9 HYPOTHESIS TESTS ABOUT THE MEAN AND PROPORTION

This chapter introduces the second topic in inferential statistics: tests of hypotheses. In a test of hypothesis, we test a certain given theory or belief about a population parameter. We may want to find out, using some sample information, whether or not a given claim (or statement) about a population parameter is true. This chapter discusses how to make such tests of hypotheses about the population mean, μ, and the population proportion, p.

As an example, a soft-drink company may claim that, on average, its cans contain 12 ounces of soda. A government agency may want to test whether or not such cans contain, on average, 12 ounces of soda. As another example, according to the U.S. Bureau of the Census, 16.3% of the population in the United States lacked health insurance in 1998. An economist may want to check if this percentage is still true for this year. In the first of these two examples we are to test a hypothesis about the population mean, μ, and in the second example we are to test a hypothesis about the population proportion, p.

9.1 HYPOTHESIS TESTS: AN INTRODUCTION

Why do we need to perform a test of hypothesis? Reconsider the example about soft-drink cans. Suppose we take a sample of 100 cans of the soft drink under investigation. We then find out that the mean amount of soda in these 100 cans is 11.89 ounces. Based on this result, can we state that, on average, all such cans contain less than 12 ounces of soda and that the company is lying to the public? Not until we perform a test of hypothesis can we make such an accusation. The reason is that the mean, $\bar{x} = 11.89$ ounces, is obtained from a sample. The difference between 12 ounces (the required average amount for the population) and 11.89 ounces (the observed average amount for the sample) may have occurred only because of the sampling error. Another sample of 100 cans may give us a mean of 12.04 ounces. Therefore, we make a test of hypothesis to find out how large the difference between 12 ounces and 11.89 ounces is and to investigate whether or not this difference has occurred as a result of chance alone. Now, if 11.89 ounces is the mean for all cans and not for just 100 cans, then we do not need to make a test of hypothesis. Instead, we can immediately state that the mean amount of soda in all such cans is less than 12 ounces. We perform a test of hypothesis only when we are making a decision about a population parameter based on the value of a sample statistic.

9.1.1 TWO HYPOTHESES

Consider a nonstatistical example of a person who has been indicted for committing a crime and is being tried in a court. Based on the available evidence, the judge or jury will make one of two possible decisions:

1. The person is not guilty.
2. The person is guilty.

At the outset of the trial, the person is presumed not guilty. The prosecutor's efforts are to prove that the person has committed the crime and, hence, is guilty.

In statistics, *the person is not guilty* is called the **null hypothesis** and *the person is guilty* is called the **alternative hypothesis**. The null hypothesis is denoted by H_0 and the alternative hypothesis is denoted by H_1. In the beginning of the trial it is assumed that the person is not guilty. The null hypothesis is usually the hypothesis that is assumed to be true to begin with. The two hypotheses for the court case are written as follows (notice the colon after H_0 and H_1):

Null hypothesis: H_0: The person is not guilty

Alternative hypothesis: H_1: The person is guilty

In a statistics example, the null hypothesis states that a given claim (or statement) about a population parameter is true. Reconsider the example of the soft-drink company's claim that, on average, its cans contain 12 ounces of soda. In reality, this claim may or may not be true. However, we will initially assume that the company's claim is true (that is, the company is not guilty of cheating and lying). To test the claim of the soft-drink company, the null hypothesis will be that the company's claim is true. Let μ be the mean amount of soda in all cans. The company's claim will be true if $\mu = 12$ ounces. Thus, the null hypothesis will be written as

H_0: $\mu = 12$ ounces (The company's claim is true)

In this example, the null hypothesis can also be written as $\mu \geq 12$ ounces because the claim of the company will still be true if the cans contain, on average, more than 12 ounces of soda. The company will be accused of cheating the public only if the cans contain, on average, less than 12 ounces of soda. However, it will not affect the test whether we use an = or a ≥ sign in the null hypothesis as long as the alternative hypothesis has a < sign. Remember that in the null hypothesis (and in the alternative hypothesis also) we use the population parameter (such as μ or p), and not the sample statistic (such as $\bar{x}$ or $\hat{p}$).

NULL HYPOTHESIS A *null hypothesis* is a claim (or statement) about a population parameter that is assumed to be true until it is declared false.

The alternative hypothesis in our statistics example will be that the company's claim is false and its soft-drink cans contain, on average, less than 12 ounces of soda—that is, $\mu < 12$ ounces. The alternative hypothesis will be written as

$$H_1\colon \mu < 12 \text{ ounces} \qquad \text{(The company's claim is false)}$$

ALTERNATIVE HYPOTHESIS An *alternative hypothesis* is a claim about a population parameter that will be true if the null hypothesis is false.

Let us return to the example of the court trial. The trial begins with the assumption that the null hypothesis is true—that is, the person is not guilty. The prosecutor assembles all the possible evidence and presents it in the court to prove that the null hypothesis is false and the alternative hypothesis is true (that is, the person is guilty). In the case of our statistics example, the information obtained from a sample will be used as evidence to decide whether or not the claim of the company is true. In the court case, the decision made by the judge (or jury) depends on the amount of evidence presented by the prosecutor. At the end of the trial, the judge (or jury) will consider whether or not the evidence presented by the prosecutor is sufficient to declare the person guilty. The amount of evidence that will be considered to be sufficient to declare the person guilty depends on the discretion of the judge (or jury).

9.1.2 REJECTION AND NONREJECTION REGIONS

In Figure 9.1, which represents the court case, the point marked 0 indicates that there is no evidence against the person being tried. The farther we move toward the right on the horizontal axis, the more convincing the evidence is that the person has committed the crime.

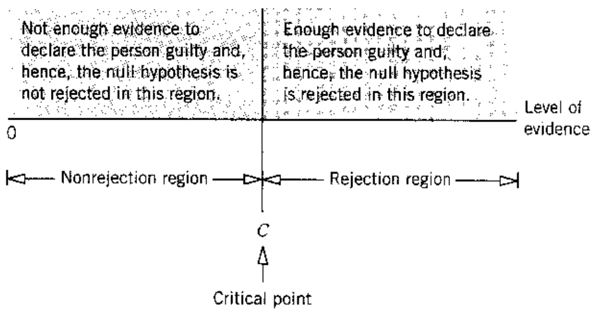


Figure 9.1 Nonrejection and rejection regions for the court case.

We have arbitrarily marked a point C on the horizontal axis. Let us assume that a judge (or jury) considers any amount of evidence to the right of point C to be sufficient and any amount of evidence to the left of C to be insufficient to declare the person guilty. Point C is called the **critical value** or **critical point** in statistics. If the amount of evidence presented by the prosecutor falls in the area to the left of point C, the verdict will reflect that there is not enough evidence to declare the person guilty. Consequently, the accused person will be declared *not guilty*. In statistics, this decision is stated as *do not reject* H_0. It is equivalent to saying that there is not enough evidence to declare the null hypothesis false. The area to the left of point C is called the *nonrejection region;* that is, this is the region where the null hypothesis is not rejected. However, if the amount of evidence falls in the area to the right of point C, the verdict will be that there is sufficient evidence to declare the person guilty. In statistics, this decision is stated as *reject* H_0 or *the null hypothesis is false.* Rejecting H_0 is equivalent to saying that *the alternative hypothesis is true.* The area to the right of point C is called the *rejection region;* that is, this is the region where the null hypothesis is rejected.

9.1.3 TWO TYPES OF ERRORS

We all know that a court's verdict is not always correct. If a person is declared guilty at the end of a trial, there are two possibilities.

1. The person has *not* committed the crime but is declared guilty (because of what may be false evidence).
2. The person *has* committed the crime and is rightfully declared guilty.

In the first case, the court has made an error by punishing an innocent person. In statistics, this kind of error is called a **Type I** or an **α *(alpha)* error**. In the second case, because the guilty person has been punished, the court has made the correct decision. The second row in the shaded portion of Table 9.1 shows these two cases. The two columns of Table 9.1, corresponding to *the person is not guilty* and *the person is guilty,* give the two actual situations. Which one of these is true is known only to the person being tried. The two rows in this table, corresponding to *the person is not guilty* and *the person is guilty,* show the two possible court decisions.

Table 9.1

		Actual Situation	
		The Person Is Not Guilty	**The Person Is Guilty**
Court's decision	The person is not guilty	Correct decision	Type II or β error
	The person is guilty	Type I or α error	Correct decision

In our statistics example, a Type I error will occur when H_0 is actually true (that is, the cans do contain, on average, 12 ounces of soda), but it just happens that we draw a sample with a mean that is much less than 12 ounces and we wrongfully reject the null hypothesis, H_0. The value of α, called the **significance level** of the test, represents the probability of making a Type I error. In other words, α is the probability of rejecting the null hypothesis, H_0, when in fact it is true.

TYPE I ERROR A *Type I error* occurs when a true null hypothesis is rejected. The value of α represents the probability of committing this type of error; that is,

$$\alpha = P(H_0 \text{ is rejected} \mid H_0 \text{ is true})$$

The value of α represents the *significance level* of the test.

The size of the rejection region in a statistics problem of a test of hypothesis depends on the value assigned to α. In a test of hypothesis, we usually assign a value to α before making the test. Although any value can be assigned to α, the commonly used values of α are .01, .025, .05, and .10. Usually the value assigned to α does not exceed .10 (or 10%).

Now, suppose that in the court trial case the person is declared not guilty at the end of the trial. Such a verdict does not indicate that the person has indeed *not* committed the crime. It is possible that the person is guilty but there is not enough evidence to prove the guilt. Consequently, in this situation there are again two possibilities.

1. The person has *not* committed the crime and is declared not guilty.
2. The person *has* committed the crime but, *because of the lack of enough evidence*, is declared not guilty.

In the first case, the court's decision is correct. But in the second case, the court has committed an error by setting a guilty person free. In statistics, this type of error is called a **Type II** or a **β** (the Greek letter *beta*) **error**. These two cases are shown in the first row of the shaded portion of Table 9.1.

In our statistics example, a Type II error will occur when the null hypothesis H_0 is actually false (that is, the soda contained in all cans, on average, is less than 12 ounces), but it happens by chance that we draw a sample with a mean that is close to or greater than 12 ounces and we wrongfully conclude *do not reject* H_0. The value of β represents the probability of making a Type II error. It represents the probability that H_0 is not rejected when actually H_0 is false. The value of $1 - \beta$ is called the **power of the test**. It represents the probability of not making a Type II error.

TYPE II ERROR A *Type II error* occurs when a false null hypothesis is not rejected. The value of β represents the probability of committing a Type II error; that is,

$$\beta = P(H_0 \text{ is not rejected} \mid H_0 \text{ is false})$$

The value of $1 - \beta$ is called the *power of the test*. It represents the probability of not making a Type II error.

The two types of errors that occur in tests of hypotheses depend on each other. We cannot lower the values of α and β simultaneously for a test of hypothesis for a fixed sample size. Lowering the value of α will raise the value of β, and lowering the value of β will raise the value of α. However, we can decrease both α and β simultaneously by increasing the sample size. The explanation of how α and β are related and the computation of β are not within the scope of this text.

Table 9.2, which is similar to Table 9.1, is written for the statistics problem of a test of hypothesis. In Table 9.2 *the person is not guilty* is replaced by H_0 *is true, the person is guilty* by H_0 *is false,* and the *court's decision* by *decision*.

Table 9.2

		Actual Situation	
		H_0 is true	H_0 is false
Decision	Do not reject H_0	Correct decision	Type II or β error
	Reject H_0	Type I or α error	Correct decision

9.1.4 TAILS OF A TEST

The statistical hypothesis-testing procedure is similar to the trial of a person in court but with two major differences. The first major difference is that in a statistical test of hypothesis, the partition of the total region into rejection and nonrejection regions is not arbitrary. Instead, it depends on the value assigned to α (Type I error). As mentioned earlier, α is also called the significance level of the test.

The second major difference relates to the rejection region. In the court case, the rejection region is on the right side of the critical point, as shown in Figure 9.1. However, in statistics, the rejection region for a hypothesis-testing problem can be on both sides, with the nonrejection region in the middle, or it can be on the left side or on the right side of the nonrejection region. These possibilities are explained in the next three parts of this section. A test with two rejection regions is called a **two-tailed test**, and a test with one rejection region is called a **one-tailed test**. The one-tailed test is called a **left-tailed test** if the rejection region is in the left tail of the distribution curve, and it is called a **right-tailed test** if the rejection region is in the right tail of the distribution curve.

TAILS OF THE TEST A *two-tailed test* has rejection regions in both tails, a *left-tailed test* has the rejection region in the left tail, and a *right-tailed test* has the rejection region in the right tail of the distribution curve.

A Two-Tailed Test

According to the U.S. Bureau of the Census, the mean family size in the United States was 3.18 in 1998. A researcher wants to check whether or not this mean has changed since 1998. The key word here is *changed*. The mean family size has changed if it has either increased or decreased during the period since 1998. This is an example of a two-tailed test. Let μ be the current mean family size for all families. The two possible decisions are

1. The mean family size has not changed—that is, $\mu = 3.18$.
2. The mean family size has changed—that is, $\mu \neq 3.18$.

We write the null and alternative hypotheses for this test as

H_0: $\mu = 3.18$ (The mean family size has not changed)

H_1: $\mu \neq 3.18$ (The mean family size has changed)

Whether a test is two-tailed or one-tailed is determined by the sign in the alternative hypothesis. If the alternative hypothesis has a *not equal to* ($\neq$) sign, as in this example, it is a two-tailed test. As shown in Figure 9.2, a two-tailed test has two rejection regions, one in

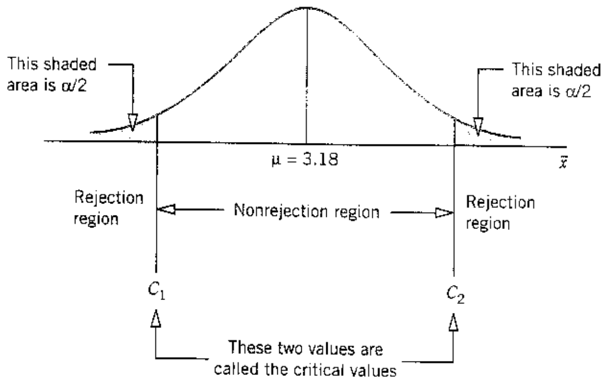


Figure 9.2 A two-tailed test.

each tail of the distribution curve. Figure 9.2 shows the sampling distribution of $\bar{x}$ for a large sample. Assuming H_0 is true, $\bar{x}$ has a normal distribution with its mean equal to 3.18 (the value of μ in H_0). In Figure 9.2, the area of each of the two rejection regions is $\alpha/2$ and the total area of both rejection regions is α (the significance level). As shown in this figure, a two-tailed test of hypothesis has two critical values that separate the two rejection regions from the nonrejection region. We will reject H_0 if the value of $\bar{x}$ obtained from the sample falls in either of the two rejection regions. We will not reject H_0 if the value of $\bar{x}$ lies in the nonrejection region. By rejecting H_0, we are saying that the difference between the value of μ stated in H_0 and the value of $\bar{x}$ obtained from the sample is too large to have occurred because of the sampling error alone. Consequently, this difference is real. By not rejecting H_0, we are saying that the difference between the value of μ stated in H_0 and the value of $\bar{x}$ obtained from the sample is small and it may have occurred because of the sampling error alone.

A Left-Tailed Test

Reconsider the example of the mean amount of soda in all soft-drink cans produced by a company. The company claims that these cans, on average, contain 12 ounces of soda. However, if these cans contain less than the claimed amount of soda, then the company can be accused of cheating. Suppose a consumer agency wants to test whether the mean amount of soda per can is less than 12 ounces. Note that the key phrase this time is *less than*, which indicates a left-tailed test. Let μ be the mean amount of soda in all cans. The two possible decisions are

1. The mean amount of soda in all cans is not less than 12 ounces—that is, $\mu = 12$ ounces.
2. The mean amount of soda in all cans is less than 12 ounces—that is, $\mu < 12$ ounces.

The null and alternative hypotheses for this test are written as

H_0: $\mu = 12$ ounces (The mean is not less than 12 ounces)

H_1: $\mu < 12$ ounces (The mean is less than 12 ounces)

In this case, we can also write the null hypothesis as H_0: $\mu \geq 12$. This will not affect the result of the test as long as the sign in H_1 is *less than* (<).

When the alternative hypothesis has a *less than* (<) sign, as in this case, the test is always left-tailed. In a left-tailed test, the rejection region is in the left tail of the distribution

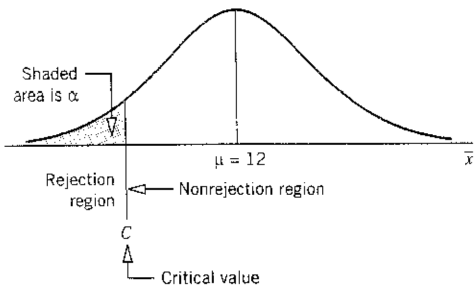


Figure 9.3 A left-tailed test.

curve, as shown in Figure 9.3, and the area of this rejection region is equal to α (the significance level). We can observe from this figure that there is only one critical value in a left-tailed test.

Assuming H_0 is true, $\bar{x}$ has a normal distribution for a large sample with its mean equal to 12 ounces (the value of μ in H_0). We will reject H_0 if the value of $\bar{x}$ obtained from the sample falls in the rejection region; we will not reject H_0 otherwise.

A Right-Tailed Test

To illustrate the third case, according to a 1999 study by the American Federation of Teachers, the mean starting salary of school teachers in the United States was \$25,735 during 1997–98. Suppose we want to test whether the current mean starting salary of all school teachers in the United States is higher than \$25,735. The key phrase in this case is *higher than*, which indicates a right-tailed test. Let μ be the current mean starting salary of school teachers in the United States. The two possible decisions this time are

1. The current mean starting salary of all school teachers in the United States is not higher than \$25,735—that is, $\mu = \$25{,}735$.
2. The current mean starting salary of all school teachers in the United States is higher than \$25,735—that is, $\mu > \$25{,}735$.

We write the null and alternative hypotheses for this test as

H_0: $\mu = \$25{,}735$ (The current mean starting salary is not higher than \$25,735)

H_1: $\mu > \$25{,}735$ (The current mean starting salary is higher than \$25,735)

In this case, we can also write the null hypothesis as H_0: $\mu \leq \$25{,}735$, which states that the current mean starting salary of all school teachers in the United States is either equal to or less than \$25,735. Again, the result of the test will not be affected whether we use an *equal to* (=) or a *less than or equal to* ($\leq$) sign in H_0 as long as the alternative hypothesis has a *greater than* (>) sign.

When the alternative hypothesis has a *greater than* (>) sign, the test is always right-tailed. As shown in Figure 9.4 on page 384, in a right-tailed test, the rejection region is in the right tail of the distribution curve. The area of this rejection region is equal to α, the significance level. Like a left-tailed test, a right-tailed test has only one critical value.

Again, assuming H_0 is true, $\bar{x}$ has a normal distribution for a large sample with its mean equal to \$25,735 (the value of μ in H_0). We will reject H_0 if the value of $\bar{x}$ obtained from the sample falls in the rejection region. Otherwise, we will not reject H_0.

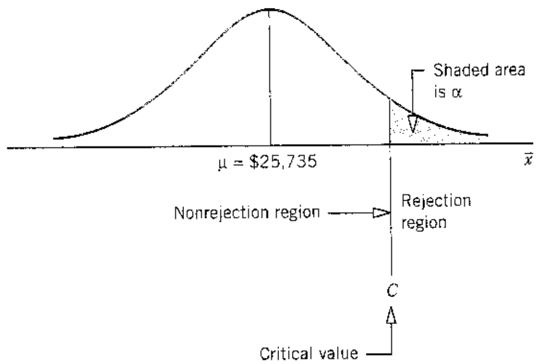


Figure 9.4 A right-tailed test.

Table 9.3 summarizes the foregoing discussion about the relationship between the signs in H_0 and H_1 and the tails of a test.

Table 9.3

	Two-Tailed Test	Left-Tailed Test	Right-Tailed Test
Sign in the null hypothesis H_0	$=$	$=$ or $\geq$	$=$ or $\leq$
Sign in the alternative hypothesis H_1	$\neq$	$<$	$>$
Rejection region	In both tails	In the left tail	In the right tail

Note that the null hypothesis always has an *equal to* ($=$) or a *greater than or equal to* ($\geq$) or a *less than or equal to* ($\leq$) sign, and the alternative hypothesis always has a *not equal to* ($\neq$) or a *less than* ($<$) or a *greater than* ($>$) sign.

A test of hypothesis involves five steps, which are listed next.

STEPS TO PERFORM A TEST OF HYPOTHESIS A statistical test of hypothesis procedure has the following five steps.

1. State the null and alternative hypotheses.
2. Select the distribution to use.
3. Determine the rejection and nonrejection regions.
4. Calculate the value of the test statistic.
5. Make a decision.

With the help of examples, these steps will be described in the next section.

EXERCISES

Concepts and Procedures

9.1 Briefly explain the meaning of each of the following terms.

a.	Null hypothesis	b.	Alternative hypothesis
c.	Critical point(s)	d.	Significance level
e.	Nonrejection region	f.	Rejection region
g.	Tails of a test	h.	Two types of errors

9.2 What are the four possible outcomes for a test of hypothesis? Show these outcomes by writing a table. Briefly describe the Type I and Type II errors.

9.3 Explain how the tails of a test depend on the sign in the alternative hypothesis. Describe the signs in the null and alternative hypotheses for a two-tailed, a left-tailed, and a right-tailed test, respectively.

9.4 Explain which of the following is a two-tailed test, a left-tailed test, or a right-tailed test.

a. H_0: $\mu = 45$, H_1: $\mu > 45$
b. H_0: $\mu = 23$, H_1: $\mu \neq 23$
c. H_0: $\mu \geq 75$, H_1: $\mu < 75$

Show the rejection and nonrejection regions for each of these cases by drawing a sampling distribution curve for the sample mean, assuming that the sample size is large in each case.

9.5 Explain which of the following is a two-tailed test, a left-tailed test, or a right-tailed test.

a. H_0: $\mu = 12$, H_1: $\mu < 12$
b. H_0: $\mu \leq 85$, H_1: $\mu > 85$
c. H_0: $\mu = 33$, H_1: $\mu \neq 33$

Show the rejection and nonrejection regions for each of these cases by drawing a sampling distribution curve for the sample mean, assuming that the sample size is large in each case.

9.6 Which of the two hypotheses (null and alternative) is initially assumed to be true in a test of hypothesis?

9.7 Consider H_0: $\mu = 20$ versus H_1: $\mu < 20$.

a. What type of error are you making if the null hypothesis is actually false and you fail to reject it?
b. What type of error are you making if the null hypothesis is actually true and you reject it?

9.8 Consider H_0: $\mu = 55$ versus H_1: $\mu \neq 55$.

a. What type of error are you making if the null hypothesis is actually false and you fail to reject it?
b. What type of error are you making if the null hypothesis is actually true and you reject it?

Applications

9.9 Write the null and alternative hypotheses for each of the following examples. Determine if each is a case of a two-tailed, a left-tailed, or a right-tailed test.

a. To test whether or not the mean price of houses in Connecticut is greater than \$143,000
b. To test if the mean number of hours spent working per week by college students who hold jobs is different from 15 hours
c. To test whether the mean life of a particular brand of auto batteries is less than 45 months
d. To test if the mean amount of time spent doing homework by all fourth-graders is different from 5 hours a week
e. To test if the mean age of all college students is different from 24 years

9.10 Write the null and alternative hypotheses for each of the following examples. Determine if each is a case of a two-tailed, a left-tailed, or a right-tailed test.

a. To test if the mean amount of time spent per week watching sports on television by all adult men is different from 9.5 hours
b. To test if the mean amount of money spent by all customers at a supermarket is less than \$85

c. To test whether the mean starting salary of college graduates is higher than $29,000 per year
d. To test if the mean GPA of all students at a university is lower than 2.9
e. To test if the mean cholesterol level of all adult men in the United States is higher than 175

9.2 HYPOTHESIS TESTS ABOUT A POPULATION MEAN: LARGE SAMPLES

From the central limit theorem discussed in Chapter 7, the sampling distribution of $\bar{x}$ is approximately normal for large samples ($n \geq 30$). Consequently, whether or not σ is known, the normal distribution is used to test hypotheses about the population mean when a sample size is large.

TEST STATISTIC In tests of hypotheses about μ for large samples, the random variable

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} \quad \text{or} \quad \frac{\bar{x} - \mu}{s_{\bar{x}}}$$

where

$$\sigma_{\bar{x}} = \sigma/\sqrt{n} \quad \text{and} \quad s_{\bar{x}} = s/\sqrt{n}$$

is called the *test statistic*. The test statistic can be defined as a rule or criterion that is used to make the decision whether or not to reject the null hypothesis.

At the end of Section 9.1, it was mentioned that a test of hypothesis procedure involves the following five steps.

1. State the null and alternative hypotheses.
2. Select the distribution to use.
3. Determine the rejection and nonrejection regions.
4. Calculate the value of the test statistic.
5. Make a decision.

Examples 9–1 through 9–3 illustrate the use of these five steps to perform tests of hypotheses about the population mean μ. Example 9–1 is concerned with a two-tailed test and Examples 9–2 and 9–3 describe one-tailed tests.

Conducting a two-tailed test of hypothesis about μ for a large sample.

EXAMPLE 9–1 The TIV Telephone Company provides long-distance telephone service in an area. According to the company's records, the average length of all long-distance calls placed through this company in 1999 was 12.44 minutes. The company's management wanted to check if the mean length of the current long-distance calls is different from 12.44 minutes. A sample of 150 such calls placed through this company produced a mean length of 13.71 minutes with a standard deviation of 2.65 minutes. Using the 5% significance level, can you conclude that the mean length of all current long-distance calls is different from 12.44 minutes?

Solution Let μ be the mean length of all current long-distance calls placed through this company and $\bar{x}$ be the corresponding mean for the sample. From the given information,

$$n = 150, \quad \bar{x} = 13.71 \text{ minutes}, \quad \text{and} \quad s = 2.65 \text{ minutes}$$

We are to test whether or not the mean length of all current long-distance calls is different from 12.44 minutes. The significance level α is .05; that is, the probability of rejecting the null hypothesis when it actually is true should not exceed .05. This is the probability of making a Type I error. We perform the test of hypothesis using the five steps.

Step 1. *State the null and alternative hypotheses.*

Notice that we are testing to find whether or not the mean length of all current long-distance calls is different from 12.44 minutes. We write the null and alternative hypotheses as follows.

H_0: $\mu = 12.44$ (The mean length of all current long-distance calls is 12.44 minutes)

H_1: $\mu \neq 12.44$ (The mean length of all current long-distance calls is different from 12.44 minutes)

Step 2. *Select the distribution to use.*

Because the sample size is large ($n > 30$), the sampling distribution of $\bar{x}$ is (approximately) normal. Consequently, we use the normal distribution to make the test.

Step 3. *Determine the rejection and nonrejection regions.*

The significance level is .05. The $\neq$ sign in the alternative hypothesis indicates that the test is two-tailed with two rejection regions, one in each tail of the normal distribution curve of $\bar{x}$. Because the total area of both rejection regions is .05 (the significance level), the area of the rejection region in each tail is .025; that is,

$$\text{Area in each tail} = \alpha/2 = .05/2 = .025$$

These areas are shown in Figure 9.5. Two critical points in this figure separate the two rejection regions from the nonrejection region. Next, we find the z values for the two critical points using the area of the rejection region. To find the z values for these critical points, we first find the area between the mean and one of the critical points. We obtain this area by subtracting .025 (the area in each tail) from .5, which gives .4750. Next we look for .4750 in the standard normal distribution table, Table VII of Appendix D. The value of z for .4750 is 1.96. Hence, the z values of the two critical points, as shown in Figure 9.5, are -1.96 and 1.96.

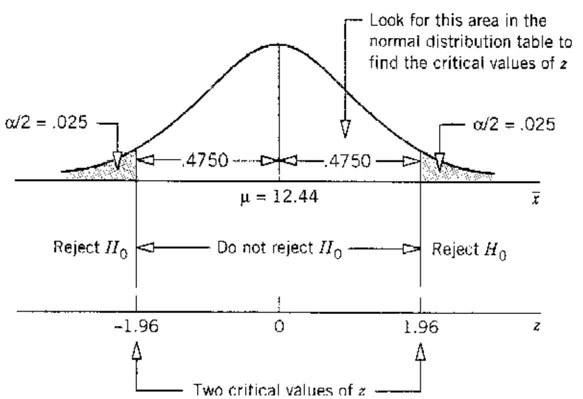


Figure 9.5

Step 4. *Calculate the value of the test statistic.*

The decision to reject or not to reject the null hypothesis will depend on whether the evidence from the sample falls in the rejection or the nonrejection region. If the value of $\bar{x}$ falls in either of the two rejection regions, we reject H_0. Otherwise, we do not reject H_0. The value of $\bar{x}$ obtained from the sample is called the *observed value of* $\bar{x}$. To locate the position of $\bar{x} = 13.71$ on the sampling distribution curve of $\bar{x}$ in Figure 9.5, we first calculate the z value for $\bar{x} = 13.71$. This is called the *value of the test statistic*. Then, we compare the value of the test statistic with the two critical values of z, -1.96 and 1.96, shown in Figure 9.5. If the value of the test statistic is between -1.96 and 1.96, we do not reject H_0. If the value of the test statistic is either greater than 1.96 or less than -1.96,we reject H_0.

CALCULATING THE VALUE OF THE TEST STATISTIC For a large sample, *the value of the test statistic* z for $\bar{x}$ for a test of hypothesis about μ is computed as follows:

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} \quad \text{if } \sigma \text{ is known}$$

$$z = \frac{\bar{x} - \mu}{s_{\bar{x}}} \quad \text{if } \sigma \text{ is not known}$$

where $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ and $s_{\bar{x}} = s/\sqrt{n}$

The value of z calculated for $\bar{x}$ using the formula is also called the **observed value of z**.

The value of $\bar{x}$ from the sample is 13.71. Because σ is not known, we calculate the z value using $s_{\bar{x}}$ as follows:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{2.65}{\sqrt{150}} = .21637159$$

From H_0

$$z = \frac{\bar{x} - \mu}{s_{\bar{x}}} = \frac{13.71 - 12.44}{.21637159} = 5.87$$

The value of μ in the calculation of the z value is substituted from the null hypothesis. The value of $z = 5.87$ calculated for $\bar{x}$ is called the *computed value of the test statistic z*. This is the value of z that corresponds to the value of $\bar{x}$ observed from the sample. It is also called the *observed value of z*.

Step 5. *Make a decision.*

In the final step we make a decision based on the location of the value of the test statistic z computed for $\bar{x}$ in Step 4. This value of $z = 5.87$ is greater than the critical value of $z = 1.96$, and it falls in the rejection region in the right tail in Figure 9.5. Hence, we reject H_0 and conclude that based on the sample information, it appears that the mean length of all such calls is not equal to 12.44 minutes.

By rejecting the null hypothesis, we are stating that the difference between the sample mean, $\bar{x} = 13.71$ minutes, and the hypothesized value of the population mean, $\mu = 12.44$ minutes, is too large and may not have occurred because of chance or sampling error alone. This difference seems to be real and, hence, the mean length of all such calls is different from 12.44 minutes. Note that the rejection of the null hypothesis does not necessarily indicate that the mean length of all such calls is definitely different from 12.44 minutes. It simply indicates that there is strong evidence (from the sample) that the mean length of such calls is not equal to 12.44 minutes. There is a possibility that the mean length of all such calls is equal to 12.44 minutes but, by the luck of the draw, we selected a sample with a mean that is too far from the hypothesized mean of 12.44 minutes. If so, we have wrongfully rejected the null hypothesis H_0. This is a Type I error and its probability is .05 in this example. ■

Conducting a right-tailed test of hypothesis about μ for a large sample.

EXAMPLE 9-2 According to an estimate, the average sale price of homes in the neighborhood with a high concentration of top-ranked professionals and executives in Stamford, Connecticut, was \$520,234 in December 1999 (*The Wall Street Journal*, December 10, 1999). A random sample of 50 homes from this neighborhood that were recently sold gave a mean sale price of \$565,750 with a standard deviation of \$75,210. Using 1% significance level, can you conclude that the current mean sale price of homes in this neighborhood is higher than \$520,234?

Solution Let μ be the current mean sale price of all homes in this neighborhood of Stamford, Connecticut, and let $\bar{x}$ be the corresponding mean for the sample. From the given information,

$$n = 50, \quad \bar{x} = \$565{,}750, \quad \text{and} \quad s = \$75{,}210$$

The significance level is $\alpha = .01$.

Step 1. *State the null and alternative hypotheses.*

We are to test whether the current mean sale price of homes in the said neighborhood of Stamford, Connecticut, is higher than \$520,234. The null and alternative hypotheses are

$$H_0\colon \mu = \$520{,}234 \quad \text{(The current mean is \$520,234)}$$

$$H_1\colon \mu > \$520{,}234 \quad \text{(The current mean is greater than \$520,234)}$$

Step 2. *Select the distribution to use.*

Because the sample size is large ($n > 30$), the sampling distribution of $\bar{x}$ is (approximately) normal. Consequently, we use the normal distribution to make the test.

Step 3. *Determine the rejection and nonrejection regions.*

The significance level is .01. The $>$ sign in the alternative hypothesis indicates that the test is right-tailed with its rejection region in the right tail of the sampling distribution curve of $\bar{x}$. Because there is only one rejection region, its area is $\alpha = .01$. As shown in Figure 9.6, the critical value of z, obtained from Table VII of Appendix D for .4900, is approximately 2.33.

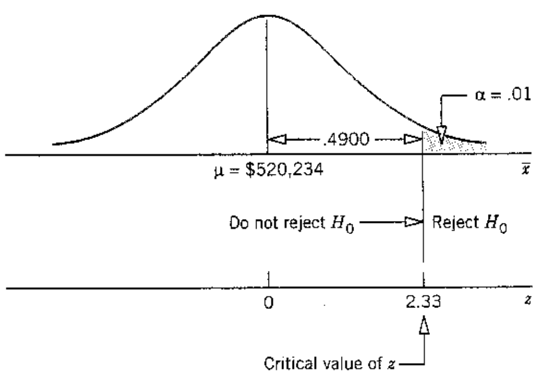


Figure 9.6

Step 4. *Calculate the value of the test statistic.*

The value of the test statistic z for $\bar{x} = \$565{,}750$ is calculated as follows:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{75{,}210}{\sqrt{50}} = 10{,}636.3002$$

From H_0

$$z = \frac{\bar{x} - \mu}{s_{\bar{x}}} = \frac{565{,}750 - 520{,}234}{10{,}636.3002} = 4.28$$

Step 5. Make a decision.

The value of the test statistic $z = 4.28$ is greater than the critical value of $z = 2.33$ and it falls in the rejection region. Consequently, we reject H_0. Therefore, we can state that the sample mean $\bar{x} = \$565{,}750$ is too far from the hypothesized population mean $\mu = \$520{,}234$. The difference between the two may not be attributed to chance or sampling error alone. Therefore, the current mean sale price of homes in the said neighborhood of Stamford, Connecticut, is higher than $520,234. ■

Conducting a left-tailed test of hypothesis about μ for a large sample.

EXAMPLE 9-3 Because couples are deciding to have fewer children, the family size in the United States has decreased over the past few decades. According to the U.S. Bureau of the Census, the mean family size was 3.18 in 1998. A researcher wanted to check if the current mean family size is less than 3.18. A sample of 900 families taken this year by this researcher produced a mean family size of 3.16 with a standard deviation of .70. Using the .025 significance level, can we conclude that the mean family size has decreased since 1998?

Solution Let μ be the current mean size of all families and $\bar{x}$ the mean family size for the sample. From the given information,

$$n = 900, \quad \bar{x} = 3.16, \quad \text{and} \quad s = .70$$

The mean family size for 1998 is given to be 3.18. The significance level α is .025.

Step 1. State the null and alternative hypotheses.

Notice that we are testing for a decrease in the mean family size. The null and alternative hypotheses are written as follows.

H_0: $\mu = 3.18$ (The mean family size has not decreased)

H_1: $\mu < 3.18$ (The mean family size has decreased)

Step 2. Select the distribution to use.

Because the sample size is large ($n > 30$), the sampling distribution of $\bar{x}$ is (approximately) normal. Consequently, we use the normal distribution to make the test.

Step 3. Determine the rejection and nonrejection regions.

The significance level is .025. The $<$ sign in the alternative hypothesis indicates that the test is left-tailed with the rejection region in the left tail of the sampling distribution curve of $\bar{x}$. The critical value of z, obtained from the normal table for .4750, is -1.96, as shown in Figure 9.7.

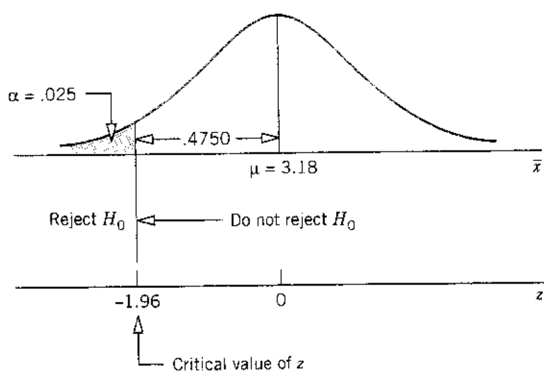


Figure 9.7

Step 4. Calculate the value of the test statistic.

The value of the test statistic z for $\bar{x} = 3.16$ is calculated as follows:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{.70}{\sqrt{900}} = .02333333$$

From H_0

$$z = \frac{\bar{x} - \mu}{s_{\bar{x}}} = \frac{3.16 - 3.18}{.02333333} = -.86$$

Step 5. Make a decision.

The value of the test statistic $z = -.86$ is greater than the critical value of $z = -1.96$, and it falls in the nonrejection region. As a result, we fail to reject H_0. Consequently, we can state that based on the sample information, it appears that the mean family size has not decreased since 1998. Note that we are not concluding that the mean family size has definitely not decreased. By not rejecting the null hypothesis, we are saying that the information obtained from the sample is not strong enough to reject the null hypothesis and to conclude that the family size has decreased since 1998. ■

In studies published in various journals, authors usually use the terms *significantly different* and *not significantly different* when deriving conclusions based on hypothesis tests. These terms are short versions of the terms *statistically significantly different* and *statistically not significantly different*. The expression *significantly different* means that the difference between the observed value of the sample mean $\bar{x}$ and the hypothesized value of the population mean μ is so large that it probably did not occur because of the sampling error alone. Consequently, the null hypothesis is rejected. In other words, the difference between $\bar{x}$ and μ is statistically significant. Thus, the statement *significantly different* is equivalent to saying that the *null hypothesis is rejected*. In Example 9–2, we can state as a conclusion that the observed value of $\bar{x} = \$565{,}750$ is significantly different from the hypothesized value of $\mu = \$520{,}234$. That is, the current mean sale price of neighborhood homes is significantly different from \$520,234.

On the other hand, the statement *not significantly different* means that the difference between the observed value of the sample mean $\bar{x}$ and the hypothesized value of the population mean μ is so small that it may have occurred just because of chance. Consequently, the null hypothesis is not rejected. Thus, the expression *not significantly different* is equivalent to saying that we *fail to reject the null hypothesis*. In Example 9–3, we can state as a conclusion that the observed value of $\bar{x} = 3.16$ is not significantly different from the hypothesized value of $\mu = 3.18$. In other words, the current mean family size does not seem to be significantly different from 3.18.

EXERCISES

■ Concepts and Procedures

9.11 What are the five steps of a test of hypothesis? Explain briefly.

9.12 What does the level of significance represent in a test of hypothesis? Explain.

9.13 By rejecting the null hypothesis in a test of hypothesis example, are you stating that the alternative hypothesis is true?

9.14 What is the difference between the critical value of z and the observed value of z?

9.15 For each of the following examples of tests of hypotheses about μ, show the rejection and nonrejection regions on the sampling distribution of the sample mean.

a. A two-tailed test with $\alpha = .05$ and $n = 40$
b. A left-tailed test with $\alpha = .01$ and $n = 67$
c. A right-tailed test with $\alpha = .02$ and $n = 55$

9.16 For each of the following examples of tests of hypotheses about μ, show the rejection and nonrejection regions on the sampling distribution of the sample mean.

a. A two-tailed test with $\alpha = .01$ and $n = 100$
b. A left-tailed test with $\alpha = .005$ and $n = 60$
c. A right-tailed test with $\alpha = .025$ and $n = 36$

9.17 Consider the following null and alternative hypotheses:

$$H_0: \mu = 25 \quad \text{versus} \quad H_1: \mu \neq 25$$

Suppose you perform this test at $\alpha = .05$ and reject the null hypothesis. Would you state that the difference between the hypothesized value of the population mean and the observed value of the sample mean is "statistically significant" or would you state that this difference is "statistically not significant"? Explain.

9.18 Consider the following null and alternative hypotheses:

$$H_0: \mu = 60 \quad \text{versus} \quad H_1: \mu > 60$$

Suppose you perform this test at $\alpha = .01$ and fail to reject the null hypothesis. Would you state that the difference between the hypothesized value of the population mean and the observed value of the sample mean is "statistically significant" or would you state that this difference is "statistically not significant"? Explain.

9.19 For each of the following significance levels, what is the probability of making a Type I error?

a. $\alpha = .025$ b. $\alpha = .05$ c. $\alpha = .01$

9.20 For each of the following significance levels, what is the probability of making a Type I error?

a. $\alpha = .10$ b. $\alpha = .02$ c. $\alpha = .005$

9.21 A random sample of 120 observations produced a sample mean of 32 and a standard deviation of 6. Find the critical and observed values of z for each of the following tests of hypotheses using $\alpha = .05$.

a. $H_0: \mu = 28$ versus $H_1: \mu > 28$
b. $H_0: \mu = 28$ versus $H_1: \mu \neq 28$

9.22 A random sample of 90 observations produced a sample mean of 15 and a standard deviation of 4. Find the critical and observed values of z for each of the following tests of hypotheses using $\alpha = .01$.

a. $H_0: \mu = 20$ versus $H_1: \mu < 20$
b. $H_0: \mu = 20$ versus $H_1: \mu \neq 20$

9.23 Consider the null hypothesis $H_0: \mu = 50$. Suppose a random sample of 120 observations is taken to perform this test. Using $\alpha = .05$, show the rejection and nonrejection regions on the sampling distribution curve of the sample mean and find the critical value(s) of z when the alternative hypothesis is

a. $H_1: \mu < 50$ b. $H_1: \mu \neq 50$ c. $H_1: \mu > 50$

9.24 Consider the null hypothesis $H_0: \mu = 35$. Suppose a random sample of 70 observations is taken to perform this test. Using $\alpha = .01$, show the rejection and nonrejection regions on the sampling distribution curve of the sample mean and find the critical value(s) of z for a

a. left-tailed test b. two-tailed test c. right-tailed test

9.25 Consider H_0: $\mu = 100$ versus H_1: $\mu \neq 100$.

a. A random sample of 64 observations produced a sample mean of 98 and a standard deviation of 12. Using $\alpha = .01$, would you reject the null hypothesis?

b. Another random sample of 64 observations taken from the same population produced a sample mean of 104 and a standard deviation of 10. Using $\alpha = .01$, would you reject the null hypothesis?

Comment on the results of parts a and b.

9.26 Consider H_0: $\mu = 45$ versus H_1: $\mu < 45$.

a. A random sample of 100 observations produced a sample mean of 43 and a standard deviation of 5. Using $\alpha = .025$, would you reject the null hypothesis?

b. Another random sample of 100 observations taken from the same population produced a sample mean of 43.8 and a standard deviation of 7. Using $\alpha = .025$, would you reject the null hypothesis?

Comment on the results of parts a and b.

9.27 Make the following tests of hypotheses.

a.	H_0: $\mu = 25$,	H_1: $\mu \neq 25$,	$n = 81$,	$\bar{x} = 28.5$,	$s = 3$,	$\alpha = .01$
b.	H_0: $\mu = 12$,	H_1: $\mu < 12$,	$n = 45$,	$\bar{x} = 11.25$,	$\sigma = 4.5$,	$\alpha = .05$
c.	H_0: $\mu = 40$,	H_1: $\mu > 40$,	$n = 100$,	$\bar{x} = 47$,	$s = 7$,	$\alpha = .10$

9.28 Make the following test of hypotheses.

a.	H_0: $\mu = 80$,	H_1: $\mu \neq 80$,	$n = 33$,	$\bar{x} = 76.5$,	$\sigma = 15$,	$\alpha = .10$
b.	H_0: $\mu = 32$,	H_1: $\mu < 32$,	$n = 75$,	$\bar{x} = 26.5$,	$s = 7.4$,	$\alpha = .01$
c.	H_0: $\mu = 55$,	H_1: $\mu > 55$,	$n = 40$,	$\bar{x} = 60.5$,	$s = 4$,	$\alpha = .05$

Applications

9.29 According to data from the U.S. Center for Health Statistics, Americans aged 18–24 made an annual average of 3.9 visits to physicians (*Statistical Abstract of the United States*, 1998). Last year a random sample of 350 Americans in this age group showed a mean of 3.7 visits per person with a standard deviation of 1.6 visits. Test at the 1% significance level whether the mean number of visits to physicians last year by Americans in this age group differed from 3.9.

9.30 According to data from the U.S. Health Care Financing Administration, the average annual expenditure on health care is \$3645 per person in the United States (*Statistical Abstract of the United States*, 1998). A random sample of 200 persons showed that they spent an average of \$3950 on health care last year with a standard deviation of \$1450. Test at the 2% significance level whether last year's mean annual health expenditure per person for all people in the United States was greater than \$3645.

9.31 According to 1999 data from Bruskin-Goldring for Goodyear/Gemini Automotive Care, the average age of the primary vehicles (which include cars, vans, pickups, and SUVs) owned by car owners in the United States was 5.6 years (*USA TODAY*, November 17, 1999). Assume that this result holds true for all primary vehicles in the United States in 1999. A recent random sample of 250 car owners from the United States showed that the average age of their primary vehicles is 6 years with a standard deviation of .75 year. Using the 2.5% significance level, can you conclude that the mean age of such cars has changed since the 1999 report?

9.32 A 1998 news article stated that the consumption of artificial sweeteners in the United States was equivalent to about 24 pounds of sugar per person per year (*Newsweek*, July 13, 1998). Suppose that a random sample of 150 Americans taken this year showed an annual average consumption of artificial sweeteners equivalent to 27 pounds of sugar with a standard deviation of 9.0 pounds. Does the sample support the alternative hypothesis that the current mean annual consumption of artificial sweeteners is more than 24 pounds of sugar? Use $\alpha = .01$. Explain your conclusion.

9.33 According to a 1997 survey of public participation in the arts, Americans who used personal computers at home during their free time averaged 5.2 hours per week on their computers (*American Demographics*, August 1998). A recent random sample of 50 Americans who use computers at home during their free time showed that the mean time these individuals spent on computers is 6.6 hours with a standard deviation of 2.3 hours.

a. At the 5% level of significance, can you conclude that the mean time spent on their computers during their free time at home by all such Americans currently exceeds 5.2 hours per week?
b. What is the Type I error in this case? Explain. What is the probability of making this error?

9.34 According to data from the College Board, the average cost per student for books and supplies for the 1997–1998 academic year was $634 at public four-year colleges in the United States (*The Chronicle of Higher Education*, August 1998). A recent random sample of 340 students at such colleges yielded a mean cost for books and supplies of $683 with a standard deviation of $254.

a. Testing at the 1% significance level, can you conclude that the mean of such costs currently differs from $634?
b. What is the Type I error in this case? Explain. What is the probability of making this error?

9.35 A study conducted a few years ago claims that adult men spend an average of 11 hours a week watching sports on television. A recent sample of 100 adult men showed that the mean time they spend per week watching sports on television is 9 hours with a standard deviation of 2.2 hours.

a. Test at the 1% significance level whether currently all adult men spend less than 11 hours per week watching sports on television.
b. What will your decision be in part a if the probability of making a Type I error is zero? Explain.

9.36 A restaurant franchise company has a policy of opening new restaurants only in those areas that have a mean household income of at least $35,000 per year. The company is currently considering an area in which to open a new restaurant. The company's research department took a sample of 150 households from this area and found that the mean income of these households is $33,400 per year with a standard deviation of $5400.

a. Using the 1% significance level, would you conclude that the company should not open a restaurant in this area?
b. What will your decision be in part a if the probability of making a Type I error is zero? Explain.

9.37 The manufacturer of a certain brand of auto batteries claims that the mean life of these batteries is 45 months. A consumer protection agency that wants to check this claim took a random sample of 36 such batteries and found that the mean life for this sample is 43.75 months with a standard deviation of 4 months.

a. Using the 2.5% significance level, would you conclude that the mean life of these batteries is less than 45 months?
b. Make the test of part a using a 5% significance level. Is your decision different from the one in part a? Comment on the results of parts a and b.

9.38 A study claims that all adults spend an average of 8 hours or more on chores during a weekend. A researcher wanted to check if this claim is true. A random sample of 200 adults taken by this researcher showed that these adults spend an average of 7.68 hours on chores during a weekend with a standard deviation of 2.1 hours.

a. Using the 1% significance level, can you conclude that the claim that all adults spend an average of 8 hours or more on chores during a weekend is false?
b. Make the test of part a using a 2.5% significance level. Is your decision different from the one in part a? Comment on the results of parts a and b.

9.39 Lazurus Steel Corporation produces iron rods that are supposed to be 36 inches long. The machine that makes these rods does not produce each rod exactly 36 inches long. The lengths of the rods vary slightly. It is known that when the machine is working properly, the mean length of the rods is 36 inches. The standard deviation of the lengths of all rods produced on this machine is always equal to .05 inch. The quality control department at the company takes a sample of 40 such rods each week, calculates the mean length of these rods, and tests the null hypothesis $\mu = 36$ inches against the alternative hypothesis $\mu \neq 36$ inches using a 1% significance level. If the null hypothesis is rejected, the machine is stopped and adjusted. A recent sample of 40 such rods produced a mean length of 36.015 inches. Based on this sample, would you conclude that the machine needs an adjustment?

9.40 At Farmer's Dairy, a machine is set to fill 32-ounce milk cartons. However, this machine does not put exactly 32 ounces of milk into each carton; the amount varies slightly from carton to carton.

It is known that when the machine is working properly, the mean net weight of these cartons is 32 ounces. The standard deviation of the milk in all such cartons is always equal to .15 ounce. The quality control inspector at this dairy takes a sample of 35 such cartons each week, calculates the mean net weight of these cartons, and tests the null hypothesis $\mu = 32$ ounces against the alternative hypothesis $\mu \neq 32$ ounces using a 2% significance level. If the null hypothesis is rejected, the machine is stopped and adjusted. A recent sample of 35 such cartons produced a mean net weight of 31.90 ounces. Based on this sample, would you conclude that the machine needs to be adjusted?

9.41 A company claims that the mean net weight of the contents of its All Taste cereal boxes is at least 18 ounces. Suppose you want to test whether or not the claim of the company is true. Explain briefly how you would conduct this test using a large sample.

9.42 A researcher claims that college students spend an average of 45 minutes per week on community service. You want to test if the mean time spent per week on community service by college students is different from 45 minutes. Explain briefly how you would conduct this test using a large sample.

9.3 HYPOTHESIS TESTS USING THE *p*-VALUE APPROACH

In the discussion of tests of hypotheses in Section 9.2, the value of the significance level α was selected before the test was performed. Sometimes we may prefer not to predetermine α. Instead, we may want to find a value such that a given null hypothesis will be rejected for any α greater than this value and it will not be rejected for any α less than this value. The **probability-value approach**, more commonly called the *p-value approach*, gives such a value. In this approach, we calculate the ***p*-value** for the test, which is defined as the smallest level of significance at which the given null hypothesis is rejected.

> ***p*-VALUE** The *p-value* is the smallest significance level at which the null hypothesis is rejected.

Using the p-value approach, we reject the null hypothesis if

$$p\text{-value} < \alpha$$

and we do not reject the null hypothesis if

$$p\text{-value} \geq \alpha$$

For a one-tailed test, the p-value is given by the area in the tail of the sampling distribution curve beyond the observed value of the sample statistic. Figure 9.8 shows the p-value for a right-tailed test about μ.

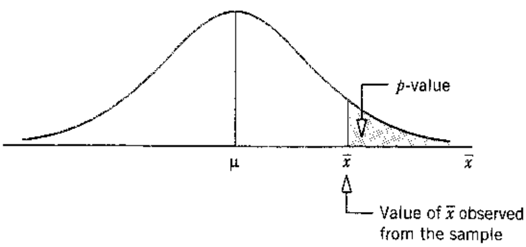


Figure 9.8 The p-value for a right-tailed test.

For a two-tailed test, the p-value is twice the area in the tail of the sampling distribution curve beyond the observed value of the sample statistic. Figure 9.9 shows the p-value for a two-tailed test. Each of the areas in the two tails gives one-half the p-value.

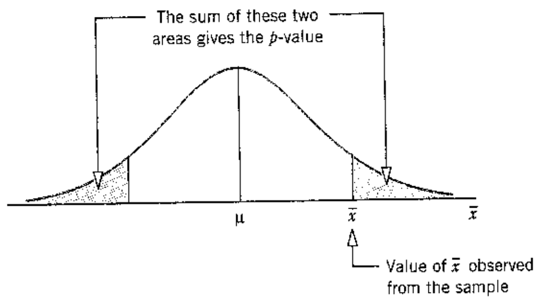


Figure 9.9 The p-value for a two-tailed test.

Examples 9–4 and 9–5 illustrate the calculation and use of the p-value.

Calculating the p-value for a one-tailed test of hypothesis.

EXAMPLE 9–4 The management of Priority Health Club claims that its members lose an average of 10 pounds or more within the first month after joining the club. A consumer agency that wanted to check this claim took a random sample of 36 members of this health club and found that they lost an average of 9.2 pounds within the first month of membership with a standard deviation of 2.4 pounds. Find the p-value for this test.

Solution Let μ be the mean weight lost during the first month of membership by all members of this health club, and let $\bar{x}$ be the corresponding mean for the sample. From the given information,

$$n = 36, \quad \bar{x} = 9.2 \text{ pounds}, \quad \text{and} \quad s = 2.4 \text{ pounds}$$

The claim of the club is that its members lose, on average, 10 pounds or more within the first month of membership. To calculate the p-value, we apply the following three steps.

Step 1. *State the null and alternative hypotheses.*

H_0: $\mu \geq 10$ (The mean weight lost is 10 pounds or more)

H_1: $\mu < 10$ (The mean weight lost is less than 10 pounds)

Step 2. *Select the distribution to use.*

Because the sample size is large, we use the normal distribution to make the test and to calculate the p-value.

Step 3. *Calculate the p-value.*

The < sign in the alternative hypothesis indicates that the test is left-tailed. The p-value is given by the area to the left of $\bar{x} = 9.2$ under the sampling distribution curve of $\bar{x}$, as shown in Figure 9.10. To find this area, we first find the z value for $\bar{x} = 9.2$ as follows:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{2.4}{\sqrt{36}} = .40$$

$$z = \frac{\bar{x} - \mu}{s_{\bar{x}}} = \frac{9.2 - 10}{.40} = -2.00$$

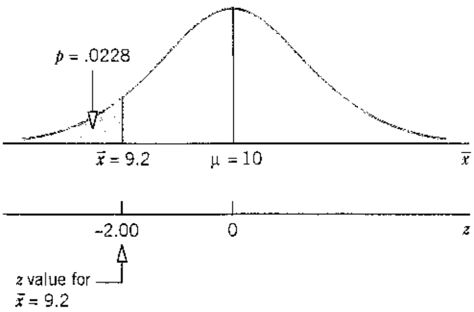


Figure 9.10 The p-value for a left-tailed test.

The area to the left of $\bar{x} = 9.2$ under the sampling distribution of $\bar{x}$ is equal to the area under the standard normal curve to the left of $z = -2.00$. From the normal distribution table, the area between the mean and $z = -2.00$ is .4772. Hence, the area to the left of $z = -2.00$ is $.5 - .4772 = .0228$. Consequently,

$$p\text{-value} = \mathbf{.0228}$$

Thus, based on the p-value of .0228 we can state that for any α (significance level) greater than .0228 we will reject the null hypothesis stated in Step 1 and for any α less than .0228 we will not reject the null hypothesis. Suppose we make the test for this example at $\alpha = .01$. Because $\alpha = .01$ is less than the p-value of .0228, we will not reject the null hypothesis. Now, suppose we make the test at $\alpha = .05$. This time, because $\alpha = .05$ is greater than the p-value of .0228, we will reject the null hypothesis. ■

The reader should make the test of hypothesis for Example 9–4 at $\alpha = .01$ and at $\alpha = .05$ by using the five steps learned in Section 9.2. The null hypothesis will not be rejected at $\alpha = .01$ (as .01 is less than $p = .0228$), and the null hypothesis will be rejected at $\alpha = .05$ (as .05 is greater than $p = .0228$).

Calculating the p-value for a two-tailed test of hypothesis.

EXAMPLE 9–5 At Canon Food Corporation, it used to take an average of 50 minutes for new workers to learn a food processing job. Recently the company installed a new food processing machine. The supervisor at the company wants to find if the mean time taken by new workers to learn the food processing procedure on this new machine is different from 50 minutes. A sample of 40 workers showed that it took, on average, 47 minutes for them to learn the food processing procedure on the new machine with a standard deviation of 7 minutes. Find the p-value for the test that the mean learning time for the food processing procedure on the new machine is different from 50 minutes.

Solution Let μ be the mean time (in minutes) taken to learn the food processing procedure on the new machine by all workers, and let $\bar{x}$ be the corresponding sample mean. From the given information,

$$n = 40, \quad \bar{x} = 47 \text{ minutes}, \quad \text{and} \quad s = 7 \text{ minutes}$$

To calculate the p-value, we apply the following three steps.

Step 1. *State the null and alternative hypotheses.*

$$H_0\text{: } \mu = 50 \text{ minutes}$$

$$H_1\text{: } \mu \neq 50 \text{ minutes}$$

Note that the null hypothesis states that the mean time for learning the food processing procedure on the new machine is 50 minutes, and the alternative hypothesis states that this time is different from 50 minutes.

Step 2. *Select the distribution to use.*

Because the sample size is large, we use the normal distribution to make the test and to calculate the p-value.

Step 3. *Calculate the p-value.*

The $\neq$ sign in the alternative hypothesis indicates that the test is two-tailed. The p-value is equal to twice the area in the tail of the sampling distribution curve of $\bar{x}$ to the left of $\bar{x} = 47$, as shown in Figure 9.11. To find this area, we first find the z value for $\bar{x} = 47$ as follows:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{7}{\sqrt{40}} = 1.10679718 \text{ minutes}$$

$$z = \frac{\bar{x} - \mu}{s_{\bar{x}}} = \frac{47 - 50}{1.10679718} = -2.71$$

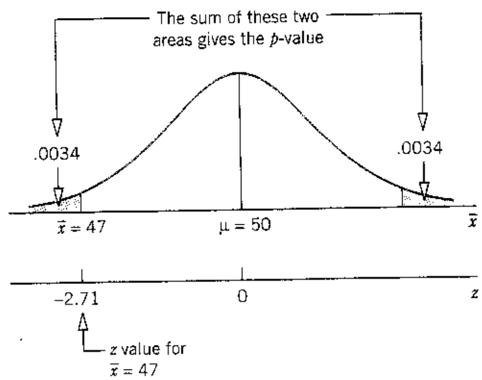


Figure 9.11 The p-value for a two-tailed test.

The area to the left of $\bar{x} = 47$ is equal to the area under the standard normal curve to the left of $z = -2.71$. From the normal distribution table, the area between the mean and $z = -2.71$ is .4966. Hence, the area to the left of $z = -2.71$ is

$$.5 - .4966 = .0034$$

Consequently, the p-value is

$$p\text{-value} = 2(.0034) = \mathbf{.0068}$$

Thus, based on the p-value of .0068, we conclude that for any α (significance level) greater than .0068 we will reject the null hypothesis and for any α less than .0068 we will not reject the null hypothesis. ∎

EXERCISES

Concepts and Procedures

9.43 Briefly explain the procedure used to calculate the p-value for a two-tailed and for a one-tailed test, respectively.

9.44 Find the p-value for each of the following hypothesis tests.

a. H_0: $\mu = 23$, H_1: $\mu \neq 23$, $n = 50$, $\bar{x} = 21.25$, $s = 5$
b. H_0: $\mu = 15$, H_1: $\mu < 15$, $n = 80$, $\bar{x} = 13.25$, $s = 5.5$
c. H_0: $\mu = 38$, H_1: $\mu > 38$, $n = 35$, $\bar{x} = 40.25$, $s = 7.2$

9.45 Find the p-value for each of the following hypothesis tests.

a. H_0: $\mu = 46$, H_1: $\mu \neq 46$, $n = 40$, $\bar{x} = 49.60$, $s = 9.7$
b. H_0: $\mu = 26$, H_1: $\mu < 26$, $n = 33$, $\bar{x} = 24.30$, $s = 4.3$
c. H_0: $\mu = 18$, H_1: $\mu > 18$, $n = 55$, $\bar{x} = 20.50$, $s = 7.8$

9.46 Consider H_0: $\mu = 29$ versus H_1: $\mu \neq 29$. A random sample of 60 observations taken from this population produced a sample mean of 31.4 and a standard deviation of 8.

a. Calculate the p-value.
b. Considering the p-value of part a, would you reject the null hypothesis if the test were made at the significance level of .05?
c. Considering the p-value of part a, would you reject the null hypothesis if the test were made at the significance level of .01?

9.47 Consider H_0: $\mu = 72$ versus H_1: $\mu > 72$. A random sample of 36 observations taken from this population produced a sample mean of 74.07 and a standard deviation of 6.

a. Calculate the p-value.
b. Considering the p-value of part a, would you reject the null hypothesis if the test were made at the significance level of .01?
c. Considering the p-value of part a, would you reject the null hypothesis if the test were made at the significance level of .025?

Applications

9.48 According to a forecast made in 1999, the average daily rate for a hotel room in the United States in the year 2000 would be \$85 (Smith Travel Research, *USA TODAY*, June 28, 1999). Suppose that a random sample of 81 hotel rooms in the year 2000 found a mean daily rate of \$88 with a standard deviation of \$12. Find the p-value for the test of hypothesis with the alternative hypothesis that the actual mean rate in the year 2000 differs from the forecast.

9.49 According to the Energy Information Administration of the U.S. Department of Energy, the average price of unleaded regular gasoline in the United States was 108.2 cents per gallon for the period January–June 1998 (*World Almanac and Book of Facts*, 1999). Suppose that a recent nationwide random sample of 200 gas stations found a mean price for unleaded regular gasoline of 106.5 cents per gallon with a standard deviation of 18.0 cents per gallon. Find the p-value for the hypothesis test with the alternative hypothesis that the mean price for unleaded gasoline currently differs from 108.2 cents per gallon.

9.50 The manufacturer of a certain brand of auto batteries claims that the mean life of these batteries is 45 months. A consumer protection agency that wants to check this claim took a random sample of 36 such batteries and found that the mean life for this sample is 43.75 months with a standard deviation of 4.5 months. Find the p-value for the test of hypothesis with the alternative hypothesis that the mean life of these batteries is less than 45 months.

9.51 A study claims that all adults spend an average of 14 hours or more on chores during a weekend. A researcher wanted to check if this claim is true. A random sample of 200 adults taken by this researcher showed that these adults spend an average of 13.75 hours on chores during a weekend with a standard deviation of 3.0 hours. Find the p-value for the hypothesis test with the alternative hypothesis that all adults spend less than 14 hours on chores during a weekend.

9.52 Data from the Consumer Expenditure Survey conducted by the U.S. Bureau of Labor Statistics showed that U.S. households spend an average of \$3028 per year on car loan or lease payments (*Money Magazine*, July 1999). Suppose that a recent random sample of 300 households yielded a mean annual expenditure for car loan or lease payments of \$3145 with a standard deviation of \$920.

a. Find the p-value for the test of hypothesis with the alternative hypothesis that the current mean annual household expenditure on car loan or lease payments exceeds \$3028.

b. If $\alpha = .01$, based on the p-value calculated in part a, would you reject the null hypothesis? Explain.
c. If $\alpha = .025$, based on the p-value calculated in part a, would you reject the null hypothesis? Explain.

9.53 A telephone company claims that the mean duration of all long-distance phone calls made by its residential customers is 10 minutes. A random sample of 100 long-distance calls made by its residential customers taken from the records of this company showed that the mean duration of calls for this sample is 9.25 minutes with a standard deviation of 3.75 minutes.

a. Find the p-value for the test that the mean duration of all long-distance calls made by residential customers is less than 10 minutes.
b. If $\alpha = .02$, based on the p-value calculated in part a, would you reject the null hypothesis? Explain.
c. If $\alpha = .05$, based on the p-value calculated in part a, would you reject the null hypothesis? Explain.

9.54 Lazurus Steel Corporation produces iron rods that are supposed to be 36 inches long. The machine that makes these rods does not produce each rod exactly 36 inches long. The lengths of the rods vary slightly. It is known that when the machine is working properly, the mean length of the rods is 36 inches. The standard deviation of the lengths of all rods produced on this machine is always equal to .05 inch. The quality control department at the company takes a sample of 40 such rods every week, calculates the mean length of these rods, and tests the null hypothesis $\mu = 36$ inches against the alternative hypothesis $\mu \neq 36$ inches. If the null hypothesis is rejected, the machine is stopped and adjusted. A recent such sample of 40 rods produced a mean length of 36.015 inches.

a. Calculate the p-value for this test of hypothesis.
b. Based on the p-value calculated in part a, will the quality control inspector decide to stop the machine and adjust it if he chooses the maximum probability of a Type I error to be .02? What if the maximum probability of a Type I error is .10?

9.55 At Farmer's Dairy, a machine is set to fill 32-ounce milk cartons. However, this machine does not put exactly 32 ounces of milk into each carton; the amount varies slightly from carton to carton. It is known that when the machine is working properly, the mean net weight of these cartons is 32 ounces. The standard deviation of the milk in all such cartons is always equal to .15 ounce. The quality control inspector at this company takes a sample of 35 such cartons every week, calculates the mean net weight of these cartons, and tests the null hypothesis $\mu = 32$ ounces against the alternative hypothesis $\mu \neq 32$ ounces. If the null hypothesis is rejected, the machine is stopped and adjusted. A recent sample of 35 such cartons produced a mean net weight of 31.90 ounces.

a. Calculate the p-value for this test of hypothesis.
b. Based on the p-value calculated in part a, will the quality control inspector decide to stop the machine and readjust it if she chooses the maximum probability of a Type I error to be .01? What if the maximum probability of a Type I error is .05?

9.4 HYPOTHESIS TESTS ABOUT A POPULATION MEAN: SMALL SAMPLES

Many times the size of a sample that is used to make a test of hypothesis about μ is small—that is, $n < 30$. This may be the case because we have limited resources and cannot afford to take a large sample or because of the nature of the experiment itself. For example, to test a new model of a car for fuel efficiency (miles per gallon), the company may prefer to use a small sample. All cars included in such a test must be sold as used cars. In the case of a small sample, if the population from which the sample is drawn is (approximately) normally distributed and the population standard deviation σ is known, we can still use the normal distribution to make a test of hypothesis about μ. However, if the population is (approximately) normally distributed, the population standard deviation σ is not known, and the sample size is small ($n < 30$), then the normal distribution is replaced by the t distribution to make a test of hypothesis about μ. In such a case the random variable

$$t = \frac{\bar{x} - \mu}{s_{\bar{x}}} \qquad \text{where } s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

has a t distribution. The t is called the **test statistic** to make a hypothesis test about a population mean for small samples.

CONDITIONS UNDER WHICH THE t DISTRIBUTION IS USED TO MAKE TESTS OF HYPOTHESIS ABOUT μ The t distribution is used to conduct a *test of hypothesis about* μ if

1. The sample size is small ($n < 30$).
2. The population from which the sample is drawn is (approximately) normally distributed.
3. The population standard deviation σ is unknown.

The procedure that is used to make hypothesis tests about μ in the case of small samples is similar to the one for large samples. We perform the same five steps with the only difference being the use of the t distribution in place of the normal distribution.

TEST STATISTIC The value of the *test statistic* t for the sample mean $\bar{x}$ is computed as

$$t = \frac{\bar{x} - \mu}{s_{\bar{x}}} \qquad \text{where } s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

The value of t calculated for $\bar{x}$ by using the above formula is also called the **observed value** of t.

Examples 9–6, 9–7, and 9–8 describe the procedure of testing hypotheses about the population mean using the t distribution.

Conducting a two-tailed test of hypothesis about μ: $n < 30$.

EXAMPLE 9–6 A psychologist claims that the mean age at which children start walking is 12.5 months. Carol wanted to check if this claim is true. She took a random sample of 18 children and found that the mean age at which these children started walking was 12.9 months with a standard deviation of .80 month. Using the 1% significance level, can you conclude that the mean age at which all children start walking is different from 12.5 months? Assume that the ages at which all children start walking have an approximately normal distribution.

Solution Let μ be the mean age at which all children start walking and $\bar{x}$ the corresponding mean for the sample. Then, from the given information,

$$n = 18, \quad \bar{x} = 12.9 \text{ months}, \quad s = .80 \text{ month}, \quad \text{and} \quad \alpha = .01$$

Step 1. *State the null and alternative hypotheses.*

We are to test if the mean age at which all children start walking is different from 12.5 months. The null and alternative hypotheses are

H_0: $\mu = 12.5$ (The mean walking age is 12.5 months)

H_1: $\mu \neq 12.5$ (The mean walking age is different from 12.5 months)

Step 2. *Select the distribution to use.*

The sample size is small and the population is approximately normally distributed. However, we do not know the population standard deviation σ. Hence, we use the t distribution to make the test.

Step 3. *Determine the rejection and nonrejection regions.*

The significance level is .01. The $\neq$ sign in the alternative hypothesis indicates that the test is two-tailed and the rejection region lies in both tails. The area of the rejection region in each tail of the t distribution curve is

$$\text{Area in each tail} = \alpha/2 = .01/2 = .005$$

$$df = n - 1 = 18 - 1 = 17$$

From the t distribution table, the critical values of t for 17 degrees of freedom and .005 area in each tail of the t distribution curve are -2.898 and 2.898. These values are shown in Figure 9.12.

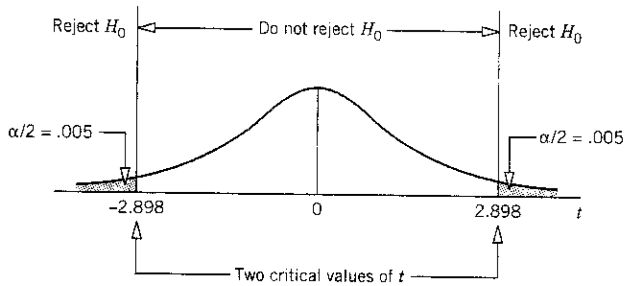


Figure 9.12

Step 4. *Calculate the value of the test statistic.*

We calculate the value of the test statistic t for $\bar{x} = 12.9$ as follows:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{.80}{\sqrt{18}} = .18856181$$

From H_0

$$t = \frac{\bar{x} - \mu}{s_{\bar{x}}} = \frac{12.9 - 12.5}{.18856181} = 2.121$$

Step 5. *Make a decision.*

The value of the test statistic $t = 2.121$ falls between the two critical points, -2.898 and 2.898, which is the nonrejection region. Consequently, we fail to reject H_0. As a result, we can state that the difference between the hypothesized population mean and the sample mean is so small that it may have occurred because of sampling error. The mean age at which children start walking is not different from 12.5 months. ∎

USING THE *p*-VALUE APPROACH IN EXAMPLE 9–6

Note that using the procedure discussed in Section 9.3, we can use the p-value approach to make a decision in all problems relating to hypothesis testing in this and succeeding chapters. For instance, we can find the p-value for $\bar{x} = 12.9$ in Example 9–6 and compare it with the given significance level to make a decision. As shown in Step 4 of Example 9–6, the t value for $\bar{x} = 12.9$ is 2.121. From the t distribution table, for $df = 17$ and $t = 2.121$, the p-value for a two-tailed test is approximately .05. (Note that this p-value for $df = 17$ and $t = 2.110$ is $2 \times .025 = .05$. However, for the same degrees of freedom but $t = 2.121$, this p-value will be slightly less than .05.) Since $\alpha = .01$ is less than the p-value of .05, we fail to reject the null hypothesis.

The same procedure can be used to obtain the p-value for the test of hypothesis problems in Section 9.5 and in succeeding chapters. To do so, we find the value of the test statistic for the given value of the sample statistic obtained from the sample and then obtain the

p-value from the corresponding probability distribution table for that value of the test statistic. Finally, we compare that p-value with the significance level and make a decision.

All the statistical software packages, including MINITAB, give the p-value in the solution to a test of hypothesis problem. Thus, if you are using a statistical software package to solve a test of hypothesis problem, you can compare the p-value given in the computer solution to the significance level and make a decision.

Conducting a left-tailed test of hypothesis about μ: $n < 30$.

EXAMPLE 9-7 Grand Auto Corporation produces auto batteries. The company claims that its top-of-the-line Never Die batteries are good, on average, for at least 65 months. A consumer protection agency tested 15 such batteries to check this claim. It found the mean life of these 15 batteries to be 63 months with a standard deviation of 2 months. At the 5% significance level, can you conclude that the claim of the company is true? Assume that the life of such a battery has an approximately normal distribution.

Solution Let μ be the mean life of all Never Die batteries and $\bar{x}$ the corresponding mean for the sample. Then, from the given information,

$$n = 15, \quad \bar{x} = 63 \text{ months}, \quad \text{and} \quad s = 2 \text{ months}$$

The significance level is $\alpha = .05$. The company's claim is that the mean life of these batteries is at least 65 months.

Step 1. *State the null and alternative hypotheses.*

We are to test whether or not the mean life of Never Die batteries is at least 65 months. The null and alternative hypotheses are as follows:

$$H_0: \mu \geq 65 \quad \text{(The mean life is at least 65 months)}$$

$$H_1: \mu < 65 \quad \text{(The mean life is less than 65 months)}$$

Step 2. *Select the distribution to use.*

The sample size is small ($n < 30$) and the life of a battery is approximately normally distributed. However, the population standard deviation is not known. Hence, we use the t distribution to make the test.

Step 3. *Determine the rejection and nonrejection regions.*

The significance level is .05. The < sign in the alternative hypothesis indicates that the test is left-tailed with the rejection region in the left tail of the t distribution curve. To find the critical value of t, we need to know the area in the left tail and the degrees of freedom.

$$\text{Area in the left tail} = \alpha = .05$$

$$df = n - 1 = 15 - 1 = 14$$

From the t distribution table, the critical value of t for 14 degrees of freedom and an area of .05 in the left tail is -1.761. This value is shown in Figure 9.13.

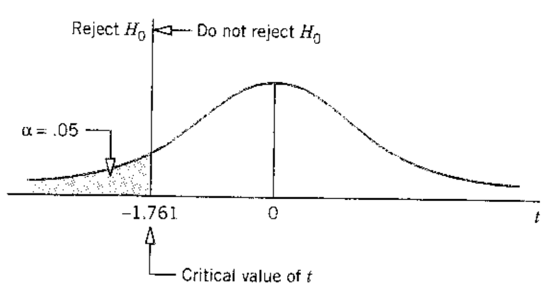


Figure 9.13

Step 4. *Calculate the value of the test statistic.*

The value of the test statistic t for $\bar{x} = 63$ is calculated as follows:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{2}{\sqrt{15}} = .51639778$$

$$t = \frac{\bar{x} - \mu}{s_{\bar{x}}} = \frac{63 - \overset{\text{From } H_0}{\overbrace{65}}}{.51639778} = -3.873$$

Step 5. *Make a decision.*

The value of the test statistic $t = -3.873$ is less than the critical value of $t = -1.761$, and it falls in the rejection region. Therefore, we reject H_0 and conclude that the sample mean is too small compared to 65 (company's claimed value of μ) and the difference between the two may not be attributed to chance alone. We can conclude that the mean life of the company's Never Die batteries is less than 65 months. ■

Conducting a right-tailed test of hypothesis about μ: $n < 30$.

EXAMPLE 9-8 The management at Massachusetts Savings Bank is always concerned about the quality of service provided to its customers. With the old computer system, a teller at this bank could serve, on average, 22 customers per hour. The management noticed that with this service rate, the waiting time for customers was too long. Recently the management of the bank installed a new computer system in the bank, expecting that it would increase the service rate and consequently make the customers happier by reducing the waiting time. To check if the new computer system is more efficient than the old system, the management of the bank took a random sample of 18 hours and found that during these hours the mean number of customers served by tellers was 28 per hour with a standard deviation of 2.5. Testing at the 1% significance level, would you conclude that the new computer system is more efficient than the old computer system? Assume that the number of customers served per hour by a teller on this computer system has an approximately normal distribution.

Solution Let μ be the mean number of customers served per hour by a teller using the new system, and let $\bar{x}$ be the corresponding mean for the sample. Then, from the given information,

$$n = 18 \text{ hours}, \quad \bar{x} = 28 \text{ customers}, \quad s = 2.5 \text{ customers}, \quad \text{and} \quad \alpha = .01$$

Step 1. *State the null and alternative hypotheses.*

We are to test whether or not the new computer system is more efficient than the old system. The new computer system will be more efficient than the old system if the mean number of customers served per hour by using the new computer system is significantly more than 22; otherwise, it will not be more efficient. The null and alternative hypotheses are

H_0: $\mu = 22$ (The new computer system is not more efficient)

H_1: $\mu > 22$ (The new computer system is more efficient)

Step 2. *Select the distribution to use.*

The sample size is small and the population is approximately normally distributed. However, we do not know the population standard deviation σ. Hence, we use the t distribution to make the test.

Step 3. *Determine the rejection and nonrejection regions.*

The significance level is .01. The > sign in the alternative hypothesis indicates that the test is right-tailed and the rejection region lies in the right tail of the t distribution curve.

$$\text{Area in the right tail} = \alpha = .01$$

$$df = n - 1 = 18 - 1 = 17$$

From the t distribution table, the critical value of t for 17 degrees of freedom and .01 area in the right tail is 2.567. This value is shown in Figure 9.14.

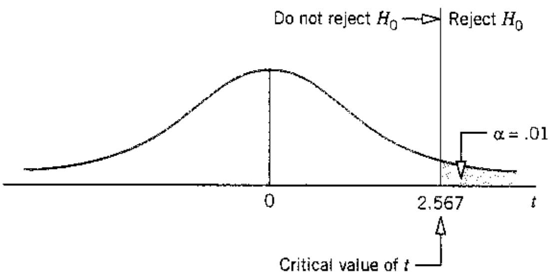


Figure 9.14

Step 4. *Calculate the value of the test statistic.*

The value of the test statistic t for $\bar{x} = 28$ is calculated as follows:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{2.5}{\sqrt{18}} = .58925565$$

$$t = \frac{\bar{x} - \mu}{s_{\bar{x}}} = \frac{28 - \overbrace{22}^{\text{From } H_0}}{.58925565} = 10.182$$

Step 5. *Make a decision.*

The value of the test statistic $t = 10.182$ is greater than the critical value of $t = 2.567$, and it falls in the rejection region. Consequently, we reject H_0. As a result, we conclude that the value of the sample mean is too large compared to the hypothesized value of the population mean, and the difference between the two may not be attributed to chance alone. The mean number of customers served per hour using the new computer system is more than 22. The new computer system is more efficient than the old computer system.

EXERCISES

Concepts and Procedures

9.56 Briefly explain the conditions that must hold true to use the t distribution to make a test of hypothesis about the population mean.

9.57 For each of the following examples of tests of hypotheses about μ, show the rejection and nonrejection regions on the t distribution curve.

a. A two-tailed test with $\alpha = .02$ and $n = 20$
b. A left-tailed test with $\alpha = .01$ and $n = 16$
c. A right-tailed test with $\alpha = .05$ and $n = 18$

9.58 For each of the following examples of tests of hypotheses about μ, show the rejection and nonrejection regions on the t distribution curve.

a. A two-tailed test with $\alpha = .01$ and $n = 15$
b. A left-tailed test with $\alpha = .005$ and $n = 25$
c. A right-tailed test with $\alpha = .025$ and $n = 22$

9.59 A random sample of 25 observations taken from a population that is normally distributed produced a sample mean of 58.5 and a standard deviation of 7.5. Find the critical and observed values of t for each of the following tests of hypotheses using $\alpha = .01$.

a. H_0: $\mu = 55$ versus H_1: $\mu > 55$
b. H_0: $\mu = 55$ versus H_1: $\mu \neq 55$

9.60 A random sample of 16 observations taken from a population that is normally distributed produced a sample mean of 42.4 and a standard deviation of 8. Find the critical and observed values of t for each of the following tests of hypotheses using $\alpha = .05$.

a. H_0: $\mu = 46$ versus H_1: $\mu < 46$
b. H_0: $\mu = 46$ versus H_1: $\mu \neq 46$

9.61 Consider the null hypothesis H_0: $\mu = 70$ about the mean of a population that is normally distributed. Suppose a random sample of 20 observations is taken from this population to make this test. Using $\alpha = .01$, show the rejection and nonrejection regions and find the critical value(s) of t for a

a. left-tailed test b. two-tailed test c. right-tailed test

9.62 Consider the null hypothesis H_0: $\mu = 35$ about the mean of a population that is normally distributed. Suppose a random sample of 22 observations is taken from this population to make this test. Using $\alpha = .05$, show the rejection and nonrejection regions and find the critical value(s) of t for a

a. left-tailed test b. two-tailed test c. right-tailed test

9.63 Consider H_0: $\mu = 80$ versus H_1: $\mu \neq 80$ for a population that is normally distributed.

a. A random sample of 25 observations taken from this population produced a sample mean of 77 and a standard deviation of 8. Using $\alpha = .01$, would you reject the null hypothesis?
b. Another random sample of 25 observations taken from the same population produced a sample mean of 86 and a standard deviation of 6. Using $\alpha = .01$, would you reject the null hypothesis?

Comment on the results of parts a and b.

9.64 Consider H_0: $\mu = 40$ versus H_1: $\mu > 40$ for a population that is normally distributed.

a. A random sample of 16 observations taken from this population produced a sample mean of 45 and a standard deviation of 5. Using $\alpha = .025$, would you reject the null hypothesis?
b. Another random sample of 16 observations taken from the same population produced a sample mean of 41.9 and a standard deviation of 7. Using $\alpha = .025$, would you reject the null hypothesis?

Comment on the results of parts a and b.

9.65 Assuming that the respective populations are normally distributed, perform the following hypothesis tests.

a.	H_0: $\mu = 24$,	H_1: $\mu \neq 24$,	$n = 25$,	$\bar{x} = 28.5$,	$s = 4.9$,	$\alpha = .01$
b.	H_0: $\mu = 30$,	H_1: $\mu < 30$,	$n = 16$,	$\bar{x} = 27.5$,	$s = 6.6$,	$\alpha = .025$
c.	H_0: $\mu = 18$,	H_1: $\mu > 18$,	$n = 20$,	$\bar{x} = 22.5$,	$s = 8$,	$\alpha = .10$

9.66 Assuming that the respective populations are normally distributed, perform the following hypothesis tests.

a.	H_0: $\mu = 60$,	H_1: $\mu \neq 60$,	$n = 14$,	$\bar{x} = 57$,	$s = 9$,	$\alpha = .05$
b.	H_0: $\mu = 35$,	H_1: $\mu < 35$,	$n = 24$,	$\bar{x} = 28$,	$s = 5.4$,	$\alpha = .005$
c.	H_0: $\mu = 47$,	H_1: $\mu > 47$,	$n = 18$,	$\bar{x} = 50$,	$s = 6$,	$\alpha = .001$

Applications

9.67 According to a basketball coach, the mean height of all female college basketball players is 69.5 inches. A random sample of 25 such players produced a mean height of 70.25 inches with a standard deviation of 2.1 inches. Assuming that the heights of all female college basketball players are normally distributed, test at the 1% significance level whether their mean height is different from 69.5 inches.

9.68 According to the College Board, in 1999 the average SAT score in mathematics for students from Rhode Island was 499. Suppose that a recent random sample of 25 mathematics SAT scores from Rhode Island had a mean of 504 with a standard deviation of 105. Using the 5% significance level, can you conclude that the current mean SAT score in mathematics for students from Rhode Island exceeds 499? Assume that such SAT scores for all students from Rhode Island are normally distributed.

9.69 The president of a university claims that the mean time spent partying by all students at this university is not more than 7 hours per week. A random sample of 20 students taken from this university showed that they spent an average of 10.50 hours partying the previous week with a standard deviation of 2.3 hours. Assuming that the time spent partying by all students at this university is approximately normally distributed, test at the 2.5% significance level whether the president's claim is true. Explain your conclusion in words.

9.70 The mean balance of all checking accounts at a bank on December 31, 1999, was \$850. A random sample of 25 checking accounts taken recently from this bank gave a mean balance of \$780 with a standard deviation of \$230. Assume that the balances of all checking accounts at this bank are normally distributed. Using the 1% significance level, can you conclude that the mean balance of such accounts has decreased during this period? Explain your conclusion in words.

9.71 A soft-drink manufacturer claims that its 12-ounce cans do not contain, on average, more than 30 calories. A random sample of 16 cans of this soft drink, which were checked for calories, contained a mean of 32 calories with a standard deviation of 3 calories. Assume that the number of calories in 12-ounce soda cans is normally distributed. Does the sample information support the alternative hypothesis that the manufacturer's claim is false? Use a significance level of 5%. Explain your conclusion in words.

9.72 According to data from the National Public Transportation Survey, commuters who use public transportation in metropolitan areas with populations of 3 million or more average 43 minutes of travel time (*USA TODAY*, April 14, 1999). A recent random sample of 20 such commuters showed a mean travel time of 52 minutes with a standard deviation of 14 minutes. Using the 2% significance level, can you conclude that the mean travel time for all such commuters currently differs from 43 minutes? Assume that the travel times for all such commuters are normally distributed.

9.73 A paint manufacturing company claims that the mean drying time for its paints is not longer than 45 minutes. A random sample of 20 gallons of paints selected from the production line of this company showed that the mean drying time for this sample is 49.50 minutes with a standard deviation of 3 minutes. Assume that the drying times for these paints have a normal distribution.

a. Using the 1% significance level, would you conclude that the company's claim is true?

b. What is the Type I error in this exercise? Explain in words. What is the probability of making such an error?

9.74 A 1999 poll of 1014 adults by the National Sleep Foundation reported that these adults have an average of (about) 7 hours sleep per night during the workweek (*USA TODAY*, March 25, 1999). A recent random sample of 18 adults showed an average of 6.75 hours of sleep per night during the workweek with a standard deviation of .60 hour. Assume that such sleep times for all adults are normally distributed.

a. Using the 2.5% significance level, can you conclude that the mean sleep duration for all adults during the workweek is currently less than 7 hours?

b. What is the Type I error in this hypothesis test? Explain. What is the probability of making such an error?

9.75 A business school claims that students who complete a three-month typing course can type, on average, at least 1200 words an hour. A random sample of 25 students who completed this course typed, on average, 1125 words an hour with a standard deviation of 85 words. Assume that the typing speeds for all students who complete this course have an approximately normal distribution.

a. Suppose the probability of making a Type I error is selected to be zero. Can you conclude that the claim of the business school is true? Answer without performing the five steps of a test of hypothesis.

b. Using the 5% significance level, can you conclude that the claim of the business school is true?

9.76 According to data from the House Ways and Means Committee, the average time required for taxpayers to prepare IRS Schedule A (itemized deductions) was (about) 4.5 hours (*USA TODAY*, April 20, 1999). A taxpayers' organization claims that the time required to prepare this form averages more than 4.5 hours. Suppose that a random sample of 26 taxpayers shows a mean preparation time of 4.75 hours with a standard deviation of .75 hour. Assume that all such preparation times are normally distributed.

a. Suppose that the probability of the Type I error is selected to be zero. Can you conclude that the claim of the taxpayers' organization is true? Answer without performing the five steps of a test of hypothesis.

b. Using the 5% level of significance, can you conclude that the claim of the taxpayers' organization is true?

9.77 A past study claims that adults in America spend an average of 18 hours a week on leisure activities. A researcher wanted to test this claim. She took a sample of 10 adults and asked them about the time they spend per week on leisure activities. Their responses (in hours) are as follows.

14 25 22 38 16 26 19 23 41 33

Assume that the times spent on leisure activities by all adults are normally distributed. Using the 5% significance level, can you conclude that the claim of the earlier study is true? (*Hint*: First calculate the sample mean and the sample standard deviation for these data using the formulas learned in Sections 3.1.1 and 3.2.2 of Chapter 3. Then make the test of hypothesis about μ.)

9.78 The past records of a supermarket show that its customers spend an average of \$65 per visit at this store. Recently the management of the store initiated a promotional campaign according to which each customer receives points based on the total money spent at the store and these points can be used to buy products at the store. The management expects that as a result of this campaign, the customers should be encouraged to spend more money at the store. To check whether this is true, the manager of the store took a sample of 12 customers who visited the store. The following data give the money (in dollars) spent by these customers at this supermarket during their visits.

88 69 141 28 106 45 32 51 78 54 110 83

Assume that the money spent by all customers at this supermarket has a normal distribution. Using the 1% significance level, can you conclude that the mean amount of money spent by all customers at this supermarket after the campaign was started is more than \$65? (*Hint*: First calculate the sample mean and the sample standard deviation for these data using the formulas learned in Sections 3.1.1 and 3.2.2 of Chapter 3. Then make the test of hypothesis about μ.)

*9.79 The manager of a service station claims that the mean amount spent on gas by its customers is \$10.90. You want to test if the mean amount spent on gas at this station is different from \$10.90. Briefly explain how you would conduct this test by taking a small sample.

*9.80 A tool manufacturing company claims that its top-of-the-line machine that is used to manufacture bolts produces an average of 88 or more bolts per hour. A company that is interested in buying this machine wants to check this claim. Suppose you are asked to conduct this test. Briefly explain how you would do so by taking a small sample.

9.5 HYPOTHESIS TESTS ABOUT A POPULATION PROPORTION: LARGE SAMPLES

Often we want to conduct a test of hypothesis about a population proportion. For example, 33% of the students listed in *Who's Who Among American High School Students* said that drugs and alcohol are the most serious problems facing their high schools. A sociologist may want to check if this percentage still holds. As another example, a mail-order company claims that 90% of all orders it receives are shipped within 72 hours. The company's management may want to determine from time to time whether or not this claim is true.

This section presents the procedure to perform tests of hypotheses about the population proportion, p, for large samples. The procedure to make such tests is similar in many respects to the one for the population mean, μ. The procedure includes the same five steps. Again, the test can be two-tailed or one-tailed. We know from Chapter 7 that when the sample size is large, the sample proportion, $\hat{p}$, is approximately normally distributed with its mean equal to p and standard deviation equal to $\sqrt{pq/n}$. Hence, we use the normal distribution to perform a test of hypothesis about the population proportion, p, for a large sample. As was mentioned in Chapters 7 and 8, in the case of a proportion, the sample size is considered to be large when np and nq are both greater than 5.

TEST STATISTIC The value of the *test statistic z* for the sample proportion, $\hat{p}$, is computed as

$$z = \frac{\hat{p} - p}{\sigma_{\hat{p}}} \qquad \text{where } \sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$$

The value of p used in this formula is the one used in the null hypothesis. The value of q is equal to $1 - p$.

The value of z calculated for $\hat{p}$ using the above formula is also called the **observed value of z.**

Examples 9–9, 9–10, and 9–11 describe the procedure to make tests of hypotheses about the population proportion, p.

Conducting a two-tailed test of hypothesis about p: large sample.

EXAMPLE 9–9 According to Hewitt Associates, 36% of the companies surveyed in 1999 paid holiday bonuses to their employees (*The Wall Street Journal*, December 22, 1999). Assume that this result holds true for all U.S. companies in 1999. A recent random sample of 400 companies showed that 33% of them pay holiday bonuses to their employees. Using the 1% significance level, can you conclude that the current percentage of companies that pay holiday bonuses to their employees is different from that for 1999?

Solution Let p be the proportion of all U.S. companies that currently pay holiday bonuses to their employees, and let $\hat{p}$ be the corresponding sample proportion. Then, from the given information,

$$n = 400, \quad \hat{p} = .33, \quad \text{and} \quad \alpha = .01$$

In 1999, 36% of the companies paid holiday bonuses to their employees. Hence,

$$p = .36 \quad \text{and} \quad q = 1 - p = 1 - .36 = .64$$

Step 1. *State the null and alternative hypotheses.*

The percentage of all U.S. companies that currently pay holiday bonuses to their employees is not different from that for 1999 if $p = .36$, and the current percentage is different from 1999 if $p \neq .36$. The null and alternative hypotheses are as follows:

$$H_0: p = .36 \qquad \text{(The current percentage is not different from 1999)}$$

$$H_1: p \neq .36 \qquad \text{(The current percentage is different from 1999)}$$

Step 2. Select the distribution to use.

The values of np and nq are

$$np = 400(.36) = 144 \quad \text{and} \quad nq = 400(.64) = 256$$

Because both np and nq are greater than 5, the sample size is large. Consequently, we use the normal distribution to make the hypothesis test about p.

Step 3. Determine the rejection and nonrejection regions.

The $\neq$ sign in the alternative hypothesis indicates that the test is two-tailed. The significance level is .01. Therefore, the total area of the two rejection regions is .01 and the rejection region in each tail of the sampling distribution of $\hat{p}$ is $\alpha/2 = .01/2 = .005$. The critical values of z, obtained from the standard normal distribution table, are -2.58 and 2.58, as shown in Figure 9.15.

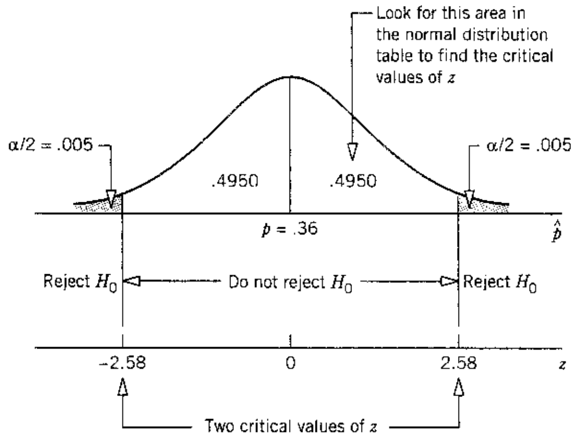


Figure 9.15

Step 4. Calculate the value of the test statistic.

The value of the test statistic z for $\hat{p} = .33$ is calculated as follows:

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(.36)(.64)}{400}} = .024$$

From H_0

$$z = \frac{\hat{p} - p}{\sigma_{\hat{p}}} = \frac{.33 - .36}{.024} = -1.25$$

Step 5. Make a decision.

The value of the test statistic $z = -1.25$ for $\hat{p}$ lies in the nonrejection region. Consequently, we fail to reject H_0. Therefore, we can state that the sample proportion is not too far from the hypothesized value of the population proportion and the difference between the two can be attributed to chance. We conclude that the percentage of U.S. companies that currently pay holiday bonuses to their employees is not different from that for 1999. ∎

Conducting a right-tailed test of hypothesis about p: large sample.

EXAMPLE 9-10 When working properly, a machine that is used to make chips for calculators does not produce more than 4% defective chips. Whenever the machine produces more than 4% defective chips, it needs an adjustment. To check if the machine is working properly, the quality control department at the company often takes samples of chips and inspects them to determine if they are good or defective. One such random sample of 200 chips taken recently from the production line contained 14 defective chips. Test at the 5% significance level whether or not the machine needs an adjustment.

Solution Let p be the proportion of defective chips in all chips produced by this machine, and let $\hat{p}$ be the corresponding sample proportion. Then, from the given information,

$$n = 200, \quad \hat{p} = 14/200 = .07, \quad \text{and} \quad \alpha = .05$$

When the machine is working properly it does not produce more than 4% defective chips. Consequently, assuming that the machine is working properly,

$$p = .04 \quad \text{and} \quad q = 1 - p = 1 - .04 = .96$$

Step 1. *State the null and alternative hypotheses.*

The machine will not need an adjustment if the percentage of defective chips is 4% or less, and it will need an adjustment if this percentage is greater than 4%. Hence, the null and alternative hypotheses are

H_0: $p \leq .04$ (The machine does not need an adjustment)

H_1: $p > .04$ (The machine needs an adjustment)

Step 2. *Select the distribution to use.*

The values of np and nq are

$$np = 200(.04) = 8 > 5 \quad \text{and} \quad nq = 200(.96) = 192 > 5$$

Because the sample size is large, we use the normal distribution to make the hypothesis test about p.

Step 3. *Determine the rejection and nonrejection regions.*

The significance level is .05. The $>$ sign in the alternative hypothesis indicates that the test is right-tailed and the rejection region lies in the right tail of the sampling distribution of $\hat{p}$ with its area equal to .05. As shown in Figure 9.16, the critical value of z, obtained from the normal distribution table for .4500, is approximately 1.65.

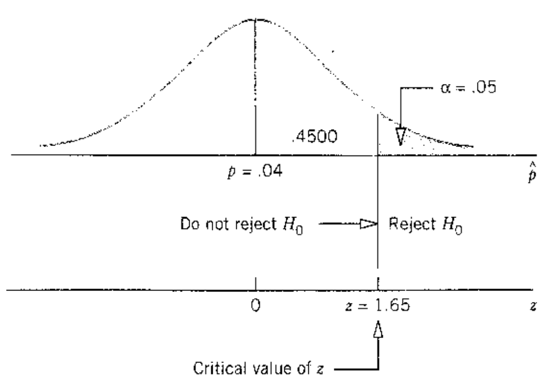


Figure 9.16

Step 4. *Calculate the value of the test statistic.*

The value of the test statistic z for $\hat{p} = .07$ is calculated as follows:

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(.04)(.96)}{200}} = .01385641$$

From H_0

$$z = \frac{\hat{p} - p}{\sigma_{\hat{p}}} = \frac{.07 - .04}{.01385641} = 2.17$$

Step 5. *Make a decision.*

Because the value of the test statistic $z = 2.17$ is greater than the critical value of $z = 1.65$ and it falls in the rejection region, we reject H_0. We conclude that the sample proportion is too far from the hypothesized value of the population proportion and the difference between the two cannot be attributed to chance alone. Therefore, based on the sample information, we conclude that the machine needs an adjustment. ■

Conducting a left-tailed test of hypothesis about p: large sample.

EXAMPLE 9-11 Direct Mailing Company sells computers and computer parts by mail. The company claims that at least 90% of all orders are mailed within 72 hours after they are received. The quality control department at the company often takes samples to check if this claim is valid. A recently taken sample of 150 orders showed that 129 of them were mailed within 72 hours. Do you think the company's claim is true? Use a 2.5% significance level.

Solution Let p be the proportion of all orders that are mailed by the company within 72 hours and $\hat{p}$ the corresponding sample proportion. Then, from the given information,

$$n = 150, \quad \hat{p} = 129/150 = .86, \quad \text{and} \quad \alpha = .025$$

The company claims that at least 90% of all orders are mailed within 72 hours. Assuming that this claim is true, the values of p and q are

$$p = .90 \quad \text{and} \quad q = 1 - p = 1 - .90 = .10$$

Step 1. *State the null and alternative hypotheses.*

The null and alternative hypotheses are

H_0: $p \geq .90$ (The company's claim is true)

H_1: $p < .90$ (The company's claim is false)

Step 2. *Select the distribution to use.*

We first check whether both np and nq are greater than 5.

$$np = 150(.90) = 135 > 5 \quad \text{and} \quad nq = 150(.10) = 15 > 5$$

Consequently, the sample size is large. Therefore, we use the normal distribution to make the hypothesis test about p.

Step 3. *Determine the rejection and nonrejection regions.*

The significance level is .025. The $<$ sign in the alternative hypothesis indicates that the test is left-tailed and the rejection region lies in the left tail of the sampling distribution of $\hat{p}$ with its area equal to .025. As shown in Figure 9.17, the critical value of z, obtained from the normal distribution table for .4750, is -1.96.

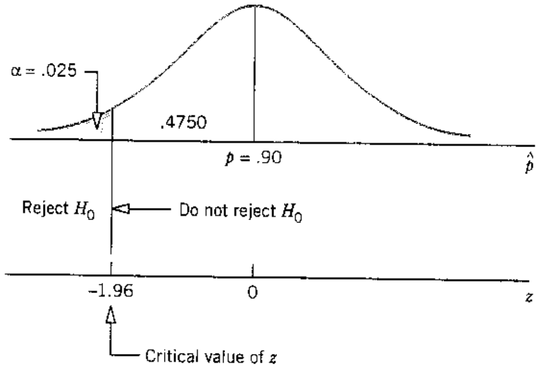


Figure 9.17

Step 4. *Calculate the value of the test statistic.*

The value of the test statistic z for $\hat{p} = .86$ is calculated as follows:

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(.90)(.10)}{150}} = .02449490$$

From H_0

$$z = \frac{\hat{p} - p}{\sigma_{\hat{p}}} = \frac{.86 - .90}{.02449490} = -1.63$$

Step 5. *Make a decision.*

The value of the test statistic $z = -1.63$ is greater than the critical value of $z = -1.96$, and it falls in the nonrejection region. Therefore, we fail to reject H_0. We can state that the difference between the sample proportion and the hypothesized value of the population proportion is small and this difference may have occurred owing to chance alone. Therefore, the proportion of all orders that are mailed within 72 hours is at least 90%, and the company's claim is true.

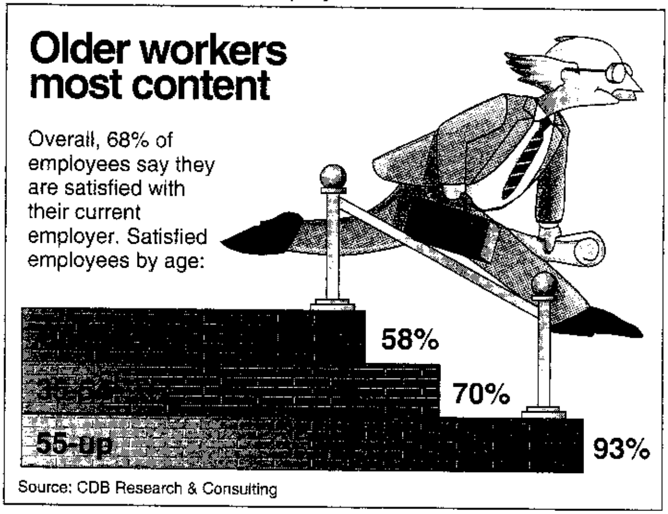
We can also use the p-value approach to make tests of hypotheses about the population proportion p. The procedure to calculate the p-value for the sample proportion is similar to the one applied to the sample mean in Section 9.3.

CASE STUDY 9-1 OLDER WORKERS MOST CONTENT

The chart on page 414 shows the percentage of employees in different age groups who are satisfied with their current employer. As we can observe from the chart, only 58% of the employees under age 35 are satisfied with their current employer. The corresponding percentage is 70% for the employees in the age group 35–54, and it jumps to 93% for employees 55 and older. Note that these percentages are based on a sample survey. Suppose that the survey included 1000 employees under 35 years of age. Suppose the overall percentage of

USA SNAPSHOTS®

A look at statistics that shape your finances



By Cindy Hall and Web Bryant, USA TODAY

68% is true for the population of all employees. We can conduct a test of hypothesis to find out whether the percentage of employees under 35 years of age who are satisfied with their current employer is less than 68%. Suppose we make this test using a 1% significance level. Below we perform this test of hypothesis.

$$H_0\colon p = .68$$

$$H_1\colon p < .68$$

Here, $n = 1000$, $\hat{p} = .58$, and $\alpha = .01$. The test is left-tailed. Using the normal distribution to perform the test, the critical value of z is -2.33. Thus, we will reject the null hypothesis if the observed value of z is -2.33 or smaller. We find the observed value of z as follows.

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(.68)(.32)}{1000}} = .01475127$$

$$z = \frac{\hat{p} - p}{\sigma_{\hat{p}}} = \frac{.58 - .68}{.01475127} = -6.78$$

The value of the test statistic $z = -6.78$ for $\hat{p}$ lies in the rejection region. Consequently, we reject H_0 and conclude that the percentage of employees under 35 years of age who are satisfied with their current employer is less than 68%.

Similarly, we can test hypotheses about the percentages of employees in the other two age groups who are satisfied with their current employers.

EXERCISES

Concepts and Procedures

9.81 Explain when a sample is large enough to use the normal distribution to make a test of hypothesis about the population proportion.

9.82 In each of the following cases, do you think the sample size is large enough to use the normal distribution to make a test of hypothesis about the population proportion? Explain why or why not.

a. $n = 40$ and $p = .11$ **b.** $n = 100$ and $p = .73$
c. $n = 80$ and $p = .05$ **d.** $n = 50$ and $p = .14$

9.83 In each of the following cases, do you think the sample size is large enough to use the normal distribution to make a test of hypothesis about the population proportion? Explain why or why not.

a. $n = 30$ and $p = .65$ **b.** $n = 70$ and $p = .05$
c. $n = 60$ and $p = .06$ **d.** $n = 900$ and $p = .17$

9.84 For each of the following examples of tests of hypotheses about the population proportion, show the rejection and nonrejection regions on the graph of the sampling distribution of the sample proportion.

a. A two-tailed test with $\alpha = .10$
b. A left-tailed test with $\alpha = .01$
c. A right-tailed test with $\alpha = .05$

9.85 For each of the following examples of tests of hypotheses about the population proportion, show the rejection and nonrejection regions on the graph of the sampling distribution of the sample proportion.

a. A two-tailed test with $\alpha = .05$
b. A left-tailed test with $\alpha = .02$
c. A right-tailed test with $\alpha = .025$

9.86 A random sample of 500 observations produced a sample proportion equal to .38. Find the critical and observed values of z for each of the following tests of hypotheses using $\alpha = .05$.

a. H_0: $p = .30$ versus H_1: $p > .30$
b. H_0: $p = .30$ versus H_1: $p \neq .30$

9.87 A random sample of 200 observations produced a sample proportion equal to .60. Find the critical and observed values of z for each of the following tests of hypotheses using $\alpha = .01$.

a. H_0: $p = .63$ versus H_1: $p < .63$
b. H_0: $p = .63$ versus H_1: $p \neq .63$

9.88 Consider the null hypothesis H_0: $p = .65$. Suppose a random sample of 1000 observations is taken to make this test about the population proportion. Using $\alpha = .05$, show the rejection and nonrejection regions and find the critical value(s) of z for a

a. left-tailed test **b.** two-tailed test **c.** right-tailed test

9.89 Consider the null hypothesis H_0: $p = .25$. Suppose a random sample of 400 observations is taken to make this test about the population proportion. Using $\alpha = .01$, show the rejection and nonrejection regions and find the critical value(s) of z for a

a. left-tailed test **b.** two-tailed test **c.** right-tailed test

9.90 Consider H_0: $p = .70$ versus H_1: $p \neq .70$.

a. A random sample of 600 observations produced a sample proportion equal to .68. Using $\alpha = .01$, would you reject the null hypothesis?
b. Another random sample of 600 observations taken from the same population produced a sample proportion equal to .76. Using $\alpha = .01$, would you reject the null hypothesis?

Comment on the results of parts a and b.

9.91 Consider H_0: $p = .45$ versus H_1: $p < .45$.

a. A random sample of 400 observations produced a sample proportion equal to .42. Using $\alpha = .025$, would you reject the null hypothesis?

b. Another random sample of 400 observations taken from the same population produced a sample proportion of .39. Using $\alpha = .025$, would you reject the null hypothesis?

Comment on the results of parts a and b.

9.92 Make the following hypothesis tests about p.

a.	H_0: $p = .45$,	H_1: $p \neq .45$,	$n = 100$,	$\hat{p} = .49$,	$\alpha = .10$
b.	H_0: $p = .72$,	H_1: $p < .72$,	$n = 700$,	$\hat{p} = .64$,	$\alpha = .05$
c.	H_0: $p = .30$,	H_1: $p > .30$,	$n = 200$,	$\hat{p} = .33$,	$\alpha = .01$

9.93 Make the following hypothesis tests about p.

a.	H_0: $p = .57$,	H_1: $p \neq .57$,	$n = 800$,	$\hat{p} = .50$,	$\alpha = .05$
b.	H_0: $p = .26$,	H_1: $p < .26$,	$n = 400$,	$\hat{p} = .23$,	$\alpha = .01$
c.	H_0: $p = .84$,	H_1: $p > .84$,	$n = 250$,	$\hat{p} = .85$,	$\alpha = .025$

Applications

9.94 Managers often tolerate poor performance by employees because the termination process is so complicated. In a survey of federal government managers who had not taken action against unsatisfactory employees, 66% stated that they failed to act because of the long process required (*USA TODAY*, July 29, 1998). Suppose that 80 such managers are randomly selected and 48 of them failed to act because of the long process required. Can you conclude that the current percentage of such managers who fail to act because of the long process involved is less than 66%? Use $\alpha = .025$.

9.95 Many traffic accidents are blamed on motorists who are distracted by cell phone use while driving. In a survey conducted by Bruskin/Goldring for Exxon, 29% of adults said that they make phone calls *sometimes* or *frequently* while driving alone (*USA TODAY*, May 20, 1999). Suppose that a recently taken random sample of 500 adults showed that 165 of them make phone calls *sometimes* or *frequently* while driving alone. At the 5% level of significance, can you conclude that the current percentage of adults who make such phone calls exceeds 29%?

9.96 According to a 1998 survey of 1200 adults by the Travel Industry Association of America, 46% made travel plans through the Internet, but only 9% actually used the Internet to make reservations (*USA TODAY*, April 26, 1999). A recent random sample of 400 travelers found that 58 of them made reservations on the Internet. Test at the 1% significance level whether the current percentage of all travelers who make reservations on the Internet exceeds 9%.

9.97 According to a telephone poll of 1049 adult Americans conducted for *Time*/CNN by Yankelovich Partners, Inc., 73% would forgive someone who told lies about them (*Time*, April 5, 1999). In a recent random sample of 1000 adult Americans, 700 indicated that they would forgive someone who lied about them. At the 5% level of significance, can you conclude that the proportion of all American adults who feel this way is less than 73%?

9.98 In a telephone poll of 428 dog owners conducted by Yankelovich Partners, Inc., for *Time*/CNN, 87% of the respondents said that they think dogs have good and bad moods as humans do (*Time*, February 1, 1999). In a recent random sample of 400 dog owners, 336 held this view.

a. Test at the 2% level of significance whether the current percentage of all dog owners who hold this view is different from 87%.

b. What is the Type I error in this case? What is the probability of making this error?

9.99 A poll of high school students released by the nonprofit Josephson Institute for Ethics found that 47% of high school students admitted stealing from a store in the past year (*USA TODAY*, October 19, 1998). Assume that this percentage was true for all such students at the time that poll was conducted. A recent survey of a random sample of 1100 high school students found that 550 of them admitted stealing from a store in the past year.

a. Test at the 5% significance level whether the percentage of all high school students who would admit to such thefts has increased since 1998.

b. What is the Type I error in this case? What is the probability of making this error?

9.100 A food company is planning to market a new type of frozen yogurt. However, before marketing this yogurt, the company wants to find what percentage of the people like it. The company's management has decided that it will market this yogurt only if at least 35% of the people like it. The company's research department selected a random sample of 400 persons and asked them to taste this yogurt. Of these 400 persons, 112 said they liked it.

a. Testing at the 2.5% significance level, can you conclude that the company should market this yogurt?

b. What will your decision be in part a if the probability of making a Type I error is zero? Explain.

9.101 A mail-order company claims that at least 60% of all orders are mailed within 48 hours. From time to time the quality control department at the company checks if this promise is fulfilled. Recently the quality control department at this company took a sample of 400 orders and found that 208 of them were mailed within 48 hours of the placement of the orders.

a. Testing at the 1% significance level, can you conclude that the company's claim is true?

b. What will your decision be in part a if the probability of making a Type I error is zero? Explain.

9.102 Brooklyn Corporation manufactures computer diskettes. The machine that is used to make these diskettes is known to produce not more than 5% defective diskettes. The quality control inspector selects a sample of 200 diskettes each week and inspects them for being good or defective. Using the sample proportion, the quality control inspector tests the null hypothesis $p \leq .05$ against the alternative hypothesis $p > .05$, where p is the proportion of diskettes that are defective. She always uses a 2.5% significance level. If the null hypothesis is rejected, the production process is stopped to make any necessary adjustments. A recent such sample of 200 diskettes contained 17 defective diskettes.

a. Using the 2.5% significance level, would you conclude that the production process should be stopped to make necessary adjustments?

b. Perform the test of part a using a 1% significance level. Is your decision different from the one in part a?

Comment on the results of parts a and b.

9.103 Shulman Steel Corporation makes bearings that are supplied to other companies. One of the machines makes bearings that are supposed to have a diameter of 4 inches. The bearings that have a diameter of either more or less than 4 inches are considered defective and are discarded. When working properly, the machine does not produce more than 7% of bearings that are defective. The quality control inspector selects a sample of 200 bearings each week and inspects them for the size of their diameters. Using the sample proportion, the quality control inspector tests the null hypothesis $p \leq .07$ against the alternative hypothesis $p > .07$, where p is the proportion of bearings that are defective. He always uses a 2% significance level. If the null hypothesis is rejected, the machine is stopped to make any necessary adjustments. One such sample of 200 bearings taken recently contained 22 defective bearings.

a. Using the 2% significance level, will you conclude that the machine should be stopped to make necessary adjustments?

b. Perform the test of part a using a 1% significance level. Is your decision different from the one in part a?

Comment on the results of parts a and b.

*9.104 Two years ago, 75% of the customers of a bank said that they were satisfied with the services provided by the bank. The manager of the bank wants to know if this percentage of satisfied customers has changed since then. She assigns this responsibility to you. Briefly explain how you would conduct such a test.

*9.105 A study claims that 65% of students at all colleges and universities hold off-campus (part-time or full-time) jobs. You want to check if the percentage of students at your school who hold off-campus jobs is different from 65%. Briefly explain how you would conduct such a test. Collect data from 40 students at your school on whether or not they hold off-campus jobs. Then, calculate the proportion of students in this sample who hold off-campus jobs. Using this information, test the hypothesis. Select your own significance level.

GLOSSARY

α The significance level of a test of hypothesis that denotes the probability of rejecting a null hypothesis when it actually is true. (The probability of committing a Type I error.)

Alternative hypothesis A claim about a population parameter that will be true if the null hypothesis is false.

β The probability of not rejecting a null hypothesis when it actually is false. (The probability of committing a Type II error.)

Critical value or **critical point** One or two values that divide the whole region under the sampling distribution of a sample statistic into rejection and nonrejection regions.

Left-tailed test A test in which the rejection region lies in the left tail of the distribution curve.

Null hypothesis A claim about a population parameter that is assumed to be true until proven otherwise.

Observed value of z or t The value of z or t calculated for a sample statistic such as the sample mean or the sample proportion.

One-tailed test A test in which there is only one rejection region, either in the left tail or in the right tail of the distribution curve.

p-value The smallest significance level at which a null hypothesis can be rejected.

Right-tailed test A test in which the rejection region lies in the right tail of the distribution curve.

Significance level The value of α that gives the probability of committing a Type I error.

Test statistic The value of z or t calculated for a sample statistic such as the sample mean or the sample proportion.

Two-tailed test A test in which there are two rejection regions, one in each tail of the distribution curve.

Type I error An error that occurs when a true null hypothesis is rejected.

Type II error An error that occurs when a false null hypothesis is not rejected.

KEY FORMULAS

1. **Value of the test statistic z for $\bar{x}$ in a test of hypothesis about μ for a large sample**

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} \quad \text{if } \sigma \text{ is known, where } \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

or

$$z = \frac{\bar{x} - \mu}{s_{\bar{x}}} \quad \text{if } \sigma \text{ is not known, where } s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

2. **Value of the test statistic t for $\bar{x}$ in a test of hypothesis about μ for a small sample**

$$t = \frac{\bar{x} - \mu}{s_{\bar{x}}} \quad \text{where} \quad s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

3. **Value of the test statistic z for $\hat{p}$ in a test of hypothesis about p for a large sample**

$$z = \frac{\hat{p} - p}{\sigma_{\hat{p}}} \quad \text{where} \quad \sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$$

SUPPLEMENTARY EXERCISES

9.106 Consider the following null and alternative hypotheses:

$$H_0: \mu = 120 \quad \text{versus} \quad H_1: \mu > 120$$

A random sample of 81 observations taken from this population produced a sample mean of 123.5 and a sample standard deviation of 15.

a. If this test is made at the 2.5% significance level, would you reject the null hypothesis?
b. What is the probability of making a Type I error in part a?
c. Calculate the p-value for the test. Based on this p-value, would you reject the null hypothesis if $\alpha = .01$? What if $\alpha = .05$?

9.107 Consider the following null and alternative hypotheses:

$$H_0: \mu = 40 \quad \text{versus} \quad H_1: \mu \neq 40$$

A random sample of 64 observations taken from this population produced a sample mean of 38.4 and a sample standard deviation of 6.

a. If this test is made at the 2% significance level, would you reject the null hypothesis?
b. What is the probability of making a Type I error in part a?
c. Calculate the p-value for the test. Based on this p-value, would you reject the null hypothesis if $\alpha = .01$? What if $\alpha = .05$?

9.108 Consider the following null and alternative hypotheses:

$$H_0: p = .82 \quad \text{versus} \quad H_1: p \neq .82$$

A random sample of 600 observations taken from this population produced a sample proportion of .86.

a. If this test is made at the 2% significance level, would you reject the null hypothesis?
b. What is the probability of making a Type I error in part a?
c. Calculate the p-value for the test. Based on this p-value, would you reject the null hypothesis if $\alpha = .025$? What if $\alpha = .005$?

9.109 Consider the following null and alternative hypotheses:

$$H_0: p = .44 \quad \text{versus} \quad H_1: p < .44$$

A random sample of 450 observations taken from this population produced a sample proportion of .39.

a. If this test is made at the 2% significance level, would you reject the null hypothesis?
b. What is the probability of making a Type I error in part a?
c. Calculate the p-value for the test. Based on this p-value, would you reject the null hypothesis if $\alpha = .01$? What if $\alpha = .025$?

9.110 A manufacturer of fluorescent lightbulbs claims that the mean life of these bulbs is at least 2500 hours. A consumer agency wanted to check whether or not this claim is true. The agency took a random sample of 36 such bulbs and tested them. The mean life for the sample was found to be 2457 hours with a standard deviation of 180 hours.

a. Do you think the sample information supports the company's claim? Use $\alpha = .025$.
b. What is the Type I error in this case? Explain. What is the probability of making this error?
c. Will your conclusion of part a change if the probability of making a Type I error is zero?

9.111 According to data from the U.S. Bureau of Justice Statistics, the average length of prison sentences for violent crimes in the United States was 90.7 months in 1997 (*Statistical Abstract of the United States,* 1998). A recent sample of 55 such sentences yielded a mean of 101 months with a standard deviation of 30 months.

a. Does the sample information support the alternative hypothesis that the current mean sentence for such crimes exceeds 90.7 months? Use $\alpha = .02$.
b. What is the Type I error in this case? Explain. What is the probability of making this error?
c. Would your conclusion for part a change if the probability of making a Type I error were zero?

9.112 According to an estimate, the average hourly wage for summer employment for high school and college students was \$6.69 in 1999 (*USA TODAY*, June 25, 1999). Suppose that a random sample of 350 such students yielded a mean hourly wage for the past summer employment of \$6.90 with a standard deviation of \$1.88.

a. Using $\alpha = .05$, can you conclude that the mean hourly wage for the past summer employment for all such students is higher than \$6.69?
b. Using $\alpha = .01$, can you conclude that the mean hourly wage for the past summer employment for all such students is higher than \$6.69?

Comment on the results of parts a and b.

9.113 According to *Family PC Magazine*, the average computer usage by readers of this magazine is 16 to 18 hours per week (*USA TODAY*, April 14, 1999). Suppose a recent random sample of 80 such readers showed an average weekly computer usage of 20 hours with a standard deviation of 9 hours.

a. Using $\alpha = .05$, does the sample information support the alternative hypothesis that current mean weekly computer usage by such readers exceeds 18 hours?
b. Using $\alpha = .01$, does the sample information support the alternative hypothesis that current mean weekly computer usage by such readers exceeds 18 hours?

Comment on the results of parts a and b.

9.114 Customers often complain about long waiting times at restaurants before the food is served. A restaurant claims that it serves food to its customers, on average, within 15 minutes after the order is placed. A local newspaper journalist wanted to check if the restaurant's claim is true. A sample of 36 customers showed that the mean time taken to serve food to them was 15.75 minutes with a standard deviation of 2.4 minutes. Using the sample mean, the journalist says that the restaurant's claim is false. Do you think the journalist's conclusion is fair to the restaurant? Use the 1% significance level to answer this question.

9.115 The customers at a bank complained about long lines and the time they had to spend waiting for service. It is known that the customers at this bank had to wait 8 minutes, on average, before being served. The management made some changes to reduce the waiting time for its customers. A sample of 32 customers taken after these changes were made produced a mean waiting time of 7.5 minutes with a standard deviation of 2.1 minutes. Using this sample mean, the bank manager displayed a huge banner inside the bank mentioning that the mean waiting time for customers has been reduced by new changes. Do you think the bank manager's claim is justifiable? Use the 2.5% significance level to answer this question.

9.116 According to data from the College Board, the average cost per student for books and supplies in the 1997–1998 academic year was \$634 at public four-year colleges in the United States (*The Chronicle of Higher Education*, August 1998). A recent random sample of 250 students at such colleges yielded a mean cost for books and supplies of \$660 with a standard deviation of \$170. Find the p-value for the test with the alternative hypothesis that the current mean for these costs differs from \$634.

9.117 The mean consumption of water per household in a city was 1245 cubic feet per month. Due to a water shortage because of a drought, the city council campaigned for water use conservation by households. A few months after the campaign was started, the mean consumption of water for a sample of 100 households was found to be 1175 cubic feet per month with a standard deviation of 250 cubic feet. Find the p-value for the hypothesis test that the mean consumption of water per household has decreased due to the campaign by the city council.

9.118 Data furnished by American Express Everyday Spending Index indicated a mean annual household expenditure of \$1044 on gasoline (*USA TODAY*, June 3, 1999). A random sample of 28 U.S. households produced a mean annual household expenditure of \$921 on gasoline with a standard deviation of \$343 for last year. Using the 1% significance level, test whether the mean gasoline expenditure for all U.S. households was less than \$1044 last year. Assume that the gasoline expenditures for all U.S. households have a normal distribution.

9.119 The administrative office of a hospital claims that the mean waiting time for patients to get treatment in its emergency ward is 25 minutes. A random sample of 16 patients who received treatment in the emergency ward of this hospital produced a mean waiting time of 27.5 minutes with a standard deviation of 4.8 minutes. Using the 1% significance level, test whether the mean waiting time at the emergency ward is different from 25 minutes. Assume that the waiting times for all patients at this emergency ward have a normal distribution.

9.120 According to data from Runzheimer International, the average cost of suburban day care (for a 3-year-old in a for-profit center for five days a week) in the Minneapolis area was \$572 per month (*USA TODAY*, August 18, 1998). A recent random sample of 25 parents in the Minneapolis area with 3-year-old children found a mean day-care cost of \$634 per month with a standard deviation of \$136. Assume that the monthly costs for all such day care are normally distributed.

a. Using $\alpha = .025$, can you conclude that the mean of all such costs currently exceeds \$572?

b. Suppose the probability of making a Type I error is zero. Can you make a decision for the test of part a without going through the five steps of hypothesis testing? If so, what is your decision? Explain.

9.121 An earlier study claims that U.S. adults spend an average of 114 minutes with their families per day. A recently taken sample of 25 adults showed that they spend an average of 109 minutes per day with their families. The sample standard deviation is 11 minutes. Assume that the times spent by adults with their families have an approximately normal distribution.

a. Using the 1% significance level, test whether the mean time spent currently by all adults with their families is less than 114 minutes a day.

b. Suppose the probability of making a Type I error is zero. Can you make a decision for the test of part a without going through the five steps of hypothesis testing? If yes, what is your decision? Explain.

9.122 A computer company that recently introduced a new software product claims that the mean time it takes to learn how to use this software is not more than 2 hours for people who are somewhat familiar with computers. A random sample of 12 such persons was selected. The following data give the times taken (in hours) by these persons to learn how to use this software.

1.75	2.25	2.40	1.90	1.50	2.75
2.15	2.25	1.80	2.20	3.25	2.60

Test at the 1% significance level whether the company's claim is true. Assume that the times taken by all persons who are somewhat familiar with computers to learn how to use this software are approximately normally distributed.

9.123 A company claims that its 8-ounce low-fat yogurt cups contain, on average, at most 150 calories per cup. A consumer agency wanted to check whether or not this claim is true. A random sample of 10 such cups produced the following data on calories.

147	159	153	146	144	161	163	153	143	158

Test at the 2.5% significance level whether the company's claim is true. Assume that the number of calories for such cups of yogurt produced by this company has an approximately normal distribution.

9.124 According to a survey conducted for *Time* magazine, 62% of the respondents said that the government should regulate gene therapy to cure or prevent disease (*Time*, January 11, 1999). In a recent random sample of 200 people, 58% held this view.

a. Test at the 2.5% significance level whether the percentage of all people who favor such government regulation is currently less than 62%.

b. How do you explain the Type I error in this case? What is the probability of making this error in part a?

9.125 According to a *USA TODAY*/Gallup telephone poll, 49% of adults expected that a worldwide collapse of the economy will occur by 2025 (*USA TODAY*, October 13, 1998). Assume that this result was true for the population of all adult Americans at the time of the poll. In a recent random sample of 800 adult Americans, 43% expected such a collapse by 2025.

a. Test at the 1% level of significance whether the current percentage of adult Americans who expect such a collapse has changed since the earlier poll.
b. How do you explain the Type I error in this case? What is the probability of making this error in part a?

9.126 More and more people are abandoning national brand products and buying store brand products to save money. The president of a company that produces national brand coffee claims that 40% of the people prefer to buy national brand coffee. A random sample of 700 people who buy coffee showed that 259 of them buy national brand coffee. Using $\alpha = .01$, can you conclude that the percentage of people who buy national brand coffee is different from 40%?

9.127 According to DATAQUEST, Inc., 26% of adults in the United States had wireless phones in 1999 (*Business Week*, May 3, 1999). A recent random sample of 200 adult Americans showed that 72 of them have wireless phones. At the 2% level of significance, can you conclude that the current percentage of U.S. adults who have wireless phones exceeds 26%?

9.128 Mong Corporation makes auto batteries. The company claims that 80% of its LL70 batteries are good for 70 months or longer. A consumer agency wanted to check if this claim is true. The agency took a random sample of 40 such batteries and found that 75% of them were good for 70 months or longer.

a. Using the 1% significance level, can you conclude that the company's claim is false?
b. What will your decision be in part a if the probability of making a Type I error is zero? Explain.

9.129 Dartmouth Distribution Warehouse makes deliveries of a large number of products to its customers. To keep its customers happy and satisfied, the company's policy is to deliver on time at least 90% of all the orders it receives from its customers. The quality control inspector at the company quite often takes samples of orders delivered and checks if this policy is maintained. A recent such sample of 90 orders taken by this inspector showed that 75 of them were delivered on time.

a. Using the 2% significance level, can you conclude that the company's policy is maintained?
b. What will your decision be in part a if the probability of making a Type I error is zero? Explain.

Advanced Exercises

9.130 Refer to Exercise 9.125. Find the p-value for the test of hypothesis mentioned in that exercise. Using this p-value, would you reject the null hypothesis at $\alpha = .05$? What if $\alpha = .01$?

9.131 Refer to Exercise 9.129. Find the p-value for the test of hypothesis mentioned in that exercise. Using this p-value, would you reject the null hypothesis at $\alpha = .05$? What if $\alpha = .01$?

9.132 Professor Hansen believes that some people have the ability to predict in advance the outcome of a spin of a roulette wheel. He takes 100 student volunteers to a casino. The roulette wheel has 38 numbers, each of which is equally likely to occur. Of these 38 numbers, 18 are red, 18 are black, and 2 are green. Each student is to place a series of five bets, choosing either a red or a black number before each spin of the wheel. Thus, a student who bets on red has an 18/38 chance of winning that bet. The same is true of betting on black.

a. Assuming random guessing, what is the probability that a particular student will win all five of his or her bets?
b. Suppose for each student we formulate the hypothesis test

H_0: The student is guessing

H_1: The student has some predictive ability

Suppose we reject H_0 only if the student wins all five bets. What is the significance level?
c. Suppose that two of the 100 students win all five of their bets. Professor Hansen says, "For these two students we can reject H_0 and conclude that we have found two students with some ability to predict." What do you make of Professor Hansen's conclusion?

9.133 Acme Bicycle Company makes derailleurs for mountain bikes. Usually no more than 4% of these parts are defective, but occasionally the machines that make them get out of adjustment and the rate of defectives exceeds 4%. To guard against this, the chief quality control inspector takes a random sample of 130 derailleurs each week and checks each one for defects. If too many of these parts are defective, the machines are shut down and adjusted. To decide how many parts must be defective in order to shut down the machines, the company's statistician has set up the hypothesis test

$$H_0\colon p \leq .04 \quad \text{versus} \quad H_1\colon p > .04$$

where p is the proportion of defectives among all derailleurs being made currently. Rejection of H_0 would call for shutting down the machines. For the inspector's convenience, the statistician would like the rejection region to have the form, "Reject H_0 if the number of defective parts is C or more." Find the value of C that will make the significance level (approximately) .05.

9.134 Alpha Airlines claims that only 15% of its flights arrive more than 10 minutes late. Let p be the proportion of all of Alpha's flights that arrive more than 10 minutes late. Consider the hypothesis test

$$H_0\colon p \leq .15 \quad \text{versus} \quad H_1\colon p > .15$$

Suppose we take a random sample of 50 flights by Alpha Airlines and agree to reject H_0 if 9 or more of them arrive late. Find the significance level for this test.

9.135 The standard therapy that is used to treat a disorder cures 60% of all patients in an average of 140 visits. A health care provider considers supporting a new therapy regime for the disorder if it is effective in reducing the number of visits while retaining the cure rate of the standard therapy. A study of 200 patients with the disorder who were treated by the new therapy regime reveals that 108 of them were cured in an average of 132 visits with a standard deviation of 38 visits. What decision should be made using a .01 level of significance?

9.136 The print on the packages of 100-watt General Electric soft-white lightbulbs states that these lightbulbs have an average life of 750 hours. Assume that the standard deviation of the lengths of lives of these lightbulbs is 50 hours. A skeptical consumer does not think these lightbulbs last as long as the manufacturer claims and she decides to test 64 randomly selected lightbulbs. She has set up the decision rule that if the average life of these 64 lightbulbs is less than 735 hours, then she will conclude that GE has printed too high an average length of life on the packages and will write them a letter to that effect. Approximately what significance level is the consumer using? Approximately what significance level is she using if she decides that GE has printed too high an average length of life on the packages if the average life of the 64 lightbulbs is less than 700 hours? Interpret the values you get.

9.137 Thirty percent of all people who are inoculated with the current vaccine that is used to prevent a disease contract the disease within a year. The developer of a new vaccine that is intended to prevent this disease wishes to test for significant evidence that the new vaccine is more effective.

a. Determine the appropriate null and alternative hypotheses.

b. The developer decides to study 100 randomly selected people by inoculating them with the new vaccine. If 84 or more of them do not contract the disease within a year, the developer will conclude that the new vaccine is superior to the old one. What significance level is the developer using for the test?

c. Suppose 20 people inoculated with the new vaccine are studied and the new vaccine is concluded to be better than the old one if fewer than 3 people contract the disease within a year. What is the significance level of the test?

9.138 The Parks and Recreation Department has determined that the thickness of the ice on a town pond must average 5 inches to be safe for ice skating. Consider the decision of whether or not to allow ice skating as a hypothesis-testing problem.

a. What is the meaning of the unknown parameter that is being tested?

b. Set up the appropriate null and alternative hypotheses for this situation.

c. What would it mean to commit a Type I error for your test in part b? What about a Type II error?
d. Would you want the significance level for your test in part b to be large or small? Suggest a significance level and explain what it means in this problem.

9.139 Since 1984, all automobiles have been manufactured with a middle taillight. You have been hired to answer the question, Is the middle taillight effective in reducing the number of rear-end collisions? You have available to you any information you could possibly want about all rear-end collisions involving cars built before 1984. How would you conduct an experiment to answer the question? In your answer, include things like: (a) the precise meaning of the unknown parameter you are testing; (b) H_0 and H_1; (c) a detailed explanation of what sample data you would collect to draw a conclusion; and (d) any assumptions you would make, particularly about the characteristics of cars built before 1984 versus those built since 1984.

9.140 Before a championship football game, the referee is given a special commemorative coin to toss to decide which team will kick the ball first. Two minutes before game time, he receives an anonymous tip that the captain of one of the teams may have substituted a biased coin that has a 70% chance of showing heads each time it is tossed. The referee has time to toss the coin 10 times to test it. He decides that if it shows 8 or more heads in 10 tosses, he will reject this coin and replace it with another coin. Let p be the probability that this coin shows heads when it is tossed once.

a. Formulate the relevant null and alternative hypotheses (in terms of p) for the referee's test.
b. Using the referee's decision rule, find α for this test.

SELF-REVIEW TEST

1. A test of hypothesis is always about
   a. a population parameter **b.** a sample statistic **c.** a test statistic
2. A Type I error is committed when
   a. a null hypothesis is not rejected when it is actually false
   b. a null hypothesis is rejected when it is actually true
   c. an alternative hypothesis is rejected when it is actually true
3. A Type II error is committed when
   a. a null hypothesis is not rejected when it is actually false
   b. a null hypothesis is rejected when it is actually true
   c. an alternative hypothesis is rejected when it is actually true
4. A critical value is the value
   a. calculated from sample data
   b. determined from a table (e.g., the normal distribution table or other such tables)
   c. neither a nor b
5. The computed value of a test statistic is the value
   a. calculated for a sample statistic
   b. determined from a table (e.g., the normal distribution table or other such tables)
   c. neither a nor b
6. The observed value of a test statistic is the value
   a. calculated for a sample statistic
   b. determined from a table (e.g., the normal distribution table or other such tables)
   c. neither a nor b

7. The significance level, denoted by α, is

   a. the probability of committing a Type I error
   b. the probability of committing a Type II error
   c. neither a nor b

8. The value of β gives the

   a. probability of committing a Type I error
   b. probability of committing a Type II error
   c. power of the test

9. The value of $1 - \beta$ gives the

   a. probability of committing a Type I error
   b. probability of committing a Type II error
   c. power of the test

10. A two-tailed test is a test with

   a. two rejection regions **b.** two nonrejection regions **c.** two test statistics

11. A one-tailed test

   a. has one rejection region **b.** has one nonrejection region **c.** both a and b

12. The smallest level of significance at which a null hypothesis is rejected is called

   a. α **b.** p-value **c.** β

13. Which of the following is not required to apply the t distribution to make a test of hypothesis about μ?

   a. $n < 30$ **b.** Population is normally distributed
   c. σ is unknown **d.** β is known

14. The sign in the alternative hypothesis in a two-tailed test is always

   a. $<$ **b.** $>$ **c.** $\neq$

15. The sign in the alternative hypothesis in a left-tailed test is always

   a. $<$ **b.** $>$ **c.** $\neq$

16. The sign in the alternative hypothesis in a right-tailed test is always

   a. $<$ **b.** $>$ **c.** $\neq$

17. A bank loan officer claims that the mean monthly mortgage payment made by all homeowners in a certain city is \$1365. A housing magazine wanted to test this claim. A random sample of 100 homeowners taken by this magazine produced the mean monthly mortgage of \$1505 with a standard deviation of \$278.

   a. Testing at the 1% significance level, would you conclude that the mean monthly mortgage payment made by all homeowners in this city is different from \$1365?
   b. What is the Type I error in part a? What is the probability of making this error?
   c. What will your decision be in part a if the probability of making a Type I error is zero? Explain.

18. An editor of a New York publishing company claims that the mean time it takes to write a textbook is at least 31 months. A sample of 16 textbook authors found that the mean time taken by them to write a textbook was 25 months with a standard deviation of 7.2 months.

   a. Using the 2.5% significance level, would you conclude that the editor's claim is true? Assume that the time taken to write a textbook is normally distributed for all textbook authors.
   b. What is the Type I error in part a? What is the probability of making this error?
   c. What will your decision be in part a if the probability of making a Type I error is .001?

19. In a survey, 66% of the respondents agreed that police should be allowed to collect DNA information from suspected criminals (*Time*, January 11, 1999). In a recently taken random sample of 500 Americans, 305 held this view.

a. Using a 5% significance level, can you conclude that the current percentage of Americans who hold this view differs from 66%?
b. What is the Type I error in part a? What is the probability of making this error?
c. What would your decision be in part a if the probability of making a Type I error were zero? Explain.

20. According to a survey, the starting salary for an entry-level position in investment banking was \$37,120 in 1998 (*Newsweek*, February 1, 1999). Assume that \$37,120 was the 1998 mean starting salary for all such positions. A recently taken random sample of 100 such positions found a mean starting salary of \$38,050 with a standard deviation of \$4420.

a. Find the p-value for the test with the alternative hypothesis that the mean starting salary for such positions exceeds \$37,120.
b. Using the p-value calculated in part a, would you reject the null hypothesis at a significance level of 1%? What if $\alpha = .05$?

***21.** Refer to Problem 19.

a. Find the p-value for the test of hypothesis mentioned in part a of that problem.
b. Using this p-value, will you reject the null hypothesis if $\alpha = .05$? What if $\alpha = .01$?

MINI-PROJECTS

Mini-Project 9–1

The mean height of players who were on the rosters of National Basketball Association teams at the beginning of the 1996–1997 season was 79.22 inches. Let μ denote the mean height of NBA players at the beginning of the 1999–2000 season.

a. Take a random sample of 15 players from the NBA data that accompany this text. Test H_0: $\mu =$ 79.22 inches against H_1: $\mu \neq 79.22$ inches using $\alpha = .05$.
b. Repeat part a for samples of 31 and 45 players, respectively.
c. Did any of the three tests in parts a and b lead to the conclusion that the mean height of NBA players in 1999–2000 is different from that in 1996–1997?

Mini-Project 9–2

A thumbtack that is tossed on a desk can land in one of the two ways shown in the illustration.

Brad and Dan cannot agree on the likelihood of obtaining a head or a tail. Brad argues that obtaining a tail is more likely than obtaining a head because of the shape of the tack. If the tack had no point at all, it would resemble a coin that has the same probability of coming up heads or tails when tossed. But, the longer the point, the less likely it is that the tack will stand up on its head when tossed. Dan believes that as the tack lands tails, the point causes the tack to jump around and come to rest in the heads position. Brad and Dan need you to settle their dispute. Do you think the tack is equally likely to land heads or tails? To investigate this question, find an ordinary thumbtack and toss it a large number of times (say, 100 times).

a. What is the meaning, in words, of the unknown parameter in this problem?
b. Set up the null and alternative hypotheses and compute the p-value based on your results from tossing the tack.
c. How would you answer the original question now? If you decide the tack is not fair, do you side with Brad or Dan?

d. What would you estimate the value of the parameter in part a to be? Find a 90% confidence interval for this parameter.
e. After doing this experiment, do you think 100 tosses are enough to infer the nature of your tack? Using your result as a preliminary estimate, determine how many tosses would be necessary to be 95% certain of having 4% accuracy; that is, the maximum error of estimate is ±4%. Have you observed enough tosses?

For instructions on using MINITAB for this chapter, please see Section B.8 of Appendix B.

See Appendix C for *Excel Adventure 9* on hypothesis testing.

COMPUTER ASSIGNMENTS

CA9.1 According to an earlier study, the mean amount spent on clothes by American women is \$575 per year. A researcher wanted to check if this result still holds true. A random sample of 39 women taken recently by this researcher produced the following data on the amounts they spent on clothes last year.

671	584	328	498	827	921	425	204	382	539
1070	854	669	328	537	849	930	1234	695	738
341	189	867	923	721	125	298	473	876	932
573	931	460	1430	391	887	958	674	782	

Using MINITAB or any other statistical software, test at the 1% significance level whether the mean expenditure on clothes for American women for last year is different from \$575. Assume that the population standard deviation is \$132.

CA9.2 The mean weight of all babies born at a hospital last year was 7.6 pounds. A random sample of 35 babies born at this hospital this year produced the following data.

8.2	9.1	6.9	5.8	6.4	10.3	12.1	9.1	5.9	7.3
11.2	8.3	6.5	7.1	8.0	9.2	5.7	9.5	8.3	6.3
4.9	7.6	10.1	9.2	8.4	7.5	7.2	8.3	7.2	9.7
6.0	8.1	6.1	8.3	6.7					

Using MINITAB or any other statistical software, test at the 2.5% significance level whether the mean weight of babies born at this hospital this year is more than 7.6 pounds.

CA9.3 The president of a large university claims that the mean time spent partying by all students at the university is not more than 7 hours per week. The following data give the times spent partying during the previous week by a random sample of 16 students taken from this university.

12	9	5	15	11	13	10	6
4	11	6	9	13	6	16	8

Using MINITAB or any other statistical software, test at the 1% significance level whether the president's claim is true. Assume that the times spent partying by all students at this university have an approximately normal distribution.

CA9.4 According to a basketball coach, the mean height of all male college basketball players is 74 inches. A random sample of 25 such players produced the following data on their heights.

68	76	74	83	77	76	69	67	71	74	79	85	69
78	75	78	68	72	83	79	82	76	69	70	81	

Using MINITAB or any other statistical software, test at the 2% significance level whether the mean height of all male college basketball players is different from 74 inches. Assume that the heights of all male college basketball players are (approximately) normally distributed.

CA9.5 A past study claims that adults in America spend an average of 18 hours a week on leisure activities. A researcher wanted to test this claim. She took a sample of 10 adults and asked them about the time they spend per week on leisure activities. Their responses (in hours) follow.

14	25	22	38	16	26	19	23	41	33

Assume that the times spent on leisure activities by all adults are normally distributed. Using the 5% significance level, can you conclude that the claim of the earlier study is true? Use MINITAB or any other statistical software to answer this question.

CA9.6 In a *Time* magazine poll, 55% of the respondents stated that smokers should pay higher insurance rates than nonsmokers (*Time*, January 11, 1999). Assume that this result was true for the population of all American adults at the time of the poll. A researcher wants to know whether this result holds true for the current population of Americans. A random sample of 700 American adults taken recently by this researcher showed that 399 of them hold this view. Using MINITAB or any other statistical software and a 2.5% significance level, can you conclude that the current percentage of American adults who think that smokers should pay more for insurance than nonsmokers is different from 55%?

CA9.7 A mail-order company claims that at least 60% of all orders it receives are mailed within 48 hours. From time to time the quality control department at the company checks if this promise is kept. Recently, the quality control department at this company took a sample of 400 orders and found that 224 of them were mailed within 48 hours of the placement of the orders. Using MINITAB or any other statistical software, test at the 1% significance level whether or not the company's claim is true.