We then define another hypothesis, called the **alternative hypothesis**, which is the opposite of what is stated in the null hypothesis. The alternative hypothesis is denoted by H_a. The hypothesis testing procedure uses data from a sample to test the two competing statements indicated by H_0 and H_a.

This chapter shows how hypothesis tests can be conducted about a population mean and a population proportion. We begin by providing examples that illustrate approaches to developing null and alternative hypotheses.

9.1 Developing Null and Alternative Hypotheses

It is not always obvious how the null and alternative hypotheses should be formulated. Care must be taken to structure the hypotheses appropriately so that the hypothesis testing conclusion provides the information the researcher or decision maker wants. The context of the situation is very important in determining how the hypotheses should be stated. All hypothesis testing applications involve collecting a sample and using the sample results to provide evidence for drawing a conclusion. Good questions to consider when formulating the null and alternative hypotheses are, What is the purpose of collecting the sample? What conclusions are we hoping to make?

to formulate
correctly will
practice. Expect
confusion
proper choice of
and alternative
The examples
section are intended
guidelines.

In the chapter introduction, we stated that the null hypothesis H_0 is a tentative assumption about a population parameter such as a population mean or a population proportion. The alternative hypothesis H_a is a statement that is the opposite of what is stated in the null hypothesis. In some situations it is easier to identify the alternative hypothesis first and then develop the null hypothesis. In other situations it is easier to identify the null hypothesis first and then develop the alternative hypothesis. We will illustrate these situations in the following examples.

The Alternative Hypothesis as a Research Hypothesis

Many applications of hypothesis testing involve an attempt to gather evidence in support of a research hypothesis. In these situations, it is often best to begin with the alternative hypothesis and make it the conclusion that the researcher hopes to support. Consider a particular automobile that currently attains a fuel efficiency of 24 miles per gallon in city driving. A product research group has developed a new fuel injection system designed to increase the miles-per-gallon rating. The group will run controlled tests with the new fuel injection system looking for statistical support for the conclusion that the new fuel injection system provides more miles per gallon than the current system.

Several new fuel injection units will be manufactured, installed in test automobiles, and subjected to research-controlled driving conditions. The sample mean miles per gallon for these automobiles will be computed and used in a hypothesis test to determine whether it can be concluded that the new system provides more than 24 miles per gallon. In terms of the population mean miles per gallon μ, the research hypothesis $\mu > 24$ becomes the alternative hypothesis. Since the current system provides an average or mean of 24 miles per gallon, we will make the tentative assumption that the new system is not any better than the current system and choose $\mu \leq 24$ as the null hypothesis. The null and alternative hypotheses are:

$$H_0\colon \mu \leq 24$$
$$H_a\colon \mu > 24$$

If the sample results lead to the conclusion to reject H_0, the inference can be made that H_a: $\mu > 24$ is true. The researchers have the statistical support to state that the new

The conclusion that the research hypothesis is true is made if the sample data provide sufficient evidence to show that the null hypothesis can be rejected.

fuel injection system increases the mean number of miles per gallon. The production of automobiles with the new fuel injection system should be considered. However, if the sample results lead to the conclusion that H_0 cannot be rejected, the researchers cannot conclude that the new fuel injection system is better than the current system. Production of automobiles with the new fuel injection system on the basis of better gas mileage cannot be justified. Perhaps more research and further testing can be conducted.

Successful companies stay competitive by developing new products, new methods, new systems, and the like that are better than what is currently available. Before adopting something new, it is desirable to conduct research to determine whether there is statistical support for the conclusion that the new approach is indeed better. In such cases, the research hypothesis is stated as the alternative hypothesis. For example, a new teaching method is developed that is believed to be better than the current method. The alternative hypothesis is that the new method is better. The null hypothesis is that the new method is no better than the old method. A new sales force bonus plan is developed in an attempt to increase sales. The alternative hypothesis is that the new bonus plan increases sales. The null hypothesis is that the new bonus plan does not increase sales. A new drug is developed with the goal of lowering blood pressure more than an existing drug. The alternative hypothesis is that the new drug lowers blood pressure more than the existing drug. The null hypothesis is that the new drug does not provide lower blood pressure than the existing drug. In each case, rejection of the null hypothesis H_0 provides statistical support for the research hypothesis. We will see many examples of hypothesis tests in research situations such as these throughout this chapter and in the remainder of the text.

The Null Hypothesis as an Assumption to Be Challenged

Of course, not all hypothesis tests involve research hypotheses. In the following discussion we consider applications of hypothesis testing where we begin with a belief or an assumption that a statement about the value of a population parameter is true. We will then use a hypothesis test to challenge the assumption and determine whether there is statistical evidence to conclude that the assumption is incorrect. In these situations, it is helpful to develop the null hypothesis first. The null hypothesis H_0 expresses the belief or assumption about the value of the population parameter. The alternative hypothesis H_a is that the belief or assumption is incorrect.

As an example, consider the situation of a manufacturer of soft drink products. The label on a soft drink bottle states that it contains 67.6 fluid ounces. We consider the label correct provided the population mean filling volume for the bottles is *at least* 67.6 fluid ounces. Without any reason to believe otherwise, we would give the manufacturer the benefit of the doubt and assume that the statement provided on the label is correct. Thus, in a hypothesis test about the population mean fluid ounces per bottle, we would begin with the assumption that the label is correct and state the null hypothesis as $\mu \geq 67.6$. The challenge to this assumption would imply that the label is incorrect and the bottles are being underfilled. This challenge would be stated as the alternative hypothesis $\mu < 67.6$. Thus, the null and alternative hypotheses are:

A manufacturer's product information is usually assumed to be true and stated as the null hypothesis. The conclusion that the information is incorrect can be made if the null hypothesis is rejected.

$$H_0\colon \mu \geq 67.6$$
$$H_a\colon \mu < 67.6$$

A government agency with the responsibility for validating manufacturing labels could select a sample of soft drink bottles, compute the sample mean fluid ounces, and use the sample results to test the preceding hypotheses. If the sample results lead to the conclusion to reject H_0, the inference that $H_a\colon \mu < 67.6$ is true can be made. With this statistical support, the agency is justified in concluding that the label is incorrect and underfilling of the bottles is occurring. Appropriate action to force the manufacturer to comply with labeling standards would be considered. However, if the sample results indicate H_0 cannot be rejected, the assumption that the manufacturer's labeling is correct cannot be rejected. With this conclusion, no action would be taken.

Let us now consider a variation of the soft drink bottle-filling example by viewing the same situation from the manufacturer's point of view. The bottle-filling operation has been designed to fill soft drink bottles with 67.6 fluid ounces as stated on the label. The company does not want to underfill the containers because that could result in an underfilling complaint from customers or, perhaps, a government agency. However, the company does not want to overfill containers either because putting more soft drink than necessary into the containers would be an unnecessary cost. The company's goal would be to adjust the bottle-filling operation so that the population mean filling weight per bottle is 67.6 fluid ounces as specified on the label.

Although this is the company's goal, from time to time any production process can get out of adjustment. If this occurs in our example, underfilling or overfilling of the soft drink bottles will occur. In either case, the company would like to know about it in order to correct the situation by readjusting the bottle-filling operation to the designed 67.6 fluid ounces. In this hypothesis testing application, we would begin with the assumption that the production process is operating correctly and state the null hypothesis as $\mu = 67.6$ fluid ounces. The alternative hypothesis that challenges this assumption is that $\mu \neq 67.6$, which indicates either overfilling or underfilling is occurring. The null and alternative hypotheses for the manufacturer's hypothesis test are

$$H_0\colon \mu = 67.6$$
$$H_a\colon \mu \neq 67.6$$

Suppose that the soft drink manufacturer uses a quality control procedure to periodically select a sample of bottles from the filling operation and computes the sample mean fluid ounces per bottle. If the sample results lead to the conclusion to reject H_0, the inference is made that $H_a\colon \mu \neq 67.6$, is true. We conclude that the bottles are not being filled properly and the production process should be adjusted to restore the population mean to 67.6 fluid ounces per bottle. However, if the sample results indicate H_0 cannot be rejected, the assumption that the manufacturer's bottle-filling operation is functioning properly cannot be rejected. In this case, no further action would be taken and the production operation would continue to run.

The two preceding forms of the soft drink manufacturing hypothesis test show that the null and alternative hypotheses may vary depending upon the point of view of the researcher or decision maker. To formulate hypotheses correctly it is important to understand the context of the situation and structure the hypotheses to provide the information the researcher or decision maker wants.

Summary of Forms for Null and Alternative Hypotheses

The hypothesis tests in this chapter involve two population parameters: the population mean and the population proportion. Depending on the situation, hypothesis tests about a population parameter may take one of three forms: Two use inequalities in the null hypothesis; the third uses an equality in the null hypothesis. For hypothesis tests involving a population mean, we let μ_0 denote the hypothesized value and we must choose one of the following three forms for the hypothesis test.

The three possible forms of potheses H_0 and H_a are wn here. Note that the uality always appears in ull hypothesis H_0.

$$H_0\colon \mu \geq \mu_0 \qquad H_0\colon \mu \leq \mu_0 \qquad H_0\colon \mu = \mu_0$$
$$H_a\colon \mu < \mu_0 \qquad H_a\colon \mu > \mu_0 \qquad H_a\colon \mu \neq \mu_0$$

For reasons that will be clear later, the first two forms are called one-tailed tests. The third form is called a two-tailed test.

In many situations, the choice of H_0 and H_a is not obvious and judgment is necessary to select the proper form. However, as the preceding forms show, the equality part of the expression (either $\geq$, $\leq$, or $=$) *always* appears in the null hypothesis. In selecting the proper form of H_0 and H_a, keep in mind that the alternative hypothesis is often what the test is attempting to establish. Hence, asking whether the user is looking for evidence to support $\mu < \mu_0$, $\mu > \mu_0$, or $\mu \neq \mu_0$ will help determine H_a. The following exercises are designed to provide practice in choosing the proper form for a hypothesis test involving a population mean.

[illegible]

1. The manager of the Danvers-Hilton Resort Hotel stated that the mean guest bill for a weekend is \$600 or less. A member of the hotel's accounting staff noticed that the total charges for guest bills have been increasing in recent months. The accountant will use a sample of future weekend guest bills to test the manager's claim.
   a. Which form of the hypotheses should be used to test the manager's claim? Explain.

$$\begin{array}{lll} H_0\colon \mu \geq 600 & H_0\colon \mu \leq 600 & H_0\colon \mu = 600 \\ H_a\colon \mu < 600 & H_a\colon \mu > 600 & H_a\colon \mu \neq 600 \end{array}$$

   b. What conclusion is appropriate when H_0 cannot be rejected?
   c. What conclusion is appropriate when H_0 can be rejected?

SELF*test*

2. The manager of an automobile dealership is considering a new bonus plan designed to increase sales volume. Currently, the mean sales volume is 14 automobiles per month. The manager wants to conduct a research study to see whether the new bonus plan increases sales volume. To collect data on the plan, a sample of sales personnel will be allowed to sell under the new bonus plan for a one-month period.
   a. Develop the null and alternative hypotheses most appropriate for this situation.
   b. Comment on the conclusion when H_0 cannot be rejected.
   c. Comment on the conclusion when H_0 can be rejected.
3. A production line operation is designed to fill cartons with laundry detergent to a mean weight of 32 ounces. A sample of cartons is periodically selected and weighed to determine whether underfilling or overfilling is occurring. If the sample data lead to a conclusion of underfilling or overfilling, the production line will be shut down and adjusted to obtain proper filling.
   a. Formulate the null and alternative hypotheses that will help in deciding whether to shut down and adjust the production line.
   b. Comment on the conclusion and the decision when H_0 cannot be rejected.
   c. Comment on the conclusion and the decision when H_0 can be rejected.
4. Because of high production-changeover time and costs, a director of manufacturing must convince management that a proposed manufacturing method reduces costs before the new method can be implemented. The current production method operates with a mean cost of \$220 per hour. A research study will measure the cost of the new method over a sample production period.
   a. Develop the null and alternative hypotheses most appropriate for this study.
   b. Comment on the conclusion when H_0 cannot be rejected.
   c. Comment on the conclusion when H_0 can be rejected

9.2 Type I and Type II Errors

The null and alternative hypotheses are competing statements about the population. Either the null hypothesis H_0 is true or the alternative hypothesis H_a is true, but not both. Ideally the hypothesis testing procedure should lead to the acceptance of H_0 when H_0 is true and the rejection of H_0 when H_a is true. Unfortunately, the correct conclusions are not always possible. Because hypothesis tests are based on sample information, we must allow for the possibility of errors. Table 9.1 illustrates the two kinds of errors that can be made in hypothesis testing.

The first row of Table 9.1 shows what can happen if the conclusion is to accept H_0. If H_0 is true, this conclusion is correct. However, if H_a is true, we make a **Type II error**; that is, we accept H_0 when it is false. The second row of Table 9.1 shows what can happen if the conclusion is to reject H_0. If H_0 is true, we make a **Type I error**; that is, we reject H_0 when it is true. However, if H_a is true, rejecting H_0 is correct.

Recall the hypothesis testing illustration discussed in Section 9.1, in which an automobile product research group developed a new fuel injection system designed to increase the miles-per-gallon rating of a particular automobile. With the current model obtaining an average of 24 miles per gallon, the hypothesis test was formulated as follows.

$$H_0\colon \mu \le 24$$
$$H_a\colon \mu > 24$$

The alternative hypothesis, $H_a\colon \mu > 24$, indicates that the researchers are looking for sample evidence to support the conclusion that the population mean miles per gallon with the new fuel injection system is greater than 24.

In this application, the Type I error of rejecting H_0 when it is true corresponds to the researchers claiming that the new system improves the miles-per-gallon rating ($\mu > 24$) when in fact the new system is not any better than the current system. In contrast, the Type II error of accepting H_0 when it is false corresponds to the researchers concluding that the new system is not any better than the current system ($\mu \le 24$) when in fact the new system improves miles-per-gallon performance.

For the miles-per-gallon rating hypothesis test, the null hypothesis is $H_0\colon \mu \le 24$. Suppose the null hypothesis is true as an equality; that is, $\mu = 24$. The probability of making a Type I error when the null hypothesis is true as an equality is called the **level of significance**. Thus, for the miles-per-gallon rating hypothesis test, the level of significance is the probability of rejecting $H_0\colon \mu \le 24$ when $\mu = 24$. Because of the importance of this concept, we now restate the definition of level of significance.

TABLE 9.1 ERRORS AND CORRECT CONCLUSIONS IN HYPOTHESIS TESTING

		Population Condition	
		H_0 True	**H_a True**
Conclusion	**Accept H_0**	Correct Conclusion	Type II Error
	Reject H_0	Type I Error	Correct Conclusion

LEVEL OF SIGNIFICANCE

The level of significance is the probability of making a Type I error when the null hypothesis is true as an equality.

The Greek symbol α (alpha) is used to denote the level of significance, and common choices for α are .05 and .01.

In practice, the person responsible for the hypothesis test specifies the level of significance. By selecting α, that person is controlling the probability of making a Type I error. If the cost of making a Type I error is high, small values of α are preferred. If the cost of making a Type I error is not too high, larger values of α are typically used. Applications of hypothesis testing that only control for the Type I error are called *significance tests*. Many applications of hypothesis testing are of this type.

If the sample data are consistent with the null hypothesis H_0, we will follow the practice of concluding "do not reject H_0." This conclusion is preferred over "accept H_0," because the conclusion to accept H_0 puts us at risk of making a Type II error.

Although most applications of hypothesis testing control for the probability of making a Type I error, they do not always control for the probability of making a Type II error. Hence, if we decide to accept H_0, we cannot determine how confident we can be with that decision. Because of the uncertainty associated with making a Type II error when conducting significance tests, statisticians usually recommend that we use the statement "do not reject H_0" instead of "accept H_0." Using the statement "do not reject H_0" carries the recommendation to withhold both judgment and action. In effect, by not directly accepting H_0, the statistician avoids the risk of making a Type II error. Whenever the probability of making a Type II error has not been determined and controlled, we will not make the statement "accept H_0." In such cases, only two conclusions are possible: *do not reject* H_0 or *reject* H_0.

Although controlling for a Type II error in hypothesis testing is not common, it can be done. More advanced texts describe procedures for determining and controlling the probability of making a Type II error.[1] If proper controls have been established for this error, action based on the "accept H_0" conclusion can be appropriate.

NOTE AND COMMENT

Walter Williams, syndicated columnist and professor of economics at George Mason University, points out that the possibility of making a Type I or a Type II error is always present in decision making (*The Cincinnati Enquirer*, August 14, 2005). He notes that the Food and Drug Administration runs the risk of making these errors in its drug approval process. The FDA must either approve a new drug or not approve it. Thus, the FDA runs the risk of making a Type I error by approving a new drug that is not safe and effective, or making a Type II error by failing to approve a new drug that is safe and effective. Regardless of the decision made, the possibility of making a costly error cannot be eliminated.

SELFtest

5. Duke Energy reported that the cost of electricity for an efficient home in a particular neighborhood of Cincinnati, Ohio was $104 per month (*Home Energy Report*, Duke Energy, March 2012). A researcher believes that the cost of electricity for a comparable neighborhood in Chicago, Illinois is higher. A sample of homes in this Chicago neighborhood will be taken

---

[1] See, for example, D. R. Anderson, D. J. Sweeney, and T. A. Williams, *Statistics for Busin* (Cincinnati: Cengage Learning, 2018).

and the sample mean monthly cost of electricity will be used to test the following null and alternative hypotheses.

$$H_0: \mu \leq 104$$
$$H_a: \mu > 104$$

a. Assume the sample data lead to rejection of the null hypothesis. What would be your conclusion about the cost of electricity in the Chicago neighborhood?
b. What is the Type I error in this situation? What are the consequences of making this error?
c. What is the Type II error in this situation? What are the consequences of making this error?

6. The label on a 3-quart container of orange juice states that the orange juice contains an average of 1 gram of fat or less. Answer the following questions for a hypothesis test that could be used to test the claim on the label.
   a. Develop the appropriate null and alternative hypotheses.
   b. What is the Type I error in this situation? What are the consequences of making this error?
   c. What is the Type II error in this situation? What are the consequences of making this error?
7. Carpetland salespersons average \$8000 per week in sales. Steve Contois, the firm's vice president, proposes a compensation plan with new selling incentives. Steve hopes that the results of a trial selling period will enable him to conclude that the compensation plan increases the average sales per salesperson.
   a. Develop the appropriate null and alternative hypotheses.
   b. What is the Type I error in this situation? What are the consequences of making this error?
   c. What is the Type II error in this situation? What are the consequences of making this error?
8. Suppose a new production method will be implemented if a hypothesis test supports the conclusion that the new method reduces the mean operating cost per hour.
   a. State the appropriate null and alternative hypotheses if the mean cost for the current production method is \$220 per hour.
   b. What is the Type I error in this situation? What are the consequences of making this error?
   c. What is the Type II error in this situation? What are the consequences of making this error?

9.3 Population Mean: σ Known

In Chapter 8, we said that the σ known case corresponds to applications in which historical data and/or other information are available that enable us to obtain a good estimate of the population standard deviation prior to sampling. In such cases the population standard deviation can, for all practical purposes, be considered known. In this section we show how to conduct a hypothesis test about a population mean for the σ known case.

The methods presented in this section are exact if the sample is selected from a population that is normally distributed. In cases where it is not reasonable to assume the population is normally distributed, these methods are still applicable if the sample size is large enough. We provide some practical advice concerning the population distribution and the sample size at the end of this section.

One-Tailed Test

One-tailed tests about a population mean take one of the following two forms.

Lower Tail Test	Upper Tail Test
$H_0: \mu \geq \mu_0$	$H_0: \mu \leq \mu_0$
$H_a: \mu < \mu_0$	$H_a: \mu > \mu_0$

Let us consider an example involving a lower tail test.

The Federal Trade Commission (FTC) periodically conducts statistical studies designed to test the claims that manufacturers make about their products. For example, the label on a large can of Hilltop Coffee states that the can contains 3 pounds of coffee. The FTC knows that Hilltop's production process cannot place exactly 3 pounds of coffee in each can, even if the mean filling weight for the population of all cans filled is 3 pounds per can. However, as long as the population mean filling weight is at least 3 pounds per can, the rights of consumers will be protected. Thus, the FTC interprets the label information on a large can of coffee as a claim by Hilltop that the population mean filling weight is at least 3 pounds per can. We will show how the FTC can check Hilltop's claim by conducting a lower tail hypothesis test.

The first step is to develop the null and alternative hypotheses for the test. If the population mean filling weight is at least 3 pounds per can, Hilltop's claim is correct. This establishes the null hypothesis for the test. However, if the population mean weight is less than 3 pounds per can, Hilltop's claim is incorrect. This establishes the alternative hypothesis. With μ denoting the population mean filling weight, the null and alternative hypotheses are as follows:

$$H_0: \mu \geq 3$$
$$H_a: \mu < 3$$

Note that the hypothesized value of the population mean is $\mu_0 = 3$.

If the sample data indicate that H_0 cannot be rejected, the statistical evidence does not support the conclusion that a label violation has occurred. Hence, no action should be taken against Hilltop. However, if the sample data indicate that H_0 can be rejected, we will conclude that the alternative hypothesis, H_a: $\mu < 3$, is true. In this case a conclusion of underfilling and a charge of a label violation against Hilltop would be justified.

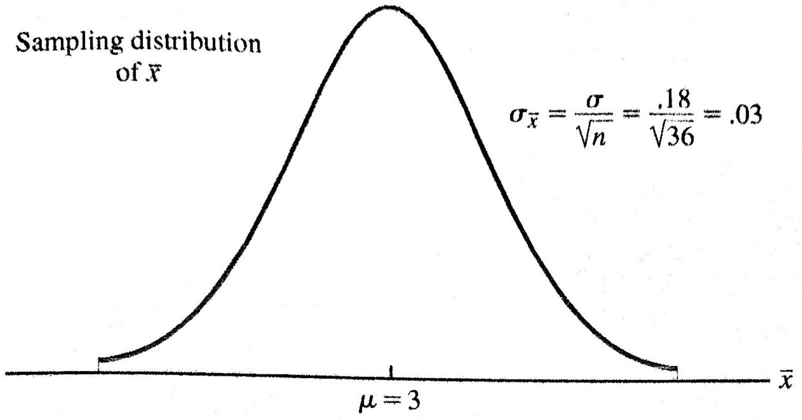
Suppose a sample of 36 cans of coffee is selected and the sample mean $\bar{x}$ is computed as an estimate of the population mean μ. If the value of the sample mean $\bar{x}$ is less than 3 pounds, the sample results will cast doubt on the null hypothesis. What we want to know is how much less than 3 pounds must $\bar{x}$ be before we would be willing to declare the difference significant and risk making a Type I error by falsely accusing Hilltop of a label violation. A key factor in addressing this issue is the value the decision maker selects for the level of significance.

As noted in the preceding section, the level of significance, denoted by α, is the probability of making a Type I error by rejecting H_0 when the null hypothesis is true as an equality. The decision maker must specify the level of significance. If the cost of making a Type I error is high, a small value should be chosen for the level of significance. If the cost is not high, a larger value is more appropriate. In the Hilltop Coffee study, the director of the FTC's testing program made the following statement: "If the company is meeting its weight specifications at $\mu = 3$, I do not want to take action against them. But, I am willing to risk a 1% chance of making such an error." From the director's statement, we set the level of significance for the hypothesis test at $\alpha = .01$. Thus, we must design the hypothesis test so that the probability of making a Type I error when $\mu = 3$ is .01.

For the Hilltop Coffee study, by developing the null and alternative hypotheses and specifying the level of significance for the test, we carry out the first two steps required in conducting every hypothesis test. We are now ready to perform the third step of hypothesis testing: collect the sample data and compute the value of what is called a test statistic.

Test statistic For the Hilltop Coffee study, previous FTC tests show that the population standard deviation can be assumed known with a value of $\sigma = .18$. In addition, these tests also show that the population of filling weights can be assumed to have a normal distribution. From the study of sampling distributions in Chapter 7 we know that if the

FIGURE 9.1 SAMPLING DISTRIBUTION OF $\bar{x}$ FOR THE HILLTOP COFFEE STUDY WHEN THE NULL HYPOTHESIS IS TRUE AS AN EQUALITY ($\mu = 3$)



population from which we are sampling is normally distributed, the sampling distribution of $\bar{x}$ will also be normally distributed. Thus, for the Hilltop Coffee study, the sampling distribution of $\bar{x}$ is normally distributed. With a known value of $\sigma = .18$ and a sample size of $n = 36$, Figure 9.1 shows the sampling distribution of $\bar{x}$ when the null hypothesis is true as an equality, that is, when $\mu = \mu_0 = 3$.[2] Note that the standard error of $\bar{x}$ is given by $\sigma_{\bar{x}} = \sigma/\sqrt{n} = .18/\sqrt{36} = .03$.

error of $\bar{x}$ is
deviation of
distribution

Because the sampling distribution of $\bar{x}$ is normally distributed, the sampling distribution of

$$z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{\bar{x} - 3}{.03}$$

is a standard normal distribution. A value of $z = -1$ means that the value of $\bar{x}$ is one standard error below the hypothesized value of the mean, a value of $z = -2$ means that the value of $\bar{x}$ is two standard errors below the hypothesized value of the mean, and so on. We can use the standard normal probability table to find the lower tail probability corresponding to any z value. For instance, the lower tail area at $z = -3.00$ is .0013. Hence, the probability of obtaining a value of z that is three or more standard errors below the mean is .0013. As a result, the probability of obtaining a value of $\bar{x}$ that is 3 or more standard errors below the hypothesized population mean $\mu_0 = 3$ is also .0013. Such a result is unlikely if the null hypothesis is true.

For hypothesis tests about a population mean in the σ known case, we use the standard normal random variable z as a **test statistic** to determine whether $\bar{x}$ deviates from the hypothesized value of μ enough to justify rejecting the null hypothesis. With $\sigma_{\bar{x}} = \sigma/\sqrt{n}$, the test statistic is as follows.

TEST STATISTIC FOR HYPOTHESIS TESTS ABOUT A POPULATION MEAN: σ KNOWN

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \tag{9.1}$$

The key question for a lower tail test is, How small must the test statistic z be before we choose to reject the null hypothesis? Two approaches can be used to answer this question: the p-value approach and the critical value approach.

p-value approach The p-value approach uses the value of the test statistic z to compute a probability called a p-value.

alue indicates
ue of the test
nusual given the
that H_0 is true.

p-VALUE

A p-value is a probability that provides a measure of the evidence against the null hypothesis provided by the sample. Smaller p-values indicate more evidence against H_0.

The p-value is used to determine whether the null hypothesis should be rejected.

Let us see how the p-value is computed and used. The value of the test statistic is used to compute the p-value. The method used depends on whether the test is a lower tail, an upper tail, or a two-tailed test. For a lower tail test, the p-value is the probability of obtaining a value for the test statistic as small as or smaller than that provided by the sample. Thus, to compute the p-value for the lower tail test in the σ known case, we must find, using the standard normal distribution, the probability that z is less than or equal to the value of the test statistic. After computing the p-value, we must then decide whether it is small enough to reject the null hypothesis; as we will show, this decision involves comparing the p-value to the level of significance.

ATA

Coffee

Let us now compute the p-value for the Hilltop Coffee lower tail test. Suppose the sample of 36 Hilltop coffee cans provides a sample mean of $\bar{x} = 2.92$ pounds. Is $\bar{x} = 2.92$ small enough to cause us to reject H_0? Because this is a lower tail test, the p-value is the area under the standard normal curve for values of $z \leq$ the value of the test statistic. Using $\bar{x} = 2.92$, $\sigma = .18$, and $n = 36$, we compute the value of the test statistic z.

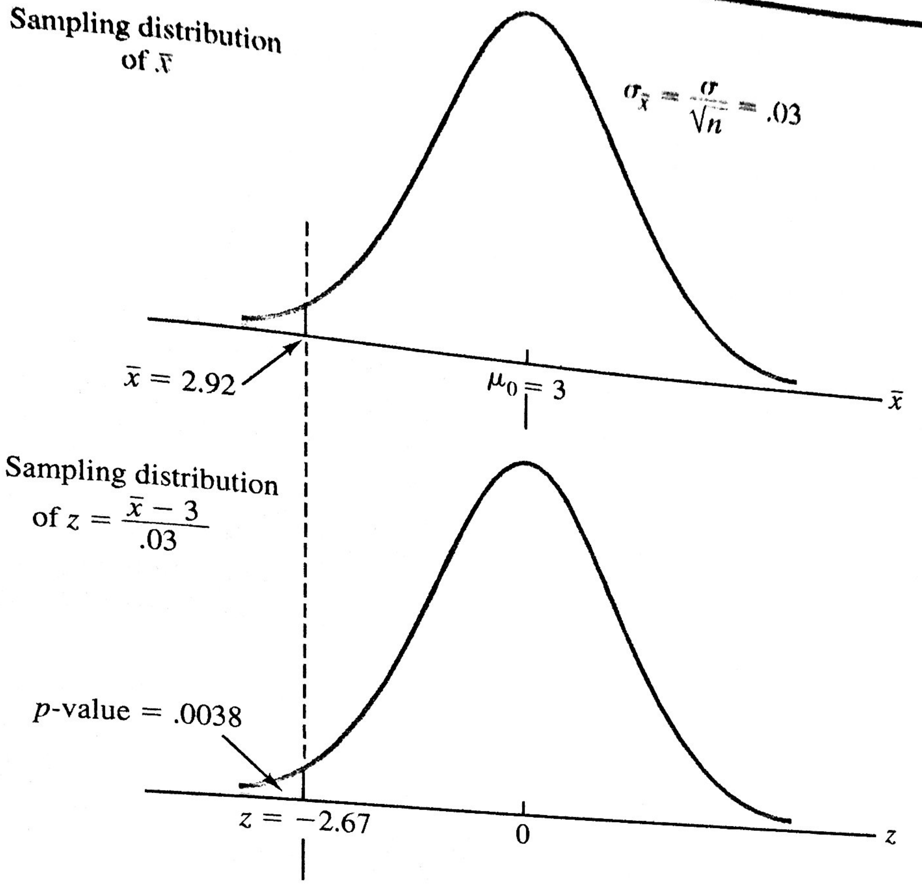
$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{2.92 - 3}{.18/\sqrt{36}} = -2.67$$

Thus, the p-value is the probability that z is less than or equal to -2.67 (the lower tail area corresponding to the value of the test statistic).

Using the standard normal probability table, we find that the lower tail area at $z = -2.67$ is .0038. Figure 9.2 shows that $\bar{x} = 2.92$ corresponds to $z = -2.67$ and a p-value $= .0038$. This p-value indicates a small probability of obtaining a sample mean of $\bar{x} = 2.92$ (and a test statistic of -2.67) or smaller when sampling from a population with $\mu = 3$. This p-value does not provide much support for the null hypothesis, but is it small enough to cause us to reject H_0? The answer depends upon the level of significance for the test.

As noted previously, the director of the FTC's testing program selected a value of .01 for the level of significance. The selection of $\alpha = .01$ means that the director is willing to tolerate a probability of .01 of rejecting the null hypothesis when it is true as an equality ($\mu_0 = 3$). The sample of 36 coffee cans in the Hilltop Coffee study resulted in a p-value $=$ .0038, which means that the probability of obtaining a value of $\bar{x} = 2.92$ or less when the null hypothesis is true as an equality is .0038. Because .0038 is less than or equal to $\alpha = .01$, we reject H_0. Therefore, we find sufficient statistical evidence to reject the null hypothesis at the .01 level of significance.

FIGURE 9.2 p-VALUE FOR THE HILLTOP COFFEE STUDY WHEN $\bar{x} = 2.92$ AND $z = -2.67$



We can now state the general rule for determining whether the null hypothesis can be rejected when using the p-value approach. For a level of significance α, the rejection rule using the p-value approach is as follows.

REJECTION RULE USING p-VALUE

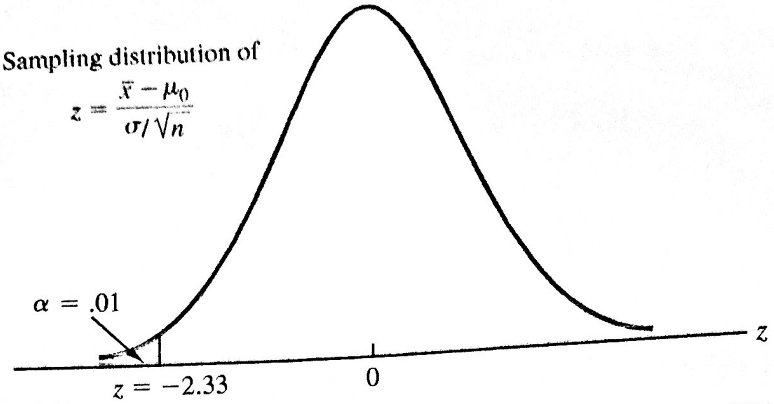
$$\text{Reject } H_0 \text{ if } p\text{-value} \leq \alpha$$

In the Hilltop Coffee test, the p-value of .0038 resulted in the rejection of the null hypothesis. Although the basis for making the rejection decision involves a comparison of the p-value to the level of significance specified by the FTC director, the observed p-value of .0038 means that we would reject H_0 for any value of $\alpha \geq .0038$. For this reason, the p-value is also called the *observed level of significance.*

Different decision makers may express different opinions concerning the cost of making a Type I error and may choose a different level of significance. By providing the p-value as part of the hypothesis testing results, another decision maker can compare the reported p-value to his or her own level of significance and possibly make a different decision with respect to rejecting H_0.

Critical value approach The critical value approach requires that we first determine a value for the test statistic called the **critical value**. For a lower tail test, the critical value serves as a benchmark for determining whether the value of the test statistic is small enough to reject the null hypothesis. It is the value of the test statistic that corresponds to an

FIGURE 9.3 CRITICAL VALUE = −2.33 FOR THE HILLTOP COFFEE HYPOTHESIS TEST



area of α (the level of significance) in the lower tail of the sampling distribution of the test statistic. In other words, the critical value is the largest value of the test statistic that will result in the rejection of the null hypothesis. Let us return to the Hilltop Coffee example and see how this approach works.

In the σ known case, the sampling distribution for the test statistic z is a standard normal distribution. Therefore, the critical value is the value of the test statistic that corresponds to an area of $\alpha = .01$ in the lower tail of a standard normal distribution. Using the standard normal probability table, we find that $z = -2.33$ provides an area of .01 in the lower tail (see Figure 9.3). Thus, if the sample results in a value of the test statistic that is less than or equal to −2.33, the corresponding p-value will be less than or equal to .01; in this case, we should reject the null hypothesis. Hence, for the Hilltop Coffee study the critical value rejection rule for a level of significance of .01 is

$$\text{Reject } H_0 \text{ if } z \leq -2.33$$

In the Hilltop Coffee example, $\bar{x} = 2.92$ and the test statistic is $z = -2.67$. Because $z = -2.67 < -2.33$, we can reject H_0 and conclude that Hilltop Coffee is underfilling cans.

We can generalize the rejection rule for the critical value approach to handle any level of significance. The rejection rule for a lower tail test follows.

REJECTION RULE FOR A LOWER TAIL TEST: CRITICAL VALUE APPROACH

$$\text{Reject } H_0 \text{ if } z \leq -z_\alpha$$

where $-z_\alpha$ is the critical value; that is, the z value that provides an area of α in the lower tail of the standard normal distribution.

Summary The p-value approach to hypothesis testing and the critical value approach will always lead to the same rejection decision; that is, whenever the p-value is less than or equal to α, the value of the test statistic will be less than or equal to the critical value. The advantage of the p-value approach is that the p-value tells us *how* significant the results are (the observed level of significance). If we use the critical value approach, we only know that the results are significant at the stated level of significance.

At the beginning of this section, we said that one-tailed tests about a population mean take one of the following two forms:

Lower Tail Test

$$H_0\colon \mu \geq \mu_0$$
$$H_a\colon \mu < \mu_0$$

Upper Tail Test

$$H_0\colon \mu \leq \mu_0$$
$$H_a\colon \mu > \mu_0$$

We used the Hilltop Coffee study to illustrate how to conduct a lower tail test. We can use the same general approach to conduct an upper tail test. The test statistic z is still computed using equation (9.1). But, for an upper tail test, the p-value is the probability of obtaining a value for the test statistic as large as or larger than that provided by the sample. Thus, to compute the p-value for the upper tail test in the σ known case, we must use the standard normal distribution to compute the probability that z is greater than or equal to the value of the test statistic. Using the critical value approach causes us to reject the null hypothesis if the value of the test statistic is greater than or equal to the critical value z_α; in other words, we reject H_0 if $z \geq z_\alpha$.

Let us summarize the steps involved in computing p-values for one-tailed hypothesis tests.

COMPUTATION OF p-VALUES FOR ONE-TAILED TESTS

1. Compute the value of the test statistic using equation (9.1).
2. **Lower tail test:** Using the standard normal distribution, compute the probability that z is less than or equal to the value of the test statistic (area in the lower tail).
3. **Upper tail test:** Using the standard normal distribution, compute the probability that z is greater than or equal to the value of the test statistic (area in the upper tail).

Two-Tailed Test

In hypothesis testing, the general form for a **two-tailed test** about a population mean is as follows:

$$H_0\colon \mu = \mu_0$$
$$H_a\colon \mu \neq \mu_0$$

In this subsection we show how to conduct a two-tailed test about a population mean for the σ known case. As an illustration, we consider the hypothesis testing situation facing MaxFlight, Inc.

The U.S. Golf Association (USGA) establishes rules that manufacturers of golf equipment must meet if their products are to be acceptable for use in USGA events. MaxFlight uses a high-technology manufacturing process to produce golf balls with a mean driving distance of 295 yards. Sometimes, however, the process gets out of adjustment and produces golf balls with a mean driving distance different from 295 yards. When the mean distance falls below 295 yards, the company worries about losing sales because the golf balls do not provide as much distance as advertised. When the mean distance passes 295 yards, MaxFlight's golf balls may be rejected by the USGA for exceeding the overall distance standard concerning carry and roll.

MaxFlight's quality control program involves taking periodic samples of 50 golf balls to monitor the manufacturing process. For each sample, a hypothesis test is conducted to determine whether the process has fallen out of adjustment. Let us develop the null and alternative hypotheses. We begin by assuming that the process is functioning correctly; that is, the golf balls being produced have a mean distance of 295 yards. This assumption establishes the null hypothesis. The alternative hypothesis is that the mean distance is not equal to 295 yards. With a hypothesized value of $\mu_0 = 295$, the null and alternative hypotheses for the MaxFlight hypothesis test are as follows:

$$H_0: \mu = 295$$
$$H_a: \mu \neq 295$$

If the sample mean $\bar{x}$ is significantly less than 295 yards or significantly greater than 295 yards, we will reject H_0. In this case, corrective action will be taken to adjust the manufacturing process. On the other hand, if $\bar{x}$ does not deviate from the hypothesized mean $\mu_0 = 295$ by a significant amount, H_0 will not be rejected and no action will be taken to adjust the manufacturing process.

The quality control team selected $\alpha = .05$ as the level of significance for the test. Data from previous tests conducted when the process was known to be in adjustment show that the population standard deviation can be assumed known with a value of $\sigma = 12$. Thus, with a sample size of $n = 50$, the standard error of $\bar{x}$ is

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{12}{\sqrt{50}} = 1.7$$

Because the sample size is large, the central limit theorem (see Chapter 7) allows us to conclude that the sampling distribution of $\bar{x}$ can be approximated by a normal distribution. Figure 9.4 shows the sampling distribution of $\bar{x}$ for the MaxFlight hypothesis test with a hypothesized population mean of $\mu_0 = 295$.

DATA
GolfTest

Suppose that a sample of 50 golf balls is selected and that the sample mean is $\bar{x} = 297.6$ yards. This sample mean provides support for the conclusion that the population mean is larger than 295 yards. Is this value of $\bar{x}$ enough larger than 295 to cause us to reject H_0 at the .05 level of significance? In the previous section we described two approaches that can be used to answer this question: the p-value approach and the critical value approach.

p-value approach Recall that the p-value is a probability used to determine whether the null hypothesis should be rejected. For a two-tailed test, values of the test statistic in *either* tail provide evidence against the null hypothesis. For a two-tailed test, the p-value is the probability of obtaining a value for the test statistic *as unlikely as or more unlikely than* that provided by the sample. Let us see how the p-value is computed for the MaxFlight hypothesis test.

FIGURE 9.4 SAMPLING DISTRIBUTION OF $\bar{x}$ FOR THE MAXFLIGHT HYPOTHESIS TEST

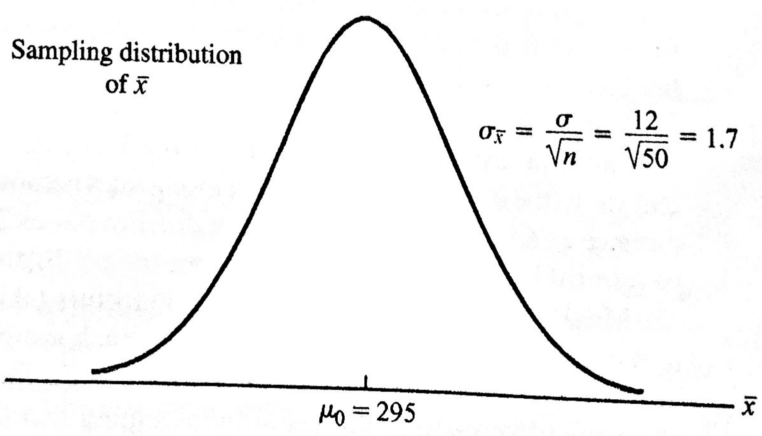
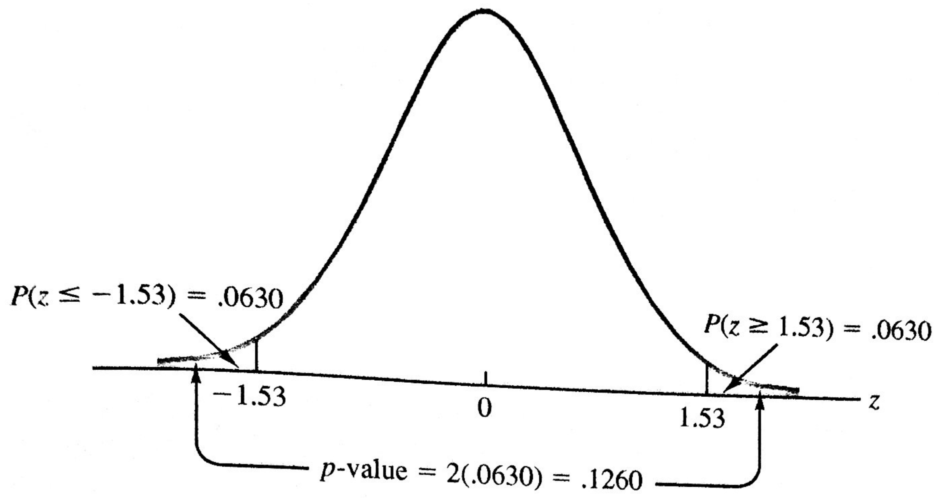


FIGURE 9.5 *p*-VALUE FOR THE MAXFLIGHT HYPOTHESIS TEST



First we compute the value of the test statistic. For the σ known case, the test statistic z is a standard normal random variable. Using equation (9.1) with $\bar{x} = 297.6$, the value of the test statistic is

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{297.6 - 295}{12/\sqrt{50}} = 1.53$$

Now to compute the *p*-value we must find the probability of obtaining a value for the test statistic *at least as unlikely as* $z = 1.53$. Clearly values of $z \geq 1.53$ are *at least as unlikely.* But, because this is a two-tailed test, values of $z \leq -1.53$ are also *at least as unlikely as* the value of the test statistic provided by the sample. In Figure 9.5, we see that the two-tailed *p*-value in this case is given by $P(z \leq -1.53) + P(z \geq 1.53)$. Because the normal curve is symmetric, we can compute this probability by finding $P(z \geq 1.53)$ and doubling it. The table for the standard normal distribution shows that $P(z < 1.53) = .9370$. Thus, the upper tail area is $P(z \geq 1.53) = 1.0000 - .9370 = .0630$. Doubling this, we find that the *p*-value for the MaxFlight two-tailed hypothesis test is *p*-value = 2(.0630) = .1260.

Next we compare the *p*-value to the level of significance to see whether the null hypothesis should be rejected. With a level of significance of $\alpha = .05$, we do not reject H_0 because the *p*-value = .1260 > .05. Because the null hypothesis is not rejected, no action will be taken to adjust the MaxFlight manufacturing process.

Let us summarize the steps involved in computing *p*-values for two-tailed hypothesis tests.

COMPUTATION OF *p*-VALUES FOR TWO-TAILED TESTS

1. Compute the value of the test statistic using equation (9.1).
2. If the value of the test statistic is in the upper tail, compute the probability that z is greater than or equal to the value of the test statistic (the upper tail area). If the value of the test statistic is in the lower tail, compute the probability that z is less than or equal to the value of the test statistic (the lower tail area).
3. Double the probability (or tail area) from step 2 to obtain the *p*-value.

Critical value approach Before leaving this section, let us see how the test statistic z can be compared to a critical value to make the hypothesis testing decision for a two-tailed test.

FIGURE 9.6 CRITICAL VALUES FOR THE MAXFLIGHT HYPOTHESIS TEST

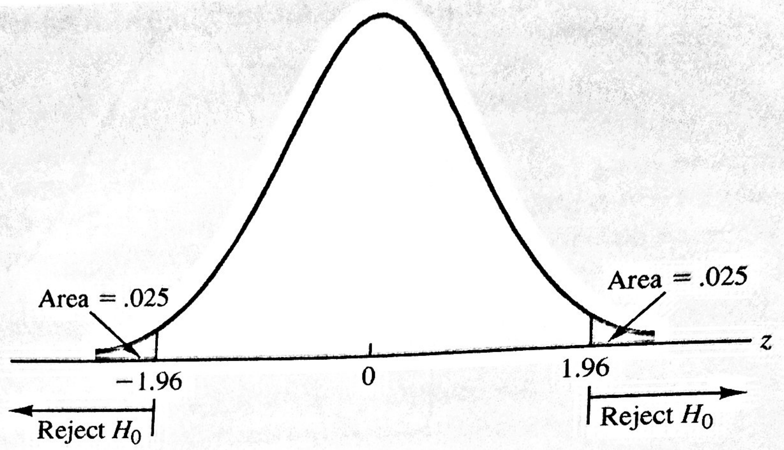


Figure 9.6 shows that the critical values for the test will occur in both the lower and upper tails of the standard normal distribution. With a level of significance of $\alpha = .05$, the area in each tail corresponding to the critical values is $\alpha/2 = .05/2 = .025$. Using the standard normal probability table, we find the critical values for the test statistic are $-z_{.025} = -1.96$ and $z_{.025} = 1.96$. Thus, using the critical value approach, the two-tailed rejection rule is

$$\text{Reject } H_0 \text{ if } z \leq -1.96 \text{ or if } z \geq 1.96$$

Because the value of the test statistic for the MaxFlight study is $z = 1.53$, the statistical evidence will not permit us to reject the null hypothesis at the .05 level of significance.

Using Excel

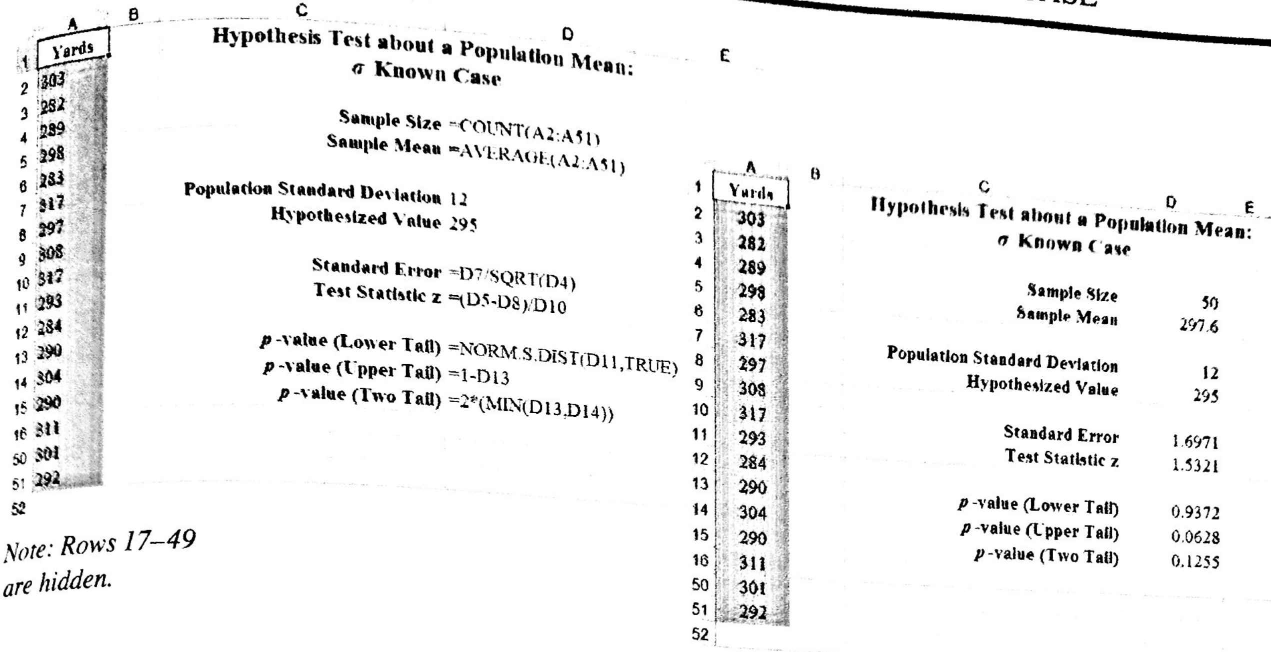
Excel can be used to conduct one-tailed and two-tailed hypothesis tests about a population mean for the σ known case using the *p*-value approach. Recall that the method used to compute a *p*-value depends upon whether the test is lower tail, upper tail, or two-tailed. Therefore, in the Excel procedure we describe we will use the sample results to compute three *p*-values: *p*-value (Lower Tail), *p*-value (Upper Tail), and *p*-value (Two Tail). The user can then choose α and draw a conclusion using whichever *p*-value is appropriate for the type of hypothesis test being conducted. We will illustrate using the MaxFlight two-tailed hypothesis test. Refer to Figure 9.7 as we describe the tasks involved. The formula worksheet is in the background; the value worksheet is in the foreground.

Enter/Access Data: Open the DATAfile named *GolfTest*. A label and the distance data for the sample of 50 golf balls are entered into cells A1:A51.

Enter Functions and Formulas: The sample size and sample mean are computed in cells D4 and D5 using Excel's COUNT and AVERAGE functions, respectively. The value worksheet shows that the sample size is 50 and the sample mean is 297.6. The value of the known population standard deviation (12) is entered into cell D7, and the hypothesized value of the population mean (295) is entered into cell D8.

The standard error is obtained in cell D10 by entering the formula =D7/SQRT(D4). The formula =(D5-D8)/D10 entered into cell D11 computes the test statistic *z*(1.5321). To compute the *p*-value for a lower tail test, we enter the formula =NORM.S.DIST(D11,TRUE) into cell D13. The *p*-value for an upper tail test is then computed in cell D14 as 1 minus the

FIGURE 9.7 EXCEL WORKSHEET: HYPOTHESIS TEST FOR THE σ KNOWN CASE



	A	B	C	D	E
1	Yards		Hypothesis Test about a Population Mean:		
2	303		σ Known Case		
3	282				
4	289		Sample Size	=COUNT(A2:A51)	
5	298		Sample Mean	=AVERAGE(A2:A51)	
6	283				
7	317		Population Standard Deviation	12	
8	297		Hypothesized Value	295	
9	308				
10	317		Standard Error	=D7/SQRT(D4)	
11	293		Test Statistic z	=(D5-D8)/D10	
12	284				
13	290		*p*-value (Lower Tail)	=NORM.S.DIST(D11,TRUE)	
14	304		*p*-value (Upper Tail)	=1-D13	
15	290		*p*-value (Two Tail)	=2*(MIN(D13,D14))	
16	311				
50	301				
51	292				
52					

Note: Rows 17–49 are hidden.

	A	B	C	D	E
1	Yards		Hypothesis Test about a Population Mean:		
2	303		σ Known Case		
3	282				
4	289		Sample Size	50	
5	298		Sample Mean	297.6	
6	283				
7	317		Population Standard Deviation	12	
8	297		Hypothesized Value	295	
9	308				
10	317		Standard Error	1.6971	
11	293		Test Statistic z	1.5321	
12	284				
13	290		*p*-value (Lower Tail)	0.9372	
14	304		*p*-value (Upper Tail)	0.0628	
15	290		*p*-value (Two Tail)	0.1255	
16	311				
50	301				
51	292				
52					

p-value for the lower tail test. Finally, the *p*-value for a two-tailed test is computed in cell D15 as two times the minimum of the two one-tailed *p*-values. The value worksheet shows that *p*-value (Lower Tail) = 0.9372, *p*-value (Upper Tail) = 0.0628, and *p*-value (Two Tail) = 0.1255.

The development of the worksheet is now complete. For the two-tailed MaxFlight problem we cannot reject H_0: $\mu = 295$ using $\alpha = .05$ because the *p*-value (Two Tail) = 0.1255 is greater than α. Thus, the quality control manager has no reason to doubt that the manufacturing process is producing golf balls with a population mean distance of 295 yards.

A template for other problems The worksheet in Figure 9.7 can be used as a template for conducting any one-tailed and two-tailed hypothesis tests for the σ known case. Just enter the appropriate data in column A, adjust the ranges for the formulas in cells D4 and D5, enter the population standard deviation in cell D7, and enter the hypothesized value in cell D8. The standard error, the test statistic, and the three *p*-values will then appear. Depending on the form of the hypothesis test (lower tail, upper tail, or two-tailed), we can then choose the appropriate *p*-value to make the rejection decision.

We can further simplify the use of Figure 9.7 as a template for other problems by eliminating the need to enter new data ranges in cells D4 and D5. To do so we rewrite the cell formulas as follows:

Cell D4: =COUNT(A:A)
Cell D5: =AVERAGE(A:A)

The DATAfile named GolfTest includes a worksheet entitled Template that uses the A:A method for entering the data ranges.

With the A:A method of specifying data ranges, Excel's COUNT function will count the number of numerical values in column A and Excel's AVERAGE function will compute the average of the numerical values in column A. Thus, to solve a new problem it is only necessary to enter the new data in column A, enter the value of the known population standard deviation in cell D7, and enter the hypothesized value of the population mean in cell D8.

The worksheet can also be used as a template for text exercises in which n, $\bar{x}$, and σ are given. Just ignore the data in column A and enter the values for n, $\bar{x}$, and σ into cells D4, D5, and D7, respectively. Then enter the appropriate hypothesized value for the population mean into cell D8. The *p*-values corresponding to lower tail, upper tail, and two-tailed hypothesis tests will then appear in cells D13:D15.

TABLE 9.2 SUMMARY OF HYPOTHESIS TESTS ABOUT A POPULATION MEAN: σ KNOWN CASE

	Lower Tail Test	Upper Tail Test	Two-Tailed Test
Hypotheses	$H_0: \mu \geq \mu_0$ $H_a: \mu < \mu_0$	$H_0: \mu \leq \mu_0$ $H_a: \mu > \mu_0$	$H_0: \mu = \mu_0$ $H_a: \mu \neq \mu_0$
Test Statistic	$z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$	$z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$	$z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$
Rejection Rule: *p*-Value Approach	Reject H_0 if p-value $\leq \alpha$	Reject H_0 if p-value $\leq \alpha$	Reject H_0 if p-value $\leq \alpha$
Rejection Rule: Critical Value Approach	Reject H_0 if $z \leq -z_\alpha$	Reject H_0 if $z \geq z_\alpha$	Reject H_0 if $z \leq -z_{\alpha/2}$ or if $z \geq z_{\alpha/2}$

Summary and Practical Advice

We presented examples of a lower tail test and a two-tailed test about a population mean. Based upon these examples, we can now summarize the hypothesis testing procedures about a population mean for the σ known case as shown in Table 9.2. Note that μ_0 is the hypothesized value of the population mean.

The hypothesis testing steps followed in the two examples presented in this section are common to every hypothesis test.

STEPS OF HYPOTHESIS TESTING

Step 1. Develop the null and alternative hypotheses.
Step 2. Specify the level of significance.
Step 3. Collect the sample data and compute the value of the test statistic.

p-Value Approach

Step 4. Use the value of the test statistic to compute the p-value.
Step 5. Reject H_0 if the p-value $\leq \alpha$.
Step 6. Interpret the statistical conclusion in the context of the application.

Critical Value Approach

Step 4. Use the level of significance to determine the critical value and the rejection rule.
Step 5. Use the value of the test statistic and the rejection rule to determine whether to reject H_0.
Step 6. Interpret the statistical conclusion in the context of the application.

Practical advice about the sample size for hypothesis tests is similar to the advice we provided about the sample size for interval estimation in Chapter 8. In most applications, a sample size of $n \geq 30$ is adequate when using the hypothesis testing procedure described in this section. In cases where the sample size is less than 30, the distribution of the population from which we are sampling becomes an important consideration. If the population is normally distributed, the hypothesis testing procedure that we described is exact and can be used for any sample size. If the population is not normally distributed but is at least roughly symmetric, sample sizes as small as 15 can be expected to provide acceptable results.

Relationship Between Interval Estimation and Hypothesis Testing

In Chapter 8 we showed how to develop a confidence interval estimate of a population mean. For the σ known case, the $(1 - \alpha)\%$ confidence interval estimate of a population mean is given by

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

In this chapter we showed that a two-tailed hypothesis test about a population mean takes the following form:

$$H_0\colon \mu = \mu_0$$
$$H_a\colon \mu \neq \mu_0$$

where μ_0 is the hypothesized value for the population mean.

Suppose that we follow the procedure described in Chapter 8 for constructing a $100(1 - \alpha)\%$ confidence interval for the population mean. We know that $100(1 - \alpha)\%$ of the confidence intervals generated will contain the population mean and $100\alpha\%$ of the confidence intervals generated will not contain the population mean. Thus, if we reject H_0 whenever the confidence interval does not contain μ_0, we will be rejecting the null hypothesis when it is true $(\mu = \mu_0)$ with probability α. Recall that the level of significance is the probability of rejecting the null hypothesis when it is true. So constructing a $100(1 - \alpha)\%$ confidence interval and rejecting H_0 whenever the interval does not contain μ_0 is equivalent to conducting a two-tailed hypothesis test with α as the level of significance. The procedure for using a confidence interval to conduct a two-tailed hypothesis test can now be summarized.

A CONFIDENCE INTERVAL APPROACH TO TESTING A HYPOTHESIS OF THE FORM

$$H_0\colon \mu = \mu_0$$
$$H_a\colon \mu \neq \mu_0$$

1. Select a simple random sample from the population and use the value of the sample mean $\bar{x}$ to develop the confidence interval for the population mean μ.

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

2. If the confidence interval contains the hypothesized value μ_0, do not reject H_0. Otherwise, reject[3] H_0.

Two-tailed hypothesis
null hypothesis
rejected if the
ence interval does
lude μ_0.

Let us illustrate by conducting the MaxFlight hypothesis test using the confidence interval approach. The MaxFlight hypothesis test takes the following form:

$$H_0\colon \mu = 295$$
$$H_a\colon \mu \neq 295$$

---

[3] To be consistent with the rule for rejecting H_0 when the p-value $\leq \alpha$, we would also reject H_0 using the confidence interval approach if μ_0 happens to be equal to one of the endpoints of the $100(1 - \alpha)\%$ confidence interval.

To test these hypotheses with a level of significance of $\alpha = .05$, we sampled 50 golf balls and found a sample mean distance of $\bar{x} = 297.6$ yards. Recall that the population standard deviation is $\sigma = 12$. Using these results with $z_{.025} = 1.96$, we find that the 95% confidence interval estimate of the population mean is

$$\bar{x} \pm z_{.025}\frac{\sigma}{\sqrt{n}}$$

$$297.6 \pm 1.96\frac{12}{\sqrt{50}}$$

$$297.6 \pm 3.3$$

or

$$294.3 \text{ to } 300.9$$

This finding enables the quality control manager to conclude with 95% confidence that the mean distance for the population of golf balls is between 294.3 and 300.9 yards. Because the hypothesized value for the population mean, $\mu_0 = 295$, is in this interval, the hypothesis testing conclusion is that the null hypothesis, H_0: $\mu = 295$, cannot be rejected.

Note that this discussion and example pertain to two-tailed hypothesis tests about a population mean. However, the same confidence interval and two-tailed hypothesis testing relationship exists for other population parameters. The relationship can also be extended to one-tailed tests about population parameters. Doing so, however, requires the development of one-sided confidence intervals, which are rarely used in practice.

NOTES AND COMMENTS

1. We have shown how to use p-values. The smaller the p-value the greater the evidence against H_0 and the more the evidence in favor of H_a. Here are some guidelines statisticians suggest for interpreting small p-values.
   - Less than .01—Overwhelming evidence to conclude that H_a is true
   - Between .01 and .05—Strong evidence to conclude that H_a is true
   - Between .05 and .10—Weak evidence to conclude that H_a is true
   - Greater than .10—Insufficient evidence to conclude that H_a is true
2. When testing a hypothesis of the population mean with a sample size that is at least 5% of the population size (that is, $n/N \geq .05$), the finite population correction factor should be used when calculating the standard error of the sampling distribution of $\bar{x}$ when σ is known, that is,

$$\sigma_{\bar{x}} = \sqrt{\frac{N-n}{N-1}}\left(\frac{\sigma}{\sqrt{n}}\right).$$

EXERCISES

Note to Student: Some of the exercises that follow ask you to use the p-value approach and others ask you to use the critical value approach. Both methods will provide the same hypothesis testing conclusion. We provide exercises with both methods to give you practice using both. In later sections and in following chapters, we will generally emphasize the p-value approach as the preferred method, but you may select either based on personal preference.

Methods

9. Consider the following hypothesis test:

$$H_0: \mu \geq 20$$
$$H_a: \mu < 20$$

A sample of 50 provided a sample mean of 19.4. The population standard deviation is 2.

a. Compute the value of the test statistic.
b. What is the p-value?
c. Using $\alpha = .05$, what is your conclusion?
d. What is the rejection rule using the critical value? What is your conclusion?

SELFtest

10. Consider the following hypothesis test:

$$H_0: \mu \leq 25$$
$$H_a: \mu > 25$$

A sample of 40 provided a sample mean of 26.4. The population standard deviation is 6.

a. Compute the value of the test statistic.
b. What is the p-value?
c. At $\alpha = .01$, what is your conclusion?
d. What is the rejection rule using the critical value? What is your conclusion?

11. Consider the following hypothesis test:

$$H_0: \mu = 15$$
$$H_a: \mu \neq 15$$

A sample of 50 provided a sample mean of 14.15. The population standard deviation is 3.

a. Compute the value of the test statistic.
b. What is the p-value?
c. At $\alpha = .05$, what is your conclusion?
d. What is the rejection rule using the critical value? What is your conclusion?

12. Consider the following hypothesis test:

$$H_0: \mu \geq 80$$
$$H_a: \mu < 80$$

A sample of 100 is used and the population standard deviation is 12. Compute the p-value and state your conclusion for each of the following sample results. Use $\alpha = .01$.

a. $\bar{x} = 78.5$
b. $\bar{x} = 77$
c. $\bar{x} = 75.5$
d. $\bar{x} = 81$

13. Consider the following hypothesis test:

$$H_0: \mu \leq 50$$
$$H_a: \mu > 50$$

A sample of 60 is used and the population standard deviation is 8. Use the critical value approach to state your conclusion for each of the following sample results. Use $\alpha = .05$.

a. $\bar{x} = 52.5$
b. $\bar{x} = 51$
c. $\bar{x} = 51.8$

14. Consider the following hypothesis test:

$$H_0: \mu = 22$$
$$H_a: \mu \neq 22$$

lation standard deviation of $300 to answer the following questions.

a. Formulate hypotheses for a test to determine whether the sample data support the conclusion that the population annual expenditure for prescription drugs per person is lower in the Midwest than in the Northeast.
b. What is the value of the test statistic?
c. What is the p-value?
d. At $\alpha = .01$, what is your conclusion?

21. Fowle Marketing Research, Inc. bases charges to a client on the assumption that telephone surveys can be completed in a mean time of 15 minutes or less. If a longer mean survey time is necessary, a premium rate is charged. A sample of 35 surveys provided the survey times shown in the DATAfile named *Fowle.* Based upon past studies, the population standard deviation is assumed known with $\sigma = 4$ minutes. Is the premium rate justified?
    a. Formulate the null and alternative hypotheses for this application.
    b. Compute the value of the test statistic.
    c. What is the p-value?
    d. At $\alpha = .01$, what is your conclusion?

22. CCN and ActMedia provided a television channel targeted to individuals waiting in supermarket checkout lines. The channel showed news, short features, and advertisements. The length of the program was based on the assumption that the population mean time a shopper stands in a supermarket checkout line is 8 minutes. A sample of actual waiting times will be used to test this assumption and determine whether actual mean waiting time differs from this standard.
    a. Formulate the hypotheses for this application.
    b. A sample of 120 shoppers showed a sample mean waiting time of 8.4 minutes. Assume a population standard deviation of $\sigma = 3.2$ minutes. What is the p-value?
    c. At $\alpha = .05$, what is your conclusion?
    d. Compute a 95% confidence interval for the population mean. Does it support your conclusion?

Population Mean: σ Unknown

In this section we describe how to conduct hypothesis tests about a population mean for the σ unknown case. Because the σ unknown case corresponds to situations in which an estimate of the population standard deviation cannot be developed prior to sampling, the sample must be used to develop an estimate of both μ and σ. Thus, to conduct a hypothesis test about a population mean for the σ unknown case, the sample mean $\bar{x}$ is used as an estimate of μ and the sample standard deviation s is used as an estimate of σ.

The steps of the hypothesis testing procedure for the σ unknown case are the same as those for the σ known case described in Section 9.3. But, with σ unknown, the computation of the test statistic and p-value is a bit different. Recall that for the σ known case, the sampling distribution of the test statistic has a standard normal distribution. For the σ unknown case, however, the sampling distribution of the test statistic follows the t distribution; it has slightly more variability because the sample is used to develop estimates of both μ and σ.

In Section 8.2 we showed that an interval estimate of a population mean for the σ unknown case is based on a probability distribution known as the t distribution. Hypothesis tests about a population mean for the σ unknown case are also based on the t distribution. For the σ unknown case, the test statistic has a t distribution with $n - 1$ degrees of freedom.

TEST STATISTIC FOR HYPOTHESIS TESTS ABOUT A POPULATION MEAN: σ UNKNOWN

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \tag{9.2}$$

In Chapter 8 we said that the t distribution is based on an assumption that the population from which we are sampling has a normal distribution. However, research shows that this assumption can be relaxed considerably when the sample size is large enough. We provide some practical advice concerning the population distribution and sample size at the end of the section.

One-Tailed Test

Let us consider an example of a one-tailed test about a population mean for the σ unknown case. A business travel magazine wants to classify transatlantic gateway airports according to the mean rating for the population of business travelers. A rating scale with a low score of 0 and a high score of 10 will be used, and airports with a population mean rating greater than 7 will be designated as superior service airports. The magazine staff surveyed a sample of 60 business travelers at each airport to obtain the ratings data. The sample for London's Heathrow Airport provided a sample mean rating of $\bar{x} = 7.25$ and a sample standard deviation of $s = 1.052$. Do the data indicate that Heathrow should be designated as a superior service airport?

DATA

AirRating

We want to develop a hypothesis test for which the decision to reject H_0 will lead to the conclusion that the population mean rating for the Heathrow Airport is *greater* than 7. Thus, an upper tail test with H_a: $\mu > 7$ is required. The null and alternative hypotheses for this upper tail test are as follows:

$$H_0: \mu \leq 7$$
$$H_a: \mu > 7$$

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

Using equation (9.2) with $\bar{x} = 7.25$, $\mu_0 = 7$, $s = 1.052$, and $n = 60$, the value of the test statistic is

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{7.25 - 7}{1.052/\sqrt{60}} = 1.84$$

The sampling distribution of t has $n - 1 = 60 - 1 = 59$ degrees of freedom. Because the test is an upper tail test, the p-value is $P(t \geq 1.84)$, that is, the upper tail area corresponding to the value of the test statistic.

The t distribution table provided in most textbooks will not contain sufficient detail to determine the exact p-value, such as the p-value corresponding to $t = 1.84$. For instance, using Table 2 in Appendix B, the t distribution with 59 degrees of freedom provides the following information.

Area in Upper Tail	**.20**	**.10**	**.05**	**.025**	**.01**	**.005**
***t* Value (59 *df*)**	.848	1.296	1.671	2.001	2.391	2.662

$t = 1.84$

We see that $t = 1.84$ is between 1.671 and 2.001. Although the table does not provide the exact p-value, the values in the "Area in Upper Tail" row show that the p-value must be less than .05 and greater than .025. With a level of significance of $\alpha = .05$, this placement is all we need to know to make the decision to reject the null hypothesis and conclude that Heathrow should be classified as a superior service airport.

It is cumbersome to use a t table to compute p-values, and only approximate values are obtained. We describe how to compute exact p-values using Excel's T.DIST function in the Using Excel subsection which follows. The exact upper tail p-value for the Heathrow Airport hypothesis test is .0354. With $.0354 < .05$, we reject the null hypothesis and conclude that Heathrow should be classified as a superior service airport.

The decision whether to reject the null hypothesis in the σ unknown case can also be made using the critical value approach. The critical value corresponding to an area of $\alpha = .05$ in the upper tail of a t distribution with 59 degrees of freedom is $t_{.05} = 1.671$. Thus the rejection rule using the critical value approach is to reject H_0 if $t \geq 1.671$. Because $t = 1.84 > 1.671$, H_0 is rejected. Heathrow should be classified as a superior service airport.

Two-Tailed Test

To illustrate how to conduct a two-tailed test about a population mean for the σ unknown case, let us consider the hypothesis testing situation facing Holiday Toys. The company manufactures and distributes its products through more than 1000 retail outlets. In planning production levels for the coming winter season, Holiday must decide how many units of each product to produce prior to knowing the actual demand at the retail level. For this year's most important new toy, Holiday's marketing director is expecting demand to average 40 units per retail outlet. Prior to making the final production decision based upon this estimate, Holiday decided to survey a sample of 25 retailers in order to develop more information about the demand for the new product. Each retailer was provided with information about the features of the new toy along with the cost and the suggested selling price. Then each retailer was asked to specify an anticipated order quantity.

With μ denoting the population mean order quantity per retail outlet, the sample data will be used to conduct the following two-tailed hypothesis test:

$$H_0\colon \mu = 40$$
$$H_a\colon \mu \neq 40$$

If H_0 cannot be rejected, Holiday will continue its production planning based on the marketing director's estimate that the population mean order quantity per retail outlet will be $\mu = 40$ units. However, if H_0 is rejected, Holiday will immediately reevaluate its production plan for the product. A two-tailed hypothesis test is used because Holiday wants to reevaluate the production plan if the population mean quantity per retail outlet is less than anticipated or greater than anticipated. Because no historical data are available (it's a new product), the population mean μ and the population standard deviation must both be estimated using $\bar{x}$ and s from the sample data.

DATA
Orders

The sample of 25 retailers provided a mean of $\bar{x} = 37.4$ and a standard deviation of $s = 11.79$ units. Before going ahead with the use of the t distribution, the analyst constructed a histogram of the sample data in order to check on the form of the population distribution. The histogram of the sample data showed no evidence of skewness or any extreme outliers, so the analyst concluded that the use of the t distribution with $n - 1 = 24$ degrees of freedom was appropriate. Using equation (9.2) with $\bar{x} = 37.4$, $\mu_0 = 40$, $s = 11.79$, and $n = 25$, the value of the test statistic is

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{37.4 - 40}{11.79/\sqrt{25}} = -1.10$$

Because we have a two-tailed test, the p-value is two times the area under the curve of the t distribution for $t \leq -1.10$. Using Table 2 in Appendix B, the t distribution table for 24 degrees of freedom provides the following information.

Area in Upper Tail	**.20**	**.10**	**.05**	**.025**	**.01**	**.005**
***t*-Value (24 *df*)**	.857	1.318	1.711	2.064	2.492	2.797

$t = 1.10$

The t distribution table contains only positive t values (corresponding to areas in the upper tail). Because the t distribution is symmetric, however, the upper tail area for $t = 1.10$ is the same as the lower tail area for $t = -1.10$. We see that $t = 1.10$ is between 0.857 and 1.318. From the "Area in Upper Tail" row, we see that the area in the tail to the right of $t = 1.10$ is between .20 and .10. When we double these amounts, we see that the p-value must be between .40 and .20. With a level of significance of $\alpha = .05$, we now know that the p-value is greater than α. Therefore, H_0 cannot be rejected. Sufficient evidence is not available to conclude that Holiday should change its production plan for the coming season.

In the Using Excel subsection which follows, we show how to compute the exact p-value for this hypothesis test using Excel. The p-value obtained is .2811. With a level of significance of $\alpha = .05$, we cannot reject H_0 because $.2811 > .05$.

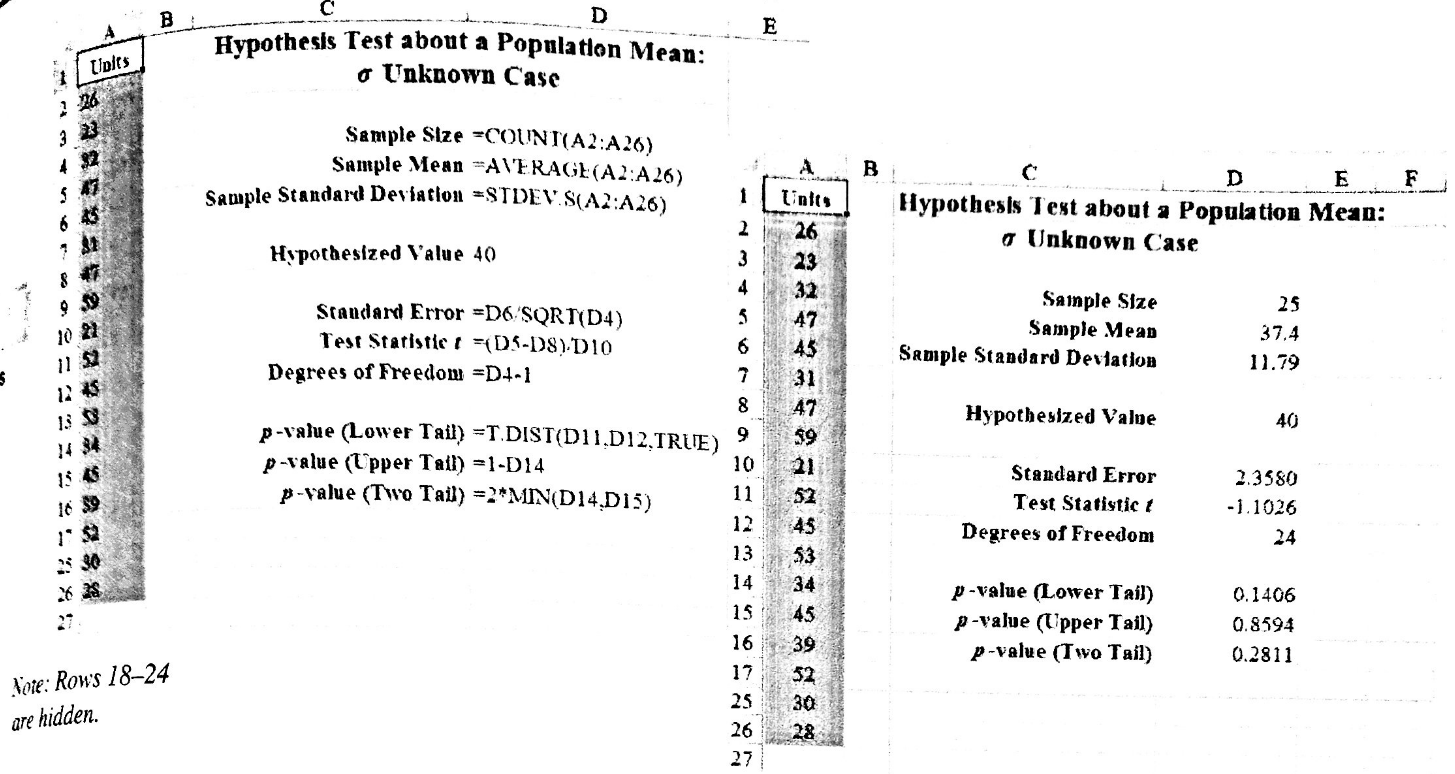
The test statistic can also be compared to the critical value to make the two-tailed hypothesis testing decision. With $\alpha = .05$ and the t distribution with 24 degrees of freedom, $-t_{.025} = -2.064$ and $t_{.025} = 2.064$ are the critical values for the two-tailed test. The rejection rule using the test statistic is

$$\text{Reject } H_0 \text{ if } t \leq -2.064 \text{ or if } t \geq 2.064$$

Based on the test statistic $t = -1.10$, H_0 cannot be rejected. This result indicates that Holiday should continue its production planning for the coming season based on the expectation that $\mu = 40$.

Using Excel

Excel can be used to conduct one-tailed and two-tailed hypothesis tests about a population mean for the σ unknown case. The approach is similar to the procedure used in the σ known case. The sample data and the test statistic (t) are used to compute three p-values: p-value (Lower Tail), p-value (Upper Tail), and p-value (Two Tail). The user can then choose α and



	A	B	C	D
1	Units		**Hypothesis Test about a Population Mean: σ Unknown Case**	
2	26			
3	23			
4	32		**Sample Size**	=COUNT(A2:A26)
5	47		**Sample Mean**	=AVERAGE(A2:A26)
6	45		**Sample Standard Deviation**	=STDEV.S(A2:A26)
7	31			
8	47		**Hypothesized Value**	40
9	59			
10	21		**Standard Error**	=D6/SQRT(D4)
11	52		**Test Statistic *t***	=(D5-D8)/D10
12	45		**Degrees of Freedom**	=D4-1
13	53			
14	34		***p*-value (Lower Tail)**	=T.DIST(D11,D12,TRUE)
15	45		***p*-value (Upper Tail)**	=1-D14
16	39		***p*-value (Two Tail)**	=2*MIN(D14,D15)
17	52			
25	30			
26	28			
27				

Note: Rows 18–24 are hidden.

	A	B	C	D	E	F
1	Units		**Hypothesis Test about a Population Mean: σ Unknown Case**			
2	26					
3	23					
4	32		**Sample Size**	25		
5	47		**Sample Mean**	37.4		
6	45		**Sample Standard Deviation**	11.79		
7	31					
8	47		**Hypothesized Value**	40		
9	59					
10	21		**Standard Error**	2.3580		
11	52		**Test Statistic *t***	-1.1026		
12	45		**Degrees of Freedom**	24		
13	53					
14	34		***p*-value (Lower Tail)**	0.1406		
15	45		***p*-value (Upper Tail)**	0.8594		
16	39		***p*-value (Two Tail)**	0.2811		
17	52					
25	30					
26	28					
27						

draw a conclusion using whichever *p*-value is appropriate for the type of hypothesis test being conducted.

Let's start by showing how to use Excel's T.DIST function to compute a lower tail *p*-value. The T.DIST function has three inputs; its general form is as follows:

T.DIST(test statistic, degrees of freedom, cumulative)

For the first input, we enter the value of the test statistic, for the second input we enter the number of degrees of freedom. For the third input, we enter TRUE if we want a cumulative probability and FALSE if we want the height of the curve. When we want to compute a lower tail *p*-value, we enter TRUE.

Once the lower tail *p*-value has been computed, it is easy to compute the upper tail and the two-tailed *p*-values. The upper tail *p*-value is just 1 minus the lower tail *p*-value. And the two-tailed *p*-value is given by two times the smaller of the lower and upper tail *p*-values.

Let us now construct an Excel worksheet to conduct the two-tailed hypothesis test for the Holiday Toys study. Refer to Figure 9.8 as we describe the tasks involved. The formula worksheet is in the background; the value worksheet is in the foreground.

Enter/Access Data: Open the DATAfile named *Orders*. A label and the order quantity data for the sample of 25 retailers are entered into cells A1:A26.

Enter Functions and Formulas: The sample size, sample mean, and sample standard deviation are computed in cells D4:D6 using Excel's COUNT, AVERAGE, and STDEV.S functions, respectively. The value worksheet shows that the sample size is 25, the sample mean is 37.4, and the sample standard deviation is 11.79. The hypothesized value of the population mean (40) is entered into cell D8.

Using the sample standard deviation as an estimate of the population standard deviation, an estimate of the standard error is obtained in cell D10 by dividing the sample standard deviation in cell D6 by the square root of the sample size in cell D4. The formula =(D5-D8)/D10 entered into cell D11 computes the test statistic $t(-1.1026)$. The degrees of freedom are computed in cell D12 as the sample size in cell D4 minus 1.

To compute the p-value for a lower tail test, we enter the following formula into cell D14:

=T.DIST(D11,D12,TRUE)

The p-value for an upper tail test is then computed in cell D15 as 1 minus the p-value for the lower tail test. Finally, the p-value for a two-tailed test is computed in cell D16 as two times the minimum of the two one-tailed p-values. The value worksheet shows that the three p-values are p-value (Lower Tail) $= 0.1406$, p-value (Upper Tail) $= 0.8594$, and p-value (Two Tail) $= 0.2811$.

The development of the worksheet is now complete. For the two-tailed Holiday Toys problem we cannot reject H_0: $\mu = 40$ using $\alpha = .05$ because the p-value (Two Tail) $= 0.2811$ is greater than α. This result indicates that Holiday should continue its production planning for the coming season based on the expectation that $\mu = 40$. The worksheet in Figure 9.8 can also be used for any one-tailed hypothesis test involving the t distribution. If a lower tail test is required, compare the p-value (Lower Tail) with α to make the rejection decision. If an upper tail test is required, compare the p-value (Upper Tail) with α to make the rejection decision.

A template for other problems The worksheet in Figure 9.8 can be used as a template for any hypothesis tests about a population mean for the σ unknown case. Just enter the appropriate data in column A, adjust the ranges for the formulas in cells D4:D6, and enter the hypothesized value in cell D8. The standard error, the test statistic, and the three p-values will then appear. Depending on the form of the hypothesis test (lower tail, upper tail, or two-tailed), we can then choose the appropriate p-value to make the rejection decision.

We can further simplify the use of Figure 9.8 as a template for other problems by eliminating the need to enter new data ranges in cells D4:D6. To do so we rewrite the cell formulas as follows:

Cell D4: =COUNT(A:A)
Cell D5: =AVERAGE(A:A)
Cell D6: =STDEV(A:A)

ATAfile named Orders
es a worksheet entitled
ate that uses the A:A
for entering the data

With the A:A method of specifying data ranges, Excel's COUNT function will count the number of numeric values in column A, Excel's AVERAGE function will compute the average of the numeric values in column A, and Excel's STDEV function will compute the standard deviation of the numeric values in Column A. Thus, to solve a new problem it is only necessary to enter the new data in column A and enter the hypothesized value of the population mean in cell D8.

Summary and Practical Advice

Table 9.3 provides a summary of the hypothesis testing procedures about a population mean for the σ unknown case. The key difference between these procedures and the ones for the σ known case is that s is used, instead of σ, in the computation of the test statistic. For this reason, the test statistic follows the t distribution.

The applicability of the hypothesis testing procedures of this section is dependent on the distribution of the population being sampled from and the sample size. When the population is normally distributed, the hypothesis tests described in this section provide exact results for any sample size. When the population is not normally distributed, the procedures are approximations. Nonetheless, we find that sample sizes of 30 or greater will provide good results in most cases. If the population is approximately normal, small sample sizes (e.g., $n < 15$) can provide acceptable results. If the population is highly skewed or contains outliers, sample sizes approaching 50 are recommended.

TABLE 9.3 SUMMARY OF HYPOTHESIS TESTS ABOUT A POPULATION MEAN: σ UNKNOWN CASE

	Lower Tail Test	Upper Tail Test	Two-Tailed Test
Hypotheses	$H_0: \mu \geq \mu_0$ $H_a: \mu < \mu_0$	$H_0: \mu \leq \mu_0$ $H_a: \mu > \mu_0$	$H_0: \mu = \mu_0$ $H_a: \mu \neq \mu_0$
Test Statistic	$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$	$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$	$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$
Rejection Rule: p-Value Approach	Reject H_0 if p-value $\leq \alpha$	Reject H_0 if p-value $\leq \alpha$	Reject H_0 if p-value $\leq \alpha$
Rejection Rule: Critical Value Approach	Reject H_0 if $t \leq -t_\alpha$	Reject H_0 if $t \geq t_\alpha$	Reject H_0 if $t \leq -t_{\alpha/2}$ or if $t \geq t_{\alpha/2}$

NOTE AND COMMENT

When testing a hypothesis of the population mean with a sample size that is at least 5% of the population size (that is, $n/N \geq .05$), the finite population correction factor should be used when calculating the standard error of the sampling distribution of $\bar{x}$ when σ is unknown, i.e., $s_{\bar{x}} = \sqrt{\dfrac{N-n}{N-1}}\left(\dfrac{s}{\sqrt{n}}\right)$.

Exercises

Methods

23. Consider the following hypothesis test:

$$H_0: \mu \leq 12$$
$$H_a: \mu > 12$$

A sample of 25 provided a sample mean $\bar{x} = 14$ and a sample standard deviation $s = 4.32$.

a. Compute the value of the test statistic.
b. Use the t distribution table (Table 2 in Appendix B) to compute a range for the p-value.
c. At $\alpha = .05$, what is your conclusion?
d. What is the rejection rule using the critical value? What is your conclusion?

SELF*test* 24. Consider the following hypothesis test:

$$H_0: \mu = 18$$
$$H_a: \mu \neq 18$$

A sample of 48 provided a sample mean $\bar{x} = 17$ and a sample standard deviation $s = 4.5$.

a. Compute the value of the test statistic.
b. Use the t distribution table (Table 2 in Appendix B) to compute a range for the p-value.
c. At $\alpha = .05$, what is your conclusion?
d. What is the rejection rule using the critical value? What is your conclusion?