9

Hypothesis Testing

Charles Lutwidge Dodgson (1832–1898)

Using the pseudonym Lewis Carroll, this English mathematician and author wrote *Alice's Adventures in Wonderland.*

For on-line student resources, visit **math.college.hmco.com/students** and follow the Statistics links to the Brase/Brase, *Understanding Basic Statistics,* 4th edition web site.

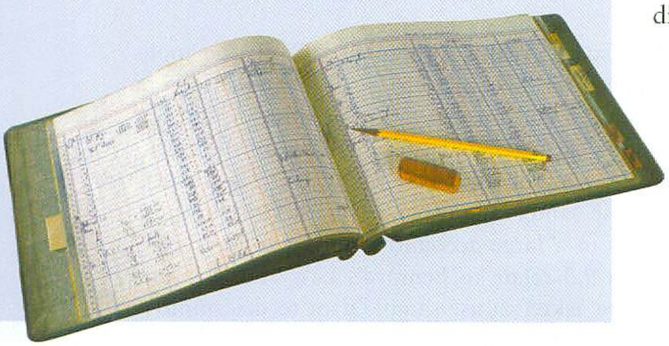
"Would you tell me, please, which way I ought to go from here?"
"That depends a good deal on where you want to get to," said the Cat.
"I don't much care where—" said Alice.
"Then it doesn't matter which way you go," said the Cat.

—Lewis Carroll
Alice's Adventures in Wonderland

Charles Dodgson was an English mathematician who loved to write children's stories in his free time. The dialogue above 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 has been made on other grounds, they can help us find the best way to get there.

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 9.1)
- What is the *P*-value of a statistical test? What does this measurement have to do with performance reliability? (SECTION 9.1)
- How do you construct statistical tests for μ? Does it make a difference whether σ is known or unknown? (SECTION 9.2)
- How do you construct statistical tests for the proportion p of successes in a binomial experiment? (SECTION 9.3)



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 occurring.

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 following probabilities of occurrence.

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) did 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 9.3 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 1 and 2 of Section 9.3 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 the 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 lowering the numbers, perhaps to hide profits from the Internal Revenue Service or to divert company profits to unscrupulous employees. (See Problems 1 and 2 of Section 9.3.)

9.1 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 parameter. When we estimate the value (or location) of a parameter, we are using methods of estimation such as those studied in Chapter 8. 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

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 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 times for a muffler installation.

Alternate hypothesis

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 one 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 greater 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 where 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 $\mu = 47$ can be rejected. Therefore, our null hypothesis is simply that $\mu = 47$. We denote the null hypothesis as

$$H_0: \mu = 47$$

(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, we are sure it is less than 47. Therefore, the alternate hypothesis is

$$H_1: \mu < 47$$

◆

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 statement 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: $\mu = 6.0$.

(b) What statement 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: $\mu \neq 6.0$ (μ is either smaller than or larger than 6.0).

◇ **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 the opposite of the symbol used 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 if H_1 states that the parameter is less than the value claimed in H_0

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

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

TABLE 9-1 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	You believe that μ is less than value stated in H_0.	You believe that μ is more than value stated in H_0.	You believe that μ is different from value stated in H_0.
H_0: $\mu = k$	H_1: $\mu < k$	H_1: $\mu > k$	H_1: $\mu \neq k$
	Left-tailed test	Right-tailed test	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 9-1 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 7.4 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*.

Test statistic for μ, given x normal and σ known

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

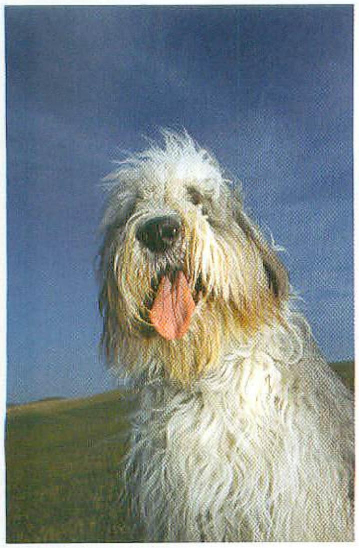
In this section, we use hypothesis tests about the mean to introduce the concepts and vocabulary of hypothesis testing. In particular, let's suppose that x has *a normal distribution* with mean μ and standard deviation σ. Then, Theorem 7.1 tells us that $\bar{x}$ has a *normal distribution* with mean μ and standard deviation $\sigma/\sqrt{n}$.

Given that x has a *normal distribution* with known standard deviation σ, then

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

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 check-ups 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

$$H_0: \mu = 115$$

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

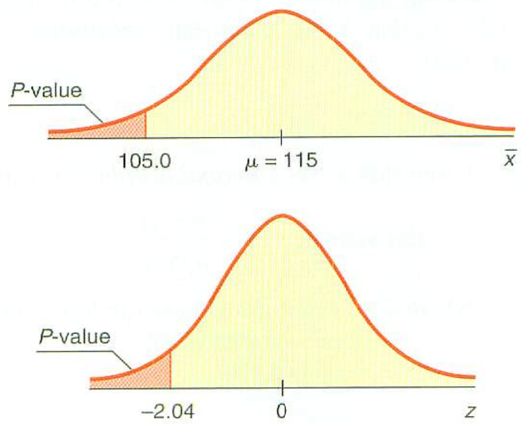
$$H_1: \mu < 115$$

(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, you 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, we accept H_1: $\mu < 115$ and conclude that Rosie's heart rate is slowing.

(c) How do you 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 9-1 shows the $\bar{x}$ distribution and the corresponding standard normal distribution with the desired probability shaded.

FIGURE 9-1

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



Since x has a normal distribution, $\bar{x}$ will also have a normal distribution for any sample size n and given σ (see Theorem 7.1). 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

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

P-value

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.

(d) What conclusion can be drawn about Rosie's average heart rate?
If H_0: $\mu = 115$ is in fact true, then 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

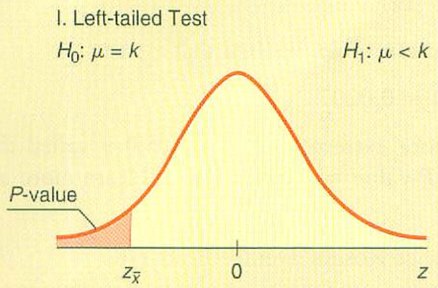
Assuming H_0 is true, the *probability* that the test statistic will take on values as extreme as or more extreme than the observed test statistic (computed from sample data) is called the ***P*-value** of the test. The smaller the P-value computed from sample data, the stronger the evidence against H_0.

The P-value is sometimes called the *probability of chance*. The P-value 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 results (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.

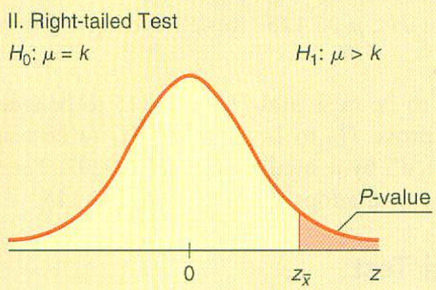
The P-value associated with the observed test statistic takes on different values depending on the alternate hypothesis and the type of test. Let's look at P-values and types of tests when the test involves the mean and standard normal distribution. Notice that in Example 2, part (c), we computed a P-value 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

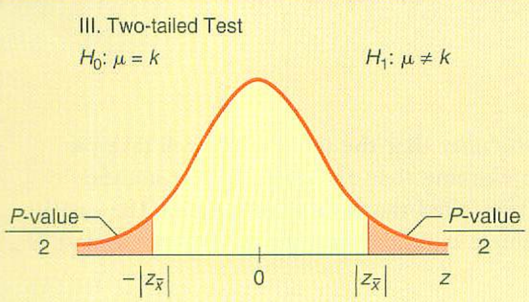
Let $z_{\bar{x}}$ represent the standardized sample test statistic for testing a mean μ using the standard normal distribution. That is, $z_{\bar{x}} = (\bar{x} - \mu)/(\sigma/\sqrt{n})$.



P-value $= P(z < z_{\bar{x}})$
This is the probability of getting a test statistic as low as or lower than $z_{\bar{x}}$.



P-value $= P(z > z_{\bar{x}})$
This is the probability of getting a test statistic as high as or higher than $z_{\bar{x}}$.



$\frac{P\text{-value}}{2} = P(z > |z_{\bar{x}}|)$; 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 9-2 indicates how these errors occur.

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

TABLE 9-2 Type I and Type II Errors

	Our Decision	
Truth of H_0	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

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.

Level of significance

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").

> The **level of significance** $\boldsymbol{\alpha}$ is the probability of rejecting H_0 when it is true. This is the probability of a type I error.

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

The quantity $1 - \beta$ is called the *power of the 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 reject the null hypothesis when it is false.

2. Using a larger value of α will increase the power of a test, 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 9-3 on the next page 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. ◇

TABLE 9-3 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
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*
H_0 is false	Type II error, with corresponding probability β	Correct decision, with corresponding probability $1 - \beta$, called the *power of the test*

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 consequences, 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 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 bearings is either too large or too small to meet specifications. Such an error would mean that the production process would not be adjusted when 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 the 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

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

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

Statistical significance

In what sense are we using the word "significant"? *Webster's Dictionary* gives two interpretations of *significance:* (1) having or signifying *meaning;* or (2) 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 housing developments are raising concern that the concentration may have changed because of increased use 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 8 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) 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, use the standard normal distribution with

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{x} - 2.3}{\frac{0.3}{\sqrt{8}}}$$

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

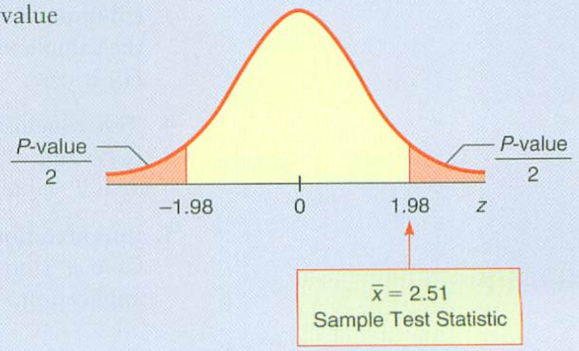
The sample of 8 measurements has mean $\bar{x} = 2.51$. Converting this measurement to z, we have

$$\text{test statistic} = z = \frac{2.51 - 2.3}{\frac{0.3}{\sqrt{8}}} \approx 1.98$$

(e) Draw a sketch showing the P-value area on the standard normal distribution. Find the P-value. *Hint:* Use Table 3 of the Appendix. Because this is a two-tailed test, we double the area in the right tail beyond $z = 1.98$.

P-value $= 2P(z > 1.98) = 2(0.0239) = 0.0478$

FIGURE 9-2 P-value



Continued

GUIDED EXERCISE 3 continued

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

Since P-value $0.0478 > 0.01$, we see that

$P\text{-value} > \alpha$

We fail to reject H_0.

(g) Summarize 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.

Meaning of accepting H_0

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. 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. When we *fail to reject* the null hypothesis, it simply means that the evidence in favor of rejection is not strong enough (see Table 9-4). 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.

Some comments about P-values and level of significance α should be made. The level of significance α should be a fixed, pre-specified 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 H_0 is in fact true.

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 $P\text{-value} \le \alpha$, then reject H_0, and if $P\text{-value} > \alpha$, then do not reject H_0.

TABLE 9-4 Meaning of the Terms *Fail to Reject* H_0 and *Reject* H_0

Term	Meaning
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	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.

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 indicated 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 in the *left* ear. Professor Sim of Sam Houston State University (Huntsville, Texas) found that emotionally loaded words had 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 9.1 PROBLEMS

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

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

5. *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 that 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 that 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 that 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.

6. *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 time in a large national houseware 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 that 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 that 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 would do the trick. 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 houseware store.
   (c) What would you use for the null and alternate hypotheses if you believe that 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 that the average shopping time is different from 4.5 minutes? Is this a right-tailed, left-tailed, or two-tailed test?

7. *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 that you do not know if the storm is getting worse or dying out. You just want to test the hypothesis that the average wave height is *different* from (either higher or lower than) 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.

8. *Chrysler Concorde: Acceleration* *Consumer Reports* stated that the mean time for a Chrysler Concorde to go from 0 to 60 miles per hour was 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 that the average time to accelerate from 0 to 60 miles per hour has been 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 9–14, 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) What sampling distribution will you use? Explain the rationale for your choice of sampling distribution. 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) State your conclusion in the context of the application.

9. *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 $\alpha = 0.01$.

10. *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 $\alpha = 0.05$.

11. *Ecology: Hummingbirds* Bill Alther is a zoologist who studies Anna's hummingbird (*Calypte anna*). (Reference: *Hummingbirds*, K. Long, 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 gm) were

3.7	2.9	3.8	4.2	4.8	3.1

The sample mean is $\bar{x} = 3.75$ gm. 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 $\sigma = 0.70$ gm. It is known that for the population of all Anna's hummingbirds, the mean weight is $\mu = 4.55$ gm. Do the data indicate that the mean weight of these birds in this part of the Grand Canyon is less than 4.55 gm? Use $\alpha = 0.01$.

12. *Finance: P/E of Stocks* The price to earnings ratio (P/E) 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 mean of 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 $\alpha = 0.05$.

13. *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 crops lost to hail in Weld County is different (either way) from the national mean of 11%? Use $\alpha = 0.01$.

14. *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,* 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.

9.2 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 9.1 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 $\bar{x}$: the standard normal distribution or a Student's *t* distribution. As discussed in Chapters 7 and 8, 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

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).

1. In the context of the application, state the *null and alternate hypotheses* and set the *level of significance* α.
2. 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 $n \geq 30$. 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*.

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

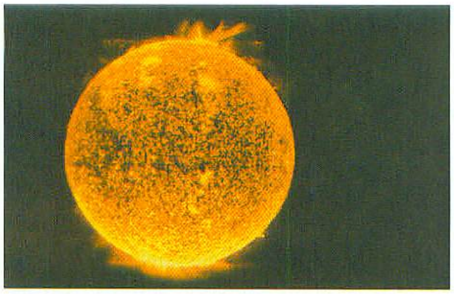
3. Use the standard normal distribution and the type of test, one-tailed or two-tailed, to find the *P-value* corresponding to the test statistic.
4. *Conclude* the test. If P-value $\leq \alpha$, then reject H_0. If P-value $> \alpha$, then do not reject H_0.
5. *State your conclusion* in the context of the application.

In Section 9.1, we examined P-value tests for normal distributions with relatively small sample size ($n < 30$). 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 number of sunspots observed in a 4-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 4-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 $\mu = 41$,

$$H_0: \mu = 41 \qquad \text{and} \qquad H_1: \mu > 41$$

(b) 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 $n = 40$,

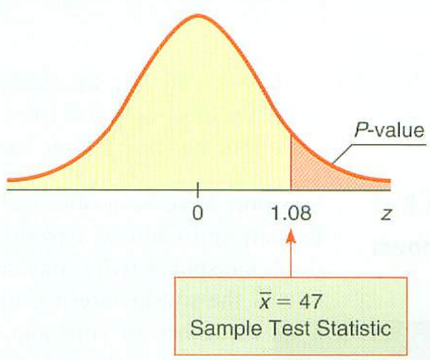
$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \approx \frac{47 - 41}{35/\sqrt{40}} \approx 1.08$$

(c) Find the P-value of the test statistic.
Figure 9-3 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 9-3. Using Table 3 of the Appendix, we find that

$$P\text{-value} = P(z > 1.08) \approx 0.1401$$

FIGURE 9-3

P-value Area



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

(e) Interpret the results.
At the 5% level of significance, the evidence is not sufficient to reject H_0. Based on the sample data, we do not think that 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

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.

1. In the context of the application, state the *null and alternate hypotheses* and set the *level of significance* α.
2. If you can assume that x 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 $n \geq 30$. Use $\bar{x}$, s, and n from the sample with μ from H_0 to compute the sample *test statistic.*

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

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. *State your conclusion* in the context of the application.

Using the Student's t table to estimate P-values

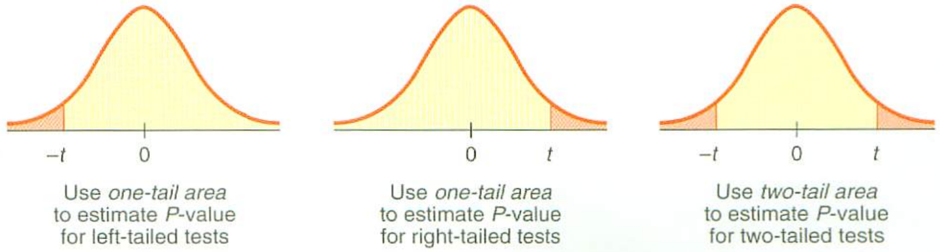
In Section 8.2, we used Table 4 of the Appendix, Student's t Distribution, to find critical values t_c for confidence intervals. The critical values are in the body of the table. We find P-values in the *rows* headed by "one-tail area" and "two-tail area," depending on whether we have a one-tailed or 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 4, 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 9-4.)

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

FIGURE 9-4

P-value for One-Tailed Tests and 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,

$$H_0: \mu = 12.5 \text{ weeks} \quad \text{and} \quad H_1: \mu \neq 12.5 \text{ weeks}$$

(b) Compute the test statistic from the sample data.
Since the x distribution is assumed to be mound-shaped and symmetric, we use the Student's t distribution. 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 9-5 shows the P-value. Using Table 4 of the Appendix, 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 9-5, Excerpt from Table 4.)

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

FIGURE 9-5

P-value

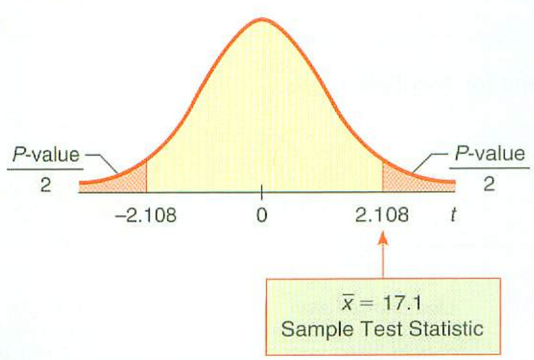


TABLE 9-5 Excerpt from Student's *t* Distribution (Table 4, Appendix)

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 4 of the Appendix does not give that specific value, but it does give a range that contains the specific P-value. As the diagram shows, the entire range is greater than α. This means that the specific P-value is greater than α, so we cannot reject H_0.

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) Interpret the results.

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									

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) What sampling distribution should you use? What is the t value of the sample test statistic? ➡ 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 $n = 61$,

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

Continued

GUIDED EXERCISE 4 continued

(c) When you use Table 4 of the Appendix 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 for the P-value.

This is a right-tailed test, so use a one-tail area (see Figure 9-6).

FIGURE 9-6 P-value

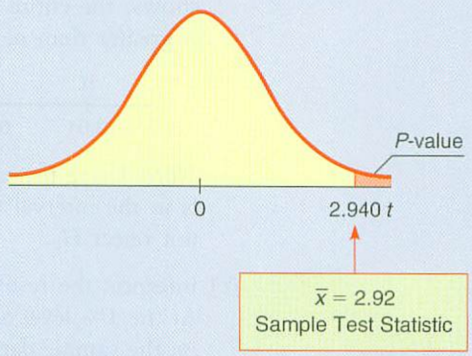


TABLE 9-6 Excerpt from Student's t Table

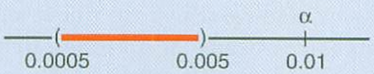
✓ one-tail area	...0.005	0.0005
two-tail area	...0.010	0.0010
$d.f.$ = 60	...2.660	3.460
	↑ Sample t = 2.940	

Using $d.f. = 60$, we find that the sample $t = 2.940$ is between the critical values 2.660 and 3.460 (see Table 9-6). 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 $P\text{-value} \approx 0.0022$. This value is in our estimated range and is less than $\alpha = 0.01$, so we reject H_0.

(e) 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 to conclude that the average projectile point length at the adjacent cliff dwelling site is longer than 2.6 cm.

TECH NOTE The TI-84Plus and TI-83Plus calculators, Excel, and Minitab all support testing of μ using the standard normal distribution. The TI-84Plus/TI-83Plus 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 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: $\mu = \mu_0$. 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.

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

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

Excel In Excel, the ZTEST function finds the P-values for a right-tailed test. (*Note:* Ignore the Excel documentation that mistakenly says ZTEST gives the P-value for a two-tailed test.) Use the menu choice **Paste Function** f_x ➤ **ZTEST.** In the 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)

Critical region 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.*

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

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

The values of $\bar{x}$ for which we reject H_0 are called the *critical region* of the $\bar{x}$ distribution. Depending on the alternate hypothesis, the critical region is located on the left side, the right side, or both sides of the $\bar{x}$ distribution. Figure 9-7 on the next page shows the relationship of the critical region to the alternate hypothesis and to 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.

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 9-8. *Critical values* are the boundaries of the critical region.

FIGURE 9-7

Critical Regions for H_0: $\mu = k$

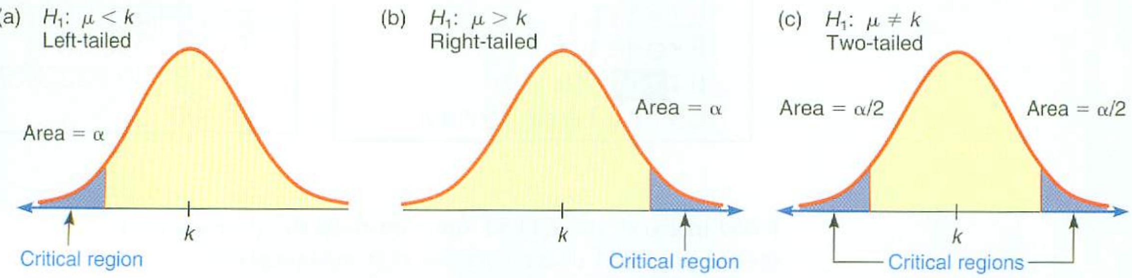
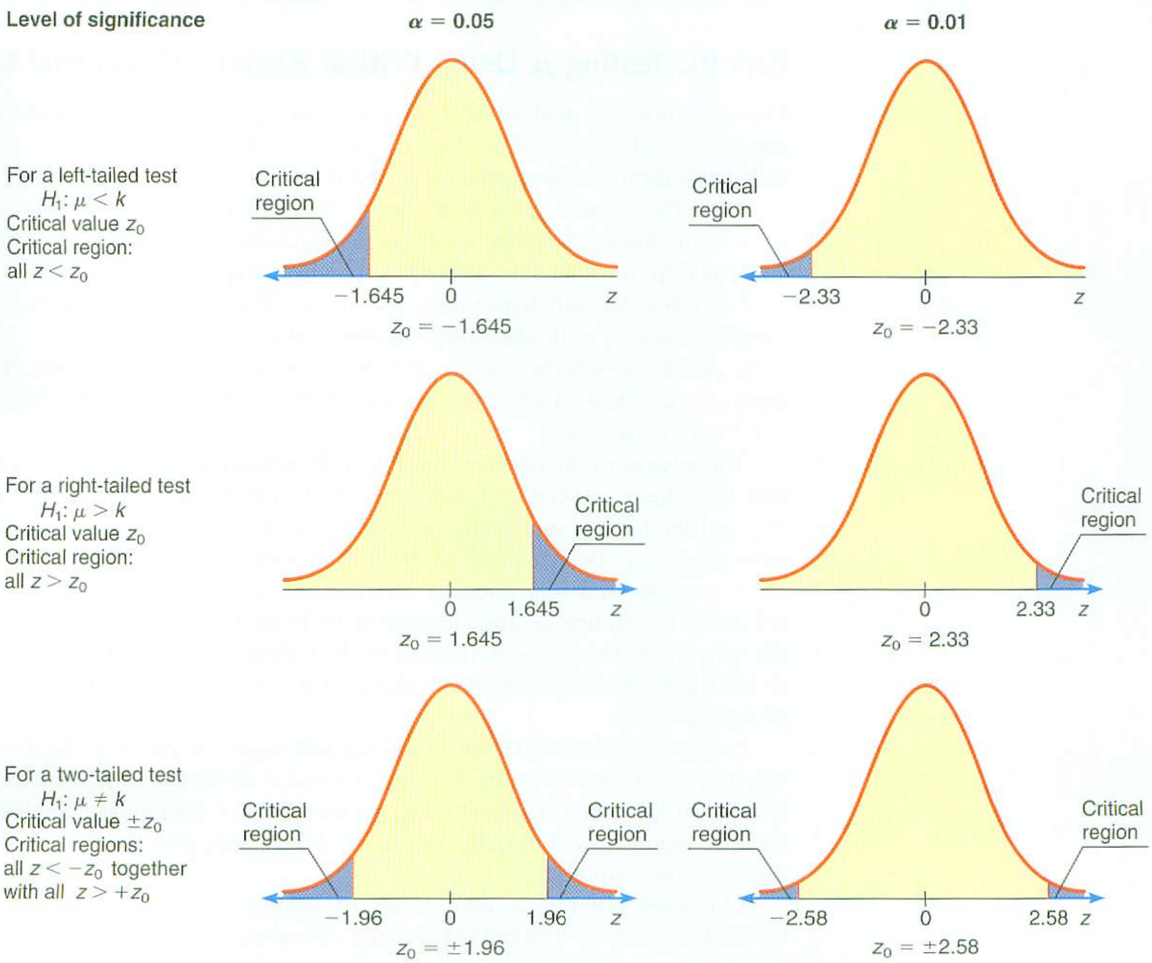


FIGURE 9-8

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



Critical values designated as z_0 for the standard normal distribution are shown in Figure 9-8. For easy reference, they are also included in Table 3(c) of the Appendix, 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 whether the sample test statistic falls in the critical region. If it does, we reject H_0. Otherwise, we do not reject H_0.

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).

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. 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 $n \geq 30$. 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. *State 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 4-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 4-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 9-8, the critical value is $z_0 = 1.645$.

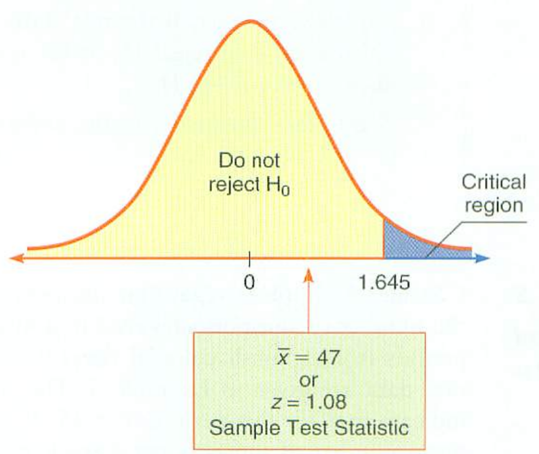
(d) Conclude the test.
We conclude the test by showing the critical region, critical value, and sample test statistic $z = 1.08$ on the standard normal curve. For a right-tailed test with $\alpha = 0.05$, the critical value is $z_0 = 1.645$. Figure 9-9 shows the critical region. As we can see, the sample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

(e) Interpret the results.
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 with the results of the P-value method?
The results, as expected, are the same. In both cases we fail to reject H_0.

FIGURE 9-9

Critical Region, $\alpha = 0.05$



◆

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 output. For example, Table 4 of the Appendix provides critical values for Student 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.
   (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 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, visit the web site **http://lib.stat.cmu.edu/** and follow the links to Datasets and then Sleep.

SECTION 9.2 PROBLEMS

Please provide the following information for Problems 1–16.

(a) What is the level of significance? State the null and alternate hypotheses.

(b) What sampling distribution will you use? Explain the rationale for your choice of sampling distribution. 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) State 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.

1. *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 $\alpha = 0.01$.

2. *Ford Taurus: Assembly Time* Let x be a random variable that represents assembly time for the Ford Taurus. *The Wall Street Journal* reported that the average assembly time is $\mu = 38$ hours. A modification to the assembly procedure has been made. Experience with this new method indicates that $\sigma = 1.2$ hours. It is thought that the average assembly time may be reduced by this modification. A random sample of 47 new Ford Taurus automobiles coming off the assembly line showed the average assembly time using the new method to be $\bar{x} = 37.5$ hours. Does this indicate that the average assembly time has been reduced? Use $\alpha = 0.01$.

3. *E-mails: Priority Lists* Message mania! A professional employee in a large corporation receives an average of $\mu = 41.7$ e-mails per day. Most of these e-mails are from other employees in the company. Because of the large number of e-mails, employees find themselves distracted and are unable to concentrate when they return to their tasks (Reference: *The Wall Street Journal*). In an effort to reduce distraction caused by such interruptions, one company established a priority list that all employees were to use before sending an e-mail. One month after the new priority list was put into place, a random sample of 45 employees showed that they were receiving an average of $\bar{x} = 36.2$ e-mails per day. The computer server through which the e-mails are routed showed that $\sigma = 18.5$. Has the new policy had any effect? Use a 5% level of significance to test the claim that there has been a change (either way) in the average number of e-mails received per day per employee.

4. *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 the average pH to be $\bar{x} = 8.1$

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.

5. *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 $s = 0.82$ year (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 $\alpha = 0.01$.

6. *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 for 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 $\alpha = 0.05$.

7. *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 that 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 $\alpha = 0.05$.

8. *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 $\alpha = 0.01$.

9. *Medical: Red Blood Cell Count* Let x be a random variable that represents red blood cell count (RBC) 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 $\alpha = 0.05$.

10. *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 9). 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 $\alpha = 0.01$.

11. *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 had an average thickness of $\mu = 67$ cm (Source: *Avalanche Handbook*, by D. McClung and P. Schaerer). The ski patrol at Vail, Colorado, is studying slab avalanches in their 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 $\bar{x} \approx 61.8$ cm 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.

12. *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.

13. *Veterinary Science: Lions* The heart rate of a healthy lion is approximately normally distributed, with mean $\mu = 40$ beats per minute (Source: *The Merck Veterinary Manual*, a reference used in most veterinary colleges). A heart rate that is too slow or too fast can indicate a health problem. A veterinarian has removed an abscessed tooth from a young, healthy zoo lion. As the animal slowly starts to come out of the anesthetic, its heart rate (in beats per minute) is taken and recorded for half an hour:

30	37	43	38	35	36

i. Use a calculator with mean and standard deviation keys to verify that $\bar{x} = 36.5$ and $s \approx 4.2$.

ii. Use a 5% level of significance to test the claim that the population average heart rate of the lion is different (either way) from 40 beats per minute.

14. *Shopping Time: Housewares* How much customers buy is a direct result of how much time they spend in a store. The mean shopping time for a woman accompanied by children in national houseware stores is 7.3 minutes (Source: *Why We Buy: The Science of Shopping* by P. Underhill). A retail research team is studying shopping habits in the Cherry Creek Mall (Denver). A random sample of women shoppers with children in a large houseware store gave the following shopping times (in minutes):

7.7	8.1	8.2	9.0	5.8	9.3	8.4	6.9	12.1	9.4
8.1	6.2	7.3	7.9	8.2	8.5	7.2	6.3	9.1	8.8

i. Use a calculator with mean and standard deviation keys to verify that $\bar{x} \approx 8.1$ min and $s \approx 1.4$ min.
ii. Assume shopping time follows an approximately normal distribution. Use a 5% level of significance to test the claim that the average shopping time for women with children in the Cherry Creek Mall is higher than the national average for this type of store.

15. *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 $\bar{x} \approx 7.36$ and sample standard deviation $s \approx 4.03$.
ii. Assuming that 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.

16. *Archaeology: Tree Rings* Tree-ring dating at archaeological excavation sites is used in conjunction with other chronologic evidence to estimate occupation dates of prehistoric Indian dwellings 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 that 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 for 1300 A.D.? Use a 1% level of significance.

17. *General: 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.

18. *General: Comparing Hypothesis Tests to U.S. Courtroom System* Compare similarities of 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. 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.

19. *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, $\mu_1 - \mu_2$, or $p_1 - p_2$, which we will study in Sections 9.3, 10.2, and 10.3.) 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 $\bar{x} = 22$ from a population with standard deviation $\sigma = 4$.

(a) What is the value of $c = 1 - \alpha$? Using the methods of Chapter 8, 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 Chapter 9, 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).

20. *Confidence Intervals and Two-Tailed Hypothesis Tests* Change the null hypothesis of Problem 19 to H_0: $\mu = 21$. Repeat parts (a) and (b).

21. *Critical Region Method: Standard Normal* Solve Problem 1 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?

22. *Critical Region Method: Standard Normal* Solve Problem 2 using the critical region method of testing. Compare your conclusion with the conclusion obtained by using the P-value method. Are they the same?

23. *Critical Region Method: Standard Normal* Solve Problem 3 using the critical region method of testing. Compare your conclusion with the conclusion obtained by using the P-value method. Are they the same?

24. *Critical Region Method: Student's t* Table 4 of the Appendix 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 4 using the critical region method of testing. Compare your conclusion with the conclusion obtained by using the P-value method. Are they the same?

25. *Critical Region Method: Student's t* Solve Problem 5 using the critical region method of testing. *Hint:* See Problem 24. Compare your conclusion with the conclusion obtained by using the P-value method. Are they the same?

26. *Critical Region Method: Student's t* Solve Problem 6 using the critical region method of testing. *Hint:* See Problem 24. Compare your conclusion with the conclusion obtained by using the P-value method. Are they the same?

9.3 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 9.1 and 9.2. 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 binomial trial, see Section 6.2). We will use $\hat{p} = r/n$ as our estimate for p, the population probability of success on each trial. The letter q 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 $nq > 5$).

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

$$\mu = p$$

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

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

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 for testing a proportion is similar to tests you have conducted before.

PROCEDURE

How to test a proportion p

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.

1. In the context of the application, state the *null and alternate hypotheses* and set the *level of significance* α.
2. 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, the $\hat{p} = r/n$ distribution can be approximated by the normal distribution using the standardized sample *test statistic*

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

   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 $\le \alpha$, then reject H_0. If P-value $> \alpha$, then do not reject H_0.
5. *State 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 (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,

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

(b) 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 $nq = 225(0.7) = 157.5$ is also greater than 5, so we can use the normal distribution for the sample statistic $\hat{p}$.

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

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

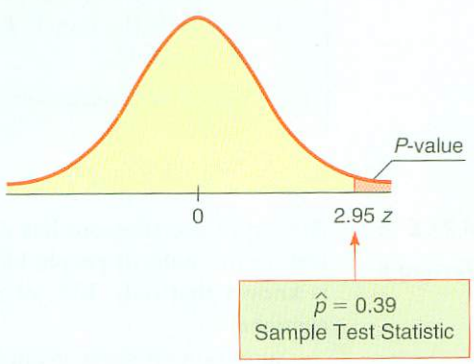
$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} \approx \frac{0.39 - 0.30}{\sqrt{\frac{0.30(0.70)}{225}}} \approx 2.95$$

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 9-10 shows the P-value. Since we have a right-tailed test, the P-value is the area to the right of $z = 2.95$ shown in Figure 9-10. Using the normal distribution (Table 3 of the Appendix), we find that P-value $= P(z > 2.95) \approx 0.0016$.

FIGURE 9-10

P-value Area



(d) Conclude the test.
Since the P-value of $0.0016 \leq 0.01$ for α, we reject H_0.

(e) Interpret the results.
At the 1% level of significance, the sample 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 germinated. 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) Calculate the sample test statistic $\hat{p}$. Using the value of p in H_0, are both $np > 5$ and $nq > 5$? Can we use the normal distribution for $\hat{p}$?

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

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

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 $\hat{p}$.

(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

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

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.

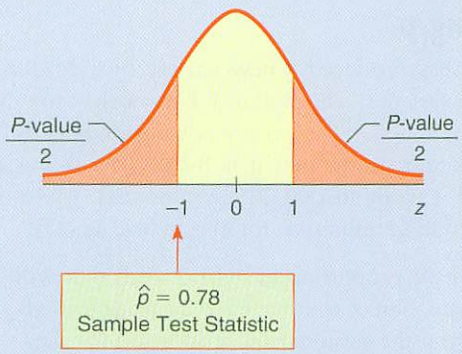
→ For a two-tailed test, using the normal distribution (Table 3(a) of the Appendix), we find that

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

Continued

GUIDED EXERCISE 5 continued

FIGURE 9-11 P-value



(e) Do we reject or fail to reject H_0? State your conclusion in the context of the application.

Since

P-value of $0.3174 > 0.05$ for α

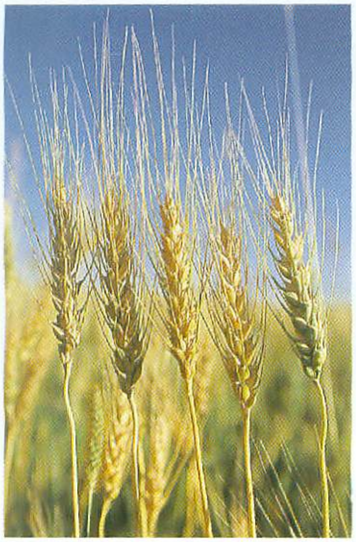
we fail to reject H_0. At the 5% level of significance, there is insufficient evidence to say that the botanist is wrong.

Critical region method

Since the $\hat{p}$ sampling distribution is approximately normal, we use Table 3 of the Appendix, Areas of a Standard Normal Distribution, 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 was 80%. Use a 5% level of significance to test the claim that the population proportion of germinating seeds for 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 the sample statistic $\hat{p}$ and the corresponding test statistic z. This was done in Guided Exercise 5, where we found that $\hat{p} = 0.78$ with corresponding $z = -1.00$.

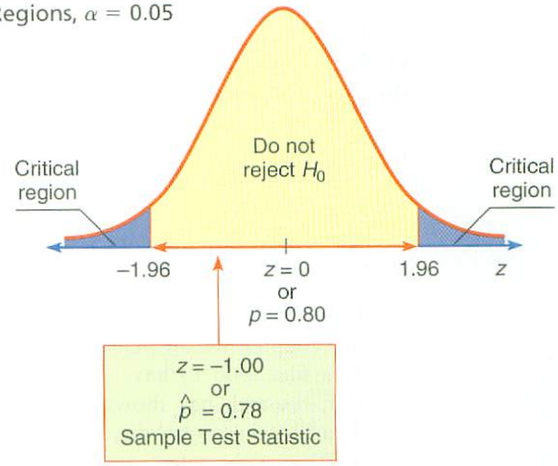
(b) Now we find the critical value z_0 for a two-tailed test using $\alpha = 0.05$. This means that we want the total area 0.05 divided between two tails, one to the right of z_0 and one to the left of $-z_0$. As shown in Figure 9-8 of Section 9.2, the critical value(s) are ± 1.96. (See also Table 3, part (c), of the Appendix for critical values of the z distribution.)

(c) Figure 9-12 shows the critical regions and the location of the sample test statistic.

(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*

FIGURE 9-12

Critical Regions, $\alpha = 0.05$



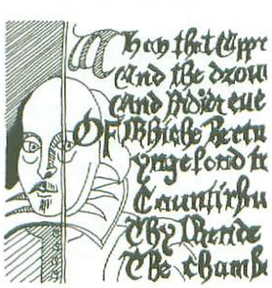
enough to reject the botanist's claim. This result, as expected, is consistent with the conclusion obtained by using the P-value method. ◆

TECH NOTE The TI-84Plus/TI-83Plus 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 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.

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 9 and 10 of this exercise set.

SECTION 9.3 PROBLEMS

Please provide the following information for Problems 1–17.

(a) What is the level of significance? State the null and alternate hypotheses.

(b) What sampling distribution will you use? Do you think that the sample size is sufficiently large? Explain. What is 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) State your conclusion in the context of the application.

1. *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 revenue 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?

2. *Focus Problem: Benford's Law* Again suppose that you are the auditor for a very large corporation. The revenue file contains millions of numbers in a large computer data bank (see Problem 1). 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 revenue 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 that 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?

3. *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 that 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%.

4. *College Athletics: Graduation Rate* Women athletes at the University of Colorado, Boulder, have a long-term graduation rate of 67% (Source: *The 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.

5. *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 $\alpha = 0.01$.

6. *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 was 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 $\alpha = 0.05$.

7. *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 territory, 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 $\alpha = 0.01$.

8. *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 $\alpha = 0.05$.

9. *Plato's* Republic: *Syllable Patterns* Prose rhythm is characterized as 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 that of the text of Plato's *Republic*? Use $\alpha = 0.01$.

10. *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 9) 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 that 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 $\alpha = 0.01$.

11. *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 that 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 $\alpha = 0.01$.

12. *Supermarket: Prices* *Harper's Index* reported that 80% of all supermarket prices end in the digit 9 or 5. Suppose that 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 $\alpha = 0.05$.

13. *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 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 occurring 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 was 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.

14. *Medical: Hypertension* Diltiazem is a commonly prescribed drug for hypertension (see source in Problem 13). 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.

15. *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 $\alpha = 0.01$.

16. *American Attitudes: NAFTA* Generally speaking, would you say that America benefits from being a member of NAFTA (North American Free Trade Agreement)? Nationally, about 28% of the U.S. population believes NAFTA benefits America (Source: *American Attitudes* by S. Mitchell, Sociology Department, Ithaca College). A random sample of 48 interstate truck drivers showed that 19 believe NAFTA benefits America. Does this indicate that the population proportion of interstate truckers who believe NAFTA benefits America is higher than 28%? Use a 5% level of significance.

17. *Careers: College Professors* If you could start your career all over again, would you still choose to be a college professor? About 76% of all U.S. college professors responded yes to this question (Source: *The Chronicle of Higher Education*). A random sample of 59 college professors in Colorado showed that 47 claimed they would choose college teaching again. Does this indicate that the proportion of professors in Colorado who would choose the career again is different from the national rate of 76%? Use a 1% level of significance.

18. *Critical Region Method: Testing Proportions* Solve Problem 3 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 9-8 or part (c) of Table 3 of the Appendix. Compare your conclusions with the conclusions obtained by using the P-value method. Are they the same?

19. *Critical Region Method: Testing Proportions* Solve Problem 5 using the critical region method of testing. *Hint:* See Problem 18. Compare your conclusions with the conclusions obtained by using the P-value method. Are they the same?

20. *Critical Region Method: Testing Proportions* Solve Problem 11 using the critical region method of testing. *Hint:* See Problem 18. Compare your conclusions with the conclusions obtained by using the P-value method. Are they the same?

SUMMARY

In this chapter, we studied statistical inference methods called *hypothesis testing*. We establish an initial claim about the value of a parameter. This claim is the null hypothesis H_0, which states that the parameter in question equals a certain value. Then we propose an alternate hypothesis H_1, which indicates that the parameter is less than, greater than, or different from the value in the null hypothesis. To determine whether or not to reject the null hypothesis, we use the evidence of the sample data and the predetermined level of significance α.

The basic steps we follow in the procedure of hypothesis testing are

1. Choose the level of significance α. State the null and alternate hypotheses H_0 and H_1.

2. Determine the sampling distribution of the sample test statistic. Convert the sample test statistic to a z value or a t value as appropriate.

3. Find (or estimate) the P-value of the sample test statistic.
4. If P-value $\leq \alpha$, reject H_0. If P-value $> \alpha$, do not reject H_0.
5. Interpret the results in the context of the application.

An alternate 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 3(c) of the Appendix. Critical values t_0 are found in Table 4 of the Appendix. If the sample test statistic falls in the critical region, reject H_0.

In this chapter, we used hypothesis testing to conduct tests involving a single mean μ (σ known or unknown) or a single proportion p. The methods of hypothesis testing are very general, and we will see them used again in later chapters with other parameters and other probability distributions.

IMPORTANT WORDS & SYMBOLS

Section 9.1

Hypothesis testing
Hypotheses
Null hypothesis H_0
Alternate hypothesis H_1
Right-tailed test
Left-tailed test
Two-tailed test
Sample test statistic
P-value
Statistical significance
Type I error
Type II error
α, the level of significance of a test and the probability of a type I error
β, the probability of a type II error
Power of a test $(1 - \beta)$

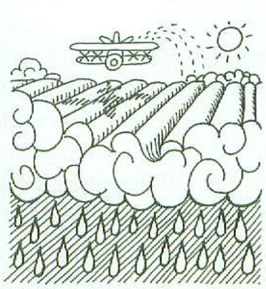
Section 9.2

$d.f.$ for testing μ when σ is unknown
Critical region
Critical value

Section 9.3

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

VIEWPOINT



Will It Rain?

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://math.college.hmco.com/students** 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

Before you solve each problem, first categorize it by answering the following question: Are we testing a single mean or a single proportion? Then provide the following information for Problems 1–10.

(a) What is the level of significance? State the null and alternate hypotheses.

(b) 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) State 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.

1. *Vehicles: Mileage* Based on information from *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.

2. *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 showed that 39 have jobs. Do the data indicate that more than 35% of the students have jobs? (Use a 5% level of significance.)

3. *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.)

4. *Unions: Salaries* The Fleetfoot Shoe Company claims that the average yearly salary of its workers is $29,800. The union believes that the average salary is much less. A random sample of 61 employees showed their average salary to be $29,500 with a sample standard deviation of $800. Use a 5% level of significance to test the company's claim.

5. *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 extensive burn 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.

6. *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 showed 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?

7. *Student Government: Opinions* The student council is thinking about discontinuing the student poetry magazine because only 20% of the students read it. A vote was taken, and it was decided to continue the magazine if more than 20% of the students are known to read it. A random sample of 200 students showed that 58 of them had read the last issue. Use a 1% level of significance to determine whether the magazine should be continued.

8. *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 in the federal civilian sector? Use $\alpha = 0.05$.

9. *Vending Machines: Coffee* A machine in the student lounge dispenses coffee. The average cup of coffee is supposed to contain 7.0 oz. Eight cups of coffee from this machine show the average content to be 7.3 oz with a standard deviation of 0.5 oz. Do you think that the machine has slipped out of adjustment and that the average amount of coffee per cup is different from 7 oz? Use a 5% level of significance.

10. *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.

"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?

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 8. 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.

(a) Review and discuss type I and type II errors associated with hypothesis testing.
(b) Review and discuss the level of significance and power of a statistical test.
(c) The following statements are very important. Give them some careful thought and discuss them.
  i. 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.
  ii. 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.
(d) 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.
(e) 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

TI-84PLUS/TI-83PLUS • *EXCEL* • *MINITAB* • *SPSS*

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

TI-84Plus/TI-83Plus

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

To draw random samples from a normal distribution with $\mu = 30$ and $\sigma = 2.5$, use the menu choices **Tools ➤ Data Analysis ➤ Random Number Generator.** In the dialogue box, the number of variables is the number of samples. Fill in the rest of the dialogue box using 50 as the number of random numbers, with 30 for the mean and 2.5 for the standard deviation.

To conduct a hypothesis test of H_0: $\mu = 30$ against H_1: $\mu > 30$, use the command **ZTEST(data range, 30, 2.5).** This command is described as returning the two-tailed P-value of the z test. However, it appears to give the P-value of a right-tailed test. Use the **ZTEST** command with caution and check your results against the table results.

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** for tests of a single mean. 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.** In the dialogue box, use Sample1 for the target variable, 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.

Cumulative Review Problems

CHAPTERS 7–9

Answers may vary due to rounding.

1. *Normal Distribution* *Oxygen demand* is a term biologists use to describe the oxygen needed by fish and other aquatic organisms for survival. The Environmental Protection Agency conducted a study of a wetland area in Marin County, California. In this wetland environment, the mean oxygen demand was $\mu = 9.9$ mg/L with 95% of the data ranging from 6.5 mg/L to 13.3 mg/L (Reference: EPA Report 832-R-93-005). Let x be a random variable that represents oxygen demand in this wetland environment. Assume x has a probability distribution that is approximately normal.
   (a) Use the 95% data range to estimate the standard deviation for oxygen demand. *Hint:* See Problem 27 of Section 7.3.
   (b) An oxygen demand below 8 indicates that some organisms in the wetland environment may be dying. What is the probability that the oxygen demand will fall below 8 mg/L?
   (c) A high oxygen demand can also indicate trouble. An oxygen demand above 12 may indicate an overabundance of organisms that endanger some types of plant life. What is the probability that the oxygen demand will exceed 12 mg/L?

2. *Normal Approximation to the Binomial* The majority of house burglars simply walk into a house that is unlocked! In fact, about 57% of all house burglars gain entrance through an unlocked window or door (Reference: *The Book of Risks* by Larry Laudan). Suppose that $n = 129$ house burglaries will occur tomorrow in Los Angeles. Let r be a binomial random variable that represents the number of burglaries that required no forced entrance.
   (a) We want to approximate the binomial random variable r by a normal random variable x. Is this appropriate? What requirements must be satisfied before we can do this? Do you think these requirements are satisfied in this case? Explain.
   (b) Compute μ and σ for the normal approximation.
   (c) What is the probability that at least 65 of the burglaries required no forced entry?

Patrol car at scene of house robbery

3. *Terminology* Please give a careful but brief answer to each of the following questions.
   (a) What is a population? How do you get a simple random sample? Give examples.
   (b) What is a sample statistic? What is a sampling distribution? Give examples.
   (c) Give a careful and complete statement of the central limit theorem.
   (d) List at least three areas of everyday life to which the above concepts can be applied. Be specific.

4. *Sampling Distribution $\bar{x}$* Workers at a large toxic cleanup project are concerned that their white blood cell counts may have been reduced. Let x be a random variable that represents white blood cell count per cubic millimeter of whole blood in a healthy adult. Then $\mu = 7500$ and $\sigma \approx 1750$ (Reference: *Diagnostic Tests with Nursing Applications*, S. Loeb). A random sample of $n = 50$ workers from the toxic cleanup site were given a blood test that showed $\bar{x} = 6820$. What is the probability that for healthy adults $\bar{x}$ will be this low or lower?
   (a) How does the central limit theorem apply? Explain.
   (b) Compute $P(\bar{x} \le 6820)$.
   (c) Based on your answer to part (b), would you recommend that additional facts be obtained, or would you recommend that the workers' claims be dismissed? Explain.

In Problems 5–7, please use the following steps (i) through (v) for all hypothesis tests.

i. What is the level of significance? State the null and alternate hypotheses.
ii. What sampling distribution will you use? What assumptions are you making? What is the value of the sample test statistic?
iii. Find (or estimate) the P-value. Sketch the sampling distribution and show the area corresponding to the P-value.
iv. Based on your answers in parts (i) to (iii), will you reject or fail to reject the null hypothesis? Are the data statistically significant at level α?
v. State 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 thereby produce a slightly more "conservative" answer.

5. *Testing and Estimating μ, σ Known* Let x be a random variable that represents micrograms of lead per liter of water (ug/l). An industrial plant discharges water into a creek. The Environmental Protection Agency has studied the discharged water and found x to have a normal distribution with $\sigma = 0.7$ ug/l (Reference: *EPA Wetlands Case Studies*).
   (a) The industrial plant says that the population mean value of x is $\mu = 2.0$ ug/l. However, a random sample of $n = 10$ water samples showed that $\bar{x} = 2.56$ ug/l. Does this indicate that the lead concentration population mean is higher than the industrial plant claims? Use $\alpha = 1\%$.
   (b) Find a 95% confidence interval for μ using the sample data and the EPA value for σ.
   (c) *Sample Size* How large a sample should be taken to be 95% confident that the sample mean $\bar{x}$ is within a margin of error $E = 0.2$ ug/l of the population mean?

6. *Testing and Estimating μ, σ Unknown* Carboxyhemoglobin is formed when hemoglobin is exposed to carbon monoxide. Heavy smokers tend to have a high percentage of carboxyhemoglobin in their blood (Reference: *Laboratory and Diagnostic Tests*, F. Fishbach). Let x be a random variable representing percentage of carboxyhemoglobin in the blood. For a person who is a regular heavy smoker, x has

a distribution that is approximately normal. A random sample of $n = 12$ blood tests given to a heavy smoker gave the following results (percent carboxyhemoglobin in the blood).

9.1 9.5 10.2 9.8 11.3 12.2 11.6 10.3 8.9 9.7 13.4 9.9

(a) Use a calculator to verify that $\bar{x} \approx 10.49$ and $s \approx 1.36$.

(b) A long-term population mean $\mu = 10\%$ is considered a health risk. However, a long-term population mean above 10% is considered a clinical alert that the person may be asymptomatic. Do the data indicate that the population mean percentage is higher than 10% for this patient? Use $\alpha = 0.05$.

(c) Use the given data to find a 99% confidence interval for μ for this patient.

7. *Testing and Estimating a Proportion p* Although older Americans are most afraid of crime, it is young people who are more likely to be the actual victims of crime. It seems that older people are more cautious about the people with whom they associate. A national survey showed that 10% of all people ages 16–19 have been victims of crime (Reference: *Bureau of Justice Statistics*). At Jefferson High School, a random sample of $n = 68$ students (ages 16–19) showed that $r = 10$ had been victims of a crime.

   (a) Do these data indicate that the population proportion of students in this school (ages 16–19) who have been victims of a crime is different (either way) from the national rate for this age group? Use $\alpha = 0.05$. Do you think the conditions $np > 5$ and $nq > 5$ are satisfied in this setting? Why is this important?

   (b) Find a 90% confidence interval for the proportion of students in this school (ages 16–19) who have been victims of a crime.

   (c) *Sample Size* How large a sample size should be used to be 95% sure that the sample proportion $\hat{p}$ is within a margin of error $E = 0.05$ of the population proportion of all students in this school (ages 16–19) who have been victims of a crime? *Hint:* Use sample data $\hat{p}$ as a preliminary estimate for p.

8. *Essay and Project* In Chapters 7, 8, and 9 you have studied sampling distributions, estimation, and hypothesis testing.

   (a) Write a brief essay in which you discuss using information from samples to infer information about populations. Be sure to include methods of estimation and hypothesis testing in your discussion. What two sampling distributions are used in estimation and hypothesis testing of population means and proportions? What are the criteria for determining the appropriate sampling distribution? What is the level of significance of a test? What is the *P*-value? How is the *P*-value related to the alternate hypothesis? How is the null hypothesis related to the sample test statistic? Explain.

   (b) Suppose you want to study the length of time devoted to commercial breaks for television programs. Identify the type of programs you want to study (e.g., sitcoms, sports events, movies, news, children's programs, etc.). Write a brief outline for your study. Discuss how to obtain random samples. How large should the sample be for a specified margin of error? Describe the protocol you will follow to measure the times of the commercial breaks. Determine whether you are going to study the average time devoted to commercials or the proportion of time devoted to commercials. What assumptions will you make regarding the population distribution? What graphics might be appropriate? What methods of estimation will you use? What methods of testing will you use?