CHAPTER 9

Hypothesis Testing

Learning Objectives

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

- **LO9-1** Set up appropriate null and alternative hypotheses.
- **LO9-2** Describe Type I and Type II errors and their probabilities.
- **LO9-3** Use critical values and *p*-values to perform a *z* test about a population mean when σ is known.
- **LO9-4** Use critical values and *p*-values to perform a *t* test about a population mean when σ is unknown.
- **LO9-5** Use critical values and *p*-values to perform a large sample *z* test about a population proportion.
- **LO9-6** Calculate Type II error probabilities and the power of a test, and determine sample size (Optional).

Chapter Outline

9.1 The Null and Alternative Hypotheses and Errors in Hypothesis Testing

9.2 *z* Tests about a Population Mean: σ Known

9.3 *t* Tests about a Population Mean: σ Unknown

9.4 *z* Tests about a Population Proportion

9.5 Type II Error Probabilities and Sample Size Determination (Optional)

Hypothesis testing is a statistical procedure used to provide evidence in favor of some statement (called a *hypothesis*). For instance, hypothesis testing might be used to assess whether a population parameter, such as a population mean, differs from a specified standard or previous value. In this chapter we discuss testing hypotheses about population means and proportions.

In order to illustrate how hypothesis testing works, we revisit several cases introduced in previous chapters and also introduce some new cases:

The e-billing Case: The consulting firm uses hypothesis testing to provide strong evidence that the new electronic billing system has reduced the mean payment time by more than 50 percent.

The Cheese Spread Case: The cheese spread producer uses hypothesis testing to supply extremely strong evidence that fewer than 10 percent of all current purchasers would stop buying the cheese spread if the new spout were used.

The Commercial Loan Case: The bank uses hypothesis testing to provide strong evidence that the mean debt-to-equity ratio for its portfolio of commercial loans is less than 1.5.

The Trash Bag Case: A marketer of trash bags uses hypothesis testing to support its claim that the mean breaking strength of its new trash bag is greater than 50 pounds. As a result, a television network approves use of this claim in a commercial.

The Valentine's Day Chocolate Case: A candy company projects that this year's sales of its special valentine box of assorted chocolates will be 10 percent higher than last year. The candy company uses hypothesis testing to assess whether it is reasonable to plan for a 10 percent increase in sales of the valentine box.

9.1 The Null and Alternative Hypotheses and Errors in Hypothesis Testing

LO9-1 Set up appropriate null and alternative hypotheses.

One of the authors' former students is employed by a major television network in the standards and practices division. One of the division's responsibilities is to reduce the chances that advertisers will make false claims in commercials run on the network. Our former student reports that the network uses a statistical methodology called **hypothesis testing** to do this.

To see how this might be done, suppose that a company wishes to advertise a claim, and suppose that the network has reason to doubt that this claim is true. The network assumes for the sake of argument that **the claim is not valid.** This assumption is called the **null hypothesis.** The statement that **the claim is valid** is called the **alternative,** or **research, hypothesis.** The network will run the commercial only if the company making the claim provides **sufficient sample evidence** to reject the null hypothesis that the claim is not valid in favor of the alternative hypothesis that the claim is valid. Explaining the exact meaning of *sufficient sample evidence* is quite involved and will be discussed as we proceed through this chapter.

The Null Hypothesis and the Alternative Hypothesis

In hypothesis testing:

1. The **null hypothesis,** denoted H_0, is the statement being tested. Often, the null hypothesis is a statement of "no difference" or "no effect." The null hypothesis is not rejected unless there is convincing sample evidence that it is false.
2. The **alternative,** or **research, hypothesis,** denoted H_a, is a statement that will be accepted only if there is convincing sample evidence that it is true.

Setting up the null and alternative hypotheses in a practical situation can be tricky. In some situations there is a statement about a population parameter (such as a population mean) for which we need to attempt to find supportive evidence. If the statement says that the parameter is

"greater than" a particular number, or if the statement says that the parameter is "less than" a particular number, then (for reasons to be discussed later) we make the statement the *alternative hypothesis,* H_a, and we make the opposite of the statement the *null hypothesis,* H_0. We illustrate these ideas in the following two cases.

EXAMPLE 9.1 The Trash Bag Case[1]: Testing An Advertising Claim

A leading manufacturer of trash bags produces the strongest trash bags on the market. The company has developed a new 30-gallon bag using a specially formulated plastic that is stronger and more biodegradable than other plastics. This plastic's increased strength allows the bag's thickness to be reduced, and the resulting cost savings will enable the company to lower its bag price by 25 percent. The company also believes the new bag is stronger than its current 30-gallon bag.

The manufacturer wants to advertise the new bag on a major television network. In addition to promoting its price reduction, the company also wants to claim the new bag is better for the environment and stronger than its current bag. The network is convinced of the bag's environmental advantages on scientific grounds. However, the network questions the company's claim of increased strength and requires statistical evidence to justify this claim. Although there are various measures of bag strength, the manufacturer and the network agree to employ "breaking strength." A bag's breaking strength is the amount of a representative trash mix (in pounds) that, when loaded into a bag suspended in the air, will cause the bag to rip or tear. Tests show that the current bag has a mean breaking strength that is very close to (but does not exceed) 50 pounds. The new bag's mean breaking strength μ is unknown and in question. Because the trash bag manufacturer wishes to show that *μ is greater than 50 pounds,* we make the statement that *μ is greater than 50 pounds the alternative hypothesis,* H_a, and we make the statement that *μ is less than or equal to 50 pounds the null hypothesis,* H_0. (Note that the *null hypothesis* says that the new trash bag is *not stronger* than the current bag.) We summarize the null and alternative hypotheses by saying that we are testing

$$H_0: \mu \leq 50 \quad \text{versus} \quad H_a: \mu > 50$$

The network will run the manufacturer's commercial if a random sample of n new bags provides sufficient evidence to reject $H_0: \mu \leq 50$ in favor of $H_a: \mu > 50$.

EXAMPLE 9.2 The e-billing Case: Reducing Mean Bill Payment Time

Recall that a management consulting firm has installed a new computer-based electronic billing system for a Hamilton, Ohio, trucking company. Because of the system's advantages, and because the trucking company's clients are receptive to using this system, the management consulting firm believes that the new system will reduce the mean bill payment time by more than 50 percent. The mean payment time using the old billing system was approximately equal to, but no less than, 39 days. Therefore, if μ denotes the mean payment time using the new system, the consulting firm believes and wishes to show that *μ is less than 19.5 days*. It follows that we make the statement that *μ is less than 19.5 days the alternative hypothesis,* H_a, and we make the statement that *μ is greater than or equal to 19.5 days the null hypothesis,* H_0. The consulting firm will randomly select a sample of n invoices and determine if their payment times provide sufficient evidence to reject $H_0: \mu \geq 19.5$ in favor of $H_a: \mu < 19.5$. If such evidence exists, the consulting firm will conclude that the new electronic billing system has reduced the Hamilton trucking company's mean bill payment time by more than 50 percent. This conclusion will be used to help demonstrate the benefits of the new billing system both to the Hamilton company and to other trucking companies that are considering using such a system.

[1]This case is based on conversations by the authors with several employees working for a leading producer of trash bags. For purposes of confidentiality, we have agreed to withhold the company's name.

In some situations we need to evaluate a statement that says that a population parameter exactly equals a particular number. It then follows that we make the statement that the population parameter *equals* the particular number the *null hypothesis,* H_0, and we make the statement that the population parameter *does not equal* the particular number the *alternative hypothesis,* H_a. We demonstrate this in the following case.

EXAMPLE 9.3 The Valentine's Day Chocolate Case[2]: Production Planning

A candy company annually markets a special 18 ounce box of assorted chocolates to large retail stores for Valentine's Day. This year the candy company has designed an extremely attractive new valentine box and will fill the box with an especially appealing assortment of chocolates. For this reason, the candy company subjectively projects—based on past experience and knowledge of the candy market—that sales of its valentine box will be 10 percent higher than last year. However, because the candy company must decide how many valentine boxes to produce, the company needs to assess whether it is reasonable to plan for a 10 percent increase in sales.

Before the beginning of each Valentine's Day sales season, the candy company sends large retail stores information about its newest valentine box of assorted chocolates. This information includes a description of the box of chocolates, as well as a preview of advertising displays that the candy company will provide to help retail stores sell the chocolates. Each retail store then places a single (nonreturnable) order of valentine boxes to satisfy its anticipated customer demand for the Valentine's Day sales season. Last year the mean order quantity of large retail stores was 300 boxes per store. If the projected 10 percent sales increase will occur, the mean order quantity, μ, of large retail stores this year will *equal* 330 boxes per store. Therefore, the candy company will test the *null hypothesis* H_0: $\mu = 330$ versus the *alternative hypothesis* H_a: $\mu \neq 330$. Here, the alternative hypothesis, H_a says that μ might be greater than or less than 330 boxes. If μ turns out to be greater than 330 boxes and the candy company bases its production on a projected mean order quantity of 330 boxes, the company will fail to satisfy demand for its valentine box. If μ turns out to be less than 330 boxes and the candy company bases its production on a projected mean order quantity of 330 boxes, the company will produce more valentine boxes than it can sell.

To perform the hypothesis test, the candy company will randomly select a sample of n large retail stores and will make an early mailing to these stores promoting this year's valentine box. The candy company will then ask each retail store to report how many valentine boxes it anticipates ordering. If the sample data do not provide sufficient evidence to reject H_0: $\mu = 330$ in favor of H_a: $\mu \neq 330$, the candy company will base its production on the projected 10 percent sales increase. On the other hand, if there is sufficient evidence to reject H_0: $\mu = 330$, the candy company will change its production plans.

We next summarize the sets of null and alternative hypotheses that we have thus far considered.

H_0: $\mu \leq 50$	H_0: $\mu \geq 19.5$	H_0: $\mu = 330$
versus	versus	versus
H_a: $\mu > 50$	H_a: $\mu < 19.5$	H_a: $\mu \neq 330$

The alternative hypothesis H_a: $\mu > 50$ is called a **one-sided, greater than alternative** hypothesis, whereas H_a: $\mu < 19.5$ is called a **one-sided, less than alternative** hypothesis, and H_a: $\mu \neq 330$ is called a **two-sided, not equal to alternative** hypothesis. Many of the alternative hypotheses we consider in this book are one of these three types. Also, note that each null hypothesis we have considered involves an **equality.** For example, the null hypothesis H_0: $\mu \leq 50$ says that μ is either less than or **equal to** 50. We will see that, in general, the approach we use to test a null hypothesis versus an alternative hypothesis requires that the null hypothesis involve an equality. For this reason, **we always formulate the null and alternative hypotheses so that the null hypothesis involves an equality.**

[2]Thanks to Krogers of Oxford, Ohio, for helpful discussions concerning this case.

The idea of a test statistic Suppose that in the trash bag case the manufacturer randomly selects a sample of $n = 40$ new trash bags. Each of these bags is tested for breaking strength, and the sample mean $\bar{x}$ of the 40 breaking strengths is calculated. In order to test H_0: $\mu \leq 50$ versus H_a: $\mu > 50$, we utilize the **test statistic**

$$z = \frac{\bar{x} - 50}{\sigma_{\bar{x}}} = \frac{\bar{x} - 50}{\sigma/\sqrt{n}}$$

The test statistic z measures the distance between $\bar{x}$ and 50. The division by $\sigma_{\bar{x}}$ says that this distance is measured in units of the standard deviation of all possible sample means. For example, a value of z equal to, say, 2.4 would tell us that $\bar{x}$ is 2.4 such standard deviations above 50. In general, a value of the test statistic that is less than or equal to zero results when $\bar{x}$ is less than or equal to 50. This provides no evidence to support rejecting H_0 in favor of H_a because the point estimate $\bar{x}$ indicates that μ is probably less than or equal to 50. However, a value of the test statistic that is greater than zero results when $\bar{x}$ is greater than 50. This provides evidence to support rejecting H_0 in favor of H_a because the point estimate $\bar{x}$ indicates that μ might be greater than 50. Furthermore, the farther the value of the test statistic is above 0 (the farther $\bar{x}$ is above 50), the stronger is the evidence to support rejecting H_0 in favor of H_a.

Hypothesis testing and the legal system If the value of the test statistic z is far enough above zero, we reject H_0 in favor of H_a. To see how large z must be in order to reject H_0, we must understand that **a hypothesis test rejects a null hypothesis H_0 only if there is strong statistical evidence against H_0.** This is similar to our legal system, which rejects the innocence of the accused only if evidence of guilt is beyond a reasonable doubt. For instance, the network will reject H_0: $\mu \leq 50$ and run the trash bag commercial only if the test statistic z is far enough above zero to show beyond a reasonable doubt that H_0: $\mu \leq 50$ is false and H_a: $\mu > 50$ is true. A test statistic that is only slightly greater than zero might not be convincing enough. However, because such a test statistic would result from a sample mean $\bar{x}$ that is slightly greater than 50, it would provide some evidence to support rejecting H_0: $\mu \leq 50$, and it certainly would not provide strong evidence supporting H_0: $\mu \leq 50$. Therefore, if the value of the test statistic is not large enough to convince us to reject H_0, **we do not say that we accept H_0. Rather we say that we do not reject H_0** because the evidence against H_0 is not strong enough. Again, this is similar to our legal system, where the lack of evidence of guilt beyond a reasonable doubt results in a verdict of **not guilty,** but does not prove that the accused is innocent.

LO9-2 Describe Type I and Type II errors and their probabilities.

Type I and Type II errors and their probabilities To determine exactly how much statistical evidence is required to reject H_0, we consider the errors and the correct decisions that can be made in hypothesis testing. These errors and correct decisions, as well as their implications in the trash bag advertising example, are summarized in Tables 9.1 and 9.2. Across the top of each table are listed the two possible "states of nature." Either H_0: $\mu \leq 50$ is true, which says the manufacturer's claim that μ is greater than 50 is false, or H_0 is false, which says the claim is true. Down the left side of each table are listed the two possible decisions we can make in the hypothesis test. Using the sample data, we will either reject H_0: $\mu \leq 50$, which implies that the claim will be advertised, or we will not reject H_0, which implies that the claim will not be advertised.

In general, the two types of errors that can be made in hypothesis testing are defined as follows:

Type I and Type II Errors

If we reject H_0 when it is true, this is a **Type I error.**
If we do not reject H_0 when it is false, this is a **Type II error.**

As can be seen by comparing Tables 9.1 and 9.2, if we commit a Type I error, we will advertise a false claim. If we commit a Type II error, we will fail to advertise a true claim.

We now let the symbol $\boldsymbol{\alpha}$ (pronounced **alpha**) **denote the probability of a Type I error,** and we let $\boldsymbol{\beta}$ (pronounced **beta**) **denote the probability of a Type II error.** Obviously, we would like both α and β to be small. A common (but not the only) procedure is to base a hypothesis test on taking a sample of a fixed size (for example, $n = 40$ trash bags) and on setting

TABLE 9.1 Type I and Type II Errors

	State of Nature	
Decision	**H_0: $\mu \leq 50$ True**	**H_0: $\mu \leq 50$ False**
Reject H_0: $\mu \leq 50$	Type I error	Correct decision
Do not reject H_0: $\mu \leq 50$	Correct decision	Type II error

TABLE 9.2 The Implications of Type I and Type II Errors in the Trash Bag Example

	State of Nature	
Decision	**Claim False**	**Claim True**
Advertise the claim	Advertise a false claim	Advertise a true claim
Do not advertise the claim	Do not advertise a false claim	Do not advertise a true claim

α equal to a small prespecified value. Setting α low means there is only a small chance of rejecting H_0 when it is true. This implies that we are requiring strong evidence against H_0 before we reject it.

We sometimes choose α as high as .10, but we usually choose α between .05 and .01. A frequent choice for α is .05. In fact, our former student tells us that the network often tests advertising claims by setting the probability of a Type I error equal to .05. That is, the network will run a commercial making a claim if the sample evidence allows it to reject a null hypothesis that says the claim is not valid in favor of an alternative hypothesis that says the claim is valid with α set equal to .05. Because a Type I error is deciding that the claim is valid when it is not, the policy of setting α equal to .05 says that, in the long run, the network will advertise only 5 percent of all invalid claims made by advertisers.

One might wonder why the network does not set α lower—say at .01. One reason is that **it can be shown that, for a fixed sample size, the lower we set α, the higher is β, and the higher we set α, the lower is β.** Setting α at .05 means that β, the probability of failing to advertise a true claim (a Type II error), will be smaller than it would be if α were set at .01. As long as (1) the claim to be advertised is plausible and (2) the consequences of advertising the claim even if it is false are not terribly serious, then it is reasonable to set α equal to .05. However, if either (1) or (2) is not true, then we might set α lower than .05. For example, suppose a pharmaceutical company wishes to advertise that it has developed an effective treatment for a disease that has been very resistant to treatment. Such a claim is (perhaps) difficult to believe. Moreover, if the claim is false, patients suffering from the disease would be subjected to false hope and needless expense. In such a case, it might be reasonable for the network to set α at .01 because this would lower the chance of advertising the claim if it is false. We usually do not set α lower than .01 because doing so often leads to an unacceptably large value of β. We explain some methods for computing the probability of a Type II error in optional Section 9.5. However, β can be difficult or impossible to calculate in many situations, and we often must rely on our intuition when deciding how to set α.

Exercises for Section 9.1

CONCEPTS

9.1 Define each of the following: Type I error, α, Type II error, β.

9.2 When testing a hypothesis, why don't we set the probability of a Type I error to be extremely small? Explain.

METHODS AND APPLICATIONS

9.3 THE VIDEO GAME SATISFACTION RATING CASE VideoGame

Recall that "very satisfied" customers give the XYZ-Box video game system a rating that is at least 42. Suppose that the manufacturer of the XYZ-Box wishes to use the 65 satisfaction ratings

to provide evidence supporting the claim that the mean composite satisfaction rating for the XYZ-Box exceeds 42.

a Letting μ represent the mean composite satisfaction rating for the XYZ-Box, set up the null and alternative hypotheses needed if we wish to attempt to provide evidence supporting the claim that μ exceeds 42.

b In the context of this situation, interpret making a Type I error; interpret making a Type II error.

9.4 THE BANK CUSTOMER WAITING TIME CASE WaitTime

Recall that a bank manager has developed a new system to reduce the time customers spend waiting for teller service during peak hours. The manager hopes the new system will reduce waiting times from the current 9 to 10 minutes to less than 6 minutes.

Suppose the manager wishes to use the 100 waiting times to support the claim that the mean waiting time under the new system is shorter than six minutes.

a Letting μ represent the mean waiting time under the new system, set up the null and alternative hypotheses needed if we wish to attempt to provide evidence supporting the claim that μ is shorter than six minutes.

b In the context of this situation, interpret making a Type I error; interpret making a Type II error.

9.5 An automobile parts supplier owns a machine that produces a cylindrical engine part. This part is supposed to have an outside diameter of three inches. Parts with diameters that are too small or too large do not meet customer requirements and must be rejected. Lately, the company has experienced problems meeting customer requirements. The technical staff feels that the mean diameter produced by the machine is off target. In order to verify this, a special study will randomly sample 100 parts produced by the machine. The 100 sampled parts will be measured, and if the results obtained cast a substantial amount of doubt on the hypothesis that the mean diameter equals the target value of three inches, the company will assign a problem-solving team to intensively search for the causes of the problem.

a The parts supplier wishes to set up a hypothesis test so that the problem-solving team will be assigned when the null hypothesis is rejected. Set up the null and alternative hypotheses for this situation.

b In the context of this situation, interpret making a Type I error; interpret making a Type II error.

9.6 The Crown Bottling Company has just installed a new bottling process that will fill 16-ounce bottles of the popular Crown Classic Cola soft drink. Both overfilling and underfilling bottles are undesirable: Underfilling leads to customer complaints and overfilling costs the company considerable money. In order to verify that the filler is set up correctly, the company wishes to see whether the mean bottle fill, μ, is close to the target fill of 16 ounces. To this end, a random sample of 36 filled bottles is selected from the output of a test filler run. If the sample results cast a substantial amount of doubt on the hypothesis that the mean bottle fill is the desired 16 ounces, then the filler's initial setup will be readjusted.

a The bottling company wants to set up a hypothesis test so that the filler will be readjusted if the null hypothesis is rejected. Set up the null and alternative hypotheses for this hypothesis test.

b In the context of this situation, interpret making a Type I error; interpret making a Type II error.

9.7 Consolidated Power, a large electric power utility, has just built a modern nuclear power plant. This plant discharges waste water that is allowed to flow into the Atlantic Ocean. The Environmental Protection Agency (EPA) has ordered that the waste water may not be excessively warm so that thermal pollution of the marine environment near the plant can be avoided. Because of this order, the waste water is allowed to cool in specially constructed ponds and is then released into the ocean. This cooling system works properly if the mean temperature of waste water discharged is 60°F or cooler. Consolidated Power is required to monitor the temperature of the waste water. A sample of 100 temperature readings will be obtained each day, and if the sample results cast a substantial amount of doubt on the hypothesis that the cooling system is working properly (the mean temperature of waste water discharged is 60°F or cooler), then the plant must be shut down and appropriate actions must be taken to correct the problem.

a Consolidated Power wishes to set up a hypothesis test so that the power plant will be shut down when the null hypothesis is rejected. Set up the null and alternative hypotheses that should be used.

b In the context of this situation, interpret making a Type I error; interpret making a Type II error.

c The EPA periodically conducts spot checks to determine whether the waste water being discharged is too warm. Suppose the EPA has the power to impose very severe penalties (for example, very heavy fines) when the waste water is excessively warm. Other things being equal, should Consolidated Power set the probability of a Type I error equal to $\alpha = .01$ or $\alpha = .05$? Explain.

9.2 z Tests about a Population Mean: σ Known

LO9-3 Use critical values and *p*-values to perform a *z* test about a population mean when σ is known.

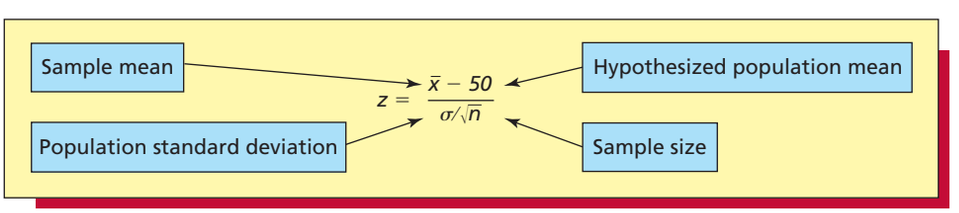
In this section we discuss hypothesis tests about a population mean that are *based on the normal distribution*. These tests are called ***z* tests,** and they require that the *true value of the population standard deviation σ is known.* Of course, in most real-world situations the true value of σ is not known. However, the concepts and calculations of hypothesis testing are most easily illustrated using the normal distribution. Therefore, in this section we will assume that—through theory or history related to the population under consideration—we know σ. When σ is unknown, we test hypotheses about a population mean by using the *t distribution*. In Section 9.3 we study ***t* tests,** and we will revisit the examples of this section assuming that σ is unknown.

Testing a "greater than" alternative hypothesis by using a critical value rule In Section 9.1 we explained how to set up appropriate null and alternative hypotheses. We also discussed how to specify a value for α, the probability of a Type I error (also called the **level of significance**) of the hypothesis test, and we introduced the idea of a test statistic. We can use these concepts to begin developing a five-step hypothesis-testing procedure. We will introduce these steps in the context of the trash bag case and testing a "greater than" alternative hypothesis.

Step 1: State the null hypothesis H_0 and the alternative hypothesis H_a. In the trash bag case, we will test H_0: $\mu \leq 50$ versus H_a: $\mu > 50$. Here, μ is the mean breaking strength of the new trash bag.

Step 2: Specify the level of significance α. The television network will run the commercial stating that the new trash bag is stronger than the current bag if we can reject H_0: $\mu \leq 50$ in favor of H_a: $\mu > 50$ by setting α equal to .05.

Step 3: Select the test statistic. In order to test H_0: $\mu \leq 50$ versus H_a: $\mu > 50$, we will test the modified null hypothesis H_0: $\mu = 50$ versus H_a: $\mu > 50$. The idea here is that if there is sufficient evidence to reject the hypothesis that μ equals 50 in favor of $\mu > 50$, then there is certainly also sufficient evidence to reject the hypothesis that μ is less than or equal to 50. In order to test H_0: $\mu = 50$ versus H_a: $\mu > 50$, we will randomly select a sample of $n = 40$ new trash bags and calculate the mean $\bar{x}$ of the breaking strengths of these bags. We will then utilize the **test statistic** illustrated below.



A positive value of this test statistic results from an $\bar{x}$ that is greater than 50 and thus provides evidence against H_0: $\mu = 50$ and in favor of H_a: $\mu > 50$. Moreover, the manufacturer has improved its trash bags multiple times in the past. Studies show that the population standard deviation σ of individual trash bag breaking strengths has remained constant for each of these updates and equals 1.65 pounds.

Step 4: Determine the critical value rule for deciding whether to reject H_0. To decide how large the test statistic z must be to reject H_0 in favor of H_a by setting the probability of a Type I error equal to α, we note that different samples would give different sample means and thus different values of z. **Because the sample size $n = 40$ is large, the Central Limit Theorem tells us that the sampling distribution of z is (approximately) a standard normal distribution if the null hypothesis H_0: $\mu = 50$ is true.** Therefore, we do the following:

- Place the probability of a Type I error, α, in the right-hand tail of the standard normal curve and use the normal table (see Table A.3, page 790) to find the normal point z_α. Here z_α, which we call a **critical value,** is the point on the horizontal axis under the standard normal curve that gives a right-hand tail area equal to α.
- **Reject H_0: $\mu = 50$ in favor of H_a: $\mu > 50$ if and only if the test statistic z is greater than the critical value z_α.** (This is called a **right tailed critical value rule.**)

FIGURE 9.1 **A Right Tailed Critical Value Rule for Testing H_0: $\mu = 50$ versus H_a: $\mu > 50$ by Setting $\alpha = .05$**

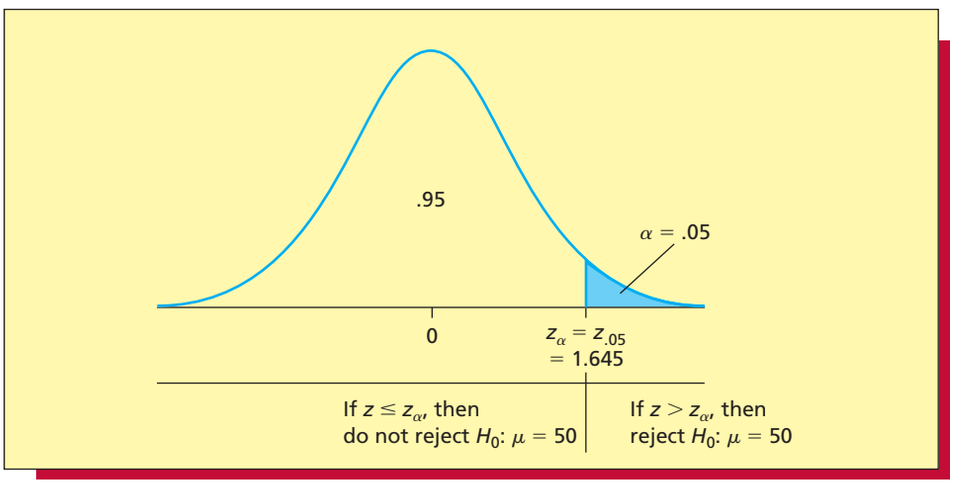


Figure 9.1 illustrates that because we have set α equal to .05, we should use the critical value $z_\alpha = z_{.05} = 1.645$ (see Table A.3). This says that we should reject H_0 if $z > 1.645$ and we should not reject H_0 if $z \leq 1.645$. Moreover, the α of .05 says that there is only a 5 percent chance of z being greater than 1.645 if H_0: $\mu = 50$ is true. Therefore, if the sample that we will actually select gives a value of z that is greater than 1.645 and thus causes us to reject H_0: $\mu = 50$, we can be intuitively confident that we have made the right decision. This is because we will have rejected H_0 by using a test that allows only a 5 percent chance of wrongly rejecting H_0. In general, if we can reject a null hypothesis in favor of an alternative hypothesis by setting the probability of a Type I error equal to α, we say that we have **statistical significance at the α level.**

Step 5: Collect the sample data, compute the value of the test statistic, and decide whether to reject H_0. Interpret the statistical results. When the sample of $n = 40$ new trash bags is randomly selected, the mean of the breaking strengths is calculated to be $\bar{x} = 50.575$ pounds. Assuming that σ is 1.65 pounds, the value of the test statistic is

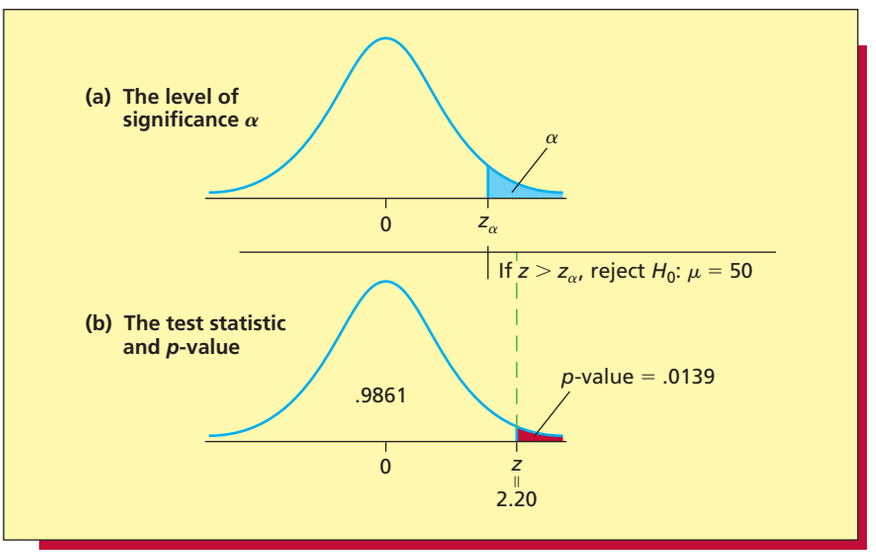
$$z = \frac{\bar{x} - 50}{\sigma/\sqrt{n}} = \frac{50.575 - 50}{1.65/\sqrt{40}} = 2.20$$

Because $z = 2.20$ is greater than the critical value $z_{.05} = 1.645$, we can reject H_0: $\mu = 50$ in favor of H_a: $\mu > 50$ by setting α equal to .05. Therefore, we conclude (at an α of .05) that the population mean breaking strength of the new trash bag exceeds 50 pounds. Furthermore, this conclusion has **practical importance** to the trash bag manufacturer because it means that the television network will approve running commercials claiming that the new trash bag is stronger than the current bag. Note, however, that the point estimate of μ, $\bar{x} = 50.575$, indicates that μ is not much larger than 50. Therefore, the trash bag manufacturer can claim only that its new bag is slightly stronger than its current bag. Of course, this might be practically important to consumers who feel that, because the new bag is 25 percent less expensive and is more environmentally sound, it is definitely worth purchasing if it has any strength advantage. However, to customers who are looking only for a substantial increase in bag strength, the statistical results would not be practically important. Notice that the point estimate of the parameter involved in a hypothesis test can help us to assess practical importance.

BI

A *p*-value for testing a "greater than" alternative hypothesis To decide whether to reject the null hypothesis H_0 at level of significance α, steps 4 and 5 of the five-step critical value procedure compare the test statistic value to a critical value. Another way to make this decision is to calculate a ***p*-value,** which measures the likelihood of the sample results if the null hypothesis H_0 is true. **Sample results that are not likely if H_0 is true are evidence that H_0 is not true.** To test H_0 by using a *p*-value procedure, we use the same first three steps used by the critical value procedure and the new steps 4 and 5:

FIGURE 9.2 The *p*-Value for Testing H_0: $\mu = 50$ versus H_a: $\mu > 50$.



Step 4: Collect the sample data, compute the value of the test statistic, and compute the *p*-value. The *p*-value for testing a null hypothesis H_0 versus an alternative hypothesis H_a is defined as follows:

The ***p*-value** is the probability, computed assuming that the null hypothesis H_0 is true, of observing a value of the test statistic that is at least as contradictory to H_0 and supportive of H_a as the value actually computed from the sample data.

In the trash bag case, the value of the test statistic computed from the sample data is $z = 2.20$. Because we are testing H_0: $\mu = 50$ versus the **greater than** alternative hypothesis H_a: $\mu > 50$, this positive test statistic value contradicts H_0 and supports H_a. A value of the test statistic that is at least as contradictory to H_0 and supportive of H_a as $z = 2.20$ is a value of the test statistic that is **greater than or equal to** $z = 2.20$. Therefore, the *p*-value is the probability, computed assuming that H_0: $\mu = 50$ is true, of observing a value of the test statistic that is greater than or equal to $z = 2.20$. As illustrated in Figure 9.2(b), this ***p*-value is the area under the standard normal curve to the right of $z = 2.20$** and equals $1 - .9861 = .0139$ (see Table A.3, page 790). The *p*-value of .0139 says that, if H_0: $\mu = 50$ is true, then only 139 in 10,000 of all possible test statistic values are at least as large, or contradictory to H_0, as the value $z = 2.20$. That is, if we are to believe that H_0 is true, we must believe that we have observed a test statistic value that can be described as having a 139 in 10,000 chance. It is difficult to believe that we have observed such a small chance, but is this evidence strong enough to reject H_0: $\mu = 50$ and run the trash bag commercial? As discussed in step 5, this depends on the level of significance α used by the television network.

Step 5: Reject H_0 if the *p*-value is less than α. Interpret the statistical results. Consider the two normal curves in Figures 9.2(a) and (b). These normal curves show that if the *p*-value of .0139 is less than a particular level of significance α, the test statistic value $z = 2.20$ is greater than the critical value z_α, and thus we can reject H_0: $\mu = 50$ at level of significance α. For example, recall that the television network has set α equal to .05. Then, **because the *p*-value of .0139 is less than the α of .05, we would reject H_0: $\mu = 50$ at level of significance .05** and thus run the trash bag commercial on the network.

Comparing the critical value and *p*-value methods Thus far we have seen that we can reject H_0: $\mu = 50$ in favor of H_a: $\mu > 50$ at level of significance α if **the test statistic z is greater than the critical value z_α,** or equivalently, **the *p*-value is less than α**. Because different television networks sometimes have different policies for evaluating an advertising claim, different

television networks sometimes use different values of α when evaluating the same advertising claim. For example, while the network of the previous example used an α value of .05 to evaluate the trash bag claim, three other networks might use three different α values (say, .04, .025, and .01) to evaluate this claim. If we use the critical value method to test H_0: $\mu = 50$ versus H_a: $\mu > 50$ at each of these α values, we would have to look up a different critical value z_α for each different α value. On the other hand, the p-value of .0139 is **more efficient** because it immediately tells us whether we can reject H_0 at each different α value. Specifically, because the p-value of .0139 is less than each of the α values .05, .04, and .025, we would reject H_0 and thus run the trash bag commercial on the networks using these α values. However, because the p-value of .0139 is greater than the α value .01, we would not reject H_0 and thus not run the trash bag commercial on the network using this α value.

A summary of the five steps of hypothesis testing In the real world, in spite of the greater efficiency of the p-value, both critical values and p-values are used to test hypotheses. For example, NBC uses critical value rules, whereas CBS uses p-values, to evaluate advertising claims statistically. Throughout this book we will (formally or informally) use the following five steps to implement the critical value and p-value approaches to hypothesis testing.

The Five Steps of Hypothesis Testing

1. State the null hypothesis H_0 and the alternative hypothesis H_a.
2. Specify the level of significance α.
3. Select the test statistic.

Using a critical value rule:

4. Determine the critical value rule for deciding whether to reject H_0.
5. Collect the sample data, compute the value of the test statistic, and decide whether to reject H_0 by using the critical value rule. Interpret the statistical results.

Using a *p*-value:

4. Collect the sample data, compute the value of the test statistic, and compute the p-value.
5. Reject H_0 at level of significance α if the p-value is less than α. Interpret the statistical results.

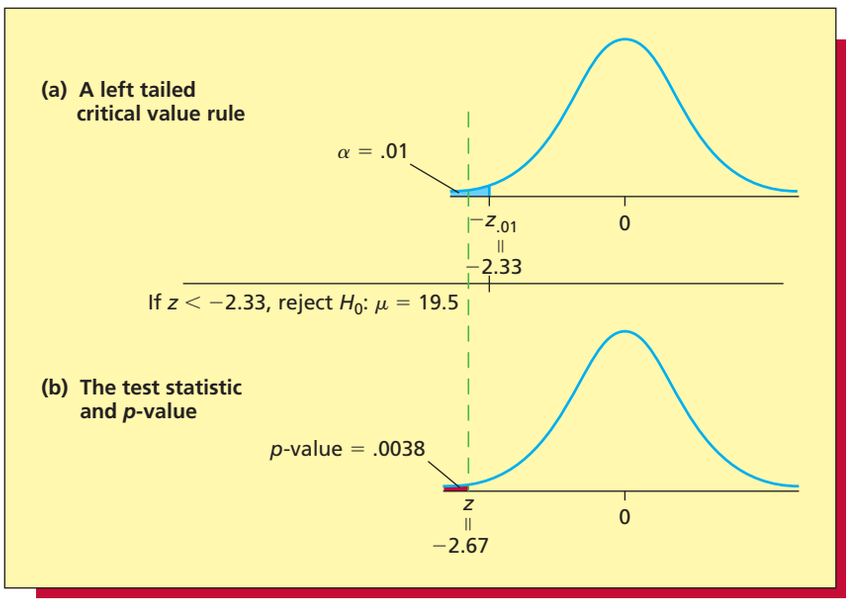
Testing a "less than" alternative hypothesis by using a critical value rule Consider the e-billing case. In order to study whether the new electronic billing system reduces the mean bill payment time by more than 50 percent, the management consulting firm will test **H_0: $\mu \geq 19.5$ versus H_a: $\mu < 19.5$ (step 1).** A Type I error (concluding that H_a: $\mu < 19.5$ is true when H_0: $\mu \geq 19.5$ is true) would result in the consulting firm overstating the benefits of the new billing system, both to the company in which it has been installed and to other companies that are considering installing such a system. Because the consulting firm desires to have only a 1 percent chance of doing this, the firm will set **α equal to .01 (step 2).**

To perform the hypothesis test, we will test the modified null hypothesis H_0: $\mu = 19.5$ versus H_a: $\mu < 19.5$. We will randomly select a sample of $n = 65$ invoices paid using the new billing system and calculate the mean $\bar{x}$ of the payment times of these invoices. Then, because the sample size is large, the Central Limit Theorem applies, and we will utilize the **test statistic (step 3)**

$$z = \frac{\bar{x} - 19.5}{\sigma/\sqrt{n}}$$

A value of the test statistic z that is less than zero results when $\bar{x}$ is less than 19.5. This provides evidence to support rejecting H_0 in favor of H_a because the point estimate $\bar{x}$ indicates that μ might be less than 19.5. To decide how much less than zero the test statistic must be to reject H_0 in favor

FIGURE 9.3 Testing H_0: $\mu = 19.5$ versus H_a: $\mu < 19.5$ by Using a Critical Value and a p-Value



of H_a at significance level α, we will use the **left tailed critical value rule** illustrated in Figure 9.3(a). Specifically, we will place α in the left-hand tail of the standard normal curve and **reject H_0: $\mu = 19.5$ in favor of H_a: $\mu < 19.5$ if and only if the test statistic z is less than the critical value $-z_\alpha$ (step 4).** Here, $-z_\alpha$ is the point on the horizontal axis under the standard normal curve that gives a left-hand tail area equal to α. Because α equals .01, the critical value $-z_\alpha$ is $-z_{.01} = -2.33$ [see Figure 9.3(a)].

When the sample of $n = 65$ invoices is randomly selected, the mean of the payment times of these invoices is calculated to be $\bar{x} = 18.1077$ days. Assuming that the population standard deviation σ of payment times for the new electronic billing system is 4.2 days (as discussed on page 280 of Chapter 7), the **value of the test statistic** is

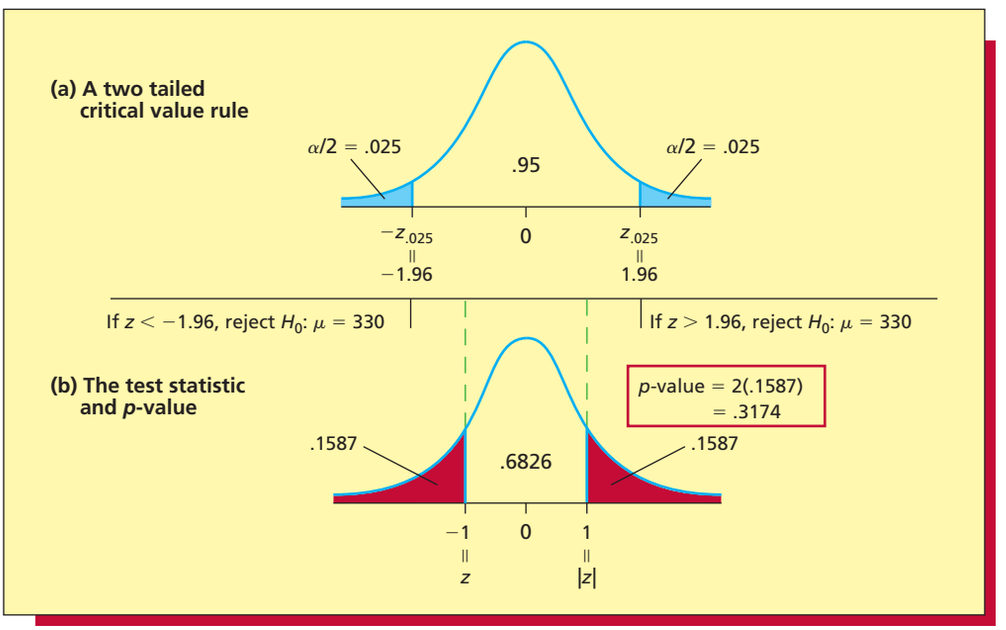
$$z = \frac{\bar{x} - 19.5}{\sigma/\sqrt{n}} = \frac{18.1077 - 19.5}{4.2/\sqrt{65}} = -2.67$$

Because $z = -2.67$ is less than the critical value $-z_{.01} = -2.33$, we can reject H_0: $\mu = 19.5$ in favor of H_a: $\mu < 19.5$ by setting α equal to .01 (step 5). Therefore, we conclude (at an α of .01) that the population mean payment time for the new electronic billing system is less than 19.5 days. This, along with the fact that the sample mean $\bar{x} = 18.1077$ is slightly less than 19.5, implies that it is reasonable for the management consulting firm to conclude (and claim) that the new electronic billing system has reduced the population mean payment time by slightly more than 50 percent.

BI

A p-value for testing a "less than" alternative hypothesis The value of the test statistic computed from the sample data is $z = -2.67$. Because we are testing H_0: $\mu = 19.5$ versus the **less than** alternative hypothesis H_a: $\mu < 19.5$, a value of the test statistic that is at least as contradictory to H_0 and supportive of H_a as $z = -2.67$ is a value of the test statistic that is **less than or equal to** $z = -2.67$. Therefore, the p-value is the probability, computed assuming that H_0: $\mu = 19.5$ is true, of observing a value of the test statistic that is less than or equal to $z = -2.67$. As illustrated in Figure 9.3(b), this ***p*-value is the area under the standard normal curve to the left of $z = -2.67$ (step 4)** and equals .0038 (see Table A.3, page 790). The p-value of .0038 says that, if H_0: $\mu = 19.5$ is true, then only 38 in 10,000 of all possible test statistic values are at least as negative, or contradictory to H_0, as the value $z = -2.67$. Moreover, recall that the management consulting firm has set α equal to .01. **Because the p-value of .0038 is less than the α of .01, we can reject H_0: $\mu = 19.5$ in favor of H_a: $\mu < 19.5$ by setting α equal to .01 (step 5).**

FIGURE 9.4 Testing H_0: $\mu = 330$ versus H_a: $\mu \neq 330$ by Using Critical Values and the *p*-Value



Testing a "not equal to" alternative hypothesis by using a critical value rule Consider the Valentine's Day chocolate case. To assess whether this year's sales of its valentine box of assorted chocolates will be 10 percent higher than last year's, the candy company will test **H_0: $\mu = 330$ versus H_a: $\mu \neq 330$ (step 1).** Here, μ is the mean order quantity of this year's valentine box by large retail stores. If the candy company does not reject H_0: $\mu = 330$ and H_0: $\mu = 330$ is false (a Type II error) the candy company will base its production of valentine boxes on a 10 percent projected sales increase that is not correct. Because the candy company wishes to have a reasonably small probability of making this Type II error, the company will set **α equal to .05 (step 2).** Setting α equal to .05 rather than .01 makes the probability of a Type II error smaller than it would be if α were set at .01. (See Section 9.5 for more on Type II errors.)

To perform the hypothesis test, the candy company will randomly select $n = 100$ large retail stores and will make an early mailing to these stores promoting this year's valentine box of assorted chocolates. The candy company will then ask each sampled retail store to report its anticipated order quantity of valentine boxes and will calculate the mean $\bar{x}$ of the reported order quantities. Because the sample size is large, the Central Limit Theorem applies, and we will utilize the **test statistic (step 3)**

$$z = \frac{\bar{x} - 330}{\sigma/\sqrt{n}}$$

A value of the test statistic that is greater than 0 results when $\bar{x}$ is greater than 330. This provides evidence to support rejecting H_0 in favor of H_a because the point estimate $\bar{x}$ indicates that μ might be greater than 330. Similarly, a value of the test statistic that is less than 0 results when $\bar{x}$ is less than 330. This also provides evidence to support rejecting H_0 in favor of H_a because the point estimate $\bar{x}$ indicates that μ might be less than 330. To decide how different from zero (positive or negative) the test statistic must be in order to reject H_0 in favor of H_a at significance level α, we will use the **two tailed critical value rule** illustrated in Figure 9.4(a). Specifically, we will divide α into two equal parts and **reject H_0: $\mu = 330$ in favor of H_a: $\mu \neq 330$ if and only if the test statistic z is greater than the critical value $z_{\alpha/2}$ or less than the critical value $-z_{\alpha/2}$ (step 4).** Note that this is equivalent to saying that we should **reject H_0 if and only if the absolute value of the test statistic, $|z|$, is greater than the critical value $z_{\alpha/2}$.** Because $\alpha/2 = .05/2 = .025$, it follows that the critical values $z_{\alpha/2}$ and $-z_{\alpha/2}$ are $z_{.025} = 1.96$ and $-z_{.025} = -1.96$ [see Figure 9.4(a)].

When the sample of $n = 100$ large retail stores is randomly selected, the mean of their reported order quantities is calculated to be $\bar{x} = 326$ boxes. Assuming that the population standard deviation σ of large retail store order quantities for this year's valentine box will be 40 boxes (the same as it was for previous years' valentine boxes), the value of the test statistic is

$$z = \frac{\bar{x} - 330}{\sigma/\sqrt{n}} = \frac{326 - 330}{40/\sqrt{100}} = -1$$

Because $z = -1$ is between the critical values $-z_{.025} = -1.96$ and $z_{.025} = 1.96$ (or, equivalently, because $|z| = 1$ is less than $z_{.025} = 1.96$), we cannot reject H_0: $\mu = 330$ in favor of H_a: $\mu \neq 330$ by setting α equal to .05 (step 5). Therefore, we cannot conclude (at an α of .05) that the population mean order quantity of this year's valentine box by large retail stores will differ from 330 boxes. It follows that the candy company will base its production of valentine boxes on the ten percent projected sales increase.

A *p*-value for testing a "not equal to" alternative hypothesis The value of the test statistic computed from the sample data is $z = -1$. Because the alternative hypothesis H_a: $\mu \neq 330$ says that μ might be greater or less than 330, both positive and negative test statistic values contradict H_0: $\mu = 330$ and support H_a: $\mu \neq 330$. It follows that a value of the test statistic that is at least as contradictory to H_0 and supportive of H_a as $z = -1$ is a value of the test statistic that is greater than or equal to 1 or less than or equal to -1. Therefore, the p-value is the probability, computed assuming that H_0: $\mu = 330$ is true, of observing a value of the test statistic that is greater than or equal to 1 or less than or equal to -1. As illustrated in Figure 9.4 (b), this p-value equals the area under the standard normal curve to the right of 1, plus the area under this curve to the left of -1. But, by the symmetry of the normal curve, the sum of these two areas, and thus the ***p*-value, is twice the area under the standard normal curve to the right of $|z| = 1$, the absolute value of the test statistic (step 4).** Because the area under the standard normal curve to the right of $|z| = 1$ is $1 - .8413 = .1587$ (see Table A.3, page 790), the p-value is $2(.1587) = .3174$. The p-value of .3174 says that, if H_0: $\mu = 330$ is true, then 31.74 percent of all possible test statistic values are at least as contradictory to H_0 as $z = -1$. Moreover, recall that the candy company has set α equal to .05. **Because the *p*-value of .3174 is greater than the α of .05, we cannot reject H_0: $\mu = 330$ in favor of H_a: $\mu \neq 330$ by setting α equal to .05 (step 5).**

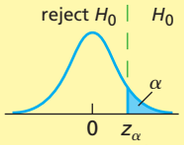
A general procedure for testing a hypothesis about a population mean In general, let μ_0 be the value for μ specified in the null hypothesis. Then, the following summary box gives the critical value and p-value procedures for performing (1) a right tailed test of H_0: $\mu = \mu_0$ versus H_a: $\mu > \mu_0$ (for example, H_0: $\mu = 50$ versus H_a: $\mu > 50$); (2) a left tailed test of H_0: $\mu = \mu_0$ versus H_a: $\mu < \mu_0$ (for example, H_0: $\mu = 19.5$ versus H_a: $\mu < 19.5$); and (3) a two tailed test of H_0: $\mu = \mu_0$ versus H_a: $\mu \neq \mu_0$ (for example, H_0: $\mu = 330$ versus H_a: $\mu \neq 330$).

Testing a Hypothesis about a Population Mean When σ Is Known

Null Hypothesis H_0: $\mu = \mu_0$ **Test Statistic** $z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ **Assumptions** Normal population or Large sample size

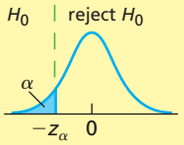
Critical Value Rule

H_a: $\mu > \mu_0$



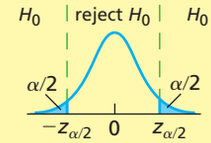
Reject H_0 if $z > z_\alpha$

H_a: $\mu < \mu_0$



Reject H_0 if $z < -z_\alpha$

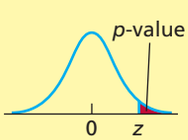
H_a: $\mu \neq \mu_0$



Reject H_0 if $|z| > z_{\alpha/2}$—that is, $z > z_{\alpha/2}$ or $z < -z_{\alpha/2}$

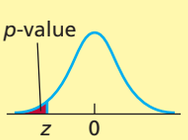
p-Value (Reject H_0 if p-Value $< \alpha$)

H_a: $\mu > \mu_0$



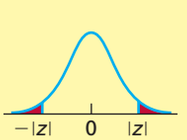
p-value = area to the right of z

H_a: $\mu < \mu_0$



p-value = area to the left of z

H_a: $\mu \neq \mu_0$



p-value = twice the area to the right of $|z|$

Using confidence intervals to test hypotheses Confidence intervals can be used to test hypotheses. Specifically, it can be proven that we can reject H_0: $\mu = \mu_0$ in favor of H_a: $\mu \neq \mu_0$ by setting the probability of a Type I error equal to α if and only if the $100(1 - \alpha)$ percent confidence interval for μ does not contain μ_0. For example, consider the Valentine's Day chocolate case and testing H_0: $\mu = 330$ versus H_a: $\mu \neq 330$ by setting α equal to .05. To do this, we use the mean $\bar{x} = 326$ of the sample of $n = 100$ reported order quantities to calculate the 95 percent confidence interval for μ to be

$$\left[\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] = \left[326 \pm 1.96\frac{40}{\sqrt{100}}\right] = [318.2, 333.8]$$

Because this interval does contain 330, we cannot reject H_0: $\mu = 330$ in favor of H_a: $\mu \neq 330$ by setting α equal to .05.

While we can use **two-sided confidence intervals** to test "not equal to" alternative hypotheses, we must use **one-sided confidence intervals** to test "greater than" or "less than" alternative hypotheses. We will not study one-sided confidence intervals in this book. However, it should be emphasized that we do not need to use confidence intervals (one-sided or two-sided) to test hypotheses. We can test hypotheses by using test statistics and critical values or *p*-values, and these are the approaches that we will feature throughout this book.

Measuring the weight of evidence against the null hypothesis As originally defined, the *p*-value is a probability that measures the likelihood of the sample results if the null hypothesis H_0 is true. **The smaller the *p*-value is, the less likely are the sample results if the null hypothesis H_0 is true. Therefore, the stronger is the evidence that H_0 is false and that the alternative hypothesis H_a is true.** Interpreted in this way, the *p*-value can be regarded as a measure of the **weight of evidence** against the null hypothesis and in favor of the alternative hypothesis. Through statistical practice, statisticians have concluded (somewhat subjectively) that:

Interpreting the Weight of Evidence against the Null Hypothesis

If the *p*-value for testing H_0 is less than

- .10, we have **some evidence** that H_0 is false.
- .05, we have **strong evidence** that H_0 is false.
- .01, we have **very strong evidence** that H_0 is false.
- .001, we have **extremely strong evidence** that H_0 is false.

We will frequently use these conclusions in future examples. Understand, however, that there are really no sharp borders between different weights of evidence. Rather, there is really only increasingly strong evidence against the null hypothesis as the *p*-value decreases. For example, the trash bag manufacturer, in addition to deciding whether H_0: $\mu = 50$ can be rejected in favor of H_a: $\mu > 50$ at each television network's chosen value of α, would almost certainly wish to know how much evidence there is that its new trash bag is stronger than its current trash bag. The *p*-value for testing H_0: $\mu = 50$ is .0139, which is less than .05 but not less than .01. Therefore, we have strong evidence, but not very strong evidence, that H_0: $\mu = 50$ is false and H_a: $\mu > 50$ is true. As another example, the *p*-value in the e-billing case is .0038. This provides very strong evidence, but not extremely strong evidence, that H_0: $\mu = 19.5$ is false and H_a: $\mu < 19.5$ is true. Finally, the *p*-value in the Valentine's Day chocolate case is .3174. This provides little evidence that H_0: $\mu = 330$ is false and H_a: $\mu \neq 330$ is true.

Exercises for Section 9.2

CONCEPTS

connect

9.8 Explain what a critical value is, and explain how it is used to test a hypothesis.

9.9 Explain what a *p*-value is, and explain how it is used to test a hypothesis.

METHODS AND APPLICATIONS

9.10 Suppose that we wish to test H_0: $\mu = 80$ versus H_a: $\mu > 80$, where σ is known to equal 20. Also, suppose that a sample of $n = 100$ measurements randomly selected from the population has a mean of $\bar{x} = 85$.

a Calculate the value of the test statistic z.
b By comparing z with a critical value, test H_0 versus H_a at $\alpha = .05$.
c Calculate the p-value for testing H_0 versus H_a.
d Use the p-value to test H_0 versus H_a at each of $\alpha = .10, .05, .01$, and $.001$.
e How much evidence is there that H_0: $\mu = 80$ is false and H_a: $\mu > 80$ is true?

9.11 Suppose that we wish to test H_0: $\mu = 20$ versus H_a: $\mu < 20$, where σ is known to equal 7. Also, suppose that a sample of $n = 49$ measurements randomly selected from the population has a mean of $\bar{x} = 18$.

a Calculate the value of the test statistic z.
b By comparing z with a critical value, test H_0 versus H_a at $\alpha = .01$.
c Calculate the p-value for testing H_0 versus H_a.
d Use the p-value to test H_0 versus H_a at each of $\alpha = .10, .05, .01$, and $.001$.
e How much evidence is there that H_0: $\mu = 20$ is false and H_a: $\mu < 20$ is true?

9.12 Suppose that we wish to test H_0: $\mu = 40$ versus H_a: $\mu \neq 40$, where σ is known to equal 18. Also, suppose that a sample of $n = 81$ measurements randomly selected from the population has a mean of $\bar{x} = 35$.

a Calculate the value of the test statistic z.
b By comparing z with critical values, test H_0 versus H_a at $\alpha = .05$.
c Calculate the p-value for testing H_0 versus H_a.
d Use the p-value to test H_0 versus H_a at each of $\alpha = .10, .05, .01$, and $.001$.
e How much evidence is there that H_0: $\mu = 40$ is false and H_a: $\mu \neq 40$ is true?

9.13 **THE VIDEO GAME SATISFACTION RATING CASE** VideoGame

Recall that "very satisfied" customers give the XYZ-Box video game system a rating that is at least 42. Suppose that the manufacturer of the XYZ-Box wishes to use the random sample of 65 satisfaction ratings to provide evidence supporting the claim that the mean composite satisfaction rating for the XYZ-Box exceeds 42.

a Letting μ represent the mean composite satisfaction rating for the XYZ-Box, set up the null hypothesis H_0 and the alternative hypothesis H_a needed if we wish to attempt to provide evidence supporting the claim that μ exceeds 42.
b The random sample of 65 satisfaction ratings yields a sample mean of $\bar{x} = 42.954$. Assuming that σ equals 2.64, use critical values to test H_0 versus H_a at each of $\alpha = .10, .05, .01$, and $.001$.
c Using the information in part b, calculate the p-value and use it to test H_0 versus H_a at each of $\alpha = .10, .05, .01$, and $.001$.
d How much evidence is there that the mean composite satisfaction rating exceeds 42?

9.14 **THE BANK CUSTOMER WAITING TIME CASE** WaitTime

Recall that a bank manager has developed a new system to reduce the time customers spend waiting for teller service during peak hours. The manager hopes the new system will reduce waiting times from the current 9 to 10 minutes to less than 6 minutes.

Suppose the manager wishes to use the random sample of 100 waiting times to support the claim that the mean waiting time under the new system is shorter than six minutes.

a Letting μ represent the mean waiting time under the new system, set up the null and alternative hypotheses needed if we wish to attempt to provide evidence supporting the claim that μ is shorter than six minutes.
b The random sample of 100 waiting times yields a sample mean of $\bar{x} = 5.46$ minutes. Assuming that $\sigma = 2.47$ minutes, use critical values to test H_0 versus H_a at each of $\alpha = .10, .05, .01$, and $.001$.
c Using the information in part b, calculate the p-value and use it to test H_0 versus H_a at each of $\alpha = .10, .05, .01$, and $.001$.
d How much evidence is there that the new system has reduced the mean waiting time to below six minutes?

9.15 Consolidated Power, a large electric power utility, has just built a modern nuclear power plant. This plant discharges waste water that is allowed to flow into the Atlantic Ocean. The Environmental Protection Agency (EPA) has ordered that the waste water may not be excessively warm so that

thermal pollution of the marine environment near the plant can be avoided. Because of this order, the waste water is allowed to cool in specially constructed ponds and is then released into the ocean. This cooling system works properly if the mean temperature of waste water discharged is 60°F or cooler. Consolidated Power is required to monitor the temperature of the waste water. A sample of 100 temperature readings will be obtained each day, and if the sample results cast a substantial amount of doubt on the hypothesis that the cooling system is working properly (the mean temperature of waste water discharged is 60°F or cooler), then the plant must be shut down and appropriate actions must be taken to correct the problem.

a Consolidated Power wishes to set up a hypothesis test so that the power plant will be shut down when the null hypothesis is rejected. Set up the null hypothesis H_0 and the alternative hypothesis H_a that should be used.

b Suppose that Consolidated Power decides to use a level of significance of $\alpha = .05$, and suppose a random sample of 100 temperature readings is obtained. If the sample mean of the 100 temperature readings is $\bar{x} = 60.482$, test H_0 versus H_a and determine whether the power plant should be shut down and the cooling system repaired. Perform the hypothesis test by using a critical value and a p-value. Assume $\sigma = 2$.

9.16 Do part *b* of Exercise 9.15 if $\bar{x} = 60.262$.

9.17 Do part *b* of Exercise 9.15 if $\bar{x} = 60.618$.

9.18 An automobile parts supplier owns a machine that produces a cylindrical engine part. This part is supposed to have an outside diameter of three inches. Parts with diameters that are too small or too large do not meet customer requirements and must be rejected. Lately, the company has experienced problems meeting customer requirements. The technical staff feels that the mean diameter produced by the machine is off target. In order to verify this, a special study will randomly sample 100 parts produced by the machine. The 100 sampled parts will be measured, and if the results obtained cast a substantial amount of doubt on the hypothesis that the mean diameter equals the target value of three inches, the company will assign a problem-solving team to intensively search for the causes of the problem.

a The parts supplier wishes to set up a hypothesis test so that the problem-solving team will be assigned when the null hypothesis is rejected. Set up the null and alternative hypotheses for this situation.

b A sample of 40 parts yields a sample mean diameter of $\bar{x} = 3.006$ inches. Assuming σ equals .016: **(1)** Use a critical value to test H_0 versus H_a by setting α equal to .05. **(2)** Should the problem-solving team be assigned? **(3)** Use a p-value to test H_0 versus H_a with $\alpha = .05$.

9.19 The Crown Bottling Company has just installed a new bottling process that will fill 16-ounce bottles of the popular Crown Classic Cola soft drink. Both overfilling and underfilling bottles are undesirable: Underfilling leads to customer complaints and overfilling costs the company considerable money. In order to verify that the filler is set up correctly, the company wishes to see whether the mean bottle fill, μ, is close to the target fill of 16 ounces. To this end, a random sample of 36 filled bottles is selected from the output of a test filler run. If the sample results cast a substantial amount of doubt on the hypothesis that the mean bottle fill is the desired 16 ounces, then the filler's initial setup will be readjusted.

a The bottling company wants to set up a hypothesis test so that the filler will be readjusted if the null hypothesis is rejected. Set up the null and alternative hypotheses for this hypothesis test.

b Suppose that Crown Bottling Company decides to use a level of significance of $\alpha = .01$, and suppose a random sample of 36 bottle fills is obtained from a test run of the filler. For each of the following four sample means—$\bar{x} = 16.05$, $\bar{x} = 15.96$, $\bar{x} = 16.02$, and $\bar{x} = 15.94$—determine whether the filler's initial setup should be readjusted. In each case, use **(1)** a critical value, **(2)** a p-value, and **(3)** a confidence interval. Assume that σ equals .1.

9.20 THE DISK BRAKE CASE

National Motors has equipped the ZX-900 with a new disk brake system. We define μ to be the mean stopping distance (from a speed of 35 mph) of all ZX-900s. National Motors would like to claim that the ZX-900 achieves a shorter mean stopping distance than the 60 ft claimed by a competitor.

a Set up the null and alternative hypotheses needed to support National Motors' claim.

b A television network will allow National Motors to advertise its claim if the appropriate null hypothesis can be rejected at $\alpha = .05$. If a random sample of 81 ZX-900s have a mean stopping distance of $\bar{x} = 57.8$ ft, will National Motors be allowed to advertise the claim? Assume that $\sigma = 6.02$ ft and justify your answer using both a critical value and a p-value.

9.3 *t* Tests about a Population Mean: σ Unknown

LO9-4 Use critical values and *p*-values to perform a *t* test about a population mean when σ is unknown.

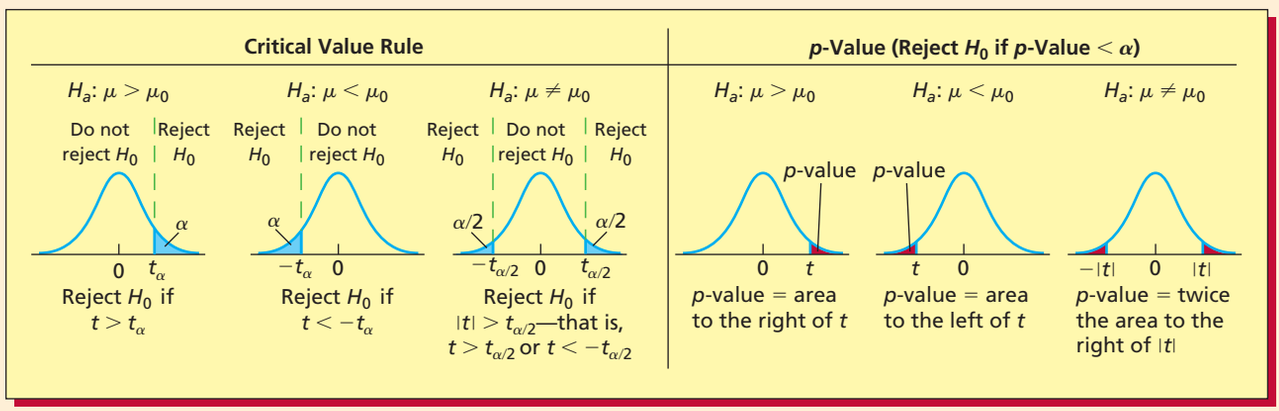
If we do not know σ (which is usually the case), we can base a hypothesis test about μ on the sampling distribution of

$$\frac{\bar{x} - \mu}{s/\sqrt{n}}$$

If the sampled population is normally distributed (or if the sample size is large—at least 30), then this sampling distribution is exactly (or approximately) a ***t* distribution having $n - 1$ degrees of freedom.** This leads to the following results:

A *t* Test about a Population Mean: σ Unknown

Null Hypothesis $H_0: \mu = \mu_0$ **Test Statistic** $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$ $df = n - 1$ **Assumptions** Normal population or Large sample size



EXAMPLE 9.4 The Commercial Loan Case: Mean Debt-to-Equity Ratio

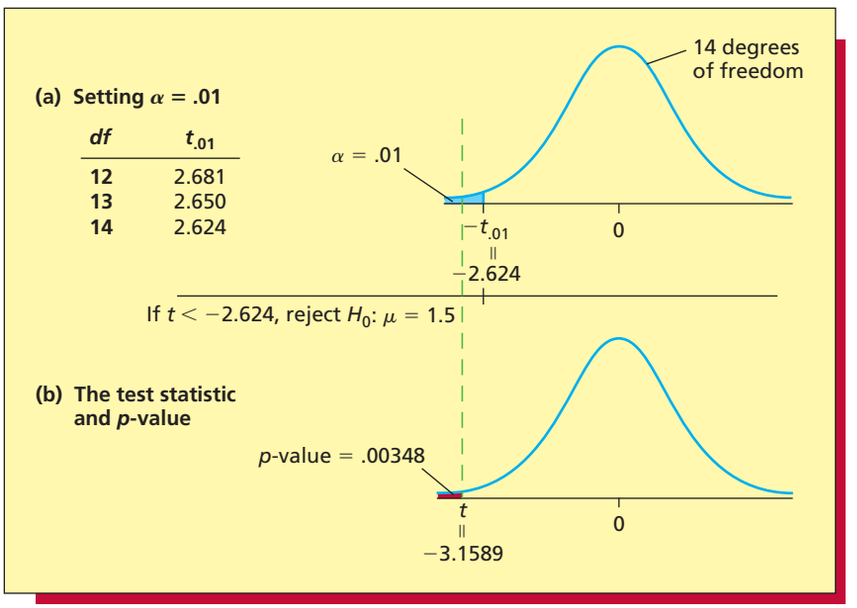
One measure of a company's financial health is its *debt-to-equity ratio.* This quantity is defined to be the ratio of the company's corporate debt to the company's equity. If this ratio is too high, it is one indication of financial instability. For obvious reasons, banks often monitor the financial health of companies to which they have extended commercial loans. Suppose that, in order to reduce risk, a large bank has decided to initiate a policy limiting the mean debt-to-equity ratio for its portfolio of commercial loans to being less than 1.5. In order to assess whether the mean debt-to-equity ratio μ of its (current) commercial loan portfolio is less than 1.5, the bank will test the **null hypothesis H_0: $\mu = 1.5$ versus the alternative hypothesis H_a: $\mu < 1.5$.** In this situation, a Type I error (rejecting H_0: $\mu = 1.5$ when H_0: $\mu = 1.5$ is true) would result in the bank concluding that the mean debt-to-equity ratio of its commercial loan portfolio is less than 1.5 when it is not. Because the bank wishes to be very sure that it does not commit this Type I error, it will test H_0 versus H_a by using a **.01 level of significance.** To perform the hypothesis test, the bank randomly selects a sample of 15 of its commercial loan accounts. Audits of these companies result in the following debt-to-equity ratios (arranged in increasing order): 1.05, 1.11, 1.19, 1.21, 1.22, 1.29, 1.31, 1.32, 1.33, 1.37, 1.41, 1.45, 1.46, 1.65, and 1.78. The mound-shaped stem-and-leaf display of these ratios is given on the page margin and indicates that the population of all debt-to-equity ratios is (approximately) normally distributed. It follows that it is appropriate to calculate the value of the **test statistic *t* in the summary box.** Furthermore, because H_a: $\mu < 1.5$ implies a left tailed test, we should **reject H_0: $\mu = 1.5$ if the value of *t* is less than the critical value $-t_\alpha = -t_{.01} = -2.624$.** Here, $-t_{.01} = -2.624$ is based on $n - 1 = 15 - 1 = 14$ degrees of freedom (see Table A.4, on page 792), and this critical value is illustrated in Figure 9.5(a). The mean and the standard deviation of the random sample of $n = 15$ debt-to-equity ratios are $\bar{x} = 1.3433$ and $s = .1921$. This implies that the **value of the test statistic** is

1.0	5
1.1	1 9
1.2	1 2 9
1.3	1 2 3 7
1.4	1 5 6
1.5	
1.6	5
1.7	8

DS DebtEq

$$t = \frac{\bar{x} - 1.5}{s/\sqrt{n}} = \frac{1.3433 - 1.5}{.1921/\sqrt{15}} = -3.1589$$

FIGURE 9.5 Testing H_0: $\mu = 1.5$ versus H_a: $\mu < 1.5$ by Using a Critical Value and the p-Value



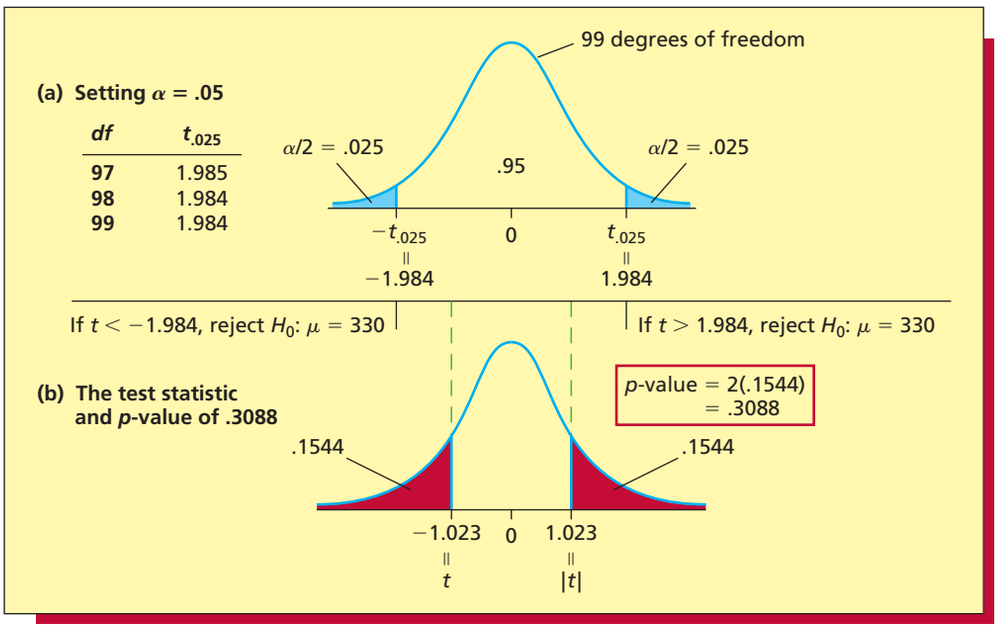

```
Test of mu = 1.5 vs < 1.5

                                                95% Upper
Variable    N     Mean    StDev   SE Mean      Bound       T       P
Ratio      15   1.3433   0.1921    0.0496     1.4307   -3.16   0.003
```

Because $t = -3.1589$ is less than $-t_{.01} = -2.624$, we reject H_0: $\mu = 1.5$ in favor of H_a: $\mu < 1.5$. That is, we conclude (at an α of .01) that the population mean debt-to-equity ratio of the bank's commercial loan portfolio is less than 1.5. This, along with the fact that the sample mean $\bar{x} = 1.3433$ is slightly less than 1.5, implies that it is reasonable for the bank to conclude that the population mean debt-to-equity ratio of its commercial loan portfolio is slightly less than 1.5.

The p-value for testing H_0: $\mu = 1.5$ versus H_a: $\mu < 1.5$ is the area under the curve of the t distribution having 14 degrees of freedom to the left of $t = -3.1589$. Tables of t points (such as Table A.4) are not complete enough to give such areas for most t statistic values, so we use computer software packages to calculate p-values that are based on the t distribution. For example, Excel tells us that the p-value for testing H_0: $\mu = 1.5$ versus H_a: $\mu < 1.5$ is .00348, which is given in the rounded form .003 on the MINITAB output at the bottom of Figure 9.5. The p-value of .00348 says that if we are to believe that H_0 is true, we must believe that we have observed a test statistic value that can be described as having a 348 in 100,000 chance. Moreover, because the p-value of .00348 is between .01 and .001, we have very strong evidence, but not extremely strong evidence, that H_0: $\mu = 1.5$ is false and H_a: $\mu < 1.5$ is true. That is, we have very strong evidence that the mean debt-to-equity ratio of the bank's commercial loan portfolio is less than 1.5.

Recall that in three cases discussed in Section 9.2 we tested hypotheses by assuming that the population standard deviation σ is known and by using z tests. If σ is actually not known in these cases (which would probably be true), we should test the hypotheses under consideration by using t tests. Furthermore, recall that in each case the sample size is large (at least 30). **In general, it can be shown that if the sample size is large, the t test is approximately valid even if the sampled population is not normally distributed (or mound shaped).** Therefore, consider the Valentine's Day chocolate case and testing **H_0: $\mu = 330$ versus H_a: $\mu \neq 330$** at the **.05 level of significance.** To perform the hypothesis test, assume that we will randomly select $n = 100$ large

FIGURE 9.6 Testing H_0: $\mu = 330$ versus H_a: $\mu \neq 330$ by Using Critical Values and the p-Value




```
Test of mu = 330 vs not = 330

Variable    N     Mean    StDev  SE Mean        95% CI            T        P
Boxes     100   326.00    39.10     3.91  (318.24, 333.76)   -1.02    0.309
```

retail stores and use their anticipated order quantities to calculate the value of the **test statistic t in the summary box**. Then, because the alternative hypothesis H_a: $\mu \neq 330$ implies a two tailed test, we will **reject H_0: $\mu = 330$ if the absolute value of t is greater than $t_{\alpha/2} = t_{.025} = 1.984$ (based on $n - 1 = 99$ degrees of freedom)**—see Figure 9.6(a). Suppose that when the sample is randomly selected, the mean and the standard deviation of the $n = 100$ reported order quantities are calculated to be $\bar{x} = 326$ and $s = 39.1$. The **value of the test statistic** is

$$t = \frac{\bar{x} - 330}{s/\sqrt{n}} = \frac{326 - 330}{39.1/\sqrt{100}} = -1.023$$

Because $|t| = 1.023$ is less than $t_{.025} = 1.984$, we cannot reject H_0: $\mu = 330$ by setting α equal to .05. It follows that we cannot conclude (at an α of .05) that this year's population mean order quantity of the valentine box by large retail stores will differ from 330 boxes. Therefore, the candy company will base its production of valentine boxes on the 10 percent projected sales increase. The p-value for the hypothesis test is twice the area under the t distribution curve having 99 degrees of freedom to the right of $|t| = 1.023$. Using a computer, we find that this p-value is .3088 (see Figure 9.6(b)), which provides little evidence against H_0: $\mu = 330$ and in favor of H_a: $\mu \neq 330$. In Exercises 9.24 and 9.25, the reader will assume that σ is unknown in the e-billing and trash bag cases and will perform appropriate t tests in these cases.

To conclude this section, note that if the sampled population is not approximately normally distributed and the sample size is not large, it might be appropriate to use a **nonparametric test about the population median** (see Chapter 18).

Exercises for Section 9.3

CONCEPTS

9.21 What assumptions must be met in order to carry out a t test about a population mean?

9.22 How do we decide whether to use a z test or a t test when testing a hypothesis about a population mean?

METHODS AND APPLICATIONS

9.23 Suppose that a random sample of nine measurements from a normally distributed population gives a sample mean of $\bar{x} = 2.57$ and a sample standard deviation of $s = .3$. Use critical values to test H_0: $\mu = 3$ versus H_a: $\mu \neq 3$ using levels of significance $\alpha = .10$, $\alpha = .05$, $\alpha = .01$, and $\alpha = .001$.

9.24 Consider the e-billing case. The mean and the standard deviation of the sample of $n = 65$ payment times are $\bar{x} = 18.1077$ and $s = 3.9612$. **(1)** Test H_0: $\mu = 19.5$ versus H_a: $\mu < 19.5$ by setting α equal to .01 and using a critical value rule. **(2)** Interpret the (computer calculated) p-value of .0031 for the test. That is, explain how much evidence there is that H_0 is false. DS PayTime

9.25 Consider the trash bag case. The mean and the standard deviation of the sample of $n = 40$ trash bag breaking strengths are $\bar{x} = 50.575$ and $s = 1.6438$. **(1)** Test H_0: $\mu = 50$ versus H_a: $\mu > 50$ by setting α equal to .05 and using a critical value rule. **(2)** Interpret the (computer calculated) p-value of .0164 for the test. That is, explain how much evidence there is that H_0 is false. DS TrashBag

9.26 THE AIR SAFETY CASE DS AlertTimes

Recall that it is hoped that the mean alert time, μ, using the new display panel is less than eight seconds. **(1)** Formulate the null hypothesis H_0 and the alternative hypothesis H_a that would be used to attempt to provide evidence that μ is less than eight seconds. **(2)** Discuss the meanings of a Type I error and a Type II error in this situation. **(3)** The mean and the standard deviation of the sample of $n = 15$ alert times are $\bar{x} = 7.4$ and $s = 1.0261$. Test H_0 versus H_a by setting α equal to .05 and using a critical value. Assume that the population of all alert times using the new display panel is approximately normally distributed. **(4)** Interpret the p-value of .02 for the test.

9.27 The *bad debt ratio* for a financial institution is defined to be the dollar value of loans defaulted divided by the total dollar value of all loans made. Suppose that a random sample of seven Ohio banks is selected and that the bad debt ratios (written as percentages) for these banks are 7%, 4%, 6%, 7%, 5%, 4%, and 9%. DS BadDebt

a Banking officials claim that the mean bad debt ratio for all Midwestern banks is 3.5 percent and that the mean bad debt ratio for Ohio banks is higher. **(1)** Set up the null and alternative hypotheses needed to attempt to provide evidence supporting the claim that the mean bad debt ratio for Ohio banks exceeds 3.5 percent. **(2)** Discuss the meanings of a Type I error and a Type II error in this situation.

b Assuming that bad debt ratios for Ohio banks are approximately normally distributed: **(1)** Use a critical value and the given sample information to test the hypotheses you set up in part *a* by setting α equal to .01. **(2)** Interpret the p-value of .006 for the test.

9.28 How might practical importance be defined for the situation in Exercise 9.27?

9.29 THE VIDEO GAME SATISFACTION RATING CASE DS VideoGame

Recall that "very satisfied" customers give the XYZ-Box video game system a composite satisfaction rating that is at least 42.

a Letting μ represent the mean composite satisfaction rating for the XYZ-Box, set up the null and alternative hypotheses needed if we wish to attempt to provide evidence supporting the claim that μ exceeds 42.

b The mean and the standard deviation of a sample of $n = 65$ customer satisfaction ratings are $\bar{x} = 42.95$ and $s = 2.6424$. **(1)** Use a critical value to test the hypotheses you set up in part *a* by setting α equal to .01. **(2)** Interpret the p-value of .0025 for the test.

9.30 THE BANK CUSTOMER WAITING TIME CASE DS WaitTime

Recall that a bank manager has developed a new system to reduce the time customers spend waiting for teller service during peak hours. The manager hopes the new system will reduce waiting times from the current 9 to 10 minutes to less than 6 minutes.

a Letting μ represent the mean waiting time under the new system, set up the null and alternative hypotheses needed if we wish to attempt to provide evidence supporting the claim that μ is shorter than six minutes.

b The mean and the standard deviation of a sample of 100 bank customer waiting times are $\bar{x} = 5.46$ and $s = 2.475$. **(1)** Use a critical value to test the hypotheses you set up in part *a* by setting α equal to .05. **(2)** Interpret the p-value of .0158 for the test.

9.31 Consider a chemical company that wishes to determine whether a new catalyst, catalyst XA-100, changes the mean hourly yield of its chemical process from the historical process mean of 750 pounds per hour. When five trial runs are made using the new catalyst, the following yields (in pounds per hour) are recorded: 801, 814, 784, 836, and 820. DS ChemYield

a Letting μ be the mean of all possible yields using the new catalyst, set up the null and alternative hypotheses needed if we wish to attempt to provide evidence that μ differs from 750 pounds.

b The mean and the standard deviation of the sample of 5 catalyst yields are $\bar{x} = 811$ and $s = 19.647$. **(1)** Using a critical value and assuming approximate normality, test the hypotheses you set up in part *a* by setting α equal to .01. **(2)** The *p*-value for the hypothesis test is given in the Excel output on the page margin. Interpret this *p*-value.

t-statistic
6.942585
p-value
0.002261

9.32 Recall from Exercise 8.12 that Bayus (1991) studied the mean numbers of auto dealers visited by early and late replacement buyers. **(1)** Letting μ be the mean number of dealers visited by all late replacement buyers, set up the null and alternative hypotheses needed if we wish to attempt to provide evidence that μ differs from 4 dealers. **(2)** A random sample of 100 late replacement buyers yields a mean and a standard deviation of the number of dealers visited of $\bar{x} = 4.32$ and $s = .67$. Use critical values to test the hypotheses you set up by setting α equal to .10, .05, .01, and .001. **(3)** Do we estimate that μ is less than 4 or greater than 4?

9.33 In 1991 the average interest rate charged by U.S. credit card issuers was 18.8 percent. Since that time, there has been a proliferation of new credit cards affiliated with retail stores, oil companies, alumni associations, professional sports teams, and so on. A financial officer wishes to study whether the increased competition in the credit card business has reduced interest rates. To do this, the officer will test a hypothesis about the current mean interest rate, μ, charged by all U.S. credit card issuers. To perform the hypothesis test, the officer randomly selects $n = 15$ credit cards and obtains the following interest rates (arranged in increasing order): 14.0, 14.6, 15.3, 15.6, 15.8, 16.4, 16.6, 17.0, 17.3, 17.6, 17.8, 18.1, 18.4, 18.7, and 19.2. A stem-and-leaf display of the interest rates is given on the page margin, and the MINITAB and Excel outputs for testing H_0: $\mu = 18.8$ versus H_a: $\mu < 18.8$ follow.

14	06
15	368
16	46
17	0368
18	147
19	2

CreditCd

```
Test of mu = 18.8 vs < 18.8

Variable      N      Mean    StDev    SE Mean        T        P
Rate         15   16.8267   1.5378     0.3971    -4.97    0.000
```

t-statistic
−4.97
p-value
0.000103

a Set up the null and alternative hypotheses needed to provide evidence that mean interest rates have decreased since 1991.

b Use the MINITAB and Excel outputs and critical values to test the hypotheses you set up in part (a) at the .05, .01, and .001 levels of significance.

c Use the MINITAB and Excel outputs and a *p*-value to test the hypotheses you set up in part (a) at the .05, .01, and .001 levels of significance.

d Based on your results in parts (b) and (c), how much evidence is there that mean interest rates have decreased since 1991?

9.4 z Tests about a Population Proportion

LO9-5 Use critical values and *p*-values to perform a large sample *z* test about a population proportion.

In this section we study a large sample hypothesis test about a population proportion (that is, about the fraction of population elements that possess some characteristic). We begin with an example.

EXAMPLE 9.5 The Cheese Spread Case: Improving Profitability

Recall that the soft cheese spread producer has decided that replacing the current spout with the new spout is profitable only if p, the true proportion of all current purchasers who would stop buying the cheese spread if the new spout were used, is less than .10. The producer feels that it is unwise to change the spout unless it has very strong evidence that p is less than .10. Therefore, the spout will be changed if and only if the null hypothesis H_0: $p = .10$ can be rejected in favor of the alternative hypothesis H_a: $p < .10$ at the .01 level of significance.

In order to see how to test this kind of hypothesis, remember that when n is large, the sampling distribution of

$$\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

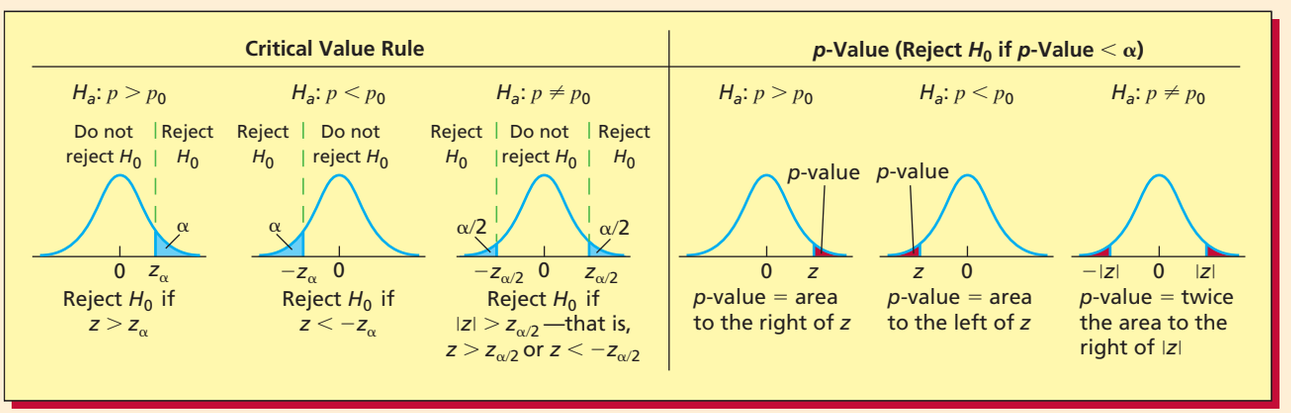
is approximately a standard normal distribution. Let p_0 denote a specified value between 0 and 1 (its exact value will depend on the problem), and consider testing the null hypothesis $H_0: p = p_0$. We then have the following result:

A Large Sample Test about a Population Proportion

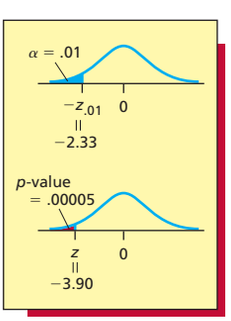
Null Hypothesis $H_0: p = p_0$

Test Statistic $$z = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1 - p_0)}{n}}}$$

Assumptions[3] $np_0 \geq 5$ and $n(1 - p_0) \geq 5$



EXAMPLE 9.6 The Cheese Spread Case: Improving Profitability



We have seen that the cheese spread producer wishes to test $\boldsymbol{H_0: p = .10}$ **versus** $\boldsymbol{H_a: p < .10}$, where p is the proportion of all current purchasers who would stop buying the cheese spread if the new spout were used. The producer will use the new spout if H_0 can be rejected in favor of H_a at the **.01 level of significance.** To perform the hypothesis test, we will randomly select $n = 1{,}000$ current purchasers of the cheese spread, find the proportion ($\hat{p}$) of these purchasers who would stop buying the cheese spread if the new spout were used, and calculate the value of the **test statistic z in the summary box.** Then, because the alternative hypothesis $H_a: p < .10$ implies a left tailed test, we will **reject $\boldsymbol{H_0: p = .10}$ if the value of z is less than $\boldsymbol{-z_\alpha = -z_{.01} = -2.33}$.** (Note that using this procedure is valid because $np_0 = 1{,}000(.10) = 100$ and $n(1 - p_0) = 1{,}000(1 - .10) = 900$ are both at least 5.) Suppose that when the sample is randomly selected, we find that 63 of the 1,000 current purchasers say they would stop buying the cheese spread if the new spout were used. Because $\hat{p} = 63/1{,}000 = .063$, the **value of the test statistic** is

$$z = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1 - p_0)}{n}}} = \frac{.063 - .10}{\sqrt{\dfrac{.10(1 - .10)}{1{,}000}}} = -3.90$$

Because $\boldsymbol{z = -3.90}$ is less than $\boldsymbol{-z_{.01} = -2.33}$, we reject $\boldsymbol{H_0: p = .10}$ in favor of $\boldsymbol{H_a: p < .10}$. That is, we conclude (at an α of .01) that the proportion of all current purchasers who would stop buying the cheese spread if the new spout were used is less than .10. It follows that the company will use the new spout. Furthermore, the point estimate $\hat{p} = .063$ says we estimate that 6.3 percent of all current customers would stop buying the cheese spread if the new spout were used.

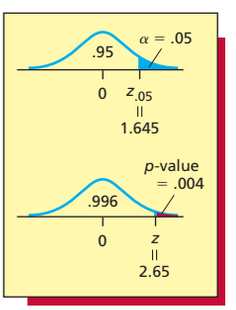
Although the cheese spread producer has made its decision by setting α equal to a single, prechosen value (.01), it would probably also wish to know the weight of evidence against H_0 and in favor of H_a. The p-value is the area under the standard normal curve to the left of $z = -3.90$. Table A.3 (page 790) tells us that this area is .00005. Because this p-value is less than .001, we

[3]Some statisticians suggest using the more conservative rule that both np_0 and $n(1 - p_0)$ must be at least 10.

have extremely strong evidence that H_a: $p < .10$ is true. That is, we have extremely strong evidence that fewer than 10 percent of all current purchasers would stop buying the cheese spread if the new spout were used.

EXAMPLE 9.7 The Phantol Case: Testing the Effectiveness of a Drug

Recent medical research has sought to develop drugs that lessen the severity and duration of viral infections. Virol, a relatively new drug, has been shown to provide relief for 70 percent of all patients suffering from viral upper respiratory infections. A major drug company is developing a competing drug called Phantol. The drug company wishes to investigate whether Phantol is more effective than Virol. To do this, the drug company will test a hypothesis about the proportion, p, of all patients whose symptoms would be relieved by Phantol. **The null hypothesis to be tested is H_0: $p = .70$, and the alternative hypothesis is H_a: $p > .70$.** If H_0 can be rejected in favor of H_a at the **.05 level of significance,** the drug company will conclude that Phantol helps more than the 70 percent of patients helped by Virol. To perform the hypothesis test, we will randomly select $n = 300$ patients having viral upper respiratory infections, find the proportion ($\hat{p}$) of these patients whose symptoms are relieved by Phantol and calculate the value of the **test statistic z in the summary box.** Then, because the alternative hypothesis H_a: $p > .70$ implies a right tailed test, we will **reject H_0: $p = .70$ if the value of z is greater than $z_\alpha = z_{.05} = 1.645$.** (Note that using this procedure is valid because $np_0 = 300(.70) = 210$ and $n(1 - p_0) = 300(1 - .70) = 90$ are both at least 5.) Suppose that when the sample is randomly selected, we find that Phantol provides relief for 231 of the 300 patients. Because $\hat{p} = 231/300 = .77$, the **value of the test statistic** is



$$z = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1 - p_0)}{n}}} = \frac{.77 - .70}{\sqrt{\dfrac{(.70)(1 - .70)}{300}}} = 2.65$$

Because $z = 2.65$ is greater than $z_{.05} = 1.645$, we reject H_0: $p = .70$ in favor of H_a: $p > .70$. That is, we conclude (at an α of .05) that Phantol will provide relief for more than 70 percent of all patients suffering from viral upper respiratory infections. More specifically, the point estimate $\hat{p} = .77$ of p says that we estimate that Phantol will provide relief for 77 percent of all such patients. Comparing this estimate to the 70 percent of all patients whose symptoms are relieved by Virol, we conclude that Phantol is somewhat more effective.

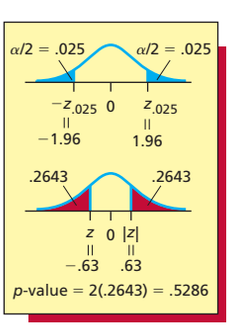
BI

The p-value for testing H_0: $p = .70$ versus H_a: $p > .70$ is the area under the standard normal curve to the right of $z = 2.65$. This p-value is $(1.0 - .9960) = .004$ (see Table A.3, page 790), and it provides very strong evidence against H_0: $p = .70$ and in favor of H_a: $p > .70$. That is, we have very strong evidence that Phantol will provide relief for more than 70 percent of all patients suffering from viral upper respiratory infections.

EXAMPLE 9.8 The Electronic Article Surveillance Case: False Alarms

C

A sports equipment discount store is considering installing an electronic article surveillance device and is concerned about the proportion, p, of all consumers who would never shop in the store again if the store subjected them to a false alarm. Suppose that industry data for general discount stores says that 15 percent of all consumers say that they would never shop in a store again if the store subjected them to a false alarm. To determine whether this percentage is different for the sports equipment discount store, the store will test the **null hypothesis H_0: $p = .15$ versus the alternative hypothesis H_a: $p \neq .15$** at the **.05 level of significance.** To perform the hypothesis test, the store will randomly select $n = 500$ consumers, find the proportion $\hat{p}$ of these consumers who say that they would never shop in the store again if the store subjected them to a false alarm, and calculate the value of the **test statistic z in the summary box.** Then, because the alternative hypothesis H_a: $p \neq .15$ implies a two tailed test, we will **reject H_0: $p = .15$ if $|z|$, the absolute value of the test statistic z, is greater than $z_{\alpha/2} = z_{.025} = 1.96$.** (Note that using this procedure is valid because $np_0 = (500)(.15) = 75$ and $n(1 - p_0) = (500)(1 - .15) = 425$ are both at least 5.)



Suppose that when the sample is randomly selected, we find that 70 out of 500 consumers say that they would never shop in the store again if the store subjected them to a false alarm. Because $\hat{p} = 70/500 = .14$, the **value of the test statistic** is

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} = \frac{.14 - .15}{\sqrt{\frac{.15(1 - .15)}{500}}} = -.63$$

BI

Because $|z| = .63$ is less than $z_{.025} = 1.96$, we cannot reject H_0: $p = .15$ in favor of H_a: $p \neq .15$. That is, we cannot conclude (at an α of .05) that the percentage of all people who would never shop in the sports discount store again if the store subjected them to a false alarm differs from the general discount store percentage of 15 percent.

The p-value for testing H_0: $p = .15$ versus H_a: $p \neq .15$ is twice the area under the standard normal curve to the right of $|z| = .63$. Because the area under the standard normal curve to the right of $|z| = .63$ is $(1 - .7357) = .2643$ (see Table A.3, page 790), the p-value is $2(.2643) = .5286$. As can be seen on the output below, MINITAB calculates this p-value to be 0.531 (this value is slightly more accurate than our (hand- and table-) calculated result.

```
Test of p = 0.15 vs p not = 0.15

  X     N   Sample p       95% CI        Z-Value   P-Value
 70   500     0.14    (0.1096, 0.1704)    −0.63     0.531
```

This p-value is large and provides little evidence against H_0: $p = .15$ and in favor of H_a: $p \neq .15$. That is, we have little evidence that the percentage of people who would never shop in the sports discount store again if the store subjected them to a false alarm differs from the general discount store percentage of 15 percent.

Exercises for Section 9.4

CONCEPTS

connect

9.34 If we wish to test a hypothesis to provide evidence supporting the claim that fewer than 5 percent of the units produced by a process are defective, formulate the null and alternative hypotheses.

9.35 What condition must be satisfied in order for the methods of this section to be appropriate?

METHODS AND APPLICATIONS

9.36 Suppose we test H_0: $p = .3$ versus H_a: $p \neq .3$ and that a random sample of $n = 100$ gives a sample proportion $\hat{p} = .20$.

a Test H_0 versus H_a at the .01 level of significance by using critical values. What do you conclude?

b Find the p-value for this test.

c Use the p-value to test H_0 versus H_a by setting α equal to .10, .05, .01, and .001. What do you conclude at each value of α?

9.37 THE MARKETING ETHICS CASE: CONFLICT OF INTEREST

Recall that a conflict of interest scenario was presented to a sample of 205 marketing researchers and that 111 of these researchers disapproved of the actions taken.

a Let p be the proportion of all marketing researchers who disapprove of the actions taken in the conflict of interest scenario. Set up the null and alternative hypotheses needed to attempt to provide evidence supporting the claim that a majority (more than 50 percent) of all marketing researchers disapprove of the actions taken.

b Assuming that the sample of 205 marketing researchers has been randomly selected, use critical values and the previously given sample information to test the hypotheses you set up in part *a* at the .10, .05, .01, and .001 levels of significance. How much evidence is there that a majority of all marketing researchers disapprove of the actions taken?

c Suppose a random sample of 1,000 marketing researchers reveals that 540 of the researchers disapprove of the actions taken in the conflict of interest scenario. Use critical values to determine how much evidence there is that a majority of all marketing researchers disapprove of the actions taken.

d Note that in parts *b* and *c* the sample proportion $\hat{p}$ is (essentially) the same. Explain why the results of the hypothesis tests in parts *b* and *c* differ.

9.38 Last year, television station WXYZ's share of the 11 P.M. news audience was approximately equal to, but no greater than, 25 percent. The station's management believes that the current audience share is higher than last year's 25 percent share. In an attempt to substantiate this belief, the station surveyed a random sample of 400 11 P.M. news viewers and found that 146 watched WXYZ.

a Let p be the current proportion of all 11 P.M. news viewers who watch WXYZ. Set up the null and alternative hypotheses needed to attempt to provide evidence supporting the claim that the current audience share for WXYZ is higher than last year's 25 percent share.

b Use critical values and the following MINITAB output to test the hypotheses you set up in part *a* at the .10, .05, .01, and .001 levels of significance. How much evidence is there that the current audience share is higher than last year's 25 percent share?

```
Test of p = 0.25 vs p > 0.25

Sample    X     N    Sample p    Z-Value    P-Value
1       146   400    0.365000       5.31      0.000
```

c Find the *p*-value for the hypothesis test in part *b*. Use the *p*-value to carry out the test by setting α equal to .10, .05, .01, and .001. Interpret your results.

d Do you think that the result of the station's survey has practical importance? Why or why not?

9.39 In the book *Essentials of Marketing Research,* William R. Dillon, Thomas J. Madden, and Neil H. Firtle discuss a marketing research proposal to study day-after recall for a brand of mouthwash. To quote the authors:

> The ad agency has developed a TV ad for the introduction of the mouthwash. The objective of the ad is to create awareness of the brand. The objective of this research is to evaluate the awareness generated by the ad measured by aided- and unaided-recall scores.
>
> A minimum of 200 respondents who claim to have watched the TV show in which the ad was aired the night before will be contacted by telephone in 20 cities.
>
> The study will provide information on the incidence of unaided and aided recall.

Suppose a random sample of 200 respondents shows that 46 of the people interviewed were able to recall the commercial without any prompting (unaided recall).

a In order for the ad to be considered successful, the percentage of unaided recall must be above the category norm for a TV commercial for the product class. If this norm is 18 percent, set up the null and alternative hypotheses needed to attempt to provide evidence that the ad is successful.

b Use the previously given sample information to: **(1)** Compute the *p*-value for the hypothesis test you set up in part *a.* **(2)** Use the *p*-value to carry out the test by setting α equal to .10, .05, .01, and .001. **(3)** How much evidence is there that the TV commercial is successful?

c Do you think the result of the ad agency's survey has practical importance? Explain your opinion.

9.40 An airline's data indicate that 50 percent of people who begin the online process of booking a flight never complete the process and pay for the flight. To reduce this percentage, the airline is considering changing its website so that the entire booking process, including flight and seat selection and payment, can be done on two simple pages rather than the current four pages. A random sample of 300 customers who begin the booking process are exposed to the new system, and 117 of them do not complete the process. **(1)** Formulate the null and alternative hypotheses needed to attempt to provide evidence that the new system has reduced the noncompletion percentage. **(2)** Use critical values and a *p*-value to perform the hypothesis test by setting α equal to .10, .05, .01, and .001.

9.41 Suppose that a national survey finds that 73 percent of restaurant employees say that work stress has a negative impact on their personal lives. A random sample of 200 employees of a large restaurant chain finds that 141 employees say that work stress has a negative impact on their personal lives. **(1)** Formulate the null and alternative hypotheses needed to attempt to provide evidence that the percentage of work-stressed employees for the restaurant chain differs from the national percentage. **(2)** Use critical values and a *p*-value to perform the hypothesis test by setting α equal to .10, .05, .01, and .001.

9.42 The manufacturer of the ColorSmart-5000 television set claims that 95 percent of its sets last at least five years without needing a single repair. In order to test this claim, a consumer group randomly selects 400 consumers who have owned a ColorSmart-5000 television set for five years. Of these 400 consumers, 316 say that their ColorSmart-5000 television sets did not need repair, while 84 say that their ColorSmart-5000 television sets did need at least one repair.

a Letting p be the proportion of ColorSmart-5000 television sets that last five years without a single repair, set up the null and alternative hypotheses that the consumer group should use to attempt to show that the manufacturer's claim is false.

b Use critical values and the previously given sample information to test the hypotheses you set up in part a by setting α equal to .10, .05, .01, and .001. How much evidence is there that the manufacturer's claim is false?

c Do you think the results of the consumer group's survey have practical importance? Explain your opinion.

LO9-6 Calculate Type II error probabilities and the power of a test, and determine sample size (Optional).

9.5 Type II Error Probabilities and Sample Size Determination (Optional)

As we have seen, we often take action (for example, advertise a claim) on the basis of having rejected the null hypothesis. In this case, we know the chances that the action has been taken erroneously because we have prespecified α, the probability of rejecting a true null hypothesis. However, sometimes we must act (for example, decide how many Valentine's Day boxes of chocolates to produce) on the basis of *not* rejecting the null hypothesis. If we must do this, it is best to know the probability of not rejecting a false null hypothesis (a Type II error). If this probability is not small enough, we may change the hypothesis testing procedure. In order to discuss this further, we must first see how to compute the probability of a Type II error.

As an example, the Federal Trade Commission (FTC) often tests claims that companies make about their products. Suppose coffee is being sold in cans that are labeled as containing three pounds, and also suppose that the FTC wishes to determine if the mean amount of coffee μ in all such cans is at least three pounds. To do this, the FTC tests H_0: $\mu \geq 3$ (or $\mu = 3$) versus H_a: $\mu < 3$ by setting $\alpha = .05$. Suppose that a sample of 35 coffee cans yields $\bar{x} = 2.9973$. Assuming that σ is known to equal .0147, we see that because

$$z = \frac{2.9973 - 3}{.0147/\sqrt{35}} = -1.08$$

is not less than $-z_{.05} = -1.645$, we cannot reject H_0: $\mu \geq 3$ by setting $\alpha = .05$. Because we cannot reject H_0, we cannot have committed a Type I error, which is the error of rejecting a true H_0. However, we might have committed a Type II error, which is the error of not rejecting a false H_0. Therefore, before we make a final conclusion about μ, we should calculate the probability of a Type II error.

A Type II error is not rejecting H_0: $\mu \geq 3$ when H_0 is false. Because any value of μ that is less than 3 makes H_0 false, there is a different Type II error (and, therefore, a different Type II error probability) associated with each value of μ that is less than 3. In order to demonstrate how to calculate these probabilities, we will calculate the probability of not rejecting H_0: $\mu \geq 3$ when in fact μ equals 2.995. This is the probability of failing to detect an average underfill of .005 pound. For a fixed sample size (for example, $n = 35$ coffee can fills), the value of β, the probability of a Type II error, depends upon how we set α, the probability of a Type I error. Because we have set $\alpha = .05$, we reject H_0 if

$$\frac{\bar{x} - 3}{\sigma/\sqrt{n}} < -z_{.05}$$

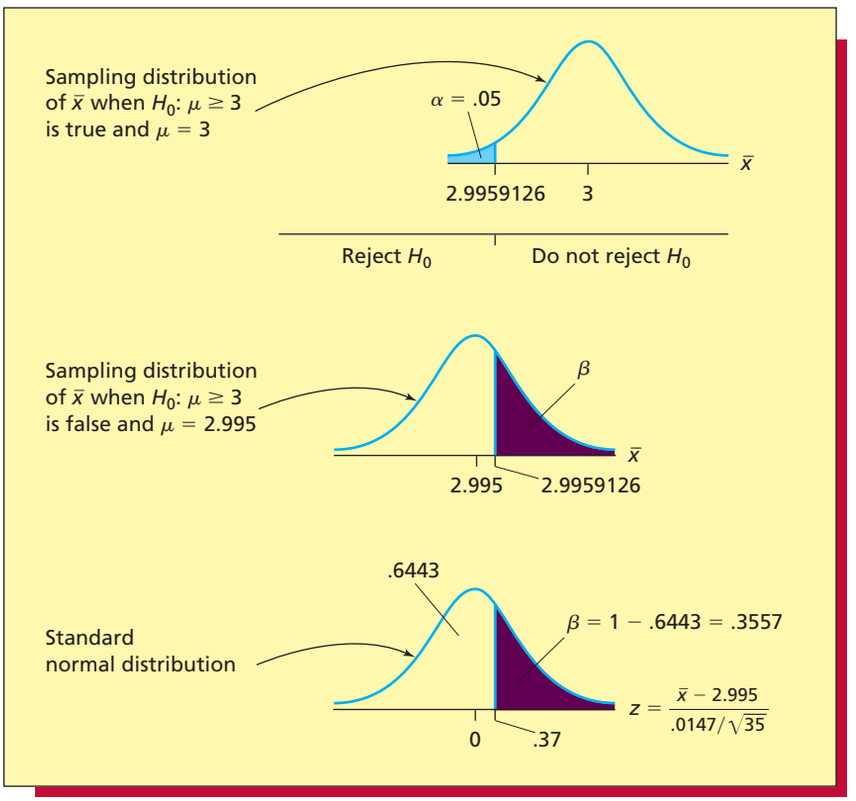
or, equivalently, if

$$\bar{x} < 3 - z_{.05}\frac{\sigma}{\sqrt{n}} = 3 - 1.645\frac{.0147}{\sqrt{35}} = 2.9959126$$

Therefore, we do not reject H_0 if $\bar{x} \geq 2.9959126$. It follows that β, the probability of not rejecting H_0: $\mu \geq 3$ when μ equals 2.995, is

$$\begin{aligned}\beta &= P(\bar{x} \geq 2.9959126 \text{ when } \mu = 2.995)\\ &= P\left(z \geq \frac{2.9959126 - 2.995}{.0147/\sqrt{35}}\right)\\ &= P(z \geq .37) = 1 - .6443 = .3557\end{aligned}$$

FIGURE 9.7 Calculating β When μ Equals 2.995



This calculation is illustrated in Figure 9.7. Similarly, it follows that β, the probability of not rejecting H_0: $\mu \geq 3$ when μ equals 2.99, is

$$\begin{aligned}\beta &= P(\bar{x} \geq 2.9959126 \text{ when } \mu = 2.99) \\ &= P\left(z \geq \frac{2.9959126 - 2.99}{.0147/\sqrt{35}}\right) \\ &= P(z \geq 2.38) = 1 - .9913 = .0087\end{aligned}$$

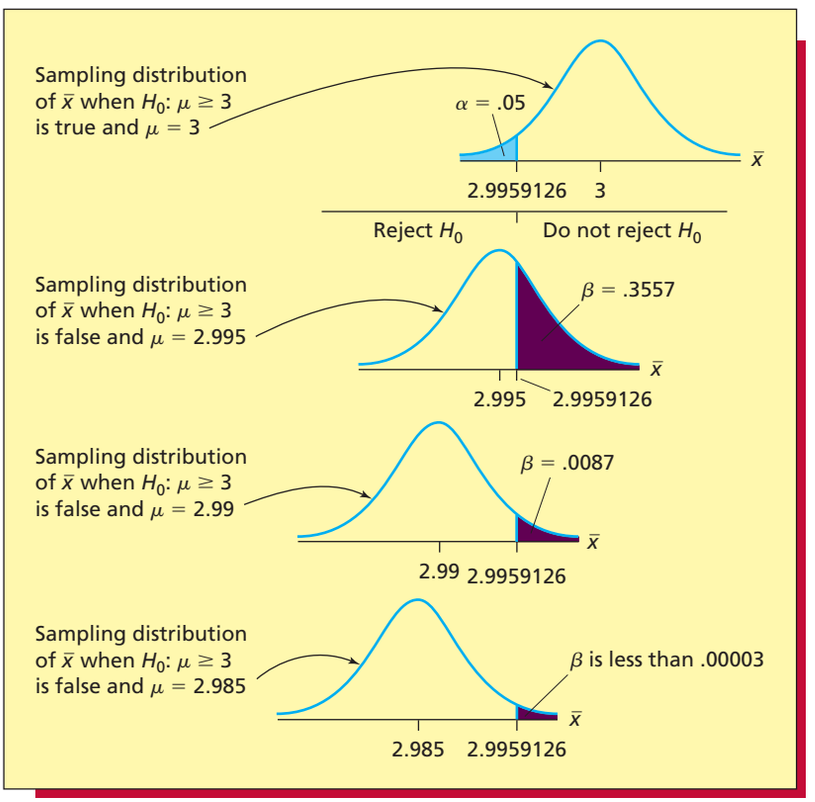
It also follows that β, the probability of not rejecting H_0: $\mu \geq 3$ when μ equals 2.985, is

$$\begin{aligned}\beta &= P(\bar{x} \geq 2.9959126 \text{ when } \mu = 2.985) \\ &= P\left(z \geq \frac{2.9959126 - 2.985}{.0147/\sqrt{35}}\right) \\ &= P(z \geq 4.39)\end{aligned}$$

This probability is less than .00003 (because z is greater than 3.99).

In Figure 9.8 we illustrate the values of β that we have calculated. Notice that the closer an alternative value of μ is to 3 (the value specified by H_0: $\mu = 3$), the larger is the associated value of β. Although alternative values of μ that are closer to 3 have larger associated probabilities of Type II errors, these values of μ have associated Type II errors with less serious consequences. For example, we are more likely not to reject H_0: $\mu = 3$ when $\mu = 2.995$ ($\beta = .3557$) than we are not to reject H_0: $\mu = 3$ when $\mu = 2.99$ ($\beta = .0087$). However, not rejecting H_0: $\mu = 3$ when $\mu = 2.995$, which means that we are failing to detect an average underfill of .005 pound, is less serious than not rejecting H_0: $\mu = 3$ when $\mu = 2.99$, which means that we are failing to detect a larger average underfill of .01 pound. In order to decide whether a particular hypothesis test adequately controls the probability of a Type II error, we must determine which Type II errors are serious, and then we must decide whether the probabilities of these errors are small enough. For

FIGURE 9.8 How β Changes as the Alternative Value of μ Changes



example, suppose that the FTC and the coffee producer agree that failing to reject H_0: $\mu = 3$ when μ equals 2.99 is a serious error, but that failing to reject H_0: $\mu = 3$ when μ equals 2.995 is not a particularly serious error. Then, because the probability of not rejecting H_0: $\mu = 3$ when μ equals 2.99 is .0087, which is quite small, we might decide that the hypothesis test adequately controls the probability of a Type II error. To understand the implication of this, recall that the sample of 35 coffee cans, which has $\bar{x} = 2.9973$, does not provide enough evidence to reject H_0: $\mu \geq 3$ by setting $\alpha = .05$. We have just shown that the probability that we have failed to detect a serious underfill is quite small (.0087), so the FTC might decide that no action should be taken against the coffee producer. Of course, this decision should also be based on the variability of the fills of the individual cans. Because $\bar{x} = 2.9973$ and $\sigma = .0147$, we estimate that 99.73 percent of all individual coffee can fills are contained in the interval $[\bar{x} \pm 3\sigma] = [2.9973 \pm 3(.0147)] = [2.9532, 3.0414]$. If the FTC believes it is reasonable to accept fills as low as (but no lower than) 2.9532 pounds, this evidence also suggests that no action against the coffee producer is needed.

Suppose, instead, that the FTC and the coffee producer had agreed that failing to reject H_0: $\mu \geq 3$ when μ equals 2.995 is a serious mistake. The probability of this Type II error is .3557, which is large. Therefore, we might conclude that the hypothesis test is not adequately controlling the probability of a serious Type II error. In this case, we have two possible courses of action. First, we have previously said that, for a fixed sample size, the lower we set α, the higher is β, and the higher we set α, the lower is β. Therefore, if we keep the sample size fixed at $n = 35$ coffee cans, we can reduce β by increasing α. To demonstrate this, suppose we increase α to .10. In this case we reject H_0 if

$$\frac{\bar{x} - 3}{\sigma/\sqrt{n}} < -z_{.10}$$

or, equivalently, if

$$\bar{x} < 3 - z_{.10}\frac{\sigma}{\sqrt{n}} = 3 - 1.282\frac{.0147}{\sqrt{35}} = 2.9968145$$

Therefore, we do not reject H_0 if $\bar{x} \geq 2.9968145$. It follows that β, the probability of not rejecting H_0: $\mu \geq 3$ when μ equals 2.995, is

$$\beta = P(\bar{x} \geq 2.9968145 \text{ when } \mu = 2.995)$$

$$= P\left(z \geq \frac{2.9968145 - 2.995}{.0147/\sqrt{35}}\right)$$

$$= P(z \geq .73) = 1 - .7673 = .2327$$

We thus see that increasing α from .05 to .10 reduces β from .3557 to .2327. However, β is still too large, and, besides, we might not be comfortable making α larger than .05. Therefore, if we wish to decrease β and maintain α at .05, we must increase the sample size. We will soon present a formula we can use to find the sample size needed to make both α and β as small as we wish.

Once we have computed β, we can calculate what we call the *power* of the test.

The **power** of a statistical test is the probability of rejecting the null hypothesis when it is false.

Just as β depends upon the alternative value of μ, so does the power of a test. In general, **the power associated with a particular alternative value of μ equals $1 - \beta$,** where β is the probability of a Type II error associated with the same alternative value of μ. For example, we have seen that, when we set $\alpha = .05$, the probability of not rejecting H_0: $\mu \geq 3$ when μ equals 2.99 is .0087. Therefore, the power of the test associated with the alternative value 2.99 (that is, the probability of rejecting H_0: $\mu \geq 3$ when μ equals 2.99) is $1 - .0087 = .9913$.

Thus far we have demonstrated how to calculate β when testing a *less than* alternative hypothesis. In the following box we present (without proof) a method for calculating the probability of a Type II error when testing a *less than,* a *greater than,* or a *not equal to* alternative hypothesis:

Calculating the Probability of a Type II Error

Assume that the sampled population is normally distributed, or that a large sample will be taken. Consider testing H_0: $\mu = \mu_0$ versus one of H_a: $\mu > \mu_0$, H_a: $\mu < \mu_0$, or H_a: $\mu \neq \mu_0$. Then, if we set the probability of a Type I error equal to α and randomly select a sample of size n, the probability, β, of a Type II error corresponding to the alternative value μ_a of μ is (exactly or approximately) equal to the area under the standard normal curve to the left of

$$z^* - \frac{|\mu_0 - \mu_a|}{\sigma/\sqrt{n}}$$

Here z^* equals z_α if the alternative hypothesis is one-sided ($\mu > \mu_0$ or $\mu < \mu_0$), in which case the method for calculating β is exact. Furthermore, z^* equals $z_{\alpha/2}$ if the alternative hypothesis is two-sided ($\mu \neq \mu_0$), in which case the method for calculating β is approximate.

EXAMPLE 9.9 The Valentine's Day Chocolate Case: Production Planning

In the Valentine's Day chocolate case we are testing H_0: $\mu = 330$ versus H_a: $\mu \neq 330$ by setting $\alpha = .05$. We have seen that the mean of the reported order quantities of a random sample of $n = 100$ large retail stores is $\bar{x} = 326$. Assuming that σ equals 40, it follows that because

$$z = \frac{326 - 330}{40/\sqrt{100}} = -1$$

is between $-z_{.025} = -1.96$ and $z_{.025} = 1.96$, we cannot reject H_0: $\mu = 330$ by setting $\alpha = .05$. Because we cannot reject H_0, we might have committed a Type II error. Suppose that the candy company decides that failing to reject H_0: $\mu = 330$ when μ differs from 330 by as many as 15 valentine boxes (that is, when μ is 315 or 345) is a serious Type II error. Because we have set α

equal to .05, β for the alternative value $\mu_a = 315$ (that is, the probability of not rejecting H_0: $\mu = 330$ when μ equals 315) is the area under the standard normal curve to the left of

$$z^* - \frac{|\mu_0 - \mu_a|}{\sigma/\sqrt{n}} = z_{.025} - \frac{|\mu_0 - \mu_a|}{\sigma/\sqrt{n}}$$
$$= 1.96 - \frac{|330 - 315|}{40/\sqrt{100}}$$
$$= -1.79$$

Here $z^* = z_{\alpha/2} = z_{.05/2} = z_{.025}$ because the alternative hypothesis ($\mu \neq 330$) is two-sided. The area under the standard normal curve to the left of -1.79 is .0367. Therefore, β for the alternative value $\mu_a = 315$ is .0367. Similarly, it can be verified that β for the alternative value $\mu_a = 345$ is .0367. It follows, because we cannot reject H_0: $\mu = 330$ by setting $\alpha = .05$, and because we have just shown that there is a reasonably small (.0367) probability that we have failed to detect a serious (that is, a 15 valentine box) deviation of μ from 330, that it is reasonable for the candy company to base this year's production of valentine boxes on the projected mean order quantity of 330 boxes per large retail store.

In the following box we present (without proof) a formula that tells us the sample size needed to make both the probability of a Type I error and the probability of a Type II error as small as we wish:

Calculating the Sample Size Needed to Achieve Specified Values of α and β

Assume that the sampled population is normally distributed, or that a large sample will be taken. Consider testing H_0: $\mu = \mu_0$ versus one of H_a: $\mu > \mu_0$, H_a: $\mu < \mu_0$, or H_a: $\mu \neq \mu_0$. Then, in order to make the probability of a Type I error equal to α and the probability of a Type II error corresponding to the alternative value μ_a of μ equal to β, we should take a sample of size

$$n = \frac{(z^* + z_\beta)^2\sigma^2}{(\mu_0 - \mu_a)^2}$$

Here z^* equals z_α if the alternative hypothesis is one-sided ($\mu > \mu_0$ or $\mu < \mu_0$), and z^* equals $z_{\alpha/2}$ if the alternative hypothesis is two-sided ($\mu \neq \mu_0$). Also, z_β is the point on the scale of the standard normal curve that gives a right-hand tail area equal to β.

EXAMPLE 9.10 Finding A Sample Size

Again consider the coffee fill example and suppose we wish to test H_0: $\mu \geq 3$ (or $\mu = 3$) versus H_a: $\mu < 3$. If we wish α to be .05 and β for the alternative value $\mu_a = 2.995$ of μ to be .05, we should take a sample of size

$$n = \frac{(z^* + z_\beta)^2\sigma^2}{(\mu_0 - \mu_a)^2} = \frac{(z_\alpha + z_\beta)^2\sigma^2}{(\mu_0 - \mu_a)^2}$$
$$= \frac{(z_{.05} + z_{.05})^2\sigma^2}{(\mu_0 - \mu_a)^2}$$
$$= \frac{(1.645 + 1.645)^2(.0147)^2}{(3 - 2.995)^2}$$
$$= 93.5592 = 94 \text{ (rounding up)}$$

Here, $z^* = z_\alpha = z_{.05} = 1.645$ because the alternative hypothesis ($\mu < 3$) is one-sided, and $z_\beta = z_{.05} = 1.645$.

Although we have set both α and β equal to the same value in the coffee fill situation, it is not necessary for α and β to be equal. As an example, again consider the Valentine's Day chocolate

case, in which we are testing H_0: $\mu = 330$ versus H_a: $\mu \neq 330$. Suppose that the candy company decides that failing to reject H_0: $\mu = 330$ when μ differs from 330 by as many as 15 valentine boxes (that is, when μ is 315 or 345) is a serious Type II error. Furthermore, suppose that it is also decided that this Type II error is more serious than a Type I error. Therefore, α will be set equal to .05 and β for the alternative value $\mu_a = 315$ (or $\mu_a = 345$) of μ will be set equal to .01. It follows that the candy company should take a sample of size

$$n = \frac{(z^* + z_\beta)^2\sigma^2}{(\mu_0 - \mu_a)^2} = \frac{(z_{\alpha/2} + z_\beta)^2\sigma^2}{(\mu_0 - \mu_a)^2}$$

$$= \frac{(z_{.025} + z_{.01})^2\sigma^2}{(\mu_0 - \mu_a)^2}$$

$$= \frac{(1.96 + 2.326)^2(40)^2}{(330 - 315)^2}$$

$$= 130.62 = 131 \text{ (rounding up)}$$

Here, $z^* = z_{\alpha/2} = z_{.05/2} = z_{.025} = 1.96$ because the alternative hypothesis ($\mu \neq 330$) is two-sided, and $z_\beta = z_{.01} = 2.326$ (see the bottom row of the t table on page 311).

To conclude this section, we point out that the methods we have presented for calculating the probability of a Type II error and determining sample size can be extended to other hypothesis tests that utilize the normal distribution. We will not, however, present the extensions in this book.

Exercises for Section 9.5

CONCEPTS

connect

9.43 We usually take action on the basis of having rejected the null hypothesis. When we do this, we know the chances that the action has been taken erroneously because we have prespecified α, the probability of rejecting a true null hypothesis. Here, it is obviously important to know (prespecify) α, the probability of a Type I error. When is it important to know the probability of a Type II error? Explain why.

9.44 Explain why we are able to compute many different values of β, the probability of a Type II error, for a single hypothesis test.

9.45 Explain what is meant by

a A serious Type II error.

b The power of a statistical test.

9.46 In general, do we want the power corresponding to a serious Type II error to be near 0 or near 1? Explain.

METHODS AND APPLICATIONS

9.47 Again consider the Consolidated Power waste water situation. Remember that the power plant will be shut down and corrective action will be taken on the cooling system if the null hypothesis H_0: $\mu \leq 60$ is rejected in favor of H_a: $\mu > 60$. In this exercise we calculate probabilities of various Type II errors in the context of this situation.

a Recall that Consolidated Power's hypothesis test is based on a sample of $n = 100$ temperature readings and assume that σ equals 2. If the power company sets $\alpha = .025$, calculate the probability of a Type II error for each of the following alternative values of μ: 60.1, 60.2, 60.3, 60.4, 60.5, 60.6, 60.7, 60.8, 60.9, 61.

b If we want the probability of making a Type II error when μ equals 60.5 to be very small, is Consolidated Power's hypothesis test adequate? Explain why or why not. If not, and if we wish to maintain the value of α at .025, what must be done?

c The **power curve** for a statistical test is a plot of the power $= 1 - \beta$ on the vertical axis versus values of μ that make the null hypothesis false on the horizontal axis. Plot the power curve for Consolidated Power's test of H_0: $\mu \leq 60$ versus H_a: $\mu > 60$ by plotting power $= 1 - \beta$ for each of the alternative values of μ in part a. What happens to the power of the test as the alternative value of μ moves away from 60?

9.48 Again consider the automobile parts supplier situation. Remember that a problem-solving team will be assigned to rectify the process producing the cylindrical engine parts if the null hypothesis H_0: $\mu = 3$ is rejected in favor of H_a: $\mu \neq 3$. In this exercise we calculate probabilities of various Type II errors in the context of this situation.

a Suppose that the parts supplier's hypothesis test is based on a sample of $n = 100$ diameters and that σ equals .023. If the parts supplier sets $\alpha = .05$, calculate the probability of a Type II error for each of the following alternative values of μ: 2.990, 2.995, 3.005, 3.010.

b If we want both the probabilities of making a Type II error when μ equals 2.995 and when μ equals 3.005 to be very small, is the parts supplier's hypothesis test adequate? Explain why or why not. If not, and if we wish to maintain the value of α at .05, what must be done?

c Plot the power of the test versus the alternative values of μ in part *a*. What happens to the power of the test as the alternative value of μ moves away from 3?

9.49 In each of the following situations, find the necessary sample size.

a In the Consolidated Power hypothesis test of H_0: $\mu \leq 60$ versus H_a: $\mu > 60$ (as discussed in Exercise 9.47) find the sample size needed to make the probability of a Type I error equal to .025 and the probability of a Type II error corresponding to the alternative value $\mu_a = 60.5$ equal to .025. Here, assume σ equals 2.

b In the automobile parts supplier's hypothesis test of H_0: $\mu = 3$ versus H_a: $\mu \neq 3$ (as discussed in Exercise 9.48) find the sample size needed to make the probability of a Type I error equal to .05 and the probability of a Type II error corresponding to the alternative value $\mu_a = 3.005$ equal to .05. Here, assume σ equals .023.

Chapter Summary

We began this chapter by learning about the two hypotheses that make up the structure of a hypothesis test. The **null hypothesis** is the statement being tested. The null hypothesis is often a statement of "no difference" or "no effect," and it is not rejected unless there is convincing sample evidence that it is false. The **alternative,** or, **research, hypothesis** is a statement that is accepted only if there is convincing sample evidence that it is true and that the null hypothesis is false. In some situations, the alternative hypothesis is a statement for which we wish to find supportive evidence. We also learned that two types of errors can be made in a hypothesis test. A **Type I error** occurs when we reject a true null hypothesis, and a **Type II error** occurs when we do not reject a false null hypothesis.

We studied two commonly used ways to conduct a hypothesis test. The first involves comparing the value of a test statistic with what is called a **critical value,** and the second employs what is called a ***p*-value.** The *p*-value measures the weight of evidence against the null hypothesis. The smaller the *p*-value, the more we doubt the null hypothesis.

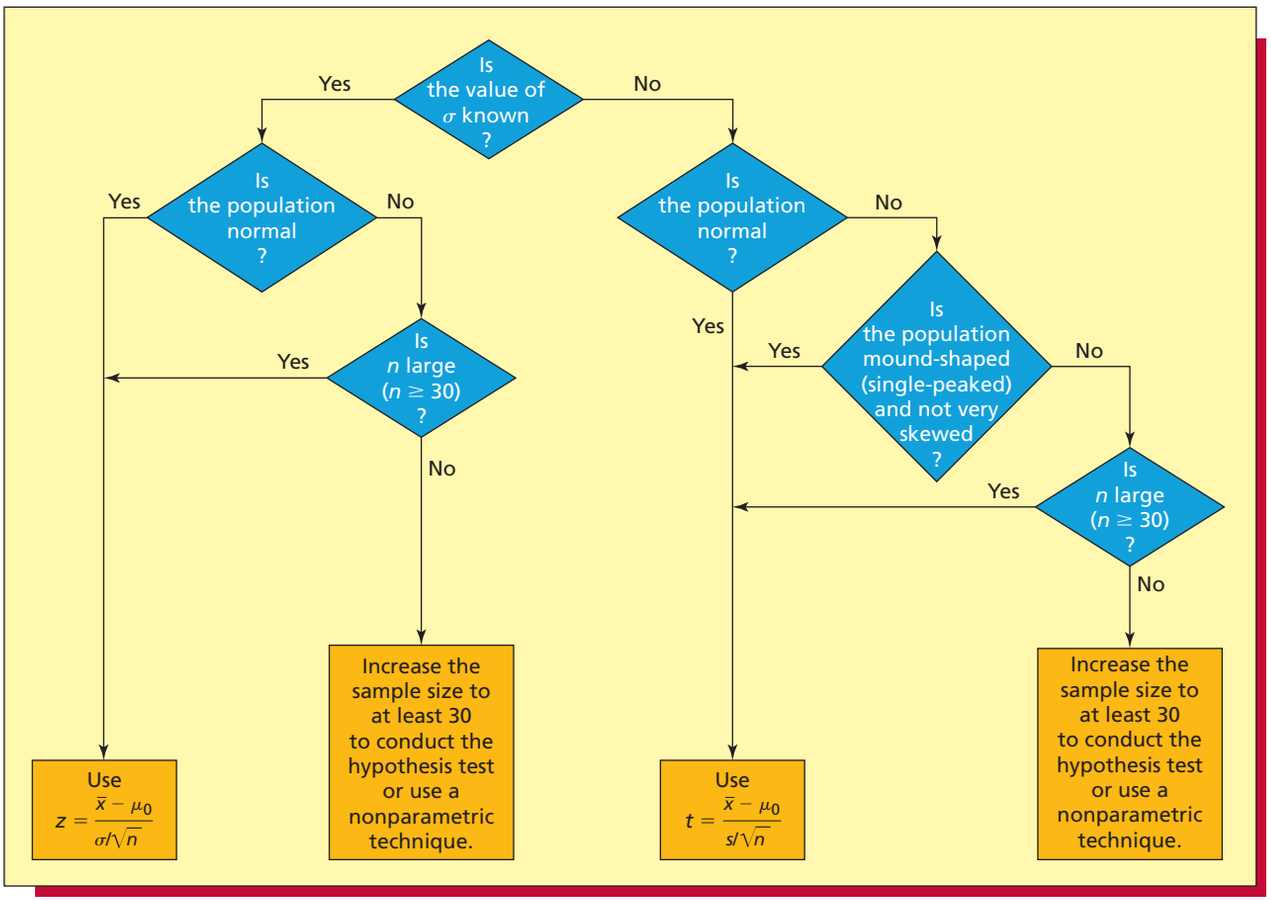
The specific hypothesis tests we covered in this chapter all dealt with a hypothesis about one population parameter. First, we studied a test about a **population mean** that is based on the assumption that the population standard deviation σ **is known.** This test employs the **normal distribution.** Second, we studied a test about a population mean that assumes that σ **is unknown.** We learned that this test is based on the ***t* distribution.** Figure 9.9 presents a flowchart summarizing how to select an appropriate test statistic to test a hypothesis about a population mean. Then we presented a test about a **population proportion** that is based on the **normal distribution.** We concluded this chapter by studying Type II error probabilities, and we showed how we can find the sample size needed to make both the probability of a Type I error and the probability of a serious Type II error as small as we wish.

Glossary of Terms

alternative (research) hypothesis: A statement that will be accepted only if there is convincing sample evidence that it is true. Sometimes it is a statement for which we wish to find supportive evidence. (page 341)

critical value: The value of the test statistic is compared with a critical value in order to decide whether the null hypothesis can be rejected. (pages 347, 350, 352)

FIGURE 9.9 Selecting an Appropriate Test Statistic to Test a Hypothesis about a Population Mean



greater than alternative: An alternative hypothesis that is stated as a *greater than* ($>$) inequality. (page 343)
less than alternative: An alternative hypothesis that is stated as a *less than* ($<$) inequality. (page 343)
not equal to alternative: An alternative hypothesis that is stated as a *not equal to* ($\neq$) inequality. (page 343)
null hypothesis: The statement being tested in a hypothesis test. It is often a statement of "no difference" or "no effect. (page 341)
***p*-value (probability value):** The probability, computed assuming that the null hypothesis H_0 is true, of observing a value of the test statistic that is at least as contradictory to H_0 and supportive of H_a as the value actually computed from the sample data. The *p*-value measures how much doubt is cast on the null hypothesis by the sample data. The smaller the *p*-value, the more we doubt the null hypothesis. (pages 349, 351, 353)
test statistic: A statistic computed from sample data in a hypothesis test. It is either compared with a critical value or used to compute a *p*-value. (page 347)
two-sided alternative hypothesis: An alternative hypothesis that is stated as a *not equal to* ($\neq$) inequality. (page 343)
Type I error: Rejecting a true null hypothesis. (page 344)
Type II error: Failing to reject a false null hypothesis. (page 344)

Important Formulas and Tests

Hypothesis Testing steps: page 350

A hypothesis test (z test) about a population mean (σ known): page 353

A hypothesis test (t test) about a population mean (σ unknown): page 357

A large sample hypothesis test about a population proportion: page 362

Calculating the probability of a Type II error: page 369

Sample size determination to achieve specified values of α and β: page 370

Supplementary Exercises

connect

9.50 The auditor for a large corporation routinely monitors cash disbursements. As part of this process, the auditor examines check request forms to determine whether they have been properly approved. Improper approval can occur in several ways. For instance, the check may have no approval, the check request might be missing, the approval might be written by an unauthorized person, or the dollar limit of the authorizing person might be exceeded.

a Last year the corporation experienced a 5 percent improper check request approval rate. Because this was considered unacceptable, efforts were made to reduce the rate of improper approvals. Letting p be the proportion of all checks that are now improperly approved, set up the null and alternative hypotheses needed to attempt to demonstrate that the current rate of improper approvals is lower than last year's rate of 5 percent.

b Suppose that the auditor selects a random sample of 625 checks that have been approved in the last month. The auditor finds that 18 of these 625 checks have been improperly approved. Use critical values and this sample information to test the hypotheses you set up in part *a* at the .10, .05, .01, and .001 levels of significance. How much evidence is there that the rate of improper approvals has been reduced below last year's 5 percent rate?

c Find the p-value for the test of part *b*. Use the p-value to carry out the test by setting α equal to .10, .05, .01, and .001. Interpret your results.

d Suppose the corporation incurs a $10 cost to detect and correct an improperly approved check. If the corporation disburses at least 2 million checks per year, does the observed reduction of the rate of improper approvals seem to have practical importance? Explain your opinion.

9.51 THE CIGARETTE ADVERTISEMENT CASE ModelAge

Recall that the cigarette industry requires that models in cigarette ads must appear to be at least 25 years old. Also recall that a sample of 50 people is randomly selected at a shopping mall. Each person in the sample is shown a "typical cigarette ad" and is asked to estimate the age of the model in the ad.

a Let μ be the mean perceived age estimate for all viewers of the ad, and suppose we consider the industry requirement to be met if μ is at least 25. Set up the null and alternative hypotheses needed to attempt to show that the industry requirement is not being met.

b Suppose that a random sample of 50 perceived age estimates gives a mean of $\bar{x} = 23.663$ years and a standard deviation of $s = 3.596$ years. Use these sample data and critical values to test the hypotheses of part *a* at the .10, .05, .01, and .001 levels of significance.

c How much evidence do we have that the industry requirement is not being met?

d Do you think that this result has practical importance? Explain your opinion.

9.52 THE CIGARETTE ADVERTISEMENT CASE ModelAge

Consider the cigarette ad situation discussed in Exercise 9.51. Using the sample information given in that exercise, the p-value for testing H_0 versus H_a can be calculated to be .0057.

a Determine whether H_0 would be rejected at each of $\alpha = .10$, $\alpha = .05$, $\alpha = .01$, and $\alpha = .001$.

b Describe how much evidence we have that the industry requirement is not being met.

9.53 In an article in the *Journal of Retailing,* Kumar, Kerwin, and Pereira study factors affecting merger and acquisition activity in retailing. As part of the study, the authors compare the characteristics of "target firms" (firms targeted for acquisition) and "bidder firms" (firms attempting to make acquisitions). Among the variables studied in the comparison were earnings per share, debt-to-equity ratio, growth rate of sales, market share, and extent of diversification.

a Let μ be the mean growth rate of sales for all target firms (firms that have been targeted for acquisition in the last five years and that have not bid on other firms), and assume growth rates are approximately normally distributed. Furthermore, suppose a random sample of 25 target firms yields a sample mean sales growth rate of $\bar{x} = 0.16$ with a standard deviation of $s = 0.12$. Use critical values and this sample information to test H_0: $\mu \leq .10$ versus H_a: $\mu > .10$ by setting α equal to .10, .05, .01, and .001. How much evidence is there that the mean growth rate of sales for target firms exceeds .10 (that is, exceeds 10 percent)?

b Now let μ be the mean growth rate of sales for all firms that are bidders (firms that have bid to acquire at least one other firm in the last five years), and again assume growth rates are approximately normally distributed. Furthermore, suppose a random sample of 25 bidders yields a sample mean sales growth rate of $\bar{x} = 0.12$ with a standard deviation of $s = 0.09$. Use critical values and this sample information to test H_0: $\mu \leq .10$ versus H_a: $\mu > .10$ by setting α equal to .10, .05, .01, and .001. How much evidence is there that the mean growth rate of sales for bidders exceeds .10 (that is, exceeds 10 percent)?

9.54 A consumer electronics firm has developed a new type of remote control button that is designed to operate longer before becoming intermittent. A random sample of 35 of the new buttons is selected and each is tested in continuous operation until becoming intermittent. The resulting lifetimes are found to have a sample mean of $\bar{x} = 1{,}241.2$ hours and a sample standard deviation of $s = 110.8$.

- **a** Independent tests reveal that the mean lifetime (in continuous operation) of the best remote control button on the market is 1,200 hours. Letting μ be the mean lifetime of the population of all new remote control buttons that will or could potentially be produced, set up the null and alternative hypotheses needed to attempt to provide evidence that the new button's mean lifetime exceeds the mean lifetime of the best remote button currently on the market.
- **b** Using the previously given sample results, use critical values to test the hypotheses you set up in part *a* by setting α equal to .10, .05, .01, and .001. What do you conclude for each value of α?
- **c** Suppose that $\bar{x} = 1{,}241.2$ and $s = 110.8$ had been obtained by testing a sample of 100 buttons. Use critical values to test the hypotheses you set up in part *a* by setting α equal to .10, .05, .01, and .001. Which sample (the sample of 35 or the sample of 100) gives a more statistically significant result? That is, which sample provides stronger evidence that H_a is true?
- **d** If we define practical importance to mean that μ exceeds 1,200 by an amount that would be clearly noticeable to most consumers, do you think that the result has practical importance? Explain why the samples of 35 and 100 both indicate the same degree of practical importance.
- **e** Suppose that further research and development effort improves the new remote control button and that a random sample of 35 buttons gives $\bar{x} = 1{,}524.6$ hours and $s = 102.8$ hours. Test your hypotheses of part *a* by setting α equal to .10, .05, .01, and .001.
  - **(1)** Do we have a highly statistically significant result? Explain.
  - **(2)** Do you think we have a practically important result? Explain.

9.55 Again consider the remote control button lifetime situation discussed in Exercise 9.54. Using the sample information given in the introduction to Exercise 9.54, the *p*-value for testing H_0 versus H_a can be calculated to be .0174.

- **a** Determine whether H_0 would be rejected at each of $\alpha = .10$, $\alpha = .05$, $\alpha = .01$, and $\alpha = .001$.
- **b** Describe how much evidence we have that the new button's mean lifetime exceeds the mean lifetime of the best remote button currently on the market.

9.56 Several industries located along the Ohio River discharge a toxic substance called carbon tetrachloride into the river. The state Environmental Protection Agency monitors the amount of carbon tetrachloride pollution in the river. Specifically, the agency requires that the carbon tetrachloride contamination must average no more than 10 parts per million. In order to monitor the carbon tetrachloride contamination in the river, the agency takes a daily sample of 100 pollution readings at a specified location. If the mean carbon tetrachloride reading for this sample casts substantial doubt on the hypothesis that the average amount of carbon tetrachloride contamination in the river is at most 10 parts per million, the agency must issue a shutdown order. In the event of such a shutdown order, industrial plants along the river must be closed until the carbon tetrachloride contamination is reduced to a more acceptable level. Assume that the state Environmental Protection Agency decides to issue a shutdown order if a sample of 100 pollution readings implies that H_0: $\mu \leq 10$ can be rejected in favor of H_a: $\mu > 10$ by setting $\alpha = .01$. If σ equals 2, calculate the probability of a Type II error for each of the following alternative values of μ: 10.1, 10.2, 10.3, 10.4, 10.5, 10.6, 10.7, 10.8, 10.9, and 11.0.

9.57 Suppose that a spokesman for U.S. travel agencies claims that more than 65 percent of all U.S. passengers who booked cruises last year used travel agencies to book their cruises. In the October 2, 2011, issue of the *Hamilton Journal News,* an actual survey of U.S. passengers who booked cruises last year found that 68 percent of those sampled used travel agencies to book their cruises. Assuming that the survey was based on 1,500 randomly selected U.S. passengers who booked cruises last year, calculate a *p*-value that shows the evidence supporting the spokesman's claim. Interpret your results.

9.58 Assume that an insurance survey is based on 1,000 randomly selected U.S. households in a particular income class and finds that 640 of these households bought life insurance last year.

- **a** If p denotes the proportion of all U.S. households in the income class that bought life insurance last year, set up the null and alternative hypotheses needed to attempt to justify the claim that more than 60 percent of U.S. households in the income class bought life insurance last year.

b Test the hypotheses you set up in part *a* by setting $\alpha = .10, .05, .01$, and $.001$. How much evidence is there that more than 60 percent of U.S. households in the income class bought life insurance last year?

9.59 How safe are child car seats? *Consumer Reports* (May 2005) tested the safety of child car seats in 30 mph crashes. They found "slim safety margins" for some child car seats. Suppose that *Consumer Reports* simulates the safety of the market-leading child car seat. Their test consists of placing the maximum claimed weight in the car seat and simulating crashes at higher and higher speeds until a problem occurs. The following data identify the speed (in miles per hour) at which a problem with the car seat (such as the strap breaking, seat shell cracked, strap adjuster broke, detached from base, etc.) first appeared: 31.0, 29.4, 30.4, 28.9, 29.7, 30.1, 32.3, 31.7, 35.4, 29.1, 31.2, 30.2. Let μ denote the true mean speed at which a problem with the car seat first appears. The following MINITAB output gives the results of using the sample data to test H_0: $\mu = 30$ versus H_a: $\mu > 30$. DS CarSeat

```
Test of mu = 30 vs > 30

Variable   N     Mean   StDev  SE Mean     T      P
mph       12  30.7833  1.7862   0.5156  1.52  0.078
```

How much evidence is there that μ exceeds 30 mph?

9.60 *Consumer Reports* (January 2005) indicates that profit margins on extended warranties are much greater than on the purchase of most products.[4] In this exercise we consider a major electronics retailer that wishes to increase the proportion of customers who buy extended warranties on digital cameras. Historically, 20 percent of digital camera customers have purchased the retailer's extended warranty. To increase this percentage, the retailer has decided to offer a new warranty that is less expensive and more comprehensive. Suppose that three months after starting to offer the new warranty, a random sample of 500 customer sales invoices shows that 152 out of 500 digital camera customers purchased the new warranty. **(1)** Letting p denote the proportion of all digital camera customers who have purchased the new warranty, calculate the p-value for testing H_0: $p = .20$ versus H_a: $p > .20$. **(2)** How much evidence is there that p exceeds .20? **(3)** Does the difference between $\hat{p}$ and .2 seem to be practically important? Explain your opinion.

9.61 *Fortune* magazine has periodically reported on the rise of fees and expenses charged by stock funds.

a Suppose that 10 years ago the average annual expense for stock funds was 1.19 percent. Let μ be the current mean annual expense for all stock funds, and assume that stock fund annual expenses are approximately normally distributed. If a random sample of 12 stock funds gives a sample mean annual expense of $\bar{x} = 1.63\%$ with a standard deviation of $s = .31\%$, use critical values and this sample information to test H_0: $\mu \leq 1.19\%$ versus H_a: $\mu > 1.19\%$ by setting α equal to .10, .05, .01, and .001. How much evidence is there that the current mean annual expense for stock funds exceeds the average of 10 years ago?

b Do you think that the result in part *a* has practical importance? Explain your opinion.

9.62 Internet Exercise

Are American consumers comfortable using their credit cards to make purchases over the Internet? Suppose that a noted authority suggests that credit cards will be firmly established on the Internet once the "80 percent barrier" is broken; that is, as soon as more than 80 percent of those who make purchases over the Internet are willing to use a credit card to pay for their transactions. A recent Gallup Poll (story, survey results, and analysis can be found at http://www.gallup.com/poll/releases/pr000223.asp) found that, out of $n = 302$ Internet purchasers surveyed, 267 have paid for Internet purchases using a credit card. Based on the results of the Gallup survey, is there sufficient evidence to conclude that the proportion of Internet purchasers willing to use a credit card now exceeds 0.80? Set up the appropriate null and alternative hypotheses, test at the 0.05 and 0.01 levels of significance, and calculate a p-value for your test.

Go to the Gallup Organization website (http://www.gallup.com). Select an interesting current poll and prepare a brief written summary of the poll or some aspect thereof. Include a statistical test for the significance of a proportion (you may have to make up your own value for the hypothesized proportion p_0) as part of your report. For example, you might select a political poll and test whether a particular candidate is preferred by a majority of voters ($p > 0.50$).

[4] *Consumer Reports,* January 2005, page 51.

Appendix 9.1 ■ One-Sample Hypothesis Testing Using Excel

Hypothesis test for a population mean in Exercise 9.33 on page 361 (data file: CreditCd.xlsx):

The Data Analysis ToolPak in Excel does not explicitly provide for one-sample tests of hypotheses. A one-sample test can be conducted using the Descriptive Statistics component of the Analysis ToolPak and a few additional computations using Excel.

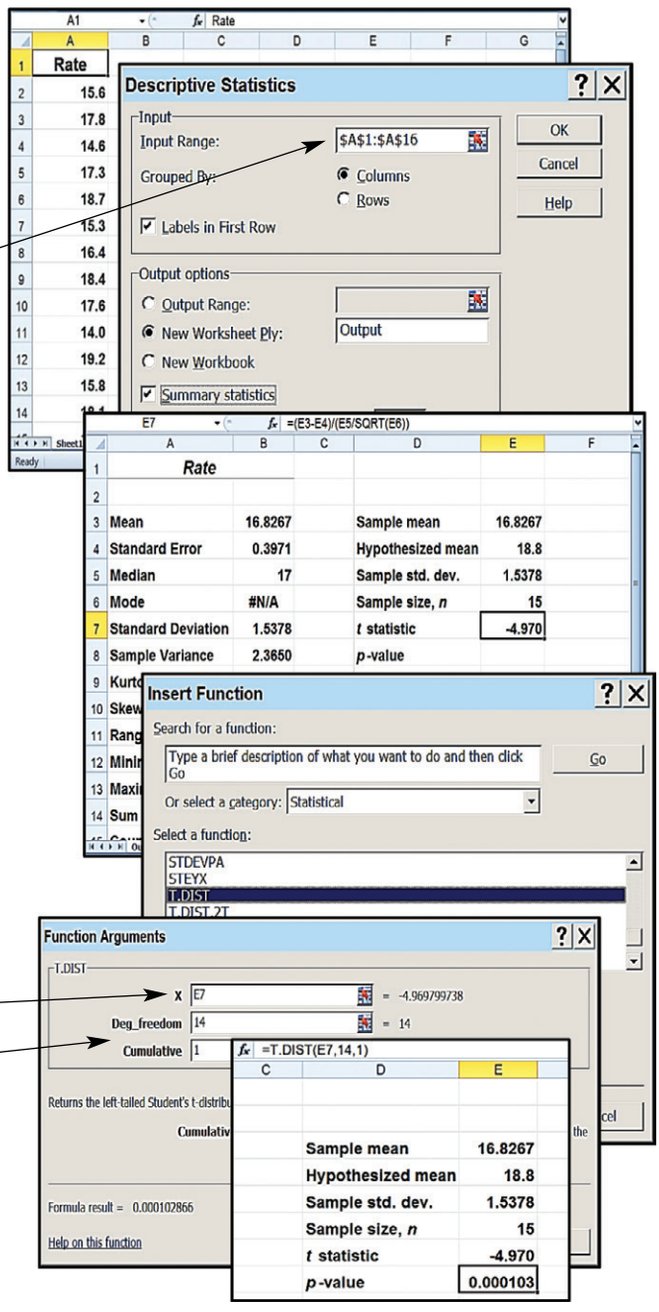
Descriptive statistics:

- Enter the interest rate data from Exercise 9.33 (page 361) into cells A2:A16 with the label Rate in cell A1.
- Select **Data : Data Analysis : Descriptive Statistics.**
- Click OK in the Data Analysis dialog box.
- In the Descriptive Statistics dialog box, enter A1:A16 into the Input Range box.
- Place checkmarks in the "Labels in first row" and Summary Statistics checkboxes.
- Under output options, select "New Worksheet Ply" and enter Output for the worksheet's name.
- Click OK in the Descriptive Statistics dialog box.

The resulting block of descriptive statistics is displayed in the Output worksheet and entries needed to carry out the *t* test have been entered into the cell range D3:E6.

Computation of the test statistic and *p*-value:

- In cell E7, type the formula
  = (E3 − E4)/(E5/SQRT(E6))
  to compute the test statistic *t* (= −4.970).
- Click on cell E8 and then select the Insert Function button f_x on the Excel toolbar.
- In the Insert Function dialog box, select Statistical from the "Or select a category:" menu, select T.DIST from the "Select a function:" menu, and click OK.
- In the T.DIST Function Arguments dialog box, enter E7 into the X window.
- Enter 14 into the Deg_freedom window and Enter 1 into the Cumulative window.
- Click OK in the T.DIST Function Arguments dialog box.
- The *p*-value for the test will be placed in cell E8.



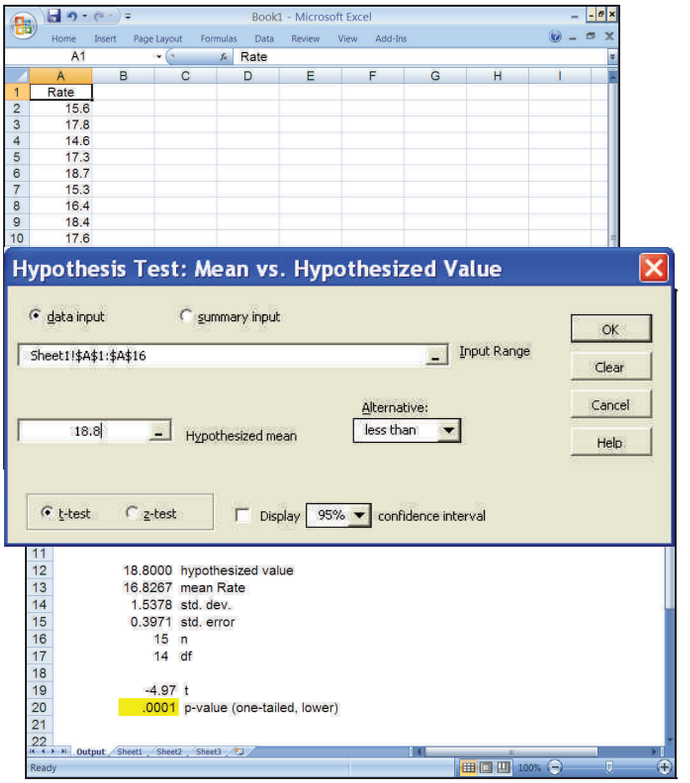
The T.DIST function returns the left tail area under the appropriate *t* curve. Because we are testing a "less than" alternative hypothesis in this example, the desired *p*-value is a left tail area. If we are testing a **"greater than" alternative,** the *p*-value (which is a right tail area) is found by using a cell formula to subtract the left tail area provided by Excel from one. In the case of a **"not equal to" alternative,** we use the T.DIST function to find the area under the *t*-curve to the left of the absolute value (abs) of the *t*-statistic. We then use cell formula(s) to subtract this area from one and to multiply the resulting right tail area by two.

Appendix 9.2 ■ One-Sample Hypothesis Testing Using MegaStat

Hypothesis test for a population mean in Exercise 9.33 on page 361 (data file: CreditCd.xlsx):

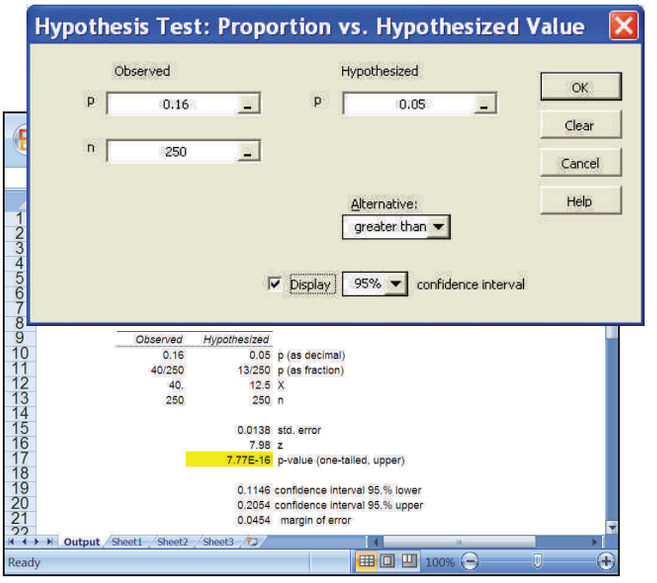
- Enter the interest rate data from Exercise 9.33 (page 361) into cells A2:A16 with the label Rate in cell A1.
- Select **Add-Ins : MegaStat : Hypothesis Tests : Mean vs. Hypothesized Value**
- In the "Hypothesis Test: Mean vs. Hypothesized Value" dialog box, click on "data input" and use the AutoExpand feature to enter the range A1: A16 into the Input Range window.
- Enter the hypothesized value (here equal to 18.8) into the Hypothesized Mean window.
- Select the desired alternative (here "less than") from the drop-down menu in the Alternative box.
- Click on t-test and click OK in the "Hypothesis Test: Mean vs. Hypothesized Value" dialog box.
- A hypothesis test employing summary data can be carried out by clicking on "summary data," and by entering a range into the Input Range window that contains the following—label; sample mean; sample standard deviation; sample size n.

A *z* test can be carried out (in the unlikely event that the population standard deviation is known) by clicking on "z-test."



Hypothesis test for a population proportion. Consider testing H_0: $p = .05$ versus H_a: $p > .05$, where $n = 250$ and $\hat{p} = .16$.

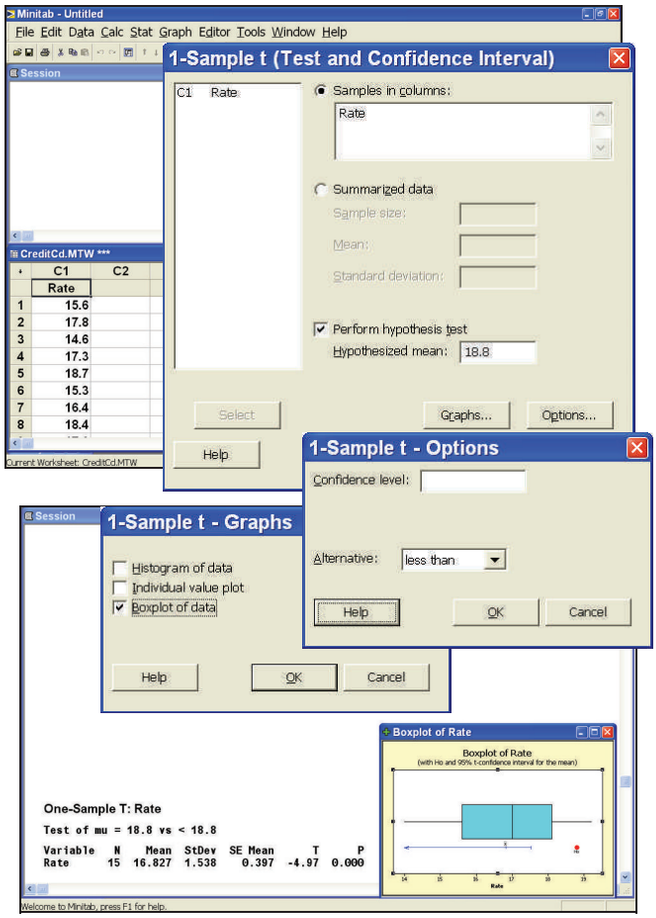
- Select **Add-Ins : MegaStat : Hypothesis Tests : Proportion vs. Hypothesized Value**
- In the "Hypothesis Test: Proportion vs. Hypothesized Value" dialog box, enter the hypothesized value (here equal to 0.05) into the "Hypothesized p" window.
- Enter the observed sample proportion (here equal to 0.16) into the "Observed p" window.
- Enter the sample size (here equal to 250) into the "n" window.
- Select the desired alternative (here "greater than") from the drop-down menu in the Alternative box.
- Check the "Display confidence interval" checkbox (if desired), and select or type the appropriate level of confidence.
- Click OK in the "Hypothesis Test: Proportion vs. Hypothesized Value" dialog box.

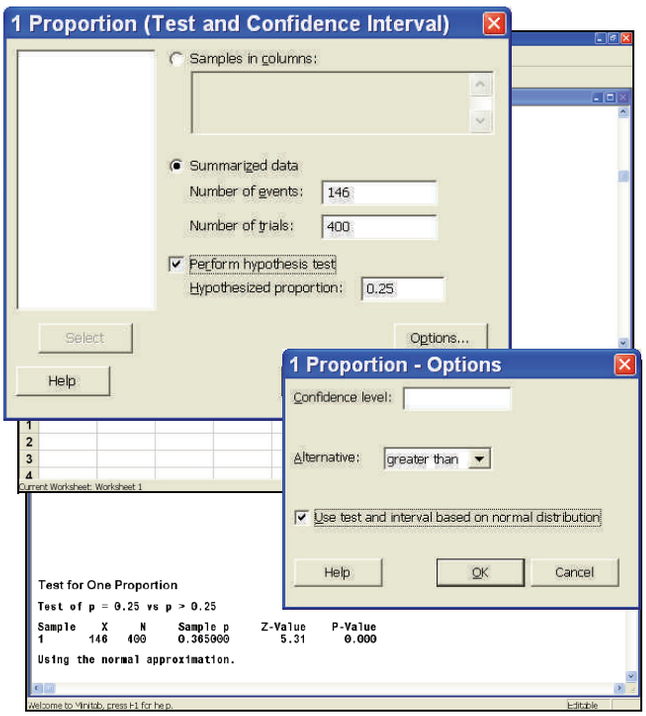


Appendix 9.3 ■ One-Sample Hypothesis Testing Using MINITAB

Hypothesis test for a population mean in Exercise 9.33 on page 361 (data file: CreditCd.MTW):

- In the Data window, enter the interest rate data from Exercise 9.33 (page 361) into a single column with variable name Rate.
- Select **Stat : Basic Statistics : 1-Sample t.**
- In the "1-Sample t (Test and Confidence Interval)" dialog box, select the "Samples in columns" option.
- Select the variable name Rate into the "Samples in columns" window.
- Place a checkmark in the "Perform hypothesis test" checkbox.
- Enter the hypothesized mean (here 18.8) into the "Hypothesized mean" window.
- Click the Options . . . button, select the desired alternative (in this case "less than") from the Alternative drop-down menu, and click OK in the "1-Sample t-Options" dialog box.
- To produce a boxplot of the data with a graphical representation of the hypothesis test, click the Graphs . . . button in the "1-Sample t (Test and Confidence Interval)" dialog box, check the "Boxplot of data" checkbox, and click OK in the "1-Sample t—Graphs" dialog box.
- Click OK in the "1-Sample t (Test and Confidence Interval)" dialog box.
- The *t* test results are given in the Session window, and the boxplot is displayed in a graphics window.



Hypothesis test for a population proportion in Exercise 9.38 on page 365:

- Select **Stat : Basic Statistics : 1 Proportion**
- In the "1 Proportion (Test and Confidence Interval)" dialog box, select the "Summarized data" option.
- Enter the sample number of successes (here equal to 146) into the "Number of events" window.
- Enter the sample size (here equal to 400) into the "Number of trials" window.
- Place a checkmark in the "Perform hypothesis test" checkbox.
- Enter the hypothesized proportion (here equal to 0.25) into the "Hypothesized proportion" window.
- Click on the Options . . . button.
- In the "1 Proportion—Options" dialog box, select the desired alternative (in this case "greater than") from the Alternative drop-down menu.
- Place a checkmark in the "Use test and interval based on normal distribution" checkbox.
- Click OK in the "1 Proportion—Options" dialog box and click OK in the "1 Proportion (Test and Confidence Interval)" dialog box.
- The hypothesis test results are given in the Session window.