Change the null hypothesis of Problem 25 to H_0: $\mu = 21$. Repeat parts (a) and (b).

27. ***Critical Region Method: Standard Normal*** Solve Problem 11 using the critical region method of testing (i.e., traditional method). Compare your conclusion with the conclusion obtained by using the *P*-value method. Are they the same?

28. ***Critical Region Method: Student's t*** Table 6 of Appendix II gives critical values for the Student's *t* distribution. Use an appropriate *d.f.* as the row header. For a *right-tailed* test, the column header is the value of α found in the *one-tail area* row. For a *left-tailed* test, the column header is the value of α found in the *one-tail area* row, but you must change the sign of the critical value *t* to $-t$. For a *two-tailed* test, the column header is the value of α from the *two-tail area* row. The critical values are the $\pm t$ values shown. Solve Problem 12 using the critical region method of testing. Compare your conclusion with the conclusion obtained by using the *P*-value method. Are they the same?

29. ***Critical Region Method: Student's t*** Solve Problem 13 using the critical region method of testing. *Hint:* See Problem 28. Compare your conclusion with the conclusion obtained by using the *P*-value method. Are they the same?

30. ***Critical Region Method: Student's t*** Solve Problem 14 using the critical region method of testing. *Hint:* See Problem 28. Compare your conclusion with the conclusion obtained by using the *P*-value method. Are they the same?

Testing a Proportion *p*

FOCUS POINTS

- Identify the components needed for testing a proportion.
- Compute the sample test statistic.
- Find the *P*-value and conclude the test.

Many situations arise that call for tests of proportions or percentages rather than means. For instance, a college registrar may want to determine if the proportion of students wanting 3-week intensive courses has increased.

How can we make such a test? In this section, we will study tests involving proportions (i.e., percentages or proportions). Such tests are similar to those in Sections 8.1 and 8.2. The main difference is that we are working with a distribution of proportions.

Tests for a single proportion

Throughout this section, we will assume that the situations we are dealing with satisfy the conditions underlying the binomial distribution. In particular, we will let *r* be a binomial random variable. This means that *r* is the number of successes out of *n* independent binomial trials (for the definition of a binomial trial, see Section 5.2). We will use $\hat{p} = r/n$ as our estimate for *p*, the population probability of success on each trial. The letter *q* again represents the population probability of failure on each trial, and so $q = 1 - p$. We also assume that the samples are large (i.e., $np > 5$ and $nq > 5$).

Criteria for using normal approximation to binomial
$np > 5$ and $nq > 5$

For large samples, $np > 5$ and $nq > 5$, the distribution of $\hat{p} = r/n$ values is well approximated by a *normal curve* with mean μ and standard deviation σ as follows:

$$\mu = p$$

$$\sigma = \sqrt{\frac{pq}{n}}$$

Hypothesis for testing *p*

The null and alternate hypotheses for tests of proportions are

Left-Tailed Test	**Right-Tailed Test**	**Two-Tailed Test**
$H_0: p = k$	$H_0: p = k$	$H_0: p = k$
$H_1: p < k$	$H_1: p > k$	$H_1: p \neq k$

depending on what is asked for in the problem. Notice that since *p* is a probability, the value *k* must be between 0 and 1.

Sample test statistic $\hat{p}$

For tests of proportions, we need to convert the sample test statistic $\hat{p}$ to a *z* value. Then we can find a *P*-value appropriate for the test. The $\hat{p}$ distribution is approximately normal, with mean *p* and standard deviation $\sqrt{pq/n}$. Therefore, the conversion of $\hat{p}$ to *z* follows the formula

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

where $\hat{p} = r/n$ is the sample test statistic
n = number of trials
p = proportion specified in H_0
$q = 1 - p$

Using this mathematical information about the sampling distribution for $\hat{p}$, the basic procedure is similar to tests you have conducted before.

PROCEDURE

How to Test a Proportion p

Requirements

Consider a binomial experiment with n trials, where p represents the population probability of success and $q = 1 - p$ represents the population probability of failure. Let r be a random variable that represents the number of successes out of the n binomial trials. The number of trials n should be sufficiently large so that both $np > 5$ and $nq > 5$ (use p from the null hypothesis). In this case, $\hat{p} = r/n$ can be approximated by the normal distribution.

Procedure

1. In the context of the application, state the *null and alternate hypotheses* and set the *level of significance α*.

2. Compute the standardized *sample test statistic*

$$z = \frac{\hat{p} - p}{\sqrt{\dfrac{pq}{n}}}$$

   where p is the value specified in H_0 and $q = 1 - p$.

3. Use the standard normal distribution and the type of test, one-tailed or two-tailed, to find the *P-value* corresponding to the test statistic.

4. *Conclude* the test. If P-value $\leq \alpha$, then reject H_0. If P-value $> \alpha$, then do not reject H_0.

5. *Interpret your conclusion* in the context of the application.

EXAMPLE 6

Testing p

A team of eye surgeons has developed a new technique for a risky eye operation to restore the sight of people blinded from a certain disease. Under the old method, it is known that only 30% of the patients who undergo this operation recover their eyesight.

Suppose that surgeons in various hospitals have performed a total of 225 operations using the new method and that 88 have been successful (i.e., the patients fully recovered their sight). Can we justify the claim that the new method is better than the old one? (Use a 1% level of significance.)

SOLUTION:

(a) Establish H_0 and H_1 and note the level of significance.

The level of significance is $\alpha = 0.01$. Let p be the probability that a patient fully recovers his or her eyesight. The null hypothesis is that p is still 0.30, even for the new method. The alternate hypothesis is that the new method has improved the chances of a patient recovering his or her eyesight. Therefore,

$$H_0\text{: } p = 0.30 \qquad \text{and} \qquad H_1\text{: } p > 0.30$$

(b) *Check Requirements* Is the sample sufficiently large to justify use of the normal distribution for $\hat{p}$? Find the sample test statistic $\hat{p}$ and convert it to a z value, if appropriate.

Using p from H_0 we note that $np = 225(0.3) = 67.5$ is greater than 5 and that $nq = 225(0.7) = 157.5$ is also greater than 5, so we can use the normal distribution for the sample statistic $\hat{p}$.

$$\hat{p} = \frac{r}{n} = \frac{88}{225} \approx 0.39$$

Ned Frisk/Blend Images/Jupiter images

The z value corresponding to $\hat{p}$ is

$$z = \frac{\hat{p} - p}{\sqrt{\dfrac{pq}{n}}} \approx \frac{0.39 - 0.30}{\sqrt{\dfrac{0.30(0.70)}{225}}} \approx 2.95$$
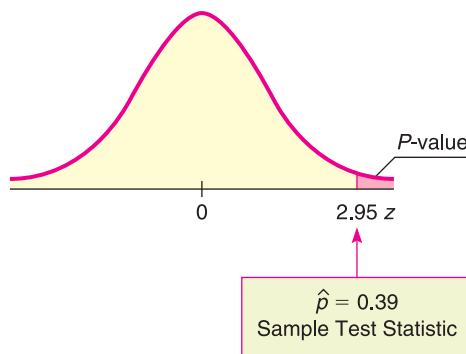
In the formula, the value for p is from the null hypothesis. H_0 specifies that $p = 0.30$, so $q = 1 - 0.30 = 0.70$.

(c) Find the P-value of the test statistic.

Figure 8-10 shows the P-value. Since we have a right-tailed test, the P-value is the area to the right of $z = 2.95$. Using the normal distribution (Table 5 of Appendix II), we find that P-value $= P(z > 2.95) \approx 0.0016$.

FIGURE 8-10

P-value Area

(d) Conclude the test.

Since the P-value of $0.0016 \leq 0.01$ for α, we reject H_0.

(e) *Interpretation* Interpret the results in the context of the problem.

At the 1% level of significance, the evidence shows that the population probability of success for the new surgery technique is higher than that of the old technique.

GUIDED EXERCISE 5 | *Testing p*

A botanist has produced a new variety of hybrid wheat that is better able to withstand drought than other varieties. The botanist knows that for the parent plants, the proportion of seeds germinating is 80%. The proportion of seeds germinating for the hybrid variety is unknown, but the botanist claims that it is 80%. To test this claim, 400 seeds from the hybrid plant are tested, and it is found that 312 germinate. Use a 5% level of significance to test the claim that the proportion germinating for the hybrid is 80%.

(a) Let p be the proportion of hybrid seeds that will germinate. Notice that we have no prior knowledge about the germination proportion for the hybrid plant. State H_0 and H_1. What is the required level of significance?

$H_0: p = 0.80$; $H_1: p \neq 0.80$; $\alpha = 0.05$

(b) *Check Requirements* Using the value of p in H_0, are both $np > 5$ and $nq > 5$? Can we use the normal distribution for $\hat{p}$?

From H_0, $p = 0.80$ and $q = 1 - p = 0.20$

$np = 400(0.8) = 320 > 5$

$nq = 400(0.2) = 80 > 5$

So, we can use the normal distribution for $\hat{p}$.

Continued

Calculate the sample test statistic $\hat{p}$.

The number of trials is $n = 400$, and the number of successes is $r = 312$. Thus,

$$\hat{p} = \frac{r}{n} = \frac{312}{400} = 0.78$$

(c) Next, we convert the sample test statistic $\hat{p} = 0.78$ to a z value. Based on our choice for H_0, what value should we use for p in our formula? Since $q = 1 - p$, what value should we use for q? Using these values for p and q, convert $\hat{p}$ to a z value.

According to H_0, $p = 0.80$. Then $q = 1 - p = 0.20$. Using these values in the following formula gives

$$z = \frac{\hat{p} - p}{\sqrt{\dfrac{pq}{n}}} = \frac{0.78 - 0.80}{\sqrt{\dfrac{0.80(0.20)}{400}}} = -1.00$$

CALCULATOR NOTE   If you evaluate the denominator separately, be sure to carry at least four digits after the decimal.

(d) Is the test right-tailed, left-tailed, or two-tailed? Find the *P*-value of the sample test statistic and sketch a standard normal curve showing the *P*-value.

For a two-tailed test, using the normal distribution (Table 5 of Appendix II), we find that

$$P\text{-value} = 2P(z < -1.00) = 2(0.1587) = 0.3174$$

FIGURE 8-11 *P*-value
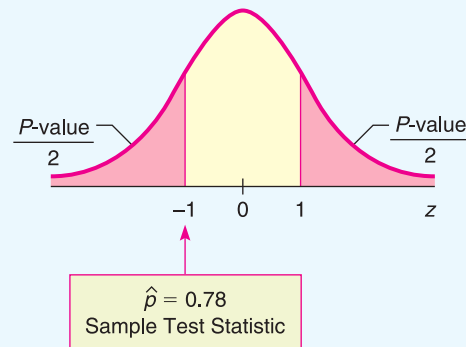
(e) Do we reject or fail to reject H_0?

We fail to reject H_0 because *P*-value of $0.3171 > 0.05$ for α.

(f) *Interpretation* Interpret your conclusion in the context of the application.

At the 5% level of significance, there is insufficient evidence to conclude that the germination rate is not 80%.

Critical region method

Since the $\hat{p}$ sampling distribution is approximately normal, we use Table 5, "Areas of a Standard Normal Distribution," in Appendix II to find critical values.

EXAMPLE 7   CRITICAL REGION METHOD FOR TESTING p

Let's solve Guided Exercise 5 using the critical region approach. In that problem, 312 of 400 seeds from a hybrid wheat variety germinated. For the parent plants, the proportion of germinating seeds is 80%. Use a 5% level of significance to test the claim that the population proportion of germinating seeds from the hybrid wheat is different from that of the parent plants.

SOLUTION:

(a) As in Guided Exercise 5, we have $\alpha = 0.05$, H_0: $p = 0.80$, and H_1: $p \neq 0.80$. The next step is to find $\hat{p}$ and the corresponding sample test statistic z. This

was done in Guided Exercise 5, where we found that $\hat{p} = 0.78$, with corresponding $z = -1.00$.

(b) Now we find the critical value z_0 for a two-tailed test using $\alpha = 0.05$. This means that we want the total area 0.05 divided between two tails, one to the right of z_0 and one to the left of $-z_0$. As shown in Figure 8-8 of Section 8.2, the critical value(s) are ± 1.96 (See also Table 5, part (c), of Appendix II for critical values of the z distribution.)

(c) Figure 8-12 shows the critical regions and the location of the sample test statistic.

FIGURE 8-12

Critical Regions, $\alpha = 0.05$
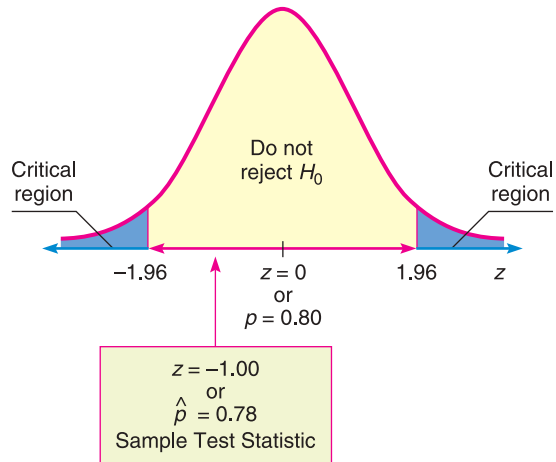
(d) Finally, we conclude the test and compare the results to Guided Exercise 5. Since the sample test statistic does not fall in the critical region, we fail to reject H_0 and conclude that, at the 5% level of significance, the evidence is not strong enough to reject the botanist's claim. This result, as expected, is consistent with the conclusion obtained by using the P-value method.

TECH NOTES   The TI-84Plus/TI-83Plus/TI-$n$spire calculators and Minitab support tests of proportions. The output for both technologies includes the sample proportion $\hat{p}$ and the P-value of $\hat{p}$. Minitab also includes the z value corresponding to $\hat{p}$.

TI-84Plus/TI-83Plus/TI-$n$spire (with TI-84Plus keypad) Press **STAT,** select **TESTS,** and use option **5:1-PropZTest.** The value of p_0 is from the null hypothesis $H_0: p = p_0$. The number of successes is the value for x.

Minitab Menu selections: **Stat ➤ Basic Statistics ➤ 1 Proportion.** Under options, set the test proportion as the value in H_0. Choose to use the normal distribution.

CRITICAL THINKING

Issues Related to Hypothesis Testing

Through our work with hypothesis tests of μ and p, we've gained experience in setting up, performing, and interpreting results of such tests.

We know that different random samples from the same population are very likely to have sample statistics $\bar{x}$ or $\hat{p}$ that differ from their corresponding parameters μ or p. Some values of a statistic from a random sample will be close to the corresponding parameter. Others may be farther away simply because we happened to draw a random sample of more extreme data values.

> The central question in hypothesis testing is whether or not you think the value of the sample test statistic is too far away from the value of the population parameter proposed in H_0 to occur by chance alone.

This is where the *P*-value of the sample test statistic comes into play. The *P*-value of the sample test statistic tells you the probability that you would get a sample statistic as far away as, or farther from, the value of the parameter as stated in the null hypothesis H_0.

If the *P*-value is very small, you reject H_0. But what does "very small" mean? It is customary to define "very small" as smaller than the preset level of significance α.

When you reject H_0, are you absolutely certain that you are making a correct decision? The answer is no! You are simply willing to take a chance that you are making a mistake (a type I error). The level of significance α describes the chance of making a mistake if you reject H_0 when it is, in fact, true.

Several issues come to mind:

1. What if the *P*-value is so close to α that we "barely" reject or fail to reject H_0? In such cases, researchers might attempt to clarify the results by
   - increasing the sample size.
   - controlling the experiment to reduce the standard deviation.
   Both actions tend to increase the magnitude of the z or t value of the sample test statistic, resulting in a smaller corresponding *P*-value.

2. How reliable is the study and the measurements in the sample?
   - When reading results of a statistical study, be aware of the source of the data and the reliability of the organization doing the study.
   - Is the study sponsored by an organization that might profit or benefit from the stated conclusions? If so, look at the study carefully to ensure that the measurements, sampling technique, and handling of data are proper and meet professional standards.

VIEWPOINT | Who Did What?

Art, music, literature, and science share a common need to classify things: Who painted that picture? Who composed that music? Who wrote that document? Who should get that patent? In statistics, such questions are called classification problems. *For example, the Federalist Papers were published anonymously in 1787–1788 by Alexander Hamilton, John Jay, and James Madison. But who wrote what? That question is addressed by F. Mosteller (Harvard University) and D. Wallace (University of Chicago) in the book* Statistics: A Guide to the Unknown, *edited by J. M. Tanur. Other scholars have studied authorship regarding Plato's* Republic *and Plato's* Dialogues, *including the* Symposium. *For more information on this topic, see the source in Problems 15 and 16 of this exercise set.*

SECTION 8.3 PROBLEMS

1. *Statistical Literacy* To use the normal distribution to test a proportion p, the conditions $np > 5$ and $nq > 5$ must be satisfied. Does the value of p come from H_0, or is it estimated by using $\hat{p}$ from the sample?

2. *Statistical Literacy* Consider a binomial experiment with n trials and r successes. For a test for a proportion p, what is the formula for the sample test statistic? Describe each symbol used in the formula.

3. *Critical Thinking* In general, if sample data are such that the null hypothesis is rejected at the $\alpha = 1\%$ level of significance based on a two-tailed test, is H_0 also rejected at the $\alpha = 1\%$ level of significance for a corresponding one-tailed test? Explain.

4. *Critical Thinking* An article in a newspaper states that the proportion of traffic accidents involving road rage is higher this year than it was last year, when it was 15%. Reconstruct the information of the study in terms of a hypothesis test. Discuss possible hypotheses, possible issues about the sample, possible levels of significance, and the "absolute truth" of the conclusion.

5. *Basic Computation: Testing p* A random sample of 30 binomials trials resulted in 12 successes. Test the claim that the population proportion of successes does not equal 0.50. Use a level of significance of 0.05.
   (a) *Check Requirements* Can a normal distribution be used for the $\hat{p}$ distribution? Explain.
   (b) State the hypotheses.
   (c) Compute $\hat{p}$ and the corresponding standardized sample test statistic.
   (d) Find the P-value of the test statistic.
   (e) Do you reject or fail to reject H_0? Explain.
   (f) *Interpretation* What do the results tell you?

6. *Basic Computation: Testing p* A random sample of 60 binomials trials resulted in 18 successes. Test the claim that the population proportion of successes exceeds 18%. Use a level of significance of 0.01.
   (a) *Check Requirements* Can a normal distribution be used for the $\hat{p}$ distribution? Explain.
   (b) State the hypotheses.
   (c) Compute $\hat{p}$ and the corresponding standardized sample test statistic.
   (d) Find the P-value of the test statistic.
   (e) Do you reject or fail to reject H_0? Explain.
   (f) *Interpretation* What do the results tell you?

For Problems 7–21, please provide the following information.
   (a) What is the level of significance? State the null and alternate hypotheses.
   (b) *Check Requirements* What sampling distribution will you use? Do you think the sample size is sufficiently large? Explain. Compute the value of the sample test statistic.
   (c) Find the P-value of the test statistic. Sketch the sampling distribution and show the area corresponding to the P-value.
   (d) Based on your answers in parts (a) to (c), will you reject or fail to reject the null hypothesis? Are the data statistically significant at level α?
   (e) *Interpret* your conclusion in the context of the application.

7. *Focus Problem: Benford's Law* Please read the Focus Problem at the beginning of this chapter. Recall that Benford's Law claims that numbers chosen from very large data files tend to have "1" as the first nonzero digit disproportionately often. In fact, research has shown that if you randomly draw a number from a very large data file, the probability of getting a number with "1" as the leading digit is about 0.301 (see the reference in this chapter's Focus Problem).

   Now suppose you are an auditor for a very large corporation. The revenue report involves millions of numbers in a large computer file. Let us say you took a random sample of $n = 215$ numerical entries from the file and $r = 46$ of the entries had a first nonzero digit of 1. Let p represent the population proportion of all numbers in the corporate file that have a first nonzero digit of 1.
   i.   Test the claim that p is less than 0.301. Use $\alpha = 0.01$.
   ii. If p is in fact less than 0.301, would it make you suspect that there are not enough numbers in the data file with leading 1's? Could this indicate that the books have been "cooked" by "pumping up" or inflating the numbers? Comment from the viewpoint of a stockholder. Comment from the perspective

of the Federal Bureau of Investigation as it looks for money laundering in the form of false profits.

iii. Comment on the following statement: "If we reject the null hypothesis at level of significance α, we have not *proved H_0* to be false. We can say that the probability is α that we made a mistake in rejecting H_0." Based on the outcome of the test, would you recommend further investigation before accusing the company of fraud?

8. *Focus Problem: Benford's Law* Again, suppose you are the auditor for a very large corporation. The revenue file contains millions of numbers in a large computer data bank (see Problem 7). You draw a random sample of $n = 228$ numbers from this file and $r = 92$ have a first nonzero digit of 1. Let p represent the population proportion of all numbers in the computer file that have a leading digit of 1.

   i. Test the claim that p is more than 0.301. Use $\alpha = 0.01$.

   ii. If p is in fact larger than 0.301, it would seem there are too many numbers in the file with leading 1's. Could this indicate that the books have been "cooked" by artificially lowering numbers in the file? Comment from the point of view of the Internal Revenue Service. Comment from the perspective of the Federal Bureau of Investigation as it looks for "profit skimming" by unscrupulous employees.

   iii. Comment on the following statement: "If we reject the null hypothesis at level of significance α, we have not *proved H_0* to be false. We can say that the probability is α that we made a mistake in rejecting H_0." Based on the outcome of the test, would you recommend further investigation before accusing the company of fraud?

9. *Sociology: Crime Rate* Is the national crime rate really going down? Some sociologists say yes! They say that the reason for the decline in crime rates in the 1980s and 1990s is demographics. It seems that the population is aging, and older people commit fewer crimes. According to the FBI and the Justice Department, 70% of all arrests are of males aged 15 to 34 years (Source: *True Odds* by J. Walsh, Merritt Publishing). Suppose you are a sociologist in Rock Springs, Wyoming, and a random sample of police files showed that of 32 arrests last month, 24 were of males aged 15 to 34 years. Use a 1% level of significance to test the claim that the population proportion of such arrests in Rock Springs is different from 70%.

10. *College Athletics: Graduation Rate* Women athletes at the University of Colorado, Boulder, have a long-term graduation rate of 67% (Source: *Chronicle of Higher Education*). Over the past several years, a random sample of 38 women athletes at the school showed that 21 eventually graduated. Does this indicate that the population proportion of women athletes who graduate from the University of Colorado, Boulder, is now less than 67%? Use a 5% level of significance.

11. *Highway Accidents: DUI* The U.S. Department of Transportation, National Highway Traffic Safety Administration, reported that 77% of all fatally injured automobile drivers were intoxicated. A random sample of 27 records of automobile driver fatalities in Kit Carson County, Colorado, showed that 15 involved an intoxicated driver. Do these data indicate that the population proportion of driver fatalities related to alcohol is less than 77% in Kit Carson County? Use $\alpha = 0.01$.

12. *Preference: Color* What is your favorite color? A large survey of countries, including the United States, China, Russia, France, Turkey, Kenya, and others, indicated that most people prefer the color blue. In fact, about 24% of the population claim blue as their favorite color (Reference: Study by J. Bunge and A. Freeman-Gallant, Statistics Center, Cornell University). Suppose a random sample of $n = 56$ college students were surveyed and $r = 12$ of them said that blue is their favorite color. Does this information imply that the color preference of all college students is different (either way) from that of the general population? Use $\alpha = 0.05$.

13. *Wildlife: Wolves* The following is based on information from *The Wolf in the Southwest: The Making of an Endangered Species* by David E. Brown (University of Arizona Press). Before 1918, the proportion of female wolves in the general population of all southwestern wolves was about 50%. However, after 1918, southwestern cattle ranchers began a widespread effort to destroy wolves. In a recent sample of 34 wolves, there were only 10 females. One theory is that male wolves tend to return sooner than females to their old territories where their predecessors were exterminated. Do these data indicate that the population proportion of female wolves is now less than 50% in the region? Use $\alpha = 0.01$.

14. *Fishing: Northern Pike* Athabasca Fishing Lodge is located on Lake Athabasca in northern Canada. In one of its recent brochures, the lodge advertises that 75% of its guests catch northern pike over 20 pounds. Suppose that last summer 64 out of a random sample of 83 guests did, in fact, catch northern pike weighing over 20 pounds. Does this indicate that the population proportion of guests who catch pike over 20 pounds is different from 75% (either higher or lower)? Use $\alpha = 0.05$.

15. *Plato's* Republic: *Syllable Patterns* Prose rhythm is characterized by the occurrence of five-syllable sequences in long passages of text. This characterization may be used to assess the similarity among passages of text and sometimes the identity of authors. The following information is based on an article by D. Wishart and S. V. Leach appearing in *Computer Studies of the Humanities and Verbal Behavior* (Vol. 3, pp. 90–99). Syllables were categorized as long or short. On analyzing Plato's *Republic,* Wishart and Leach found that about 26.1% of the five-syllable sequences are of the type in which two are short and three are long. Suppose that Greek archaeologists have found an ancient manuscript dating back to Plato's time (about 427–347 B.C.). A random sample of 317 five-syllable sequences from the newly discovered manuscript showed that 61 are of the type two short and three long. Do the data indicate that the population proportion of this type of five-syllable sequence is different (either way) from the text of Plato's *Republic*? Use $\alpha = 0.01$.

16. *Plato's* Dialogues: *Prose Rhythm* *Symposium* is part of a larger work referred to as Plato's *Dialogues.* Wishart and Leach (see source in Problem 15) found that about 21.4% of five-syllable sequences in *Symposium* are of the type in which four are short and one is long. Suppose an antiquities store in Athens has a very old manuscript that the owner claims is part of Plato's *Dialogues.* A random sample of 493 five-syllable sequences from this manuscript showed that 136 were of the type four short and one long. Do the data indicate that the population proportion of this type of five-syllable sequence is higher than that found in Plato's *Symposium*? Use $\alpha = 0.01$.

17. *Consumers: Product Loyalty* *USA Today* reported that about 47% of the general consumer population in the United States is loyal to the automobile manufacturer of their choice. Suppose Chevrolet did a study of a random sample of 1006 Chevrolet owners and found that 490 said they would buy another Chevrolet. Does this indicate that the population proportion of consumers loyal to Chevrolet is more than 47%? Use $\alpha = 0.01$.

18. *Supermarket: Prices* *Harper's Index* reported that 80% of all supermarket prices end in the digit 9 or 5. Suppose you check a random sample of 115 items in a supermarket and find that 88 have prices that end in 9 or 5. Does this indicate that less than 80% of the prices in the store end in the digits 9 or 5? Use $\alpha = 0.05$.

19. *Medical: Hypertension* This problem is based on information taken from *The Merck Manual* (a reference manual used in most medical and nursing schools). Hypertension is defined as a blood pressure reading over 140 mm Hg systolic and/or over 90 mm Hg diastolic. Hypertension, if not corrected, can cause long-term health problems. In the college-age population (18–24 years), about 9.2% have hypertension. Suppose that a blood donor program is taking place in a college dormitory this week (final exams week). Before each student gives blood, the nurse takes a blood pressure reading. Of 196 donors, it is found that 29 have hypertension. Do these data indicate that the population proportion of students with hypertension during final exams week is higher than 9.2%? Use a 5% level of significance.

20. *Medical: Hypertension* Diltiazem is a commonly prescribed drug for hypertension (see source in Problem 19). However, diltiazem causes headaches in about 12% of patients using the drug. It is hypothesized that regular exercise might help reduce the headaches. If a random sample of 209 patients using diltiazem exercised regularly and only 16 had headaches, would this indicate a reduction in the population proportion of patients having headaches? Use a 1% level of significance.

21. *Myers–Briggs: Extroverts* Are most student government leaders extroverts? According to Myers–Briggs estimates, about 82% of college student government leaders are extroverts (Source: *Myers–Briggs Type Indicator Atlas of Type Tables*). Suppose that a Myers–Briggs personality preference test was given to a random sample of 73 student government leaders attending a large national leadership conference and that 56 were found to be extroverts. Does this indicate that the population proportion of extroverts among college student government leaders is different (either way) from 82%? Use $\alpha = 0.01$.

22. *Critical Region Method: Testing Proportions* Solve Problem 9 using the critical region method of testing. Since the sampling distribution of $\hat{p}$ is the normal distribution, you can use critical values from the standard normal distribution as shown in Figure 8-8 or part (c) of Table 5, Appendix II. Compare your conclusions with the conclusions obtained by using the *P*-value method. Are they the same?

23. *Critical Region Method: Testing Proportions* Solve Problem 11 using the critical region method of testing. *Hint:* See Problem 22. Compare your conclusions with the conclusions obtained by using the *P*-value method. Are they the same?

24. *Critical Region Method: Testing Proportions* Solve Problem 17 using the critical region method of testing. *Hint:* See Problem 22. Compare your conclusions with the conclusions obtained by using the *P*-value method. Are they the same?

Tests Involving Paired Differences (Dependent Samples)

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

Paired data

Paired data
Creating data pairs

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

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

EXAMPLE 8   PAIRED DATA

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

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

GUIDED EXERCISE 6   Paired data

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

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

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

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

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

Testing the differences d

When testing *paired* data, we take the difference d of the data pairs *first* and look at the mean difference $\bar{d}$. Then we use a test on $\bar{d}$. Theorem 8.1 provides the basis for our work with paired data.

THEOREM 8.1   Consider a random sample of n data pairs. Suppose the differences d between the first and second members of each data pair are (approximately) normally distributed, with population mean μ_d. Then the t values

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

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

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

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

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

Hypotheses for testing the mean of paired differences

$$H_0: \mu_d = 0$$

The alternate hypothesis depends on the problem and can be

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

Sample test statistic

For paired difference tests, we make our decision regarding H_0 according to the evidence of the sample mean $\bar{d}$ of the differences of measurements. By Theorem 8.1, we convert the sample test statistic $\bar{d}$ to a t value using the formula

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

where s_d = sample standard deviation of the differences d

n = number of data pairs

μ_d = 0, as specified in H_0

P-values from Table 6 of Appendix II

To find the *P*-value (or an interval containing the *P*-value) corresponding to the test statistic t computed from $\bar{d}$, we use the Student's t distribution table

(Table 6, Appendix II). Recall from Section 8.2 that we find the test statistic t (or, if t is negative, $|t|$) in the row headed by $d.f. = n - 1$, where n is the number of data pairs. The P-value for the test statistic is the column entry in the *one-tail area* row for one-tailed tests (right or left). For two-tailed tests, the P-value is the column entry in the *two-tail area* row. Usually the exact test statistic t is not in the table, so we obtain an interval that contains the P-value by using adjacent entries in the table. Table 8-7 gives the basic structure for using the Student's t distribution table to find the P-value or an interval containing the P-value.

TABLE 8-7	Using Student's *t* Distribution Table for *P*-values			
For one-tailed tests:	one-tail area	*P*-value	*P*-value	
For two-tailed tests:	two-tail area	*P*-value	*P*-value	
Use row header	$d.f. = n - 1$		Find *t* value ↑	

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

PROCEDURE

HOW TO TEST PAIRED DIFFERENCES USING THE STUDENT'S t DISTRIBUTION

Requirements

Obtain a simple random sample of n matched data pairs A, B. Let d be a random variable representing the difference between the values in a matched data pair. Compute the sample mean $\overline{d}$ and sample standard deviation s_d. If you can assume that d has a normal distribution or simply has a mound-shaped, symmetric distribution, then any sample size n will work. If you cannot assume this, then use a sample size $n \geq 30$.

Procedure

1. Use the *null hypothesis* of *no difference*, $H_0: \mu_d = 0$. In the context of the application, choose the *alternate hypothesis* to be $H_1: \mu_d > 0$, $H_1: \mu_d < 0$, or $H_1: \mu_d \neq 0$. Set the *level of significance* α.

2. Use $\overline{d}$, s_d, the sample size n, and $\mu_d = 0$ from the null hypothesis to compute the *sample test statistic*

$$t = \frac{\overline{d} - 0}{\frac{s_d}{\sqrt{n}}} = \frac{\overline{d}\sqrt{n}}{s_d}$$

with degrees of freedom $d.f. = n - 1$.

3. Use the Student's t distribution and the type of test, one-tailed or two-tailed, to find (or estimate) the *P-value* corresponding to the test statistic.

4. *Conclude* the test. If P-value $\leq \alpha$, then reject H_0. If P-value $> \alpha$, then do not reject H_0.

5. *Interpret your conclusion* in the context of the application.

Problem 22 at the end of this section shows how to find confidence intervals for μ_d.

EXAMPLE 9   PAIRED DIFFERENCE TEST

A team of heart surgeons at Saint Ann's Hospital knows that many patients who undergo corrective heart surgery have a dangerous buildup of anxiety before their scheduled operations. The staff psychiatrist at the hospital has started a new

The Image Works

counseling program intended to reduce this anxiety. A test of anxiety is given to patients who know they must undergo heart surgery. Then each patient participates in a series of counseling sessions with the staff psychiatrist. At the end of the counseling sessions, each patient is retested to determine anxiety level. Table 8-8 indicates the results for a random sample of nine patients. Higher scores mean higher levels of anxiety. Assume the distribution of differences is mound-shaped and symmetric.

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

SOLUTION: Before we answer this question, let us notice two important points: (1) We have a *random sample* of nine patients, and (2) we have a *pair* of measurements taken on the same patient before and after counseling sessions. In our problem, the sample size is $n = 9$ pairs (i.e., patients), and the d values are found in the fourth column of Table 8-8.

(a) Note the level of significance and set the hypotheses.

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

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

(b) *Check Requirements* Is it appropriate to use a Student's t distribution for the sample test statistic? Explain. What degrees of freedom are used?

We have a random sample of paired differences d. Under the assumption that the d distribution is mound-shaped and symmetric, we use a Student's t distribution with $d.f. = n - 1 = 9 - 1 = 8$.

(c) Find the sample test statistic $\bar{d}$ and convert it to a corresponding test statistic t.

First we need to compute $\bar{d}$ and s_d. Using formulas or a calculator and the d values shown in Table 8-8, we find that

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

TABLE 8-8

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

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

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

(d) Find the *P*-value for the test statistic and sketch the *P*-value on the t distribution. Since we have a right-tailed test, the *P*-value is the area to the right

of $t = 4.363$, as shown in Figure 8-13. In Table 6 of Appendix II, we find an interval containing the P-value. Use entries from the row headed by $d.f. = n - 1 = 9 - 1 = 8$. The test statistic $t = 4.363$ falls between 3.355 and 5.041. The P-value for the sample t falls between the corresponding one-tail areas 0.005 and 0.0005. (See Table 8-9, excerpt from Table 6, Appendix II.)
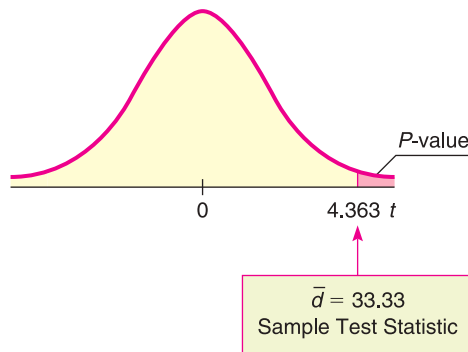
$$0.0005 < P\text{-value} < 0.005$$

TABLE 8-9   Excerpt from Student's t Distribution Table (Table 6, Appendix II)

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

Sample $t = 4.363$

FIGURE 8-13

P-value

P-value

0   4.363 t

$\bar{d} = 33.33$
Sample Test Statistic

(e) Conclude the test.

α

0.0005   0.005   0.01

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

(f) *Interpretation* Interpret the results in the context of the application.
At the 1% level of significance, we conclude that the counseling sessions reduce the average anxiety level of patients about to undergo corrective heart surgery.

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

GUIDED EXERCISE 7   Paired difference test

Do educational toys make a difference in the age at which a child learns to read? To study this question, researchers designed an experiment in which one group of preschool children spent 2 hours each day (for 6 months) in a room well supplied with "educational" toys such as alphabet blocks, puzzles, ABC readers, coloring books featuring letters, and so forth. A control group of children spent 2 hours a day for 6 months in a "noneducational" toy room.

Continued

It was anticipated that IQ differences and home environment might be uncontrollable factors unless identical twins could be used. Therefore, six pairs of identical twins of preschool age were randomly selected. From each pair, one member was randomly selected to participate in the experimental (i.e., educational toy room) group and the other in the control (i.e., noneducational toy room) group. For each twin, the data item recorded is the age in months at which the child began reading at the primary level (Table 8-10). Assume the distribution of differences is mound-shaped and symmetric.

TABLE 8-10   Reading Ages for Identical Twins (in Months)

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

(a) Compute the entries in the $d = B - A$ column of Table 8-10. Using formulas for the mean and sample standard deviation or a calculator with mean and sample standard deviation keys, compute $\bar{d}$ and s_d.

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

$\bar{d} \approx -2.33$
$s_d \approx 2.16$

(b) What is the null hypothesis?

$H_0: \mu_d = 0$

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

$H_1: \mu_d \neq 0$

(d) *Check Requirements* What distribution does the sample test statistic t follow? Find the degrees of freedom.

Because the d distribution is mound-shaped and symmetric and we have a random sample of $n = 6$ paired differences, the sample test statistic follows a Student's t distribution with $d.f. = n - 1 = 6 - 1 = 5$.

(e) Convert the sample test statistic $\bar{d}$ to a t value.

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

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

(f) When we use Table 6 of Appendix II to find an interval containing the P-value, do we use one-tail or two-tail areas? Why? Sketch a figure showing the P-value. Find an interval containing the P-value.

This is a two-tailed test, so we use two-tail areas.
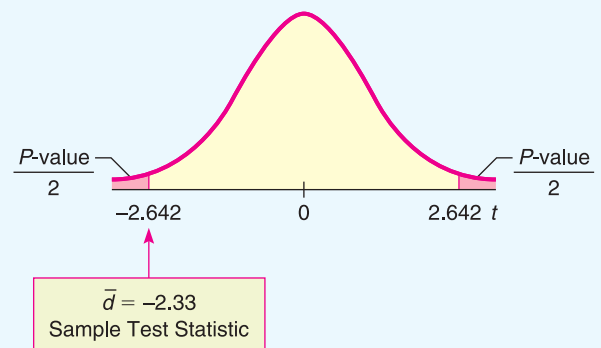
FIGURE 8-14   *P*-value

TABLE 8-11   Excerpt from Student's *t* Table

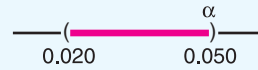|		
one-tail area	0.025	0.010
✓ two-tail area	0.050	0.020
$d.f. = 5$	2.571	3.365

↑
Sample t = 2.642

Continued

GUIDED EXERCISE 7   *continued*

The sample t is between 2.571 and 3.365.
$0.020 < P\text{-value} < 0.050$
Note: Using the raw data and software,
$P\text{-value} \approx 0.0457$.

(g) Using $\alpha = 0.05$, do we reject or fail to reject H_0?

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

(h) *Interpretation* What do the results mean in the context of this application?

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

LOOKING FORWARD

The test for paired differences used in this section is called a *parametric test*. Such tests usually require certain assumptions, such as a normal distribution or a large sample size. In Section 11.1, you will find the sign test for matched pairs. This is a *nonparametric test*. Such tests are useful when you cannot make assumptions about the population distribution. The disadvantage of nonparametric tests is that they tend to accept the null hypothesis more often than they should. That is, they are less sensitive tests. Which type of test should you use? If it is reasonable to assume that the underlying population is normal (or at least mound-shaped and symmetric), or if you have a large sample size, then use the more powerful parametric test described in this section.

TECH NOTES   Both Excel 2007 and Minitab support paired difference tests directly. On the TI-84Plus/TI-83Plus/TI-*n*spire calculators, construct a column of differences and then do a t test on the data in that column. For each technology, be sure to relate the alternate hypothesis to the "before and after" assignments. All the displays show the results for the data of Guided Exercise 7.

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

Excel 2007 Enter the data in two columns. On the home ribbon, click the **Data** tab. In the Analysis Group, select **Data Analysis**. In the dialogue box, select **t-Test: Paired Two-Sample for Means.** Fill in the dialogue box with the hypothesized mean difference of 0. Set alpha.

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

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

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

EXAMPLE 10   CRITICAL REGION METHOD

Let's revisit Guided Exercise 7 regarding educational toys and reading age and conclude the test using the critical region method. Recall that there were six pairs of twins. One twin of each set was given educational toys and the other was not. The difference *d* in reading ages for each pair of twins was measured, and $\alpha = 0.05$.

SOLUTION: From Guided Exercise 7, we have

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

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

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

(b) Sketch the critical regions and place the *t* value of the sample test statistic $\bar{d}$ on the sketch. Conclude the test. Compare the result to the result given by the *P*-value method of Guided Exercise 7.
Since the sample test statistic falls in the critical region (see Figure 8-15), we reject H_0 at the 5% level of significance. At this level, educational toys seem to make a difference in reading age. Notice that this conclusion is consistent with the conclusion obtained using the *P*-value.

FIGURE 8-15

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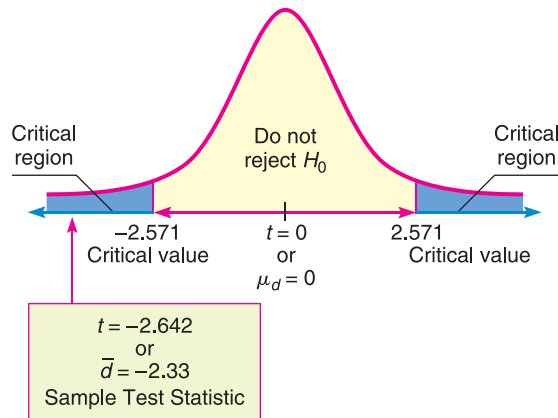
VIEWPOINT | DUI

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

SECTION 8.4 PROBLEMS

1. *Statistical Literacy* Are data that can be paired independent or dependent?

2. *Statistical Literacy* Consider a set of data pairs. What is the first step in processing the data for a paired differences test? What is the formula for the sample test statistic t? Describe each symbol used in the formula.

3. *Statistical Literacy* When testing the difference of means for paired data, what is the null hypothesis?

4. *Statistical Literacy* When conducting a paired differences test, what is the value of n?

5. *Statistical Literacy* When using a Student's t distribution for a paired differences test with n data pairs, what value do you use for the degrees of freedom?

6. *Critical Thinking* Alisha is conducting a paired differences test for a "before (B score) and after (A score)" situation. She is interested in testing whether the average of the "before" scores is higher than that of the "after" scores.
   (a) To use a right-tailed test, how should Alisha construct the differences between the "before" and "after" scores?
   (b) To use a left-tailed test, how should she construct the differences between the "before" and "after" scores?

7. *Basic Computation: Paired Differences Test* For a random sample of 36 data pairs, the sample mean of the differences was 0.8. The sample standard deviation of the differences was 2. At the 5% level of significance, test the claim that the population mean of the differences is different from 0.
   (a) *Check Requirements* Is it appropriate to use a Student's t distribution for the sample test statistic? Explain. What degrees of freedom are used?
   (b) State the hypotheses.

(c) Compute the sample test statistic.
(d) Estimate the *P*-value of the sample test statistic.
(e) Do we reject or fail to reject the null hypothesis? Explain.
(f) *Interpretation* What do your results tell you?

8. | *Basic Computation: Paired Differences Test* For a random sample of 20 data pairs, the sample mean of the differences was 2. The sample standard deviation of the differences was 5. Assume that the distribution of the differences is mound-shaped and symmetric. At the 1% level of significance, test the claim that the population mean of the differences is positive.
(a) *Check Requirements* Is it appropriate to use a Student's *t* distribution for the sample test statistic? Explain. What degrees of freedom are used?
(b) State the hypotheses.
(c) Compute the sample test statistic.
(d) Estimate the *P*-value of the sample test statistic.
(e) Do we reject or fail to reject the null hypothesis? Explain.
(f) *Interpretation* What do your results tell you?

For Problems 9–21 assume that the distribution of differences *d* is mound-shaped and symmetric.

Please provide the following information for Problems 9–21.
(a) What is the level of significance? State the null and alternate hypotheses. Will you use a left-tailed, right-tailed, or two-tailed test?
(b) *Check Requirements* What sampling distribution will you use? What assumptions are you making? Compute the value of the sample test statistic.
(c) Find (or estimate) the *P*-value. Sketch the sampling distribution and show the area corresponding to the *P*-value.
(d) Based on your answers in parts (a) to (c), will you reject or fail to reject the null hypothesis? Are the data statistically significant at level *α*?
(e) *Interpret* your conclusion in the context of the application.

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

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

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

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

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

10. | *Fishing: Shore or Boat?* Is fishing better from a boat or from the shore? Pyramid Lake is located on the Paiute Indian Reservation in Nevada. Presidents, movie stars, and people who just want to catch fish go to Pyramid Lake for really large cutthroat trout. Let row *B* represent hours per fish caught fishing from the shore, and let row *A* represent hours per fish caught using a boat. The following data

are paired by month from October through April (Source: *Pyramid Lake Fisheries,* Paiute Reservation, Nevada).

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

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

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

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

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

12. *Wildlife: Highways* The western United States has a number of four-lane interstate highways that cut through long tracts of wilderness. To prevent car accidents with wild animals, the highways are bordered on both sides with 12-foot-high woven wire fences. Although the fences prevent accidents, they also disturb the winter migration pattern of many animals. To compensate for this disturbance, the highways have frequent wilderness underpasses designed for exclusive use by deer, elk, and other animals.

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

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

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

Gender Study of Large Wolf Packs

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

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

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

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

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

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

15. *Navajo Reservation: Hogans* The following data are based on information taken from the book *Navajo Architecture: Forms, History, Distributions* by S. C. Jett and V. E. Spencer (University of Arizona Press). A survey of houses and traditional hogans was made in a number of different regions of the modern Navajo Indian Reservation. The following table is the result of a random sample of eight regions on the Navajo Reservation.

Area on Navajo Reservation	Number of Inhabited Houses	Number of Inhabited Hogans
Bitter Springs	18	13
Rainbow Lodge	16	14
Kayenta	68	46
Red Mesa	9	32
Black Mesa	11	15
Canyon de Chelly	28	47
Cedar Point	50	17
Burnt Water	50	18

Does this information indicate that the population mean number of inhabited houses is greater than that of hogans on the Navajo Reservation? Use a 5% level of significance.

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

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

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

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

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

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

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

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

i.   Let d be the random variable $d = A - B$. Use a calculator to verify that $\bar{d} \approx 2.472$ and $s_d \approx 12.124$.
ii. Do the data indicate that the U.S. population mean cost of living index for housing is higher than that for groceries in these areas? Use $\alpha = 0.05$.

18. *Economics: Cost of Living Index* In the following data pairs, *A* represents the cost of living index for utilities and *B* represents the cost of living index for transportation. The data are paired by metropolitan areas in the United States. A random sample of 46 metropolitan areas gave the following information (Reference: *Statistical Abstract of the United States,* 121st Edition).

A:	90	84	85	106	83	101	89	125	105
B:	100	91	103	103	109	109	94	114	113

A:	118	133	104	84	80	77	90	92	90
B:	120	130	117	109	107	104	104	113	101

A:	106	95	110	112	105	93	119	99	109
B:	96	109	103	107	103	102	101	86	94

A:	109	113	90	121	120	85	91	91	97
B:	88	100	104	119	116	104	121	108	86

A:	95	115	99	86	88	106	80	108	90	87
B:	100	83	88	103	94	125	115	100	96	127

i. Let d be the random variable $d = A - B$. Use a calculator to verify that $\bar{d} \approx -5.739$ and $s_d \approx 15.910$.

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

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

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

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

20. *Psychology: Training Rats* The following data are based on information from the Regis University Psychology Department. In an effort to determine if rats perform certain tasks more quickly if offered larger rewards, the following experiment was performed. On day 1, a group of three rats was given a reward of one food pellet each time they ran a maze. A second group of three rats was given a reward of five food pellets each time they ran the maze. On day 2, the groups were reversed, so the first group now got five food pellets for running the maze and the second group got only one pellet for running the same maze. The average times in seconds for each rat to run the maze 30 times are shown in the following table.

Rat	A	B	C	D	E	F
Time with one food pellet	3.6	4.2	2.9	3.1	3.5	3.9
Time with five food pellets	3.0	3.7	3.0	3.3	2.8	3.0

Do these data indicate that rats receiving larger rewards tend to run the maze in less time? Use a 5% level of significance.

21. *Psychology: Training Rats* The same experimental design discussed in Problem 20 was used to test rats trained to climb a sequence of short ladders. Times in seconds for eight rats to perform this task are shown in the following table.

Rat	A	B	C	D	E	F	G	H
Time 1 pellet	12.5	13.7	11.4	12.1	11.0	10.4	14.6	12.3
Time 5 pellets	11.1	12.0	12.2	10.6	11.5	10.5	12.9	11.0

Do these data indicate that rats receiving larger rewards tend to climb the ladder in less time? Use a 5% level of significance.

22. *Expand Your Knowledge: Confidence Intervals for μ_d* Using techniques from Section 7.2, we can find a confidence interval for μ_d. Consider a random sample of n matched data pairs A, B. Let $d = B - A$ be a random variable representing the difference between the values in a matched data pair. Compute the sample mean $\overline{d}$ of the differences and the sample standard deviation s_d. If d has a normal distribution or is mound-shaped, or if $n \geq 30$, then a confidence interval for μ_d is

$$\overline{d} - E < \mu_d < \overline{d} + E$$

where $E = t_c \dfrac{s_d}{\sqrt{n}}$

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

t_c = critical value for confidence level c and $d.f. = n - 1$

(a) Using the data of Problem 9, find a 95% confidence interval for the mean difference between percentage increase in company revenue and percentage increase in CEO salary.

(b) Use the confidence interval method of hypothesis testing outlined in Problem 25 of Section 8.2 to test the hypothesis that population mean percentage increase in company revenue is different from that of CEO salary. Use a 5% level of significance.

23. *Critical Region Method: Student's t* Solve Problem 9 using the critical region method of testing. Compare your conclusions with the conclusion obtained by using the P-value method. Are they the same?

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

SECTION 8.5

Testing $\mu_1 - \mu_2$ and $p_1 - p_2$ (Independent Samples)

FOCUS POINTS

- Identify independent samples and sampling distributions.
- Compute the sample test statistic and P-value for $\mu_1 - \mu_2$ and conclude the test.
- Compute the sample test statistic and P-value for $p_1 - p_2$ and conclude the test.

Independent Samples

Many practical applications of statistics involve a comparison of two population means or two population proportions. In Section 8.4, we considered tests of differences of means for *dependent samples*. With dependent samples, we could pair the data and then consider the difference of data measurements d. In this section, we will turn our attention to tests of differences of means from *independent samples*. We will see new techniques for testing the difference of means from independent samples.

Independent Samples

First, let's consider independent samples. We say that two sampling distributions are *independent* if there is no relationship whatsoever between specific values of the two distributions.

EXAMPLE 11   INDEPENDENT SAMPLE

A teacher wishes to compare the effectiveness of two teaching methods. Students are randomly divided into two groups: The first group is taught by method 1; the second group, by method 2. At the end of the course, a comprehensive exam is given to all students, and the mean score $\overline{x}_1$ for group 1 is compared with the mean score $\overline{x}_2$ for group 2. Are the samples independent or dependent?

SOLUTION: Because the students were *randomly* divided into two groups, it is reasonable to say that the $\overline{x}_1$ distribution is independent of the $\overline{x}_2$ distribution.

EXAMPLE 12   DEPENDENT SAMPLE

In Section 8.4, we considered a situation in which a shoe manufacturer claimed that for the general population of adult U.S. citizens, the average length of the left foot is longer than the average length of the right foot. To study this claim, the manufacturer gathers data in this fashion: Sixty adult U.S. citizens are drawn at random, and for these 60 people, both their left and right feet are measured. Let $\overline{x}_1$ be the mean length of the left feet and $\overline{x}_2$ be the mean length of the right feet. Are the $\overline{x}_1$ and $\overline{x}_2$ distributions independent for this method of collecting data?

SOLUTION: In this method, there is only *one* random sample of people drawn, and both the left and right feet are measured from this sample. The length of a person's left foot is usually related to the length of the person's right foot, so in this case the $\overline{x}_1$ and $\overline{x}_2$ distributions are *not* independent. In fact, we could pair the data and consider the distribution of the differences, left foot length minus right foot length. Then we would use the techniques of paired difference tests as found in Section 8.4.

GUIDED EXERCISE 8 | Independent sample

Suppose the shoe manufacturer of Example 12 gathers data in the following way: Sixty adult U.S. citizens are drawn at random and their left feet are measured; then another 60 adult U.S. citizens are drawn at random and their right feet are measured. Again, $\overline{x}_1$ is the mean of the left foot measurements and $\overline{x}_2$ is the mean of the right foot measurements.

Are the $\overline{x}_1$ and $\overline{x}_2$ distributions independent for this method of collecting data?

For this method of gathering data, two random samples are drawn: one for the left-foot measurements and one for the right-foot measurements. The first sample is not related to the second sample. The $\overline{x}_1$ and $\overline{x}_2$ distributions are independent.

Part I: Testing $\mu_1-\mu_2$ When σ_1 and σ_2 Are Known

Properties of $\overline{x}_1 - \overline{x}_2$ distribution, σ_1 and σ_2 known

In this part, we will use distributions that arise from a difference of means from independent samples. How do we obtain such distributions? If we have two statistical variables x_1 and x_2, each with its own distribution, we take independent random samples of size n_1 from the x_1 distribution and of size n_2 from the x_2 distribution. Then we can compute the respective means $\overline{x}_1$ and $\overline{x}_2$. Consider the difference $\overline{x}_1 - \overline{x}_2$. This represents a difference of means. If we repeat the sampling process over and over, we will come up with lots of $\overline{x}_1 - \overline{x}_2$ values. These values can be arranged in a frequency table, and we can make a histogram for the

distribution of $\bar{x}_1 - \bar{x}_2$ values. This will give us an experimental idea of the theoretical distribution of $\bar{x}_1 - \bar{x}_2$.

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

THEOREM 8.2   Let x_1 have a normal distribution with mean μ_1 and standard deviation σ_1. Let x_2 have a normal distribution with mean μ_2 and standard deviation σ_2. If we take independent random samples of size n_1 from the x_1 distribution and of size n_2 from the x_2 distribution, then the variable $\bar{x}_1 - \bar{x}_2$ has

1. A normal distribution

2. Mean $\mu_1 - \mu_2$

3. Standard deviation

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

COMMENT   Theorem 8.2 requires that x_1 and x_2 have normal distributions. However, if both n_1 and n_2 are 30 or larger, then for most practical applications, the central limit theorem assures us that $\bar{x}_1$ and $\bar{x}_2$ are approximately normally distributed. In this case, the conclusions of the theorem are again valid, even if the original x_1 and x_2 distributions are not normal.

Hypotheses for testing difference of means

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

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

As mentioned in Section 8.1, the null hypothesis is set up to see if it can be rejected. When testing the difference of means, we first set up the hypothesis H_0 that there is no difference. The alternate hypothesis could then be any of the ones listed in Table 8-12. The alternate hypothesis and consequent type of test used depend on the particular problem. Note that μ_1 is always listed first.

TABLE 8-12   **Alternate Hypotheses and Type of Test: Difference of Two Means**

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

Using Theorem 8.2 and the central limit theorem (Section 6.5), we can summarize the procedure for testing $\mu_1 - \mu_2$ when both σ_1 and σ_2 are known.

PROCEDURE

HOW TO TEST $\mu_1 - \mu_2$ WHEN BOTH σ_1 AND σ_2 ARE KNOWN

Requirements

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

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

n_1 and n_2 are sample sizes from populations 1 and 2

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

Procedure

1. In the context of the application, state the *null and alternate hypotheses* and set the *level of significance* α. It is customary to use
   $H_0: \mu_1 - \mu_2 = 0$.

2. Use $\mu_1 - \mu_2 = 0$ from the null hypothesis together with $\bar{x}_1, \bar{x}_2, \sigma_1, \sigma_2,$ n_1, and n_2 to compute the *sample test statistic*.

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

3. Use the standard normal distribution and the type of test, one-tailed or two-tailed, to find the *P-value* corresponding to the sample test statistic.

4. *Conclude the test.* If P-value $\leq \alpha$, then reject H_0. If P-value $> \alpha$, then do not reject H_0.

5. *Interpret your conclusion* in the context of the application.

EXAMPLE 13   TESTING THE DIFFERENCE OF MEANS (σ_1 AND σ_2 KNOWN)

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

   Model 1: Mean time $\bar{x}_1 = 11.4$ min; $\sigma_1 = 2.5$ min; $n_1 = 10$
   Model 2: Mean time $\bar{x}_2 = 9.9$ min; $\sigma_2 = 3.0$ min; $n_2 = 12$

Assume that the time required to bring water to a boil is normally distributed for each stove. Is there any difference (either way) between the performances of these two models? Use a 5% level of significance.

SOLUTION:

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

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

   The alternate hypothesis states that there is a difference:

   $H_1: \mu_1 \neq \mu_2$ or $H_1: \mu_1 - \mu_2 \neq 0$

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

(b) *Check Requirements* What distribution does the sample test statistic follow?
   The sample test statistic z follows a standard normal distribution because the original distributions from which the samples are drawn are normal. In addition, the population standard deviations of the original distributions are known and the samples are independent.

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

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \frac{1.5}{\sqrt{\dfrac{2.5^2}{10} + \dfrac{3.0^2}{12}}} \approx 1.28$$
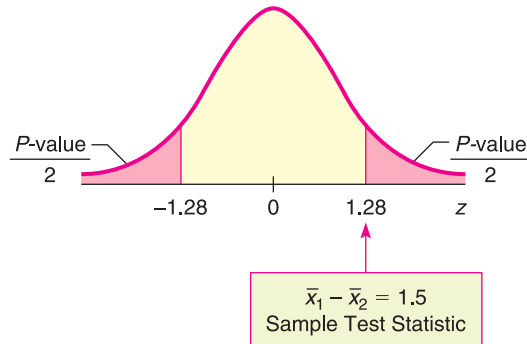
(d) Find the *P*-value and sketch the area on the standard normal curve. Figure 8-16 shows the *P*-value. Use the standard normal distribution (Table 5 of Appendix II) and the fact that we have a two-tailed test.

$$P\text{-value} \approx 2(0.1003) = 0.2006$$

FIGURE 8-16

P-value

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

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

GUIDED EXERCISE 9   **Testing the difference of means (σ_1 and σ_2 known)**

Let us return to Example 11 at the beginning of this section. A teacher wishes to compare the effectiveness of two teaching methods for her students. Students are randomly divided into two groups. The first group is taught by method 1; the second group, by method 2. At the end of the course, a comprehensive exam is given to all students.

The first group consists of $n_1 = 49$ students, with a mean score of $\bar{x}_1 = 74.8$ points. The second group has $n_2 = 50$ students, with a mean score of $\bar{x}_2 = 81.3$ points. The teacher claims that the second method will increase the mean score on the comprehensive exam. Is this claim justified at the 5% level of significance? Earlier research for the two methods indicates that $\sigma_1 = 14$ points and $\sigma_2 = 15$ points.

Let μ_1 and μ_2 be the mean scores of the distribution of all scores using method 1 and method 2, respectively.

(a) What is the null hypothesis?

$\Rightarrow$ $H_0\colon \mu_1 = \mu_2$ or $H_0\colon \mu_1 - \mu_2 = 0$

(b) *Check Requirements* What distribution does the sample test statistic follow? Explain.

$\Rightarrow$ The sample test statistic follows a standard normal distribution because the samples are large, both σ_1 and σ_2 are known, and the samples are independent.

(c) To examine the validity of the teacher's claim, what should the alternate hypothesis be? What is α?

$\Rightarrow$ $H_1\colon \mu_1 < \mu_2$ (the second method gives a higher average score) or $H_1\colon \mu_1 - \mu_2 < 0$.

$\alpha = 0.05$

(d) Compute $\bar{x}_1 - \bar{x}_2$.

$\Rightarrow$ $\bar{x}_1 - \bar{x}_2 = 74.8 - 81.3 = -6.5$

(e) Convert $\bar{x}_1 - \bar{x}_2 = -6.5$ to the sample test statistic *z*.

$\Rightarrow$ Using $\sigma_1 = 14$, $\sigma_2 = 15$, $n_1 = 49$, $n_2 = 50$, and $\mu_1 - \mu_2 = 0$ from H_0, we have

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \frac{-6.5 - 0}{\sqrt{\dfrac{14^2}{49} + \dfrac{15^2}{50}}} \approx -2.23$$
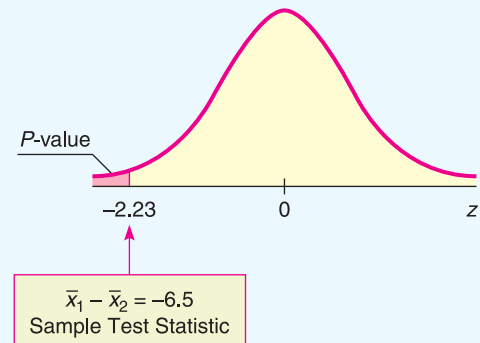
Continued

GUIDED EXERCISE 9   *continued*

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

➡ Figure 8-17 shows the *P*-value. It is the area to the left of $z = -2.23$. Using Table 5 of Appendix II, we find *P*-value $= P(z < -2.23) \approx 0.0129$.

FIGURE 8-17 *P*-value

P-value

-2.23 0 z

$\bar{x}_1 - \bar{x}_2 = -6.5$
Sample Test Statistic

(g) Conclude the test.

➡ Since *P*-value of $0.0129 \leq 0.05$ for α, reject H_0.

(h) *Interpret* the results in the context of the application.

➡ At the 5% level of significance, there is sufficient evidence to show that the second teaching method increased the population mean score on the exam.

Part II: Testing $\mu_1 - \mu_2$ When σ_1 and σ_2 Are Unknown

To test $\mu_1 - \mu_2$ when σ_1 and σ_2 are unknown, we use distribution methods similar to those in Chapter 7 for estimating $\mu_1 - \mu_2$. In particular, if the two distributions are normal or approximately mound-shaped, or if both sample sizes are large (≥ 30), we use a Student's t distribution. Let's summarize the method of testing $\mu_1 - \mu_2$.

PROCEDURE

HOW TO TEST $\mu_1 - \mu_2$ WHEN σ_1 AND σ_2 ARE UNKNOWN

Requirements

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

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

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

n_1 and n_2 are sample sizes from populations 1 and 2

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

Procedure

1. In the context of the application, state the *null and alternate hypotheses* and set the *level of significance* α. It is customary to use $H_0: \mu_1 - \mu_2 = 0$.

2. Use $\mu_1 - \mu_2 = 0$ from the null hypothesis together with $\bar{x}_1, \bar{x}_2, s_1, s_2,$ $n_1,$ and n_2 to compute the *sample test statistic*.

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

Continued

Problem 27 at the end of this section gives the formula for Satterthwaite's approximation for degrees of freedom.

The sample test statistic distribution is approximately that of a Student's t with *degrees of freedom d.f.* = *smaller* of $n_1 - 1$ and $n_2 - 1$.

Note that statistical software gives a more accurate and larger *d.f.* based on Satterthwaite's approximation (see Problem 27).

3. Use a Student's t distribution and the type of test, one-tailed or two-tailed, to find the *P-value* corresponding to the sample test statistic.

4. *Conclude the test.* If P-value $\leq \alpha$, then reject H_0. If P-value $> \alpha$, then do not reject H_0.

5. *Interpret your conclusion* in the context of the application.

EXAMPLE 14

TESTING THE DIFFERENCE OF MEANS (σ_1 AND σ_2 UNKNOWN)

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

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

$$\text{Brand A: } \bar{x}_1 = 21.8 \text{ min;} \qquad s_1 = 8.7 \text{ min;} \qquad n_1 = 12$$
$$\text{Brand B: } \bar{x}_2 = 18.9 \text{ min;} \qquad s_2 = 7.5 \text{ min;} \qquad n_2 = 12$$

Past experience with the drug composition of the two remedies permits researchers to assume that both distributions are approximately normal. Let us use a 5% level of significance to test the claim that there is no difference between the two brands in the mean time required for bodily absorption. Also, find or estimate the *P*-value of the sample test statistic.

SOLUTION:

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

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

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

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

(b) *Check Requirements* What distribution does the sample test statistic follow? We use a Student's t distribution with *d.f.* = smaller sample size $- 1 = 12 - 1 = 11$. Note that both samples are of size 12. Degrees of freedom can be computed also by Satterthwaite's approximation. A Student's t distribution is appropriate because the original populations are approximately normal, the population standard deviations are not known, and the samples are independent.

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

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

(d) Estimate the *P*-value and sketch the area on a *t* graph.

Figure 8-18 shows the *P*-value. The degrees of freedom are *d.f.* = 11 (since both samples are of size 12). Because the test is a two-tailed test, the *P*-value is the area to the right of 0.875 together with the area to the left of −0.875. In the Student's *t* distribution table (Table 6 of Appendix II), we find an interval containing the *P*-value. Find 0.875 in the row headed by *d.f.* = 11. The test statistic 0.875 falls between the entries 0.697 and 1.214. Because this is a two-tailed test, we use the corresponding *P*-values 0.500 and 0.250 from the *two-tail* area row (see Table 8-13, Excerpt from Table 6). The *P*-value for the sample *t* is in the interval

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

(e) Conclude the test.

$$
\begin{array}{c}
\alpha \\
\end{array}
$$
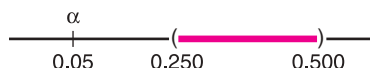
Since the interval containing the *P*-value lies to the right of $\alpha = 0.05$, we fail to reject H_0.

Note: Using the raw data and a calculator with Satterthwaite's approximation for the degrees of freedom *d.f.* ≈ 21.53, the *P*-value ≈ 0.3915. This value is in the interval we computed.
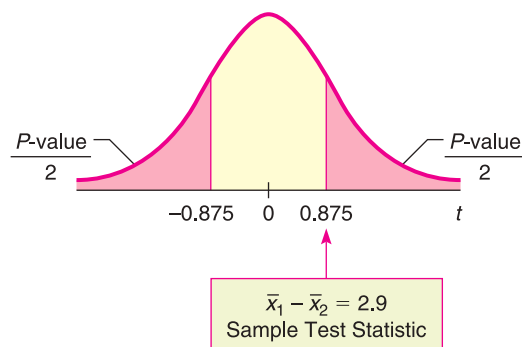
FIGURE 8-18

P-value

TABLE 8-13	**Excerpt from Table 6, Appendix II**	
one-tail area	0.250	0.125
✓ two-tail area	0.500	0.250
d.f. = 11	0.697	1.214
	Sample *t* = 0.875	

(f) *Interpretation* Interpret the results.

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

GUIDED EXERCISE 10 | **Testing the difference of means (σ_1 and σ_2 unknown)**

Suppose the experiment to measure the times in minutes for the headache remedies to enter the bloodstream (Example 14) yielded sample means, sample standard deviations, and sample sizes as follows:

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

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

Brand B claims to be faster. Is this claim justified at the 1% level of significance? (Use the following steps to obtain the answer.)

Continued

GUIDED EXERCISE 10   *continued*

(a) What is α? State H_0 and H_1.

⮕ $\alpha = 0.01$

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

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

(b) *Check Requirements* What distribution does the sample test statistic follow? Explain

⮕ Student's t distribution with $d.f.$ $n_2 - 1 = 8 - 1 = 7$ since $n_2 < n_1$. From Example 14, we know that the original distributions are approximately normal. Both σ_1 and σ_2 are unknown and the samples are independent.

(c) Compute $\bar{x}_1 - \bar{x}_2$ and the corresponding sample test statistic t.

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

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

(d) What degrees of freedom do you use? To find an interval containing the P-value, do you use one-tail or two-tail areas of Table 6, Appendix II? Sketch a figure showing the P-value. Find an interval for the P-value.

⮕ Since $n_2 < n_1$, $d.f. = n_2 - 1 = 8 - 1 = 7$. Use *one-tail area* of Table 6.
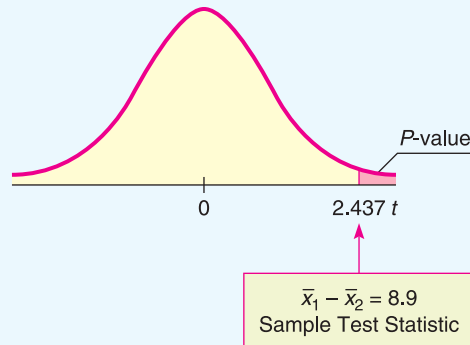
FIGURE 8-19   *P*-value

TABLE 8-14   **Excerpt from Student's *t* Table**

✓ one-tail area	0.025	0.010
two-tail area	0.050	0.020
$d.f. = 7$	2.365	2.998

↑
Sample t = 2.437

The sample t is between 2.365 and 2.998.

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

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

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

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

(f) *Interpret* the results in the context of the application.

⮕ At the 5% level of significance, there is sufficient evidence to conclude that the mean time for brand B to enter the bloodstream is less than that for brand A.
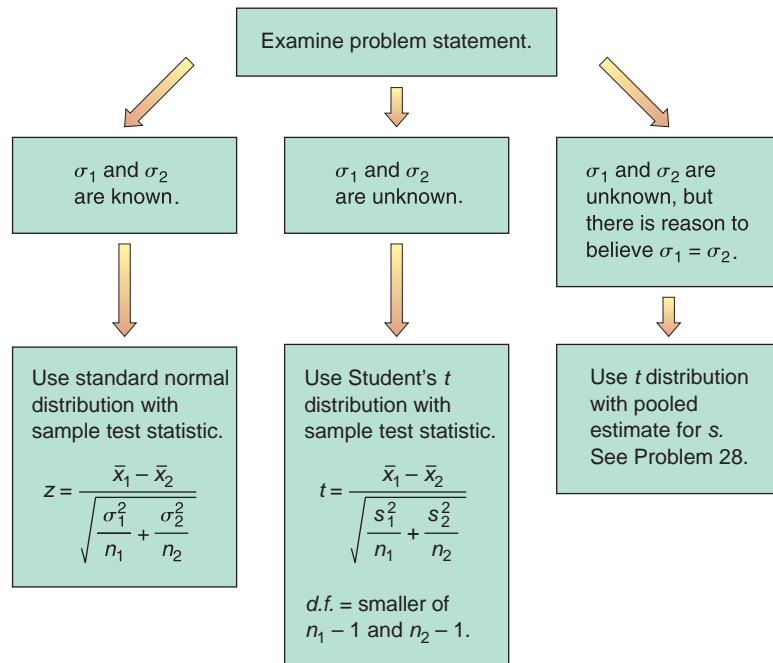
Problem 28 at the end of this section discusses the pooled standard deviation method for testing the difference of means when population standard deviations are equal. Section 10.4 has techniques for testing whether two standard deviations or two variances are equal.

Pooled standard deviation

There is another method of testing $\mu_1 - \mu_2$ when σ_1 and σ_2 are unknown. Suppose that the sample values s_1 and s_2 are sufficiently close and that there is reason to believe $\sigma_1 = \sigma_2$ (or the standard deviations are approximately equal). This situation can happen when you make a slight change or alteration to a known process or method of production. The standard deviation may not change much, but the outputs or means could be very different. When there is reason to believe that $\sigma_1 = \sigma_2$, it is best to use a *pooled standard deviation*. The sample test statistic $\bar{x}_1 - \bar{x}_2$ has a corresponding t variable with an *exact* Student's t distribution and degrees of freedom $d.f. = n_1 + n_2 - 2$. Problem 28 at the end of the section provides the details.

Under the null hypothesis H_0: $\mu_1 = \mu_2$ which distribution should you use for the sample test statistic $\bar{x}_1 - \bar{x}_2$?

Examine problem statement.

σ_1 and σ_2 are known.	σ_1 and σ_2 are unknown.	σ_1 and σ_2 are unknown, but there is reason to believe $\sigma_1 = \sigma_2$.
Use standard normal distribution with sample test statistic. $$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$	Use Student's t distribution with sample test statistic. $$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$ $d.f.$ = smaller of $n_1 - 1$ and $n_2 - 1$.	Use t distribution with pooled estimate for s. See Problem 28.

Part III: Testing a Difference of Proportions $p_1 - p_2$

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

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

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

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

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

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

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

For most practical problems involving a comparison of two binomial populations, the experimenters will want to test the null hypothesis $H_0: p_1 = p_2$. Consequently, this is the type of test we shall consider. Since the values of p_1 and p_2 are unknown, and since specific values are not assumed under the null hypothesis $H_0: p_1 = p_2$, the best estimate for the common value is the total number of successes $(r_1 + r_2)$ divided by the total number of trials $(n_1 + n_2)$. If we denote this *pooled estimate of proportion* by $\bar{p}$ (read "p bar"), then

Pooled estimate of proportion $\bar{p}$

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

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

Criteria for using the normal approximation to the binomial

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

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

is larger than 5 (see Section 6.6).

As stated earlier, the sample statistic $\hat{p}_1 - \hat{p}_2$ has a normal distribution with mean $\mu = p_1 - p_2$ and standard deviation $\sigma = \sqrt{p_1q_1/n_1 + p_2q_2/n_2}$. Under the null hypothesis, we assume that $p_1 = p_2$ and then use the pooled estimate $\bar{p}$ in place of each p. Using all this information, we find that the *sample test statistic* is

Sample test statistic

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

where $\bar{p} = \dfrac{r_1 + r_2}{n_1 + n_2}$ and $\bar{q} = 1 - \bar{p}$

$$\hat{p}_1 = \frac{r_1}{n_1} \quad \text{and} \quad \hat{p}_2 = \frac{r_2}{n_2}$$

Using this information, we summarize the procedure for testing $p_1 - p_2$.

PROCEDURE

HOW TO TEST A DIFFERENCE OF PROPORTIONS $P_1 - P_2$

Procedure

Consider two independent binomial experiments.

Binomial Experiment 1

n_1 = number of trials

r_1 = number of successes out of n_1 trials

$\hat{p}_1 = \dfrac{r_1}{n_1}$

p_1 = population probability of success on a single trial

Binomial Experiment 2

n_2 = number of trials

r_2 = number of successes out of n_2 trials

$\hat{p}_2 = \dfrac{r_2}{n_2}$

p_2 = population probability of success on a single trial

1. Use the *null hypothesis* of no difference, $H_0: p_1 - p_2 = 0$. In the context of the application, choose the *alternate hypothesis*. Set the *level of significance* α.

2. The null hypothesis claims that $p_1 = p_2$; therefore, *pooled best estimates* for the population probabilities of success and failure are

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

Requirements

The number of trials should be sufficiently large so that each of the four quantities $n_1\bar{p}$, $n_1\bar{q}$, $n_2\bar{p}$, and $n_2\bar{q}$ is larger than 5.

In this case, you compute the *sample test statistic*

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

3. Use the standard normal distribution and a type of test, one-tailed or two-tailed, to find the *P-value* corresponding to the sample test statistic.

4. *Conclude the test.* If P-value $\leq \alpha$, then reject H_0. If P-value $> \alpha$, then do not reject H_0.

5. *Interpret your conclusion* in the context of the application.

EXAMPLE 15   TESTING THE DIFFERENCE OF PROPORTIONS

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

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

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

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

SOLUTION:

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

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

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

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

(b) Calculate the *pooled estimates* $\bar{p}$ and $\bar{q}$.
Under the null hypothesis that $p_1 = p_2$ we find

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

(c) *Check Requirements* What distribution does the sample test statistic follow?
The sample test statistic follows a standard normal distribution because the number of binomial trials is large enough that each of the products $n_1\bar{p}$, $n_1\bar{q}$, $n_2\bar{p}$, $n_2\bar{q}$ exceeds 5.

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

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

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

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

then

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

We computed the pooled estimate for $\bar{p}$ and $\bar{q}$ in part (b).
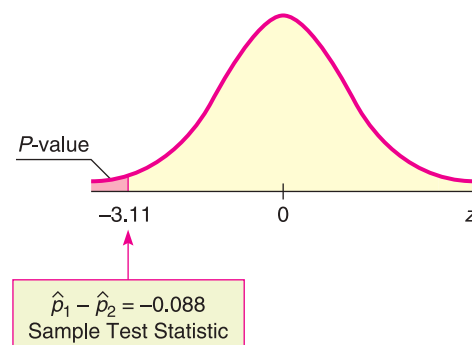
$$\bar{p} = 0.516 \quad \text{and} \quad \bar{q} = 1 - \bar{p} = 0.484$$

Using these values, we find the sample test statistic

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

(e) Find the *P*-value and sketch the area on the standard normal curve.
Figure 8-20 shows the *P*-value. This is a left-tailed test, so the *P*-value is the area to the left of -3.11. Using the standard normal distribution (Table 5 of Appendix II), we find P-value $= P(z < -3.11) \approx 0.0009$.

FIGURE 8-20

P-value

Copyright 2010 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s).
Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

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

(g) *Interpretation* Interpret the results in the context of the application.

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

GUIDED EXERCISE 11 | **Testing the difference of proportions**

In Example 15 about voter registration, suppose that a random sample of 1100 potential voters was randomly divided into two groups.

>   *Group 1*: 500 potential voters; no registration reminders sent; 248 registered to vote
>   *Group 2*: 600 potential voters; registration reminders sent; 332 registered to vote

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

(a) What is α? State H_0 and H_1.

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

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

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

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

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

(c) *Check Requirements* What distribution does the sample test statistic follow?

⟹ The sample test statistic z follows a standard normal distribution because the number of binomial trials is large enough that each of the products $n_1\bar{p}$, $n_1\bar{q}$, $n_2\bar{p}$, $n_2\bar{q}$ exceeds 5.

(d) Compute $\hat{p}_1 - \hat{p}_2$.

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

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

(e) Find the sample test statistic.

⟹ $$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\dfrac{\bar{p}\,\bar{q}}{n_1} + \dfrac{\bar{p}\,\bar{q}}{n_2}}} = \frac{-0.057}{\sqrt{\dfrac{(0.527)(0.473)}{500} + \dfrac{(0.572)(0.473)}{600}}}$$

$$\approx -1.89$$

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

⟹ Figure 8-21 shows the *P*-value. It is the area to the left of $z = -1.89$. Using Table 5 of Appendix II, we find *P*-value $= P(z < -1.89) = 0.0294$.

FIGURE 8-21 *P*-value

P-value

-1.89 0 z

$\hat{p}_1 - \hat{p}_2 = -0.057$
Sample Test Statistic
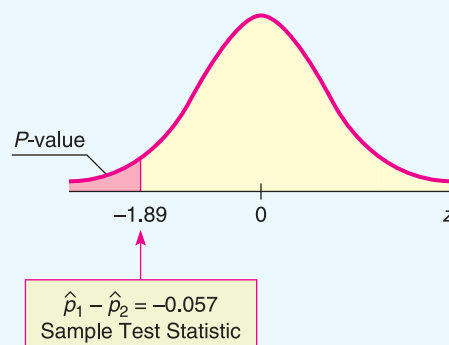
Continued

(g) Conclude the test.

⟹   Since *P*-value of 0.0294 > 0.01 for α, we cannot reject H_0.

(h) *Interpret* the results in the context of the application.

⟹   At the 1% level of significance, the data do not support the claim that the reminders increased the proportion of registered voters.

TECH NOTES   The TI-84Plus/TI-83Plus/TI-*n*spire calculators, Excel 2007, and Minitab all support testing the difference of means for independent samples. The TI-84Plus/TI-83Plus/TI-*n*spire calculators and Minitab also support testing the difference of proportions. When testing the difference of means using the normal distribution, the technologies require the population standard deviations for the distributions of the two samples. When testing the difference of means using a Student's *t* distribution, the technologies give the option of using the pooled standard deviation. As discussed in Problem 28 of this section, the pooled standard deviation is appropriate when the standard deviations of the two populations are approximately equal. If the pooled standard deviation option is not selected, the technologies compute the sample test statistic for $\bar{x}_1 - \bar{x}_2$ using the procedures described in this section. However, Satterthwaite's approximation for the degrees of freedom is used.

TI-84Plus/TI-83Plus/TI-*n*spire (with TI-84Plus keypad) Enter the data. Press **STAT** and select **TESTS**. Options **3: 2-SampZTest, 4: 2-SampTTest, 6: 2-PropZTest** perform tests for the difference of means using the normal distribution, difference of means using a Student's *t* distribution, and difference of proportions, respectively. The calculator uses the symbol $\hat{p}$ to designate the pooled estimate for *p*.

Excel 2007 Enter the data in columns. On the home ribbon, click the **Data** tab. Then in the Analysis Group, select **Data Analysis**. The choice **z-Test Two Sample Means** conducts a test for the difference of means using the normal distribution. The choice **t-Test: Two-Sample Assuming Unequal Variances** conducts a test using a Student's *t* distribution with Satterthwaite's approximation for the degrees of freedom. The choice **t-Test: Two-Sample Assuming Equal Variances** conducts a test using the pooled standard deviation.

Minitab Enter the data in two columns. Use the menu choices **Stat ➤ Basic Statistics.** The choice **2 sample t** tests the difference of means using a Student's *t* distribution. In the dialogue box, leaving the box **Assume equal variances** unchecked produces a test using Satterthwaite's approximation for the degrees of freedom. Checking the box **Assume equal variances** produces a test using the pooled standard deviation. Minitab does not support testing the difference of means using the normal distribution. The menu item **2 Proportions** tests the difference of proportions. Under **Options**, select the null and alternate hypotheses.

Part IV: Testing $\mu_1 - \mu_2$ and $p_1 - p_2$ Using Critical Regions

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

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

Test conclusions based on critical values

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

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

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

Otherwise, in each case, do not reject H_0.

Critical values z_0 for tests using the standard normal distribution can be found in Table 5(c) of Appendix II. Critical values t_0 for tests using a Student's t distribution are found in Table 6 of Appendix II. Use the row headed by the appropriate degrees of freedom and the column that includes the value of α (level of significance) in the *one-tail area* row for one-tailed tests or the *two-tail area* row for two-tailed tests.

EXAMPLE 16   CRITICAL REGION METHOD

Use the critical region method to solve the application in Example 14 (testing $\mu_1 - \mu_2$ when σ_1 and σ_2 are unknown).

SOLUTION: Example 14 involves testing the difference in average time for two headache remedies to reach the bloodstream. For brand A, $\bar{x}_1 = 21.8$ min, $s_1 = 8.7$ min, and $n_1 = 12$; for brand B, $\bar{x}_2 = 18.9$ min, $s_2 = 7.5$ min, and $n_2 = 12$. The level of significance α is 0.05. The Student's t distribution is appropriate because both populations are approximately normal.

(a) To use the critical region method to test for a difference in average time, we use the same hypotheses and the same sample test statistic as in Example 14.

$$H_0: \mu_1 = \mu_2; H_1: \mu_1 \neq \mu_2;$$
$$\bar{x}_1 - \bar{x}_2 = 2.9 \text{ min; sample test statistic } t = 0.875$$

(b) Instead of finding the P-value of the sample test statistic, we use α and H_1 to find the critical values in Table 6 of Appendix II. We have $d.f. = 11$ (since both samples are of size 12). We find $\alpha = 0.05$ in the *two-tail area* row, since we have a two-tailed test. The critical values are $\pm t_0 = \pm 2.201$.

Next compare the sample test statistic $t = 0.875$ to the critical values. Figure 8-22 shows the critical regions and the sample test statistic. We see that the sample test statistic falls in the "do not reject H_0" region. At the 5% level of significance, sample evidence does not show a difference in the time for the drugs to reach the bloodstream. The result is consistent with the result obtained by the P-value method of Example 14.

FIGURE 8-22

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

$t = 0.875$
or
$\bar{x}_1 - \bar{x}_2 = 2.9$
Sample Test Statistic
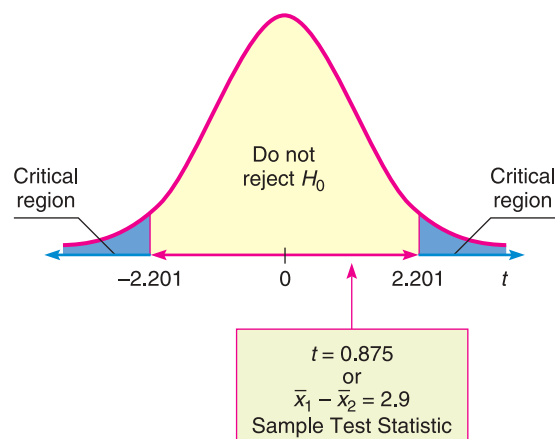
VIEWPOINT | Temper! Temper!

In her book Red Ink Behaviors, *Jean Hollands discusses inappropriate, problem behaviors of professional employees in the corporate business world. Temper tantrums, flaming e-mails, omitting essential information, sabotaging fellow workers, and the arrogant opinion that others are "dumb and dispensable" create personnel problems that cost companies a lot in the form of wasted time, reduced productivity, and lost revenues. A study of major industries in the Silicon Valley area gave Hollands data for estimating just how much time and money are wasted by such "red ink behaviors." For more information, see Problems 25 and 26 in this section.*

SECTION 8.5 PROBLEMS

1. *Statistical Literacy* Consider a hypothesis test of difference of means for two independent populations x_1 and x_2.
   (a) What does the null hypothesis say about the relationship between the two population means?
   (b) In the case that the sample test statistic is z, give the formula for z.
   (c) In the case that the sample test statistic is t, give the formula for t.

2. *Statistical Literacy* Consider a hypothesis test of difference of means for two independent populations x_1 and x_2. Suppose that both sample sizes are greater than 30 and that you know σ_1 but not σ_2. Is it standard practice to use the normal distribution or a Student's t distribution?

3. *Statistical Literacy* Consider a hypothesis test of difference of means for two independent populations x_1 and x_2. What are two ways of expressing the null hypothesis?

4. *Statistical Literacy* Consider a hypothesis test of difference of proportions for two independent populations. Suppose random samples produce r_1 successes out of n_1 trials for the first population and r_2 successes out of n_2 trials for the second population.
   (a) What does the null hypothesis claim about the relationship between the proportions of successes in the two populations?
   (b) What is the formula for the sample test statistic?

5. *Statistical Literacy* Consider a hypothesis test of difference of proportions for two independent populations. Suppose random samples produce r_1 successes out of n_1 trials for the first population and r_2 successes out of n_2 trials for the second population. What is the best pooled estimate $\bar{p}$ for the population probability of success using $H_0: p_1 = p_2$?

6. *Critical Thinking* Consider use of a Student's t distribution to test the difference of means for independent populations using random samples of sizes n_1 and n_2.
   (a) Which process gives the larger degrees of freedom, Satterthwaite's approximation or using the smaller of $n_1 - 1$ and $n_2 - 1$? Which method is more conservative? What do we mean by "conservative"? Note that computer programs and other technologies commonly use Satterthwaite's approximation.
   (b) Using the same hypotheses and sample data, is the P-value smaller for larger degrees of freedom? How might a larger P-value impact the significance of a test?

7. *Critical Thinking* When conducting a test for the difference of means for two independent populations x_1 and x_2, what alternate hypothesis would indicate that the mean of the x_2 population is smaller than that of the x_1 population? Express the alternate hypothesis in two ways.

8. | *Critical Thinking* When conducting a test for the difference of means for two independent populations x_1 and x_2, what alternate hypothesis would indicate that the mean of the x_2 population is larger than that of the x_1 population? Express the alternate hypothesis in two ways.

9. | *Basic Computation: Testing $\mu_1 - \mu_2$* A random sample of 49 measurements from one population had a sample mean of 10, with sample standard deviation 3. An independent random sample of 64 measurements from a second population had a sample mean of 12, with sample standard deviation 4. Test the claim that the population means are different. Use level of significance 0.01.
   (a) *Check Requirements* What distribution does the sample test statistic follow? Explain.
   (b) State the hypotheses.
   (c) Compute $\bar{x}_1 - \bar{x}_2$ and the corresponding sample test statistic.
   (d) Estimate the *P*-value of the sample test statistic.
   (e) Conclude the test.
   (f) *Interpret* the results.

10. | *Basic Computation: Testing $\mu_1 - \mu_2$* Two populations have mound-shaped, symmetric distributions. A random sample of 16 measurements from the first population had a sample mean of 20, with sample standard deviation 2. An independent random sample of 9 measurements from the second population had a sample mean of 19, with sample standard deviation 3. Test the claim that the population mean of the first population exceeds that of the second. Use a 5% level of significance.
   (a) *Check Requirements* What distribution does the sample test statistic follow? Explain.
   (b) State the hypotheses.
   (c) Compute $\bar{x}_1 - \bar{x}_2$ and the corresponding sample test statistic.
   (d) Estimate the *P*-value of the sample test statistic.
   (e) Conclude the test.
   (f) *Interpret* the results.

11. | *Basic Computation: Testing $\mu_1 - \mu_2$* A random sample of 49 measurements from a population with population standard deviation 3 had a sample mean of 10. An independent random sample of 64 measurements from a second population with population standard deviation 4 had a sample mean of 12. Test the claim that the population means are different. Use level of significance 0.01.
   (a) *Check Requirements* What distribution does the sample test statistic follow? Explain.
   (b) State the hypotheses.
   (c) Compute $\bar{x}_1 - \bar{x}_2$ and the corresponding sample test statistic.
   (d) Find the *P*-value of the sample test statistic.
   (e) Conclude the test.
   (f) *Interpret* the results.

12. | *Basic Computation: Testing $\mu_1 - \mu_2$* Two populations have normal distributions. The first has population standard deviation 2 and the second has population standard deviation 3. A random sample of 16 measurements from the first population had a sample mean of 20. An independent random sample of 9 measurements from the second population had a sample mean of 19. Test the claim that the population mean of the first population exceeds that of the second. Use a 5% level of significance.
   (a) *Check Requirements* What distribution does the sample test statistic follow? Explain.
   (b) State the hypotheses.
   (c) Compute $\bar{x}_1 - \bar{x}_2$ and the corresponding sample test statistic.
   (d) Find the *P*-value of the sample test statistic.
   (e) Conclude the test
   (f) *Interpret* the results.

13. *Basic Computation: Test* $p_1 - p_2$ For one binomial experiment, 75 binomial trials produced 45 successes. For a second independent binomial experiment, 100 binomial trials produced 70 successes. At the 5% level of significance, test the claim that the probabilities of success for the two binomial experiments differ.
    (a) Compute the pooled probability of success for the two experiments.
    (b) *Check Requirements* What distribution does the sample test statistic follow? Explain.
    (c) State the hypotheses.
    (d) Compute $\hat{p}_1 - \hat{p}_2$ and the corresponding sample test statistic.
    (e) Find the *P*-value of the sample test statistic.
    (f) Conclude the test.
    (g) *Interpret* the results.

14. *Basic Computation: Test* $p_1 - p_2$ For one binomial experiment, 200 binomial trials produced 60 successes. For a second independent binomial experiment, 400 binomial trials produced 156 successes. At the 5% level of significance, test the claim that the probability of success for the second binomial experiment is greater than that for the first.
    (a) Compute the pooled probability of success for the two experiments.
    (b) *Check Requirements* What distribution does the sample test statistic follow? Explain.
    (c) State the hypotheses.
    (d) Compute $\hat{p}_1 - \hat{p}_2$ and the corresponding sample test statistic.
    (e) Find the *P*-value of the sample test statistic.
    (f) Conclude the test.
    (g) *Interpret* the results.

Please provide the following information for Problems 15–26 and 29–35.
(a) What is the level of significance? State the null and alternate hypotheses.
(b) *Check Requirements* What sampling distribution will you use? What assumptions are you making? Compute the sample test statistic.
(c) Find (or estimate) the *P*-value. Sketch the sampling distribution and show the area corresponding to the *P*-value.
(d) Based on your answers in parts (a) to (c), will you reject or fail to reject the null hypothesis? Are the data statistically significant at level α?
(e) *Interpret* your conclusion in the context of the application.
   Note: For degrees of freedom *d.f.* not in the Student's *t* table, use the closest *d.f.* that is *smaller.* In some situations, this choice of *d.f.* may increase the *P*-value a small amount, and therefore produce a slightly more "conservative" answer.
Answers may vary due to rounding.

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

16. *Environment: Pollution Index* Based on information from *The Denver Post*, a random sample of $n_1 = 12$ winter days in Denver gave a sample mean pollution index of $\bar{x}_1 = 43$. Previous studies show that $\sigma_1 = 21$. For Englewood (a suburb of Denver), a random sample of $n_2 = 14$ winter days gave a sample mean pollution index of $\bar{x}_2 = 36$. Previous studies show that $\sigma_2 = 15$. Assume the pollution index is normally distributed in both Englewood and

Denver. Do these data indicate that the mean population pollution index of Englewood is different (either way) from that of Denver in the winter? Use a 1% level of significance.

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

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

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

x_1: **New England Crime Rate**

| 3.5 | 3.7 | 4.0 | 3.9 | 3.3 | 4.1 | 1.8 | 4.8 | 2.9 | 3.1 |

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

x_2: **Rocky Mountain States**

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

(Reference: *Crime in the United States*, Federal Bureau of Investigation.) Assume that the crime rate distribution is approximately normal in both regions.

i. Use a calculator to verify that $\bar{x}_1 \approx 3.51$, $s_1 \approx 0.81$, $\bar{x}_2 \approx 3.87$, and $s_2 \approx 0.94$.

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

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

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

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

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

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

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

(Reference: National Center for Health Statistics.)

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

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

21. *Education: Tutoring*   In the journal *Mental Retardation*, an article reported the results of a peer tutoring program to help mildly mentally retarded children learn to read. In the experiment, the mildly retarded children were randomly divided into two groups: the experimental group received peer tutoring along with regular instruction, and the control group received regular instruction with no peer tutoring. There were $n_1 = n_2 = 30$ children in each group. The Gates–MacGintie Reading Test was given to both groups before instruction began. For the experimental group, the mean score on the vocabulary portion of the test was $\bar{x}_1 = 344.5$, with sample standard deviation $s_1 = 49.1$. For the control group, the mean score on the same test was $\bar{x}_2 = 354.2$, with sample standard deviation $s_2 = 50.9$. Use a 5% level of significance to test the hypothesis that there was no difference in the vocabulary scores of the two groups before the instruction began.

22. *Education: Tutoring*   In the article cited in Problem 21, the results of the following experiment were reported. Form 2 of the Gates–MacGintie Reading Test was administered to both an experimental group and a control group after 6 weeks of instruction, during which the experimental group received peer tutoring and the control group did not. For the experimental group $n_1 = 30$ children, the mean score on the vocabulary portion of the test was $\bar{x}_1 = 368.4$, with sample standard deviation $s_1 = 39.5$. The average score on the vocabulary portion of the test for the $n_2 = 30$ subjects in the control group was $\bar{x}_2 = 349.2$, with sample standard deviation $s_2 = 56.6$. Use a 1% level of significance to test the claim that the experimental group performed better than the control group.

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

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

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

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

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

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

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

Field A samples, x_1:

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

Field B samples, x_2:

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

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

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

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

x_1:	1	5	8	4	2	4	10

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

x_2:	10	5	4	7	9	4	10	3

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

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

26. *Management: Intimidators and Stressors* This problem is based on information regarding productivity in leading Silicon Valley companies (see reference in Problem 25). In large corporations, an "intimidator" is an employee who tries to stop communication, sometimes sabotages others, and, above all, likes to listen to him- or herself talk. Let x_1 be a random variable representing productive hours per week lost by peer employees of an intimidator.

x_1:	8	3	6	2	2	5	2

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

x_2:	3	3	10	7	6	2	5	8

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

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

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

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

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

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

(a) In Problem 19, we tested whether the population average crime rate μ_2 in the Rocky Mountain region is higher than that in New England, μ_1. The data were $n_1 = 10$, $\bar{x}_1 \approx 3.51$, $s_1 \approx 0.81$, $n_2 = 12$, $\bar{x}_2 \approx 3.87$ and $s_2 \approx 0.94$. Use Satterthwaite's formula to compute the degrees of freedom for the Student's t distribution.

(b) When you did Problem 19, you followed the convention that degrees of freedom $d.f. = smaller$ of $n_1 = 1$ and $n_2 = 1$. Compare this value of $d.f.$ with that found with Satterthwaite's formula.

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

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

and where the **pooled standard deviation s** is

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

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

(a) There are many situations in which we want to compare means from populations having standard deviations that are equal. This method applies even if the standard deviations are known to be only approximately equal (see Section 10.4 for methods to test that $\sigma_1 = \sigma_2$). Consider Problem 23 regarding average incidence of fox rabies in two regions. For region I, $n_1 = 16$, $\bar{x}_1 = 4.75$, and $s_1 \approx 2.82$ and for region II, $n_2 = 15$, $\bar{x}_2 \approx 3.93$, and $s_2 \approx 2.43$. The two sample standard deviations are sufficiently close that we can assume $\sigma_1 = \sigma_2$. Use the method of pooled standard deviation to redo Problem 23, where we tested if there was a difference in population mean average incidence of rabies at the 5% level of significance.

(b) Compare the t value calculated in part (a) using the pooled standard deviation with the t value calculated in Problem 23 using the unpooled standard deviation. Compare the degrees of freedom for the sample test statistic. Compare the conclusions.

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

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

31. *Sociology: High School Dropouts* This problem is based on information taken from *Life in America's Fifty States* by G. S. Thomas. A random sample of $n_1 = 153$ people ages 16 to 19 was taken from the island of Oahu, Hawaii, and 12 were found to be high school dropouts. Another random sample of $n_2 = 128$ people ages 16 to 19 was taken from Sweetwater County, Wyoming, and 7 were found to be high school dropouts. Do these data indicate that the population proportion of high school dropouts on Oahu is different (either way) from that of Sweetwater County? Use a 1% level of significance.

32. *Political Science: Voters* A random sample of $n_1 = 228$ voters registered in the state of California showed that 141 voted in the last general election. A random sample of $n_2 = 216$ registered voters in the state of Colorado showed that 125 voted in the most recent general election. (See reference in Problem 31.) Do these data indicate that the population proportion of voter turnout in Colorado is higher than that in California? Use a 5% level of significance.

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

34. *Art: Politics* Do you prefer paintings in which the people are fully clothed? This question was asked by a professional survey group on behalf of the National Arts Society (see reference in Problem 30). A random sample of $n_1 = 59$ people who are conservative voters showed that $r_1 = 45$ said yes. Another random sample of $n_2 = 62$ people who are liberal voters showed that $r_2 = 36$ said yes. Does this indicate that the population proportion of conservative voters who prefer art with fully clothed people is higher than that of liberal voters? Use $\alpha = 0.05$.

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

36.   *Critical Region Method: Testing*   Redo Problem 15
       using the critical region method and compare your results to those obtained
       using the *P*-value method.

37.   *Critical Region Method: Testing*   Redo Problem 19
       using the critical region method and compare your results to those obtained
       using the *P*-value method.

38.   *Critical Region Method: Testing $p_1 - p_2$* Redo Problem 29 using the critical
       region method and compare your results to those obtained using the *P*-value
       method.

Chapter Review

SUMMARY

Hypothesis testing is a major component of inferential statistics. In hypothesis testing, we propose a specific value for the population parameter in question. Then we use sample data from a random sample and probability to determine whether or not to reject this specific value for the parameter.

Basic components of a hypothesis test are:

- The *null hypothesis* H_0 states that a parameter equals a specific value.

- The *alternate hypothesis* H_1 states that the parameter is greater than, less than, or simply not equal to the value specified in H_0.

- The *level of significance* α of the test is the probability of rejecting H_0 when it is true.

- The *sample test statistic* corresponding to the parameter in H_0 is computed from a random sample and appropriate sampling distribution.

- Assuming H_0 is true, the probability that a sample test statistic will take on a value as extreme as, or more extreme than, the observed sample test statistic is the *P-value* of the test. The *P*-value is computed by using the sample test statistic, the corresponding sampling distribution, H_0, and H_1.

- If *P*-value $\leq \alpha$, we reject H_0. If *P*-value $> \alpha$, we fail to reject H_0.

- We say that sample data are *significant* if we can reject H_0.

An alternative way to conclude a test of hypotheses is to use critical regions based on the alternate hypothesis and α. Critical values z_0 are found in Table 5(c) of Appendix II. Critical values t_0 are found in Table 6 of Appendix II. If the sample test statistic falls beyond the critical values—that is, in the critical region—we reject H_0.

The methods of hypothesis testing are very general, and we will see them used again in later chapters. In this chapter, we looked at tests involving

- Parameter μ. Use standard normal or Student's t distribution. See procedure displays in Section 8.2.

- Parameter p. Use standard normal distribution. See procedure displays in Section 8.3.

- Paired difference test for difference of means from dependent populations. Use Student's t distribution. See procedure displays in Section 8.4.

- Parameter $\mu_1 - \mu_2$ from independent populations. Use standard normal or Student's t distribution. See procedure displays in Section 8.5.

- Parameter $p_1 - p_2$ from independent populations. Use standard normal distribution. See procedure displays in Section 8.5.

FINDING THE *P*-VALUE CORRESPONDING TO A SAMPLE TEST STATISTIC

Use the appropriate sampling distribution as described in procedure displays for each of the various tests.
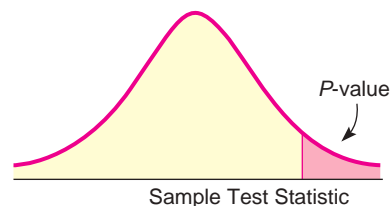
Left-Tailed Test
P-value = area to the left of the sample test statistic.
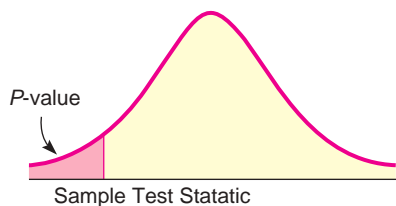
Right-Tailed Test
P-value = area to the right of the sample test statistic.

P-value

Sample Test Statatic

P-value

Sample Test Statistic

Two-Tailed Test

Sample test statistic lies to *left* of center *P*-value = twice the area to the left of sample test statistic

Sample test statistic lies to *right* of center *P*-value = twice area to the right of sample test statistic

P-value = twice the area

Sample Test Statistic

P-value = twice this area

Sample Test Statatic
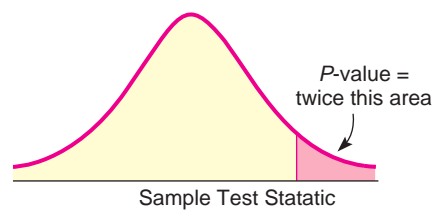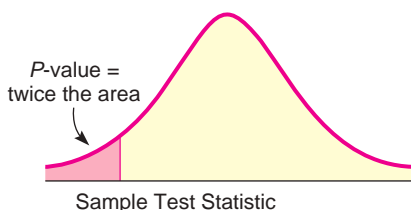
Sampling Distributions for Inferences Regarding μ or p

Parameter	Condition	Sampling Distribution
μ	σ is known and x has a normal distribution or $n \geq 30$	Normal Distribution
μ	σ is not known and x has a normal or mound-shaped symmetric distribution or $n \geq 30$	Student's t Distribution with $d.f. = n - 1$
p	$np > 5$ and $n(1 - p) > 5$	Normal Distribution

IMPORTANT WORDS & SYMBOLS

Section 8.1
Hypothesis testing 411
Hypothesis 411
Null hypothesis H_0 411
Alternate hypothesis H_1 411
right-tailed test 412
two-tailed test 412
left-tailed test 412
sample test statistic 413
P-value 414
Type I error 416
Type II error 416
Level of significance of a test and the probability of a type I error, α 416
Probability of a type II error, β 417
Power of a test $(1 - \beta)$ 417
statistical significance 418

Section 8.2
$d.f.$ for testing μ when σ is unknown 428
Critical region method 432
Critical value 433

Section 8.3
Criteria for using normal approximation to binomial, $np > 5$ and $nq > 5$ 442

Section 8.4
Paired data 452
Dependent samples 452

Section 8.5
Independent samples 466
$d.f.$ for testing $\mu_1 - \mu_2$ when σ_1 and σ_2 are unknown 472
Pooled standard deviation 475
Pooled estimate of proportion $\bar{p}$ 476

Do cloud-seed experiments ever work? If you seed the clouds, will it rain? If it does rain, who will benefit? Who will be displeased by the rain? If you seed the clouds and nothing happens, will taxpayers (who support the effort) complain or rejoice? Maybe this should be studied over a remote island—such as Tasmania (near Australia). Using what you already know about statistical testing, you can conduct your own tests, given the appropriate data. Remember, there are sociological questions (pleased/displeased with result) as well as technical questions (number of inches of rain produced). For data regarding cloud-seeding experiments over Tasmania, visit the Brase/Brase statistics site at **http://www.cengage.com/statistics/brase** *and find the link to DASL, the Carnegie Mellon University Data and Story Library. From the DASL site, look under Datasets for Cloud.*

CHAPTER REVIEW PROBLEMS

1. *Statistical Literacy* When testing μ or the difference of means $\mu_1 - \mu_2$ from independent populations, how do we decide whether to use the standard normal distribution or a Student's t distribution?

2. *Statistical Literacy* What do we mean when we say a test is *significant*? Does this necessarily mean the results are important?

3. *Critical Thinking* All other conditions being equal, does a larger sample size increase or decrease the corresponding magnitude of the z or t value of the sample test statistic?

4. *Critical Thinking* All other conditions being equal, does a z or t value with larger magnitude have a larger or smaller corresponding P-value?

Before you solve each problem below, first categorize it by answering the following question: Are we testing a single mean, a difference of means, a paired difference, a single proportion, or a difference of proportions? Assume underlying population distributions are mound-shaped and symmetric for problems with small samples that involve testing a mean or difference of means. Then provide the following information for Problems 5–18.

   (a) What is the level of significance? State the null and alternate hypotheses.
   (b) *Check Requirements* What sampling distribution will you use? What assumptions are you making? What is the value of the sample test statistic?
   (c) Find (or estimate) the P-value. Sketch the sampling distribution and show the area corresponding to the P-value.
   (d) Based on your answers in parts (a) to (c), will you reject or fail to reject the null hypothesis? Are the data statistically significant at level α?
   (e) *Interpret* your conclusion in the context of the application.
       Note: For degrees of freedom *d.f.* not in the Student's t table, use the closest *d.f.* that is *smaller*. In some situations, this choice of *d.f.* may increase the P-value by a small amount and therefore produce a slightly more "conservative" answer. Answers may vary due to rounding.

5. *Vehicles: Mileage* Based on information in *Statistical Abstract of the United States* (116th Edition), the average annual miles driven per vehicle in the United States is 11.1 thousand miles, with $\sigma \approx 600$ miles. Suppose that a random sample of 36 vehicles owned by residents of Chicago showed that the average mileage driven last year was 10.8 thousand miles. Does this indicate that the average miles driven per vehicle in Chicago is different from (higher or lower than) the national average? Use a 0.05 level of significance.

6. *Student Life: Employment* Professor Jennings claims that only 35% of the students at Flora College work while attending school. Dean Renata thinks that

the professor has underestimated the number of students with part-time or full-time jobs. A random sample of 81 students shows that 39 have jobs. Do the data indicate that more than 35% of the students have jobs? (Use a 5% level of significance.)

7. *Toys: Electric Trains* The Toylot Company makes an electric train with a motor that it claims will draw an average of only 0.8 ampere (A) under a normal load. A sample of nine motors was tested, and it was found that the mean current was $\bar{x} = 1.4$ A, with a sample standard deviation of $s = 0.41$ A. Do the data indicate that the Toylot claim of 0.8 A is too low? (Use a 1% level of significance.)

8. *Highways: Reflective Paint* The highway department is testing two types of reflecting paint for concrete bridge end pillars. The two kinds of paint are alike in every respect except that one is orange and the other is yellow. The orange paint is applied to 12 bridges, and the yellow paint is applied to 12 bridges. After a period of 1 year, reflectometer readings were made on all these bridge end pillars. (A higher reading means better visibility.) For the orange paint, the mean reflectometer reading was $\bar{x}_1 = 9.4$, with standard deviation $s_1 = 2.1$. For the yellow paint, the mean was $\bar{x}_2 = 6.9$, with standard deviation $s_2 = 2.0$. Based on these data, can we conclude that the yellow paint has less visibility after 1 year? (Use a 1% level of significance.)

9. *Medical: Plasma Compress* A hospital reported that the normal death rate for patients with extensive burns (more than 40% of skin area) has been significantly reduced by the use of new fluid plasma compresses. Before the new treatment, the mortality rate for extensively burned patients was about 60%. Using the new compresses, the hospital found that only 40 of 90 patients with extensive burns died. Use a 1% level of significance to test the claim that the mortality rate has dropped.

10. *Bus Lines: Schedules* A comparison is made between two bus lines to determine if arrival times of their regular buses from Denver to Durango are off schedule by the same amount of time. For 51 randomly selected runs, bus line A was observed to be off schedule an average time of 53 minutes, with standard deviation 19 minutes. For 60 randomly selected runs, bus line B was observed to be off schedule an average of 62 minutes, with standard deviation 15 minutes. Do the data indicate a significant difference in average off-schedule times? Use a 5% level of significance.

11. *Matches: Number per Box* The Nero Match Company sells matchboxes that are supposed to have an average of 40 matches per box, with $\sigma = 9$. A random sample of 94 Nero matchboxes shows the average number of matches per box to be 43.1. Using a 1% level of significance, can you say that the average number of matches per box is more than 40?

12. *Magazines: Subscriptions* A study is made of residents in Phoenix and its suburbs concerning the proportion of residents who subscribe to *Sporting News*. A random sample of 88 urban residents showed that 12 subscribed, and a random sample of 97 suburban residents showed that 18 subscribed. Does this indicate that a higher proportion of suburban residents subscribe to *Sporting News*? (Use a 5% level of significance.)

13. *Archaeology: Arrowheads* The Wind Mountain archaeological site is in southwest New Mexico. Prehistoric Native Americans called Anasazi once lived and hunted small game in this region. A stemmed projectile point is an arrowhead that has a notch on each side of the base. Both stemmed and stemless projectile points were found at the Wind Mountain site. A random sample of $n_1 = 55$ stemmed projectile points showed the mean length to be $\bar{x}_1 = 3.0$ cm, with sample standard deviation $s_1 = 0.8$ cm. Another random sample of $n_2 = 51$ stemless projectile

points showed the mean length to be $\bar{x}_2 = 2.7$ cm, with $s_2 = 0.9$ cm (Source: *Mimbres Mogollon Archaeology*, by A. I. Woosley and A. J. McIntyre, University of New Mexico Press). Do these data indicate a difference (either way) in the population mean length of the two types of projectile points? Use a 5% level of significance.

14. *Civil Service: College Degrees* The Congressional Budget Office reports that 36% of federal civilian employees have a bachelor's degree or higher (*The Wall Street Journal*). A random sample of 120 employees in the private sector showed that 33 have a bachelor's degree or higher. Does this indicate that the percentage of employees holding bachelor's degrees or higher in the private sector is less than that in the federal civilian sector? Use $\alpha = 0.05$.

15. *Vending Machines: Coffee* A machine in the student lounge dispenses coffee. The average cup of coffee is supposed to contain 7.0 ounces. Eight cups of coffee from this machine show the average content to be 7.3 ounces with a standard deviation of 0.5 ounce. Do you think that the machine has slipped out of adjustment and that the average amount of coffee per cup is different from 7 ounces? Use a 5% level of significance.

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

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

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

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

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

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

18. *Sports Car: Fuel Injection* The manufacturer of a sports car claims that the fuel injection system lasts 48 months before it needs to be replaced. A consumer group tests this claim by surveying a random sample of 10 owners who had the fuel injection system replaced. The ages of the cars at the time of replacement were (in months):

29   42   49   48   53   46   30   51   42   52

i.   Use your calculator to verify that the mean age of a car when the fuel injection system fails is $\bar{x} = 44.2$ months, with standard deviation $s \approx 8.61$ months.

ii. Test the claim that the fuel injection system lasts less than an average of 48 months before needing replacement. Use a 5% level of significance.

DATA HIGHLIGHTS: GROUP PROJECTS

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

1. "With Sampling, There Is Too a Free Lunch"—This is a headline that appeared in the *Wall Street Journal*. The article is about food product samples available at grocery stores. Giving out food samples is expensive and labor-intensive. It clogs supermarket aisles. It is risky. What if a customer tries an item and spits it out on the floor or says the product is awful? It creates litter. Some customers drop toothpicks or small paper cups on the floor or spill the product. However, the budget that companies are willing to spend to have their products sampled is growing. The director of communications for Bigg's "hypermarket" (a combination grocery and general-merchandise store) says that more than 60% of customers sample products and about 37% of those who sample buy the product.

   (a) Let's test the hypothesis that 60% of customers sample a particular product. What is the null hypothesis? Do you believe that the percentage of customers who sample products is less than, more than, or just different from 60%? What will you use for the alternate hypothesis?

   (b) Choose a level of significance α.

   (c) Go to a grocery store when special products are being sampled (not just the usual in-house store samples often available at the deli or bakery). Count the number of customers going by the display when a sample is available and the number of customers who try the sample. Be sure the number of customers n is large enough to use the normal distribution to approximate the binomial.

   (d) Using your sample data, conclude the hypothesis test. What is your conclusion?

   (e) Do you think different food products might have a higher or lower percentage of customers trying them? For instance, does a higher percentage of customers try samples of pizza than samples of yogurt? How could you use statistics to justify your answer?

   (f) Do you want to include young children in your sample? Do they pick up items to include in the customer's basket, or do they just munch the samples?

2. "Sweets May Not Be Culprit in Hyper Kids"—This is a *USA Today* headliner reporting results of a study that appeared in the *New England Journal of Medicine*. In this study, the subjects were 25 normal preschoolers, aged 3 to 5, and 23 kids, aged 6 to 10, who had been described as "sensitive to sugar." The kids and their families were put on three different diets for 3 weeks each. One

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

(a) This research involved comparing several means, not just two. (An introduction to such methods, called *analysis of variance*, is found in Chapter 10.) However, let us take a simplified view of the problem and consider the difference of behavior when children consumed the diet with sugar compared with their behavior when they consumed the diet with aspartame and low sugar. List some variables that might be measured to reflect the behavior of the children.

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

LINKING CONCEPTS: WRITING PROJECTS

Discuss each of the following topics in class or review the topics on your own. Then write a brief but complete essay in which you summarize the main points. Please include formulas and graphs as appropriate.

The most important questions in life usually cannot be answered with absolute certainty. Many important questions are answered by giving an estimate and a measure of confidence in the estimate. This was the focus of Chapter 7. However, sometimes important questions must be answered in a more straightforward manner by a simple yes or no. Hypothesis testing is the statistical process of answering questions with a straightforward yes or no *and* providing an estimate of the risk in accepting the answer.

1. Review and discuss type I and type II errors associated with hypothesis testing.

2. Review and discuss the level of significance and power of a statistical test.

3. The following statements are very important. Give them some careful thought and discuss them.

   (a) When we fail to reject the null hypothesis, we do not claim that it is absolutely true. We simply claim that at the given level of significance, the data were not sufficient to reject the null hypothesis.

   (b) When we accept the alternate hypothesis, we do not claim that the null hypothesis is absolutely false. We do claim that at the given level of significance, the data presented enough evidence to reject the null hypothesis.

4. In the text, it is said that a statistical test is a package of five basic ingredients. List these ingredients, discuss them in class, and write a short description of how these ingredients relate to the above discussion questions.

5. As access to computers becomes more and more prevalent, we see the *P*-value reported in hypothesis testing more frequently. Review the use of the *P*-value in hypothesis testing. What is the difference between the level of significance of a test and the *P*-value? Considering both the *P*-value and level of significance, under what conditions do we reject or fail to reject the null hypothesis?

Using Technology

Simulation

Recall that the level of significance α is the probability of mistakenly rejecting a true null hypothesis. If $\alpha = 0.05$, then we expect to mistakenly reject a true null hypothesis about 5% of the time. The following simulation conducted with Minitab demonstrates this phenomenon.

We draw 40 random samples of size 50 from a population that is normally distributed with mean $\mu = 30$ and standard deviation $\sigma = 2.5$. The display shows the results of a hypothesis test with

$$H_0: \mu = 30 \qquad H_1: \mu > 30$$

for each of the 40 samples labeled C1 through C40. Because each of the 40 samples is drawn from a population with mean $\mu = 30$, the null hypothesis $H_0: \mu = 30$ is true for the test based on each sample. However, as the display shows, for some samples we reject the true null hypothesis.

(a) How many of the 40 samples have a sample mean $\bar{x}$ above $\mu = 30$? below $\mu = 30$?

(b) Look at the P-value of the sample statistic $\bar{x}$ in each of the 40 samples. How many P-values are less than or equal to $\alpha = 0.05$? What percent of the P-values are less than or equal to α? What percent of the samples have us reject H_0 when, in fact, each of the samples was drawn from a normal distribution with $\mu = 30$, as hypothesized in the null hypothesis?

(c) If you have access to computer or calculator technology that creates random samples from a normal distribution with a specified mean and standard deviation, repeat this simulation. Do you expect to get the same results? Why or why not?

Minitab Display: Random samples of size 50 from a normal population with $\mu = 30$ and $\sigma = 2.5$

```
Z-Test
Test of mu = 30.000 vs. mu > 30.000
The assumed sigma = 2.50
Variable   N    Mean    StDev   SE Mean      Z      P
C1        50   30.002   2.776    0.354    0.01   0.50
C2        50   30.120   2.511    0.354    0.34   0.37
C3        50   30.032   2.721    0.354    0.09   0.46
C4        50   30.504   2.138    0.354    1.43   0.077
C5        50   29.901   2.496    0.354   -0.28   0.61
C6        50   30.059   2.836    0.354    0.17   0.43
C7        50   30.443   2.519    0.354    1.25   0.11
C8        50   29.775   2.530    0.354   -0.64   0.74
C9        50   30.188   2.204    0.354    0.53   0.30
C10       50   29.907   2.302    0.354   -0.26   0.60
C11       50   30.036   2.762    0.354    0.10   0.46
C12       50   30.656   2.399    0.354    1.86   0.032
C13       50   30.158   2.884    0.354    0.45   0.33
C14       50   29.830   3.129    0.354   -0.48   0.68
C15       50   30.308   2.241    0.354    0.87   0.19
C16       50   29.751   2.165    0.354   -0.70   0.76
C17       50   29.833   2.358    0.354   -0.47   0.68
C18       50   29.741   2.836    0.354   -0.73   0.77
C19       50   30.441   2.194    0.354    1.25   0.11
C20       50   29.820   2.156    0.354   -0.51   0.69
C21       50   29.611   2.360    0.354   -1.10   0.86
C22       50   30.569   2.659    0.354    1.61   0.054
C23       50   30.294   2.302    0.354    0.83   0.20
C24       50   29.978   2.298    0.354   -0.06   0.53
C25       50   29.836   2.438    0.354   -0.46   0.68
C26       50   30.102   2.322    0.354    0.29   0.39
C27       50   30.066   2.266    0.354    0.19   0.43
C28       50   29.071   2.219    0.354   -2.63   1.00
C29       50   30.597   2.426    0.354    1.69   0.046
C30       50   30.092   2.296    0.354    0.26   0.40
C31       50   29.803   2.495    0.354   -0.56   0.71
C32       50   29.546   2.335    0.354   -1.28   0.90
C33       50   29.702   1.902    0.354   -0.84   0.80
C34       50   29.233   2.657    0.354   -2.17   0.98
C35       50   30.097   2.472    0.354    0.28   0.39
C36       50   29.733   2.588    0.354   -0.76   0.78
C37       50   30.379   2.976    0.354    1.07   0.14
C38       50   29.424   2.827    0.354   -1.63   0.95
C39       50   30.288   2.396    0.354    0.81   0.21
C40       50   30.195   3.051    0.354    0.55   0.29
```

Technology Hints

Press **STAT** and select **EDIT.** Highlight the list name, such as L1. Then press **MATH,** select **PRB,** and highlight **6:randNorm(μ, σ, sample size**). Press enter. Fill in the values of $\mu = 30$, $\sigma = 2.5$, and sample size $= 50$. Press enter. Now list L1 contains a random sample from the normal distribution specified.

To test the hypothesis H_0: $\mu = 30$ against H_1: $\mu > 30$, press **STAT,** select **TESTS,** and use option **1:Z-Test.** Fill in the value 30 for μ_0, 2.5 for σ, and $> \mu_0$. The output provides the value of the sample statistic $\bar{x}$, its corresponding z value, and the P-value of the sample statistic.

Excel 2007

To draw random samples from a normal distribution with $\mu = 30$ and $\sigma = 2.5$, click the **Data** tab on the home ribbon and select **Data Analysis** in the Analysis Group. In the dialogue box, select **Random Number Generator.** In the first dialogue box, select **Normal** for the Distribution. In the resulting dialogue box shown below, the number of variables is the number of samples. Fill in the rest of the dialogue box as shown.

Random Number Generation		? ✕
Number of Variables:	40	OK
Number of Random Numbers:	50	Cancel
Distribution:	Normal ▼	Help
Parameters		
Mean =	30	
Standard Deviation =	2.5	
Random Seed:		
Output options		
○ Output Range:		
● New Worksheet Ply:		
○ New Workbook		

To conduct a hypothesis test of H_0: $\mu = 30$ against H_1: $\mu > 30$, click the **Insert Function** $\boxed{f_x}$ on the home screen. For the category, select **Statistical** and then select the function **ZTEST.** Fill in the dialogue box with the array containing the random numbers, 30 for X, the value in H_0, and 2.5 for sigma.

Minitab

To generate random samples from a normal distribution, use the menu choices **Calc ➤ Random Data ➤ Normal.** In the dialogue box, the number of rows refers to the sample size. Use 50 rows. Then designate the columns for the samples. Using C1–C40 will generate 40 random samples and put the samples in columns C1 through C40.

To test the hypothesis H_0: $\mu = 30$ against H_1: $\mu > 30$, use the menu choices **Stat ➤ Basic Statistics ➤ 1-SampleZ.** Use columns C1–C40 as the variables. Fill in 30 for the test mean, use "greater than" for the alternate hypothesis, and use 2.5 for sigma.

SPSS

SPSS uses a Student's t distribution to test the mean and difference of means. SPSS uses the sample standard deviation s even if the population σ is known. Use the menu choices **Analyze ➤ Compare Means** and then **One-Sample T Test** or **Independent-Sample T Tests** for tests of a single mean or a difference of means, respectively. In the dialogue box, fill in the test value of the null hypothesis.

To generate 40 random samples of size $n = 50$ from a normal distribution with $\mu = 30$ and $\sigma = 2.5$, first enter consecutive integers from 1 to 50 in a column of the data editor. Then, under variable view, enter the variable names Sample1 through Sample40. Use the menu choices **Transform ➤ Compute Variable.** In the dialogue box, use Sample1 for the target variable, select **Random Numbers** for the Function Group and then select the function **Rv.Normal(mean, stddev).** Use 30 for the mean and 2.5 for the standard deviation. Continue until you have 40 samples. To sample from other distributions, use appropriate functions in the Compute dialogue box.

The SPSS display shows the test results ($H_0: \mu = 30$; $H_1: \mu \neq 30$) for a sample of size $n = 50$ drawn from a normal distribution with $\mu = 30$ and $\sigma = 2.5$. The P-value is given as the significance for a two-tailed test.

For a one-tailed test, divide the significance by 2. In the display, the significance is 0.360 for a two-tailed test. So, for a one-tailed test, the P-value is 0.360/2, or 0.180.

SPSS Display

T-Test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
SAMPLE1	50	30.3228	2.46936	.34922

One-Sample Test

	Test Value = 30					
					90% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
SAMPLE1	.924	49	.360	.3228	−.2627	.9083