CHAPTER 9

Hypothesis Tests

CONTENTS

STATISTICS *in* PRACTICE

JOHN MORRELL & COMPANY*
CINCINNATI, OHIO

John Morrell & Company, which began in England in 1827, is considered the oldest continuously operating meat manufacturer in the United States. It is a wholly owned and independently managed subsidiary of Smithfield Foods, Smithfield, Virginia. John Morrell & Company offers an extensive product line of processed meats and fresh pork to consumers under 13 regional brands including John Morrell, E-Z-Cut, Tobin's First Prize, Dinner Bell, Hunter, Kretschmar, Rath, Rodeo, Shenson, Farmers Hickory Brand, Iowa Quality, and Peyton's. Each regional brand enjoys high brand recognition and loyalty among consumers.

Fully-cooked entrees allow consumers to heat and serve in the same microwaveable tray. © Courtesy of John Morrell's Convenient Cuisine products.

Market research at Morrell provides management with up-to-date information on the company's various products and how the products compare with competing brands of similar products. A recent study compared a Beef Pot Roast made by Morrell to similar beef products from two major competitors. In the three-product comparison test, a sample of consumers was used to indicate how the products rated in terms of taste, appearance, aroma, and overall preference.

One research question concerned whether the Beef Pot Roast made by Morrell was the preferred choice of more than 50% of the consumer population. Letting p indicate the population proportion preferring Morrell's product, the hypothesis test for the research question is as follows:

$$H_0: p \leq .50$$
$$H_a: p > .50$$

The null hypothesis H_0 indicates the preference for Morrell's product is less than or equal to 50%. If the sample data support rejecting H_0 in favor of the alternative hypothesis H_a, Morrell will draw the research conclusion that in a three-product comparison, their Beef Pot Roast is preferred by more than 50% of the consumer population.

In an independent taste test study using a sample of 224 consumers in Cincinnati, Milwaukee, and Los Angeles, 150 consumers selected the Beef Pot Roast made by Morrell as the preferred product. Using statistical hypothesis testing procedures, the null hypothesis H_0 was rejected. The study provided statistical evidence supporting H_a and the conclusion that the Morrell product is preferred by more than 50% of the consumer population.

The point estimate of the population proportion was $\bar{p} = 150/224 = .67$. Thus, the sample data provided support for a food magazine advertisement showing that in a three-product taste comparison, Beef Pot Roast made by Morrell was "preferred 2 to 1 over the competition."

In this chapter we will discuss how to formulate hypotheses and how to conduct tests like the one used by Morrell. Through the analysis of sample data, we will be able to determine whether a hypothesis should or should not be rejected.

*The authors are indebted to Marty Butler, vice president of Marketing, John Morrell, for providing this Statistics in Practice.

In Chapters 7 and 8 we showed how a sample could be used to develop point and interval estimates of population parameters. In this chapter we continue the discussion of statistical inference by showing how hypothesis testing can be used to determine whether a statement about the value of a population parameter should or should not be rejected.

In hypothesis testing we begin by making a tentative assumption about a population parameter. This tentative assumption is called the **null hypothesis** and is denoted by H_0. 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

Learning to formulate hypotheses correctly will take practice. Expect some initial confusion over the proper choice for H_0 and H_a. The examples in this section show a variety of forms for H_0 and H_a depending upon the application.

In some applications it may not be 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. Guidelines for establishing the null and alternative hypotheses will be given for three types of situations in which hypothesis testing procedures are commonly employed.

Testing Research Hypotheses

Consider a particular automobile model that currently attains an average fuel efficiency of 24 miles per gallon. A product research group developed a new fuel injection system specifically designed to increase the miles-per-gallon rating. To evaluate the new system, several will be manufactured, installed in automobiles, and subjected to research-controlled driving tests. Here the product research group is looking for evidence to conclude that the new system *increases* the mean miles-per-gallon rating. In this case, the research hypothesis is that the new fuel injection system will provide a mean miles-per-gallon rating exceeding 24; that is, $\mu > 24$. As a general guideline, a research hypothesis should be stated as the *alternative hypothesis.* Hence, the appropriate null and alternative hypotheses for the study are:

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

→ Rej. H_0.

The conclusion that the research hypothesis is true is made if the sample data contradict the null hypothesis.

If the sample results indicate that H_0 cannot be rejected, researchers cannot conclude that the new fuel injection system is better. Perhaps more research and subsequent testing should be conducted. However, if the sample results indicate that H_0 can be rejected, researchers can make the inference that $H_a\colon \mu > 24$ is true. With this conclusion, the researchers gain the statistical support necessary to state that the new system increases the mean number of miles per gallon. Production with the new system should be considered.

In research studies such as these, the null and alternative hypotheses should be formulated so that the rejection of H_0 supports the research conclusion. The research hypothesis therefore should be expressed as the alternative hypothesis.

Testing the Validity of a Claim

As an illustration of testing the validity of a claim, consider the situation of a manufacturer of soft drinks who states that two-liter soft drink containers are filled with an average of at least 67.6 fluid ounces. A sample of two-liter containers will be selected, and the contents will be measured to test the manufacturer's claim. In this type of hypothesis testing situation, we generally assume that the manufacturer's claim is true unless the sample evidence is contradictory. Using this approach for the soft-drink example, we would state the null and alternative hypotheses as follows.

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

A manufacturer's claim is usually given the benefit of the doubt and stated as the null hypothesis. The conclusion that the claim is false can be made if the null hypothesis is rejected.

If the sample results indicate H_0 cannot be rejected, the manufacturer's claim will not be challenged. However, if the sample results indicate H_0 can be rejected, the inference will be made that $H_a\colon \mu < 67.6$ is true. With this conclusion, statistical evidence indicates that the manufacturer's claim is incorrect and that the soft-drink containers are being filled with a mean less than the claimed 67.6 ounces. Appropriate action against the manufacturer may be considered.

In any situation that involves testing the validity of a claim, the null hypothesis is generally based on the assumption that the claim is true. The alternative hypothesis is then formulated so that rejection of H_0 will provide statistical evidence that the stated assumption is incorrect. Action to correct the claim should be considered whenever H_0 is rejected.

Testing in Decision-Making Situations

This type of hypothesis test is employed in the quality control procedure called lot-acceptance sampling.

In testing research hypotheses or testing the validity of a claim, action is taken if H_0 is rejected. In some instances, however, action must be taken both when H_0 cannot be rejected and when H_0 can be rejected. In general, this type of situation occurs when a decision maker must choose between two courses of action, one associated with the null hypothesis and another associated with the alternative hypothesis. For example, on the basis of a sample of parts from a shipment just received, a quality control inspector must decide whether to accept the shipment or to return the shipment to the supplier because it does not meet specifications. Assume that specifications for a particular part require a mean length of two inches per part. If the mean length is greater or less than the two-inch standard, the parts will cause quality problems in the assembly operation. In this case, the null and alternative hypotheses would be formulated as follows.

$$H_0\colon \mu = 2$$
$$H_a\colon \mu \ne 2$$

If the sample results indicate H_0 cannot be rejected, the quality control inspector will have no reason to doubt that the shipment meets specifications, and the shipment will be accepted. However, if the sample results indicate H_0 should be rejected, the conclusion will be that the parts do not meet specifications. In this case, the quality control inspector will have sufficient evidence to return the shipment to the supplier. Thus, we see that for these types of situations, action is taken both when H_0 cannot be rejected and when H_0 can be rejected.

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 hypotheses H_0 and H_a are shown here. Note that the equality always appears in the null hypothesis H_0.

$$H_0\colon \mu \ge \mu_0 \qquad H_0\colon \mu \le \mu_0 \qquad H_0\colon \mu = \mu_0$$
$$H_a\colon \mu < \mu_0 \qquad H_a\colon \mu > \mu_0 \qquad H_a\colon \mu \ne \mu_0$$

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 $\ge$, $\le$, 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.

Exercises

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 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.

   $$H_0: \mu \geq 600 \qquad H_0: \mu \leq 600 \qquad H_0: \mu = 600$$
   $$H_a: \mu < 600 \qquad H_a: \mu > 600 \qquad H_a: \mu \neq 600$$

   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 research 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

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

نقبل H_0 وهي خطأ

نرفض H_0 وهي صح

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.

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

The alternative hypothesis, $H_a: \mu > 24$, 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 ($\mu \leq 24$) when in fact the new system improves miles-per-gallon performance.

For the miles-per-gallon rating hypothesis test, the null hypothesis is $H_0: \mu \leq 24$. 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: $\mu \leq 24$ when $\mu = 24$. Because of the importance of this concept, we now restate the definition of level of significance.

α 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.

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.

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," unless we have specifically controlled for the Type II error.

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.* If proper controls have been established for this error, action based on the "accept H_0" conclusion can be appropriate.

NOTES AND COMMENTS

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 their drug approval process. With a Type I error, the FDA fails to approve a drug that is safe and effective. A Type II error means the FDA approves a drug that has unanticipated dangerous side effects. Regardless of the decision made, the possibility of making a costly error cannot be eliminated.

Exercises

SELF test

5. Nielsen reported that young men in the United States watch 56.2 minutes of prime-time TV daily (*The Wall Street Journal Europe,* November 18, 2003). A researcher believes that young men in Germany spend more time watching prime-time TV. A sample of German young men will be selected by the researcher and the time they spend watching TV in one day will be recorded. The sample results will be used to test the following null and alternative hypotheses.

$$H_0: \mu \le 56.2$$
$$H_a: \mu > 56.2$$

   a. What is the Type I error in this situation? What are the consequences of making this error?
   b. 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 claims 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.

*See, for example, D. R. Anderson, D. J. Sweeney, and T. A. Williams, *Statistics for Business and Economics,* 10th ed. (Cincinnati: South-Western, 2008).

   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 is 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 Tests

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

Lower Tail Test	Upper Tail Test
$H_0: \mu \ge \mu_0$	$H_0: \mu \le \mu_0$
$H_a: \mu < \mu_0$	$H_a: \mu > \mu_0$

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 outcome 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 outcome establishes the alternative hypothesis. With μ denoting the population mean filling weight, the null and alternative hypotheses are as follows:

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

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 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 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 $\mu = \mu_0 = 3$.* Note that the standard error of $\bar{x}$ is given by $\sigma_{\bar{x}} = \sigma/\sqrt{n} = .18/\sqrt{36} = .03$.

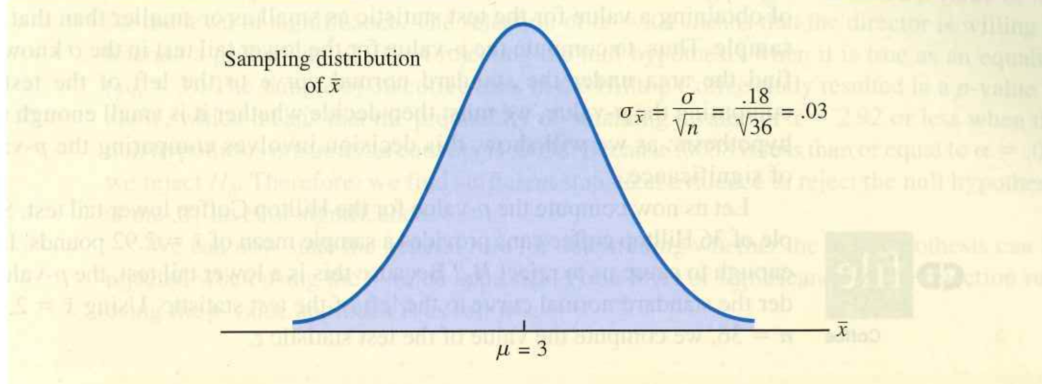
The standard error of $\bar{x}$ is the standard deviation of the sampling distribution of $\bar{x}$.

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

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

*In constructing sampling distributions for hypothesis tests, it is assumed that H_0 is satisfied as an equality.

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



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**.

A small p-value indicates the value of the test statistic is unusual given the assumption 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 the area under the standard normal curve to the left 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.

CD file Coffee

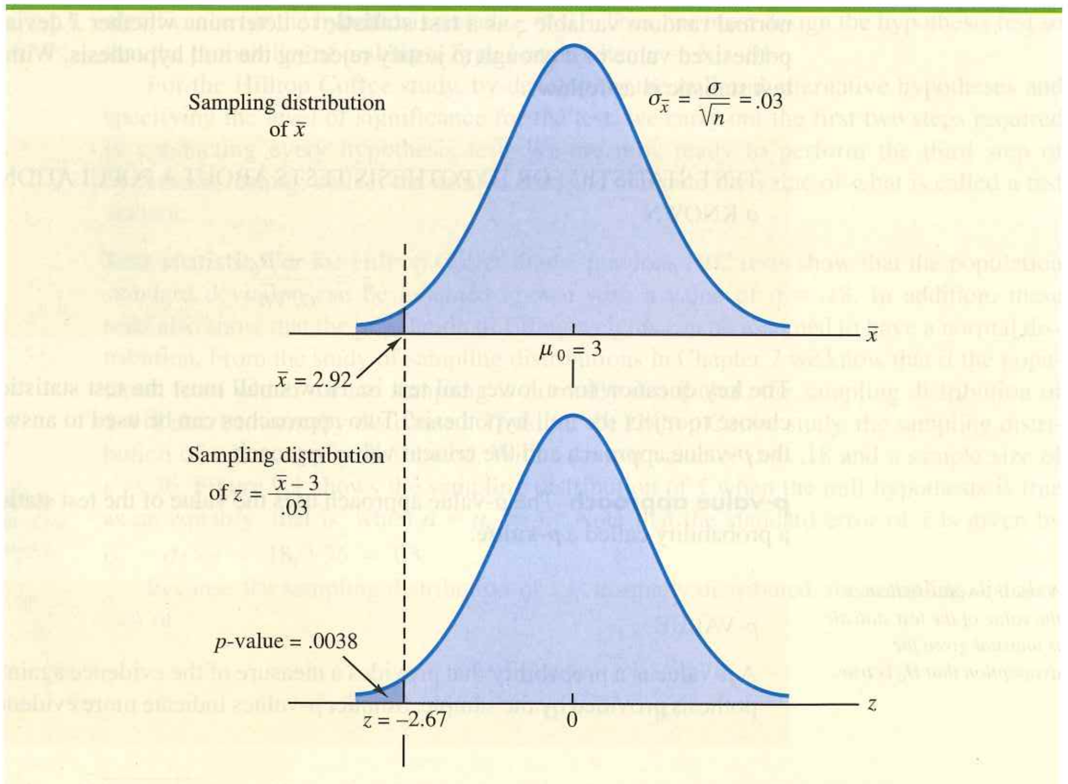
Let us now compute the p-value for the Hilltop Coffee lower tail test. Suppose the sample of 36 Hilltop coffee cans provides a sample mean of $\bar{x} = 2.92$ pounds. Is $\bar{x} = 2.92$ small enough to cause us to reject H_0? Because this is a lower tail test, the p-value is the area under the standard normal curve to the left 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 the test statistic z is less than or equal to -2.67 (the area under the standard normal curve to the left 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

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



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.

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

Reject H_0 if p-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.

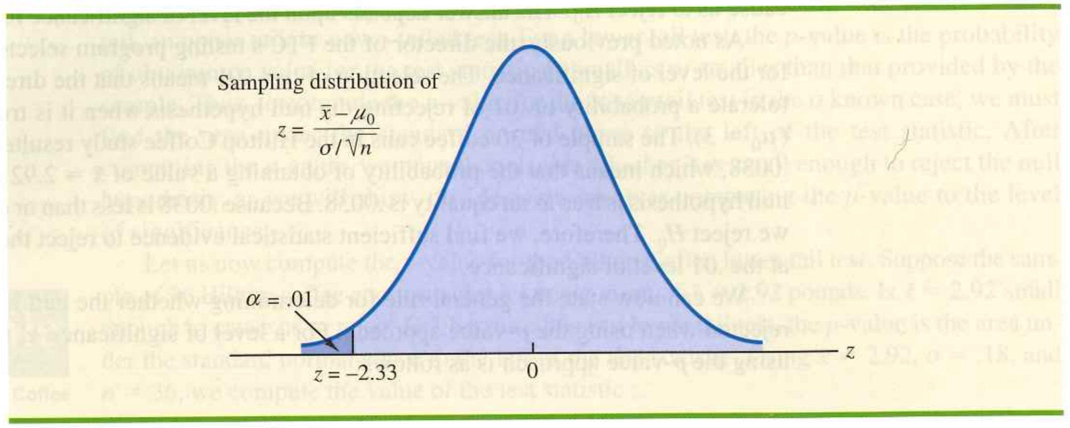
Critical value approach The critical value approach requires that we first determine a value for the test statistic called the **critical value**. For a lower tail test, the critical value serves as a benchmark for determining whether the value of the test statistic is small enough to reject the null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

Reject H_0 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.

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



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

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 find the area under the standard normal curve to the right 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 $z \geq z_\alpha$.

The computation of p-values can be confusing. 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 z.
2. **Lower tail test:** Compute the area under the standard normal curve to the left of the test statistic.
3. **Upper tail test:** Compute the area under the standard normal curve to the right of the test statistic.

Two-Tailed Test

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

$$H_0\text{: } \mu = \mu_0$$
$$H_a\text{: } \mu \neq \mu_0$$

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:

$$H_0\text{: } \mu = 295$$
$$H_a\text{: } \mu \neq 295$$

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

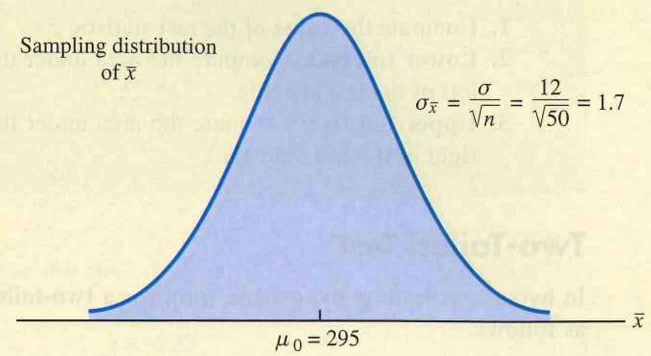


Figure 9.4 shows the sampling distribution of $\bar{x}$ for the MaxFlight hypothesis test with a hypothesized population mean of $\mu_0 = 295$.

CD file 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.

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

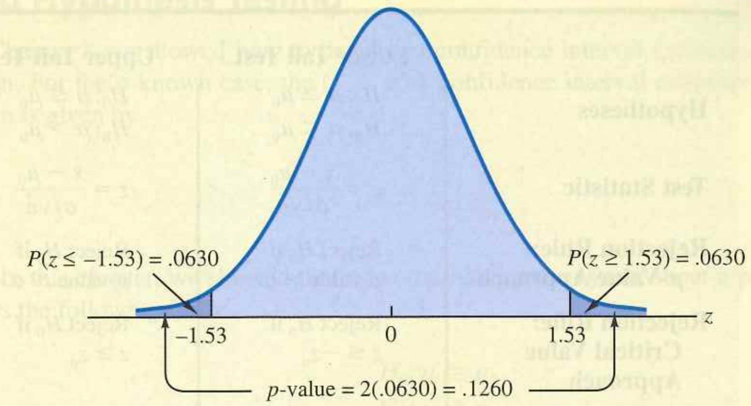
$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{297.6 - 295}{12/\sqrt{50}} = 1.53$$

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 the area under the standard normal curve to the right of $z = 1.53$ and doubling it. The table for the standard normal distribution shows that the area to the left of $z = 1.53$ is .9370. Thus, the area under the standard normal curve to the right of the test statistic $z = 1.53$ is $1.0000 - .9370 = .0630$. Doubling this probability, we find 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.

The computation of the p-value for a two-tailed test may seem a bit confusing as compared to the computation of the p-value for a one-tailed test. But it can be simplified by following three steps.

FIGURE 9.5 p-VALUE FOR THE MAXFLIGHT HYPOTHESIS TEST



COMPUTATION OF p-VALUE FOR A TWO-TAILED TEST

1. Compute the value of the test statistic z.
2. If the value of the test statistic is in the upper tail ($z > 0$), find the area under the standard normal curve to the right of z. If the value of the test statistic is in the lower tail ($z < 0$), find the area under the standard normal curve to the left of z.
3. Double the tail area, or probability, obtained in 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 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 beyond 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.

FIGURE 9.6 CRITICAL VALUES FOR THE MAXFLIGHT HYPOTHESIS TEST

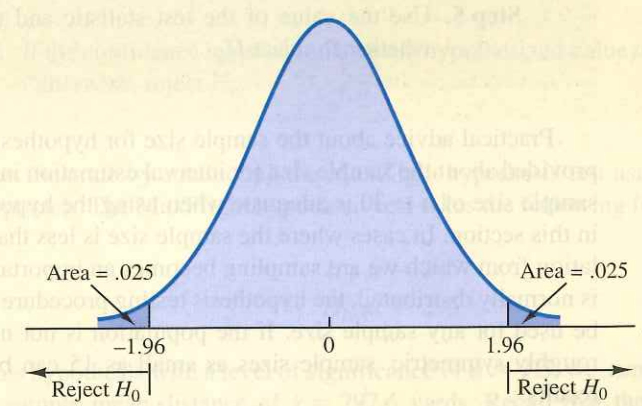


TABLE 9.2 SUMMARY OF HYPOTHESIS TESTS ABOUT A POPULATION MEAN: σ KNOWN CASE

	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	$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$	$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$	$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$
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 $\leq \alpha$.

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.

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:

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

where μ_0 is the hypothesized value for the population mean.

Suppose that we follow the procedure described in Chapter 8 for constructing a $(1 - \alpha)\%$ confidence interval for the population mean. We know that $(1 - \alpha)\%$ of the confidence intervals generated will contain the population mean and $\alpha\%$ 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 $(1 - \alpha)\%$ 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

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

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 H_0.

For a two-tailed hypothesis test, the null hypothesis can be rejected if the confidence interval does not include μ_0.

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

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

To test this hypothesis 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

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 H_a is true.
- Between .01 and .05—Strong evidence to conclude H_a is true.
- Between .05 and .10—Weak evidence to conclude H_a is true.
- Greater than .10—Insufficient evidence to conclude H_a is true.

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?

SELF test

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?

SELF test

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:

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

A sample of 75 is used and the population standard deviation is 10. Compute the p-value and state your conclusion for each of the following sample results. Use $\alpha = .01$.

a. $\bar{x} = 23$
b. $\bar{x} = 25.1$
c. $\bar{x} = 20$

Applications

SELF test

15. Individuals filing federal income tax returns prior to March 31 received an average refund of $1056. Consider the population of "last-minute" filers who mail their tax return during the last five days of the income tax period (typically April 10 to April 15).
   a. A researcher suggests that a reason individuals wait until the last five days is that on average these individuals receive lower refunds than do early filers. Develop appropriate hypotheses such that rejection of H_0 will support the researcher's contention.
   b. For a sample of 400 individuals who filed a tax return between April 10 and 15, the sample mean refund was $910. Based on prior experience a population standard deviation of $\sigma = \$1600$ may be assumed. What is the p-value?
   c. At $\alpha = .05$, what is your conclusion?
   d. Repeat the preceding hypothesis test using the critical value approach.

CD file RentalRates

16. Reis, Inc., a New York real estate research firm, tracks the cost of apartment rentals in the United States. In mid-2002, the nationwide mean apartment rental rate was $895 per month (*The Wall Street Journal,* July 8, 2002). Assume that, based on the historical quarterly surveys, a population standard deviation of $\sigma = \$225$ is reasonable. In a current study of apartment rental rates, a sample of 180 apartments nationwide provided the apartment rental rates shown in the CD file named RentalRates. Do the sample data enable Reis to conclude that the population mean apartment rental rate now exceeds the level reported in 2002?
   a. State the null and alternative hypotheses.
   b. What is the p-value?
   c. At $\alpha = .01$, what is your conclusion?
   d. What would you recommend Reis consider doing at this time?
17. Wall Street securities firms paid out record year-end bonuses of $125,500 per employee for 2005 (*Fortune,* February 6, 2006). Suppose we would like to take a sample of employees at the Jones & Ryan securities firm to see whether the mean year-end bonus is different from the reported mean of $125,500 for the population.
   a. State the null and alternative hypotheses you would use to test whether the year-end bonuses paid by Jones & Ryan were different from the population mean.
   b. Suppose a sample of 40 Jones & Ryan employees showed a sample mean year-end bonus of $118,000. Assume a population standard deviation of $\sigma = \$30{,}000$ and compute the p-value.
   c. With $\alpha = .05$ as the level of significance, what is your conclusion?
   d. Repeat the preceding hypothesis test using the critical value approach.
18. The average annual total return for U.S. Diversified Equity mutual funds from 1999 to 2003 was 4.1% (*BusinessWeek,* January 26, 2004). A researcher would like to conduct a hypothesis test to see whether the returns for mid-cap growth funds over the same period are significantly different from the average for U.S. Diversified Equity funds.
   a. Formulate the hypotheses that can be used to determine whether the mean annual return for mid-cap growth funds differ from the mean for U.S. Diversified Equity funds.
   b. A sample of 40 mid-cap growth funds provides a mean return of $\bar{x} = 3.4\%$. Assume the population standard deviation for mid-cap growth funds is known from previous studies to be $\sigma = 2\%$. Use the sample results to compute the test statistic and p-value for the hypothesis test.
   c. At $\alpha = .05$, what is your conclusion?

19. In 2001, the U.S. Department of Labor reported the average hourly earnings for U.S. production workers to be $14.32 per hour (*The World Almanac,* 2003). A sample of 75 production workers during 2003 showed a sample mean of $14.68 per hour. Assuming the population standard deviation $\sigma = \$1.45$, can we conclude that an increase occurred in the mean hourly earnings since 2001? Use $\alpha = .05$.
20. For the United States, the mean monthly Internet bill is $32.79 per household (CNBC, January 18, 2006). A sample of 50 households in a southern state showed a sample mean of $30.63. Use a population standard deviation of $\sigma = \$5.60$.
   a. Formulate hypotheses for a test to determine whether the sample data support the conclusion that the mean monthly Internet bill in the southern state is less than the national mean of $32.79.
   b. What is the value of the test statistic?
   c. What is the p-value?
   d. At $\alpha = .01$, what is your conclusion?

CD file Fowle

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 CD file 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 8.5 minutes. Assume a population standard deviation $\sigma = 3.2$ 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?

9.4 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 Tests

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?

CD file
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:

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

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 the area under the curve of the t distribution to the right of $t = 1.84$.

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.

Appendix F shows how to compute p-values using Excel or Minitab.

Because it is cumbersome to use a t table to compute p-values, and only approximate values are obtained, we show how to compute the exact p-value using Excel or Minitab. The directions can be found in Appendix F at the end of this text. Using Excel or Minitab with $t = 1.84$ provides the upper tail p-value of .0354 for the Heathrow Airport hypothesis test. With $.0354 < .05$, we reject the null hypothesis and conclude that 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:

$$H_0: \mu = 40$$
$$H_a: \mu \ne 40$$

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.

CD file
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 for the t distribution to the left of $t = -1.10$. 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 only contains positive t values. Because the t distribution is symmetric, however, the area under the curve to the right of $t = 1.10$ is the same as the area under the curve to the left of $t = -1.10$. 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.

Appendix F shows how the p-value for this test can be computed using Excel or Minitab. The p-value obtained is .2822. With a level of significance of $\alpha = .05$, we cannot reject H_0 because $.2822 > .05$.

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 $\mu = 40$.

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	$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$	$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$	$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$
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 $t \geq t_{\alpha/2}$

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?

25. Consider the following hypothesis test:

$$H_0: \mu \geq 45$$
$$H_a: \mu < 45$$

A sample of 36 is used. Identify the p-value and state your conclusion for each of the following sample results. Use $\alpha = .01$.

a. $\bar{x} = 44$ and $s = 5.2$
b. $\bar{x} = 43$ and $s = 4.6$
c. $\bar{x} = 46$ and $s = 5.0$

26. Consider the following hypothesis test:

$$H_0: \mu = 100$$
$$H_a: \mu \neq 100$$

A sample of 65 is used. Identify the p-value and state your conclusion for each of the following sample results. Use $\alpha = .05$.

a. $\bar{x} = 103$ and $s = 11.5$
b. $\bar{x} = 96.5$ and $s = 11.0$
c. $\bar{x} = 102$ and $s = 10.5$

Applications

SELF test

27. The Employment and Training Administration reported the U.S. mean unemployment insurance benefit of \$238 per week (*The World Almanac,* 2003). A researcher in the state of Virginia anticipated that sample data would show evidence that the mean weekly unemployment insurance benefit in Virginia was below the national level.
a. Develop appropriate hypotheses such that rejection of H_0 will support the researcher's contention.
b. For a sample of 100 individuals, the sample mean weekly unemployment insurance benefit was \$231 with a sample standard deviation of \$80. What is the p-value?
c. At $\alpha = .05$, what is your conclusion?
d. Repeat the preceding hypothesis test using the critical value approach.

28. A shareholders' group, in lodging a protest, claimed that the mean tenure for a chief executive officer (CEO) was at least nine years. A survey of companies reported in *The Wall Street Journal* found a sample mean tenure of $\bar{x} = 7.27$ years for CEOs with a standard deviation of $s = 6.38$ years (*The Wall Street Journal,* January 2, 2007).
a. Formulate hypotheses that can be used to test the validity of the claim made by the shareholders' group.
b. Assume 85 companies were included in the sample. What is the p-value for your hypothesis test?
c. At $\alpha = .01$, what is your conclusion?

CD file
Diamonds

29. The cost of a one-carat VS2 clarity, H color diamond from Diamond Source USA is \$5600 (http://www.diasource.com, March 2003). A midwestern jeweler makes calls to contacts in the diamond district of New York City to see whether the mean price of diamonds there differs from \$5600.
a. Formulate hypotheses that can be used to determine whether the mean price in New York City differs from \$5600.
b. A sample of 25 New York City contacts provided the prices shown in the CD file named Diamonds. What is the p-value?
c. At $\alpha = .05$, can the null hypothesis be rejected? What is your conclusion?
d. Repeat the preceding hypothesis test using the critical value approach.

30. AOL Time Warner Inc.'s CNN has been the longtime ratings leader of cable television news. Nielsen Media Research indicated that the mean CNN viewing audience was 600,000 viewers per day during 2002 (*The Wall Street Journal,* March 10, 2003). Assume that for a sample of 40 days during the first half of 2003, the daily audience was 612,000 viewers with a sample standard deviation of 65,000 viewers.
a. What are the hypotheses if CNN management would like information on any change in the CNN viewing audience?
b. What is the p-value?
c. Select your own level of significance. What is your conclusion?
d. What recommendation would you make to CNN management in this application?

31. Raftelis Financial Consulting reported that the mean quarterly water bill in the United States is \$47.50 (*U.S. News & World Report,* August 12, 2002). Some water systems are operated by public utilities, whereas other water systems are operated by private companies. An economist pointed out that privatization does not equal competition and that monopoly powers provided to public utilities are now being transferred to private companies. The concern is that consumers end up paying higher-than-average rates for water provided by private companies. The water system for Atlanta, Georgia, is provided by a private company. A sample of 64 Atlanta consumers showed a mean quarterly water bill of \$51 with a sample standard deviation of \$12. At $\alpha = .05$, does the Atlanta sample support the conclusion that above-average rates exist for this private water system? What is your conclusion?

CD file
UsedCars

32. According to the National Automobile Dealers Association, the mean price for used cars is \$10,192. A manager of a Kansas City used car dealership reviewed a sample of 50 recent used car sales at the dealership in an attempt to determine whether the population mean price for used cars at this particular dealership differed from the national mean. The prices for the sample of 50 cars are shown in the CD file named UsedCars.
a. Formulate the hypotheses that can be used to determine whether a difference exists in the mean price for used cars at the dealership.
b. What is the p-value?
c. At $\alpha = .05$, what is your conclusion?

33. Annual per capita consumption of milk is 21.6 gallons (*Statistical Abstract of the United States: 2006*). Being from the Midwest, you believe milk consumption is higher there and wish to support your opinion. A sample of 16 individuals from the midwestern town of Webster City showed a sample mean annual consumption of 24.1 gallons with a standard deviation of $s = 4.8$.
a. Develop a hypothesis test that can be used to determine whether the mean annual consumption in Webster City is higher than the national mean.
b. What is a point estimate of the difference between mean annual consumption in Webster City and the national mean?
c. At $\alpha = .05$, test for a significant difference. What is your conclusion?

34. Joan's Nursery specializes in custom-designed landscaping for residential areas. The estimated labor cost associated with a particular landscaping proposal is based on the number of plantings of trees, shrubs, and so on to be used for the project. For cost-estimating purposes, managers use two hours of labor time for the planting of a medium-sized tree. Actual times from a sample of 10 plantings during the past month follow (times in hours).

1.7	1.5	2.6	2.2	2.4	2.3	2.6	3.0	1.4	2.3

With a .05 level of significance, test to see whether the mean tree-planting time differs from two hours.
a. State the null and alternative hypotheses.
b. Compute the sample mean.
c. Compute the sample standard deviation.
d. What is the p-value?
e. What is your conclusion?

9.5 Population Proportion

In this section we show how to conduct a hypothesis test about a population proportion p. Using p_0 to denote the hypothesized value for the population proportion, the three forms for a hypothesis test about a population proportion are as follows.

$$H_0: p \geq p_0 \qquad H_0: p \leq p_0 \qquad H_0: p = p_0$$
$$H_a: p < p_0 \qquad H_a: p > p_0 \qquad H_a: p \neq p_0$$

The first form is called a lower tail test, the second form is called an upper tail test, and the third form is called a two-tailed test.

Hypothesis tests about a population proportion are based on the difference between the sample proportion $\bar{p}$ and the hypothesized population proportion p_0. The methods used to conduct the hypothesis test are similar to those used for hypothesis tests about a population mean. The only difference is that we use the sample proportion and its standard error to compute the test statistic. The p-value approach or the critical value approach is then used to determine whether the null hypothesis should be rejected.

Let us consider an example involving a situation faced by Pine Creek golf course. Over the past year, 20% of the players at Pine Creek were women. In an effort to increase the proportion of women players, Pine Creek implemented a special promotion designed to attract women golfers. One month after the promotion was implemented, the course manager requested a statistical study to determine whether the proportion of women players at Pine Creek had increased. Because the objective of the study is to determine whether the proportion of women golfers increased, an upper tail test with $H_a: p > .20$ is appropriate. The null and alternative hypotheses for the Pine Creek hypothesis test are as follows:

$$H_0: p \le .20$$
$$H_a: p > .20$$

If H_0 can be rejected, the test results will give statistical support for the conclusion that the proportion of women golfers increased and the promotion was beneficial. The course manager specified that a level of significance of $\alpha = .05$ be used in carrying out this hypothesis test.

The next step of the hypothesis testing procedure is to select a sample and compute the value of an appropriate test statistic. To show how this step is done for the Pine Creek upper tail test, we begin with a general discussion of how to compute the value of the test statistic for any form of a hypothesis test about a population proportion. The sampling distribution of $\bar{p}$, the point estimator of the population parameter p, is the basis for developing the test statistic.

When the null hypothesis is true as an equality, the expected value of $\bar{p}$ equals the hypothesized value p_0; that is, $E(\bar{p}) = p_0$. The standard error of $\bar{p}$ is given by

$$\sigma_{\bar{p}} = \sqrt{\frac{p_0(1 - p_0)}{n}}$$

In Chapter 7 we said that if $np \ge 5$ and $n(1 - p) \ge 5$, the sampling distribution of $\bar{p}$ can be approximated by a normal distribution.* Under these conditions, which usually apply in practice, the quantity

$$z = \frac{\bar{p} - p_0}{\sigma_{\bar{p}}} \quad \textbf{(9.3)}$$

has a standard normal probability distribution. With $\sigma_{\bar{p}} = \sqrt{p_0(1 - p_0)/n}$, the standard normal random variable z is the test statistic used to conduct hypothesis tests about a population proportion.

*In most applications involving hypothesis tests of a population proportion, sample sizes are large enough to use the normal approximation. The exact sampling distribution of $\bar{p}$ is discrete with the probability for each value of $\bar{p}$ given by the binomial distribution. So hypothesis testing is a bit more complicated for small samples when the normal approximation cannot be used.

TEST STATISTIC FOR HYPOTHESIS TESTS ABOUT A POPULATION PROPORTION

$$z = \frac{\bar{p} - p_0}{\sqrt{\dfrac{p_0(1 - p_0)}{n}}} \quad \textbf{(9.4)}$$

CD file
WomenGolf

We can now compute the test statistic for the Pine Creek hypothesis test. Suppose a random sample of 400 players was selected, and that 100 of the players were women. The proportion of women golfers in the sample is

$$\bar{p} = \frac{100}{400} = .25$$

Using equation (9.4), the value of the test statistic is

$$z = \frac{\bar{p} - p_0}{\sqrt{\dfrac{p_0(1 - p_0)}{n}}} = \frac{.25 - .20}{\sqrt{\dfrac{.20(1 - .20)}{400}}} = \frac{.05}{.02} = 2.50$$

Because the Pine Creek hypothesis test is an upper tail test, the p-value is the probability that z is greater than or equal to $z = 2.50$; that is, it is the area under the standard normal curve to the right of $z = 2.50$. Using the standard normal probability table, we find that the area to the left of $z = 2.50$ is .9938. Thus, the p-value for the Pine Creek test is $1.0000 - .9938 = .0062$. Figure 9.7 shows this p-value calculation.

Recall that the course manager specified a level of significance of $\alpha = .05$. A p-value $= .0062 < .05$ gives sufficient statistical evidence to reject H_0 at the .05 level of significance. Thus, the test provides statistical support for the conclusion that the special promotion increased the proportion of women players at the Pine Creek golf course.

The decision whether to reject the null hypothesis can also be made using the critical value approach. The critical value corresponding to an area of .05 in the upper tail of a normal probability distribution is $z_{.05} = 1.645$. Thus, the rejection rule using the critical value approach is to reject H_0 if $z \ge 1.645$. Because $z = 2.50 > 1.645$, H_0 is rejected.

Again, we see that the p-value approach and the critical value approach lead to the same hypothesis testing conclusion, but the p-value approach provides more information. With a

FIGURE 9.7 CALCULATION OF THE p-VALUE FOR THE PINE CREEK HYPOTHESIS TEST

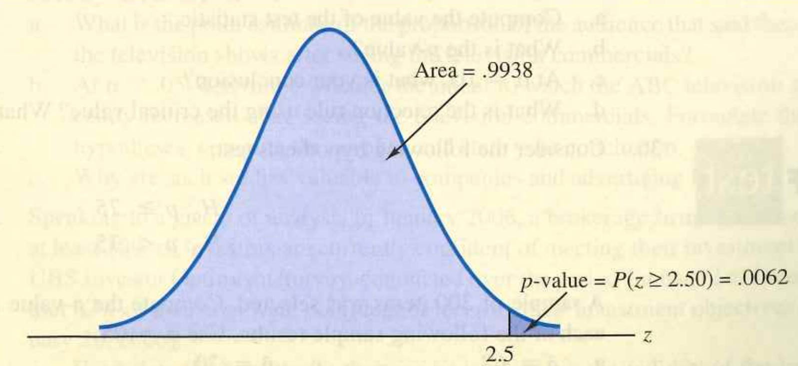


TABLE 9.4 SUMMARY OF HYPOTHESIS TESTS ABOUT A POPULATION PROPORTION

	Lower Tail Test	Upper Tail Test	Two-Tailed Test
Hypotheses	$H_0: p \geq p_0$ $H_a: p < p_0$	$H_0: p \leq p_0$ $H_a: p > p_0$	$H_0: p = p_0$ $H_a: p \neq p_0$
Test Statistic	$z = \dfrac{\bar{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$	$z = \dfrac{\bar{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$	$z = \dfrac{\bar{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$
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 $z \geq z_{\alpha/2}$

p-value = .0062, the null hypothesis would be rejected for any level of significance greater than or equal to .0062.

Summary

The procedure used to conduct a hypothesis test about a population proportion is similar to the procedure used to conduct a hypothesis test about a population mean. Although we only illustrated how to conduct a hypothesis test about a population proportion for an upper tail test, similar procedures can be used for lower tail and two-tailed tests. Table 9.4 provides a summary of the hypothesis tests about a population proportion. We assume that $np \geq 5$ and $n(1-p) \geq 5$; thus the normal probability distribution can be used to approximate the sampling distribution of $\bar{p}$.

Exercises

Methods

35. Consider the following hypothesis test:

$$H_0: p = .20$$
$$H_a: p \neq .20$$

A sample of 400 provided a sample proportion $\bar{p} = .175$.

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?

SELF test

36. Consider the following hypothesis test:

$$H_0: p \geq .75$$
$$H_a: p < .75$$

A sample of 300 items was selected. Compute the *p*-value and state your conclusion for each of the following sample results. Use $\alpha = .05$.

a. $\bar{p} = .68$
b. $\bar{p} = .72$
c. $\bar{p} = .70$
d. $\bar{p} = .77$

Applications

37. A study found that, in 2005, 12.5% of U.S. workers belonged to unions (*The Wall Street Journal*, January 21, 2006). Suppose a sample of 400 U.S. workers is collected in 2006 to determine whether union efforts to organize have increased union membership.
   a. Formulate the hypotheses that can be used to determine whether union membership increased in 2006.
   b. If the sample results show that 52 of the workers belonged to unions, what is the *p*-value for your hypothesis test?
   c. At $\alpha = .05$, what is your conclusion?

SELF test

38. A study by *Consumer Reports* showed that 64% of supermarket shoppers believe supermarket brands to be as good as national name brands. To investigate whether this result applies to its own product, the manufacturer of a national name-brand ketchup asked a sample of shoppers whether they believed that supermarket ketchup was as good as the national brand ketchup.
   a. Formulate the hypotheses that could be used to determine whether the percentage of supermarket shoppers who believe that the supermarket ketchup was as good as the national brand ketchup differed from 64%.
   b. If a sample of 100 shoppers showed 52 stating that the supermarket brand was as good as the national brand, what is the *p*-value?
   c. At $\alpha = .05$, what is your conclusion?
   d. Should the national brand ketchup manufacturer be pleased with this conclusion? Explain.
39. The National Center for Health Statistics released a report that stated 70% of adults do not exercise regularly (Associated Press, April 7, 2002). A researcher decided to conduct a study to see whether the claim made by the National Center for Health Statistics differed on a state-by-state basis.
   a. State the null and alternative hypotheses assuming the intent of the researcher is to identify states that differ from the 70% reported by the National Center for Health Statistics.
   b. At $\alpha = .05$, what is the research conclusion for the following states:

Wisconsin:	252 of 350 adults did not exercise regularly
California:	189 of 300 adults did not exercise regularly

40. Before the 2003 Super Bowl, ABC predicted that 22% of the Super Bowl audience would express an interest in seeing one of its forthcoming new television shows, including *8 Simple Rules, Are You Hot?,* and *Dragnet.* ABC ran commercials for these television shows during the Super Bowl. The day after the Super Bowl, Intermediate Advertising Group of New York sampled 1532 viewers who saw the commercials and found that 414 said that they would watch one of the ABC advertised television shows (*The Wall Street Journal,* January 30, 2003).
   a. What is the point estimate of the proportion of the audience that said they would watch the television shows after seeing the television commercials?
   b. At $\alpha = .05$, determine whether the intent to watch the ABC television shows significantly increased after seeing the television commercials. Formulate the appropriate hypotheses, compute the *p*-value, and state your conclusion.
   c. Why are such studies valuable to companies and advertising firms?
41. Speaking to a group of analysts in January 2006, a brokerage firm executive claimed that at least 70% of investors are currently confident of meeting their investment objectives. A UBS Investor Optimism Survey, conducted over the period January 2 to January 15, found that 67% of investors were confident of meeting their investment objectives (CNBC, January 20, 2006).
   a. Formulate the hypotheses that can be used to test the validity of the brokerage firm executive's claim.

b. Assume the UBS Investor Optimism Survey collected information from 300 investors. What is the p-value for the hypothesis test?

c. At $\alpha = .05$, should the executive's claim be rejected?

42. According to the University of Nevada Center for Logistics Management, 6% of all merchandise sold in the United States gets returned (*BusinessWeek*, January 15, 2007). A Houston department store sampled 80 items sold in January and found that 12 of the items were returned.

    a. Construct a point estimate of the proportion of items returned for the population of sales transactions at the Houston store.

    b. Construct a 95% confidence interval for the proportion of returns at the Houston store.

    c. Is the proportion of returns at the Houston store significantly different from the returns for the nation as a whole? Provide statistical support for your answer.

CD file Eagle

43. Eagle Outfitters is a chain of stores specializing in outdoor apparel and camping gear. They are considering a promotion that involves mailing discount coupons to all their credit card customers. This promotion will be considered a success if more than 10% of those receiving the coupons use them. Before going national with the promotion, coupons were sent to a sample of 100 credit card customers.

    a. Develop hypotheses that can be used to test whether the population proportion of those who will use the coupons is sufficient to go national.

    b. The file Eagle contains the sample data. Develop a point estimate of the population proportion.

    c. Use $\alpha = .05$ to conduct your hypothesis test. Should Eagle go national with the promotion?

CD file Drowsy

44. In a cover story, *BusinessWeek* published information about sleep habits of Americans (*BusinessWeek*, January 26, 2004). The article noted that sleep deprivation causes a number of problems, including highway deaths. Fifty-one percent of adult drivers admit to driving while drowsy. A researcher hypothesized that this issue was an even bigger problem for night shift workers.

    a. Formulate the hypotheses that can be used to help determine whether more than 51% of the population of night shift workers admit to driving while drowsy.

    b. A sample of 500 night shift workers identified those who admitted to driving while drowsy. What is the sample proportion? What is the p-value?

    c. At $\alpha = .01$, what is your conclusion?

45. Many investors and financial analysts believe the Dow Jones Industrial Average (DJIA) provides a good barometer of the overall stock market. On January 31, 2006, 9 of the 30 stocks making up the DJIA increased in price (*The Wall Street Journal*, February 1, 2006). On the basis of this fact, a financial analyst claims we can assume that 30% of the stocks traded on the New York Stock Exchange (NYSE) went up the same day.

    a. Formulate null and alternative hypotheses to test the analyst's claim.

    b. A sample of 50 stocks traded on the NYSE that day showed that 24 went up. What is your point estimate of the population proportion of stocks that went up?

    c. Conduct your hypothesis test using $\alpha = .01$ as the level of significance. What is your conclusion?

Summary

Hypothesis testing is a statistical procedure that uses sample data to determine whether a statement about the value of a population parameter should or should not be rejected. The hypotheses are two competing statements about a population parameter. One statement is called the null hypothesis (H_0), and the other statement is called the alternative hypothesis (H_a). In Section 9.1 we provided guidelines for developing hypotheses for three situations frequently encountered in practice.

Whenever historical data or other information provides a basis for assuming that the population standard deviation is known, the hypothesis testing procedure for the population mean is based on the standard normal distribution. Whenever σ is unknown, the sample standard deviation s is used to estimate σ and the hypothesis testing procedure is based on the t distribution. In both cases, the quality of results depends on both the form of the population distribution and the sample size. If the population has a normal distribution, both hypothesis testing procedures are applicable, even with small sample sizes. If the population is not normally distributed, larger sample sizes are needed. General guidelines about the sample size were provided in Sections 9.3 and 9.4. In the case of hypothesis tests about a population proportion, the hypothesis testing procedure uses a test statistic based on the standard normal distribution.

In all cases, the value of the test statistic can be used to compute a p-value for the test. A p-value is a probability used to determine whether the null hypothesis should be rejected. If the p-value is less than or equal to the level of significance α, the null hypothesis can be rejected.

Hypothesis testing conclusions can also be made by comparing the value of the test statistic to a critical value. For lower tail tests, the null hypothesis is rejected if the value of the test statistic is less than or equal to the critical value. For upper tail tests, the null hypothesis is rejected if the value of the test statistic is greater than or equal to the critical value. Two-tailed tests consist of two critical values: one in the lower tail of the sampling distribution and one in the upper tail. In this case, the null hypothesis is rejected if the value of the test statistic is less than or equal to the critical value in the lower tail or greater than or equal to the critical value in the upper tail.

Glossary

Null hypothesis The hypothesis tentatively assumed true in the hypothesis testing procedure.
Alternative hypothesis The hypothesis concluded to be true if the null hypothesis is rejected.
Type I error The error of rejecting H_0 when it is true.
Type II error The error of accepting H_0 when it is false.
Level of significance The probability of making a Type I error when the null hypothesis is true as an equality.
One-tailed test A hypothesis test in which rejection of the null hypothesis occurs for values of the test statistic in one tail of its sampling distribution.
Test statistic A statistic whose value helps determine whether a null hypothesis should be rejected.
***p*-value** 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. 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. 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. For a two-tailed test, the p-value is the probability of obtaining a value for the test statistic at least as unlikely as or more unlikely than that provided by the sample.
Critical value A value that is compared with the test statistic to determine whether H_0 should be rejected.
Two-tailed test A hypothesis test in which rejection of the null hypothesis occurs for values of the test statistic in either tail of its sampling distribution.

Key Formulas

Test Statistic for Hypothesis Tests About a Population Mean: σ Known

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

Test Statistic for Hypothesis Tests About a Population Mean: σ Unknown

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

Test Statistic for Hypothesis Tests About a Population Proportion

$$z = \frac{\bar{p} - p_0}{\sqrt{\dfrac{p_0(1 - p_0)}{n}}} \tag{9.4}$$

Supplementary Exercises

46. A production line operates with a mean filling weight of 16 ounces per container. Overfilling or underfilling presents a serious problem and when detected requires the operator to shut down the production line to readjust the filling mechanism. From past data, a population standard deviation $\sigma = .8$ ounces is assumed. A quality control inspector selects a sample of 30 items every hour and at that time makes the decision of whether to shut down the line for readjustment. The level of significance is $\alpha = .05$.
    a. State the hypothesis test for this quality control application.
    b. If a sample mean of $\bar{x} = 16.32$ ounces were found, what is the p-value? What action would you recommend?
    c. If a sample mean of $\bar{x} = 15.82$ ounces were found, what is the p-value? What action would you recommend?
    d. Use the critical value approach. What is the rejection rule for the preceding hypothesis testing procedure? Repeat parts (b) and (c). Do you reach the same conclusion?

47. At Western University the historical mean of scholarship examination scores for freshman applications is 900. A historical population standard deviation $\sigma = 180$ is assumed known. Each year, the assistant dean uses a sample of applications to determine whether the mean examination score for the new freshman applications has changed.
    a. State the hypotheses.
    b. What is the 95% confidence interval estimate of the population mean examination score if a sample of 200 applications provided a sample mean of $\bar{x} = 935$?
    c. Use the confidence interval to conduct a hypothesis test. Using $\alpha = .05$, what is your conclusion?
    d. What is the p-value?

48. *Playbill* is a magazine distributed around the country to people attending musicals and other theatrical productions. The mean annual household income for the population of *Playbill* readers is $119,155 (*Playbill,* January 2006). Assume the standard deviation is $\sigma = \$20{,}700$. A San Francisco civic group has asserted that the mean for theater goers in the Bay Area is higher. A sample of 60 theater attendees in the Bay Area showed a sample mean household income of $126,100.
    a. Develop hypotheses that can be used to determine whether the sample data support the conclusion that theater attendees in the Bay Area have a higher mean household income than that for all *Playbill* readers.

    b. What is the p-value based on the sample of 60 theater attendees in the Bay Area?
    c. Use $\alpha = .01$ as the level of significance. What is your conclusion?

49. On Friday, Wall Street traders were anxiously awaiting the federal government's release of numbers on the January increase in nonfarm payrolls. The early consensus estimate among economists was for a growth of 250,000 new jobs (CNBC, February 3, 2006). However, a sample of 20 economists taken Thursday afternoon provided a sample mean of 266,000 with a sample standard deviation of 24,000. Financial analysts often call such a sample mean, based on late-breaking news, the *whisper number.* Treat the "consensus estimate" as the population mean. Conduct a hypothesis test to determine whether the whisper number justifies a conclusion of a statistically significant increase in the consensus estimate of economists. Use $\alpha = .01$ as the level of significance.

50. The College Board reported that the average number of freshman class applications to public colleges and universities is 6000 (*USA Today,* December 26, 2002). During a recent application/enrollment period, a sample of 32 colleges and universities showed that the sample mean number of freshman class applications was 5812 with a sample standard deviation of 1140. Do the data indicate a change in the mean number of applications? Use $\alpha = .05$.

51. An extensive study of the cost of health care in the United States presented data showing that the mean spending per Medicare enrollee in 2003 was $6883 (*Money,* Fall 2003). To investigate differences across the country, a researcher took a sample of 40 Medicare enrollees in Indianapolis. For the Indianapolis sample, the mean 2003 Medicare spending was $5980 and the standard deviation was $2518.
    a. State the hypotheses that should be used if we would like to determine whether the mean annual Medicare spending in Indianapolis is lower than the national mean.
    b. Use the preceding sample results to compute the test statistic and the p-value.
    c. Use $\alpha = .05$. What is your conclusion?
    d. Repeat the hypothesis test using the critical value approach.

52. The chamber of commerce of a Florida Gulf Coast community advertises that area residential property is available at a mean cost of $125,000 or less per lot. Suppose a sample of 32 properties provided a sample mean of $130,000 per lot and a sample standard deviation of $12,500. Use a .05 level of significance to test the validity of the advertising claim.

CD file
Gasoline

53. The U.S. Energy Administration reported that the mean price for a gallon of regular gasoline in the United States was $2.357 (U.S. Energy Administration, January 30, 2006). Data for a sample of regular gasoline prices at 50 service stations in the Lower Atlantic states are contained in the data file named Gasoline. Conduct a hypothesis test to determine whether the mean price for a gallon of gasoline in the Lower Atlantic states is different from the national mean. Use $\alpha = .05$ for the level of significance, and state your conclusion.

54. A study by the Centers for Disease Control (CDC) found that 23.3% of adults are smokers and that roughly 70% of those who do smoke indicate that they want to quit (Associated Press, July 26, 2002). CDC reported that, of people who smoked at some point in their lives, 50% have been able to kick the habit. Part of the study suggested that the success rate for quitting rose by education level. Assume that a sample of 100 college graduates who smoked at some point in their lives showed that 64 had been able to successfully stop smoking.
    a. State the hypotheses that can be used to determine whether the population of college graduates has a success rate higher than the overall population when it comes to breaking the smoking habit.
    b. Given the sample data, what is the proportion of college graduates who, having smoked at some point in their lives, were able to stop smoking?
    c. What is the p-value? At $\alpha = .01$, what is your hypothesis testing conclusion?

55. An airline promotion to business travelers is based on the assumption that two-thirds of business travelers use a laptop computer on overnight business trips.
    a. State the hypotheses that can be used to test the assumption.
    b. What is the sample proportion from an American Express sponsored survey that found 355 of 546 business travelers use a laptop computer on overnight business trips?
    c. What is the *p*-value?
    d. Use $\alpha = .05$. What is your conclusion?

56. Virtual call centers are staffed by individuals working out of their homes. Most home agents earn $10 to $15 per hour without benefits versus $7 to $9 per hour with benefits at a traditional call center (*BusinessWeek,* January 23, 2006). Regional Airways is considering employing home agents, but only if a level of customer satisfaction greater than 80% can be maintained. A test was conducted with home service agents. In a sample of 300 customers 252 reported that they were satisfied with service.
    a. Develop hypotheses for a test to determine whether the sample data support the conclusion that customer service with home agents meets the Regional Airways criterion.
    b. What is your point estimate of the percentage of satisfied customers?
    c. What is the *p*-value provided by the sample data?
    d. What is your hypothesis testing conclusion? Use $\alpha = .05$ as the level of significance.

57. During the 2004 election year, new polling results were reported daily. In an IBD/TIPP poll of 910 adults, 503 respondents reported that they were optimistic about the national outlook, and President Bush's leadership index jumped 4.7 points to 55.3 (*Investor's Business Daily,* January 14, 2004).
    a. What is the sample proportion of respondents who are optimistic about the national outlook?
    b. A campaign manager wants to claim that this poll indicates that the majority of adults are optimistic about the national outlook. Construct a hypothesis test so that rejection of the null hypothesis will permit the conclusion that the proportion optimistic is greater than 50%.
    c. Use the polling data to compute the *p*-value for the hypothesis test in part (b). Explain to the manager what this *p*-value means about the level of significance of the results.

58. A radio station in Myrtle Beach announced that at least 90% of the hotels and motels would be full for the Memorial Day weekend. The station advised listeners to make reservations in advance if they planned to be in the resort over the weekend. On Saturday night a sample of 58 hotels and motels showed 49 with a no-vacancy sign and 9 with vacancies. What is your reaction to the radio station's claim after seeing the sample evidence? Use $\alpha = .05$ in making the statistical test. What is the *p*-value?

59. According to the federal government, 24% of workers covered by their company's health care plan were not required to contribute to the premium (*Statistical Abstract of the United States: 2006*). A recent study found that 81 out of 400 workers sampled were not required to contribute to their company's health care plan.
    a. Develop hypotheses that can be used to test whether the percent of workers not required to contribute to their company's health care plan has declined.
    b. What is a point estimate of the proportion receiving free company-sponsored health care insurance?
    c. Has a statistically significant decline occurred in the proportion of workers receiving free company-sponsored health care insurance? Use $\alpha = .05$.

Case Problem 1 Quality Associates, Inc.

Quality Associates, Inc., a consulting firm, advises its clients about sampling and statistical procedures that can be used to control their manufacturing processes. In one particular application, a client gave Quality Associates a sample of 800 observations taken during a

time in which that client's process was operating satisfactorily. The sample standard deviation for these data was .21; hence, with so much data, the population standard deviation was assumed to be .21. Quality Associates then suggested that random samples of size 30 be taken periodically to monitor the process on an ongoing basis. By analyzing the new samples, the client could quickly learn whether the process was operating satisfactorily. When the process was not operating satisfactorily, corrective action could be taken to eliminate the problem. The design specification indicated the mean for the process should be 12. The hypothesis test suggested by Quality Associates follows.

$$H_0: \mu = 12$$
$$H_a: \mu \neq 12$$

Corrective action will be taken any time H_0 is rejected.

The samples below were collected at hourly intervals during the first day of operation of the new statistical process control procedure. These data are available in the data set Quality.

Managerial Report

1. Conduct a hypothesis test for each sample at the .01 level of significance and determine what action, if any, should be taken. Provide the test statistic and *p*-value for each test.

CD file
Quality

Sample 1	Sample 2	Sample 3	Sample 4
11.55	11.62	11.91	12.02
11.62	11.69	11.36	12.02
11.52	11.59	11.75	12.05
11.75	11.82	11.95	12.18
11.90	11.97	12.14	12.11
11.64	11.71	11.72	12.07
11.80	11.87	11.61	12.05
12.03	12.10	11.85	11.64
11.94	12.01	12.16	12.39
11.92	11.99	11.91	11.65
12.13	12.20	12.12	12.11
12.09	12.16	11.61	11.90
11.93	12.00	12.21	12.22
12.21	12.28	11.56	11.88
12.32	12.39	11.95	12.03
11.93	12.00	12.01	12.35
11.85	11.92	12.06	12.09
11.76	11.83	11.76	11.77
12.16	12.23	11.82	12.20
11.77	11.84	12.12	11.79
12.00	12.07	11.60	12.30
12.04	12.11	11.95	12.27
11.98	12.05	11.96	12.29
12.30	12.37	12.22	12.47
12.18	12.25	11.75	12.03
11.97	12.04	11.96	12.17
12.17	12.24	11.95	11.94
11.85	11.92	11.89	11.97
12.30	12.37	11.88	12.23
12.15	12.22	11.93	12.25

2. Compute the standard deviation for each of the four samples. Does the assumption of .21 for the population standard deviation appear reasonable?
3. Compute limits for the sample mean $\bar{x}$ around $\mu = 12$ such that, as long as a new sample mean is within those limits, the process will be considered to be operating satisfactorily. If $\bar{x}$ exceeds the upper limit or if $\bar{x}$ is below the lower limit, corrective action will be taken. These limits are referred to as upper and lower control limits for quality control purposes.
4. Discuss the implications of changing the level of significance to a larger value. What mistake or error could increase if the level of significance is increased?

Case Problem 2 Unemployment Study

Each month the U.S. Bureau of Labor Statistics publishes a variety of unemployment statistics, including the number of individuals who are unemployed and the mean length of time the individuals have been unemployed. For November 1998, the Bureau of Labor Statistics reported that the national mean length of time of unemployment was 14.6 weeks.

The mayor of Philadelphia requested a study on the status of unemployment in the Philadelphia area. A sample of 50 unemployed residents of Philadelphia included data on their age and the number of weeks without a job. A portion of the data collected in November 1998 follows. The complete data set is available in the data file BLS.

CD file BLS

Age	Weeks	Age	Weeks	Age	Weeks
56	22	22	11	25	12
35	19	48	6	25	1
22	7	48	22	59	33
57	37	25	5	49	26
40	18	40	20	33	13

Managerial Report

1. Use descriptive statistics to summarize the data.
2. Develop a 95% confidence interval estimate of the mean age of unemployed individuals in Philadelphia.
3. Conduct a hypothesis test to determine whether the mean duration of unemployment in Philadelphia is greater than the national mean duration of 14.6 weeks. Use a .01 level of significance. What is your conclusion?
4. Is there a relationship between the age of an unemployed individual and the number of weeks of unemployment? Explain.

Appendix 9.1 Hypothesis Testing with Minitab

We describe the use of Minitab to conduct hypothesis tests about a population mean and a population proportion.

Population Mean: σ Known

We illustrate using the MaxFlight golf ball distance example in Section 9.3. The data are in column C1 of a Minitab worksheet. The population standard deviation $\sigma = 12$ is assumed

known and the level of significance is $\alpha = .05$. The following steps can be used to test the hypothesis $H_0: \mu = 295$ versus $H_a: \mu \neq 295$.

CD file GolfTest

Step 1. Select the **Stat** menu
Step 2. Choose **Basic Statistics**
Step 3. Choose **1-Sample Z**
Step 4. When the 1-Sample Z dialog box appears:
Enter C1 in the **Samples in columns** box
Enter 12 in the **Standard deviation** box
Select **Perform Hypothesis Test**
Enter 295 in the **Hypothesized mean** box
Select **Options**
Step 5. When the 1-Sample Z-Options dialog box appears:
Enter 95 in the **Confidence level** box*
Select **not equal** in the **Alternative** box
Click **OK**
Step 6. Click **OK**

In addition to the hypothesis testing results, Minitab provides a 95% confidence interval for the population mean.

The procedure can be easily modified for a one-tailed hypothesis test by selecting the less than or greater than option in the **Alternative** box in step 5.

Population Mean: σ Unknown

CD file AirRating

The ratings that 60 business travelers gave for Heathrow Airport are entered in column C1 of a Minitab worksheet. The level of significance for the test is $\alpha = .05$, and the population standard deviation σ will be estimated by the sample standard deviation s. The following steps can be used to test the hypothesis $H_0: \mu \leq 7$ against $H_a: \mu > 7$.

Step 1. Select the **Stat** menu
Step 2. Choose **Basic Statistics**
Step 3. Choose **1-Sample t**
Step 4. When the 1-Sample t dialog box appears:
Enter C1 in the **Samples in columns** box
Select **Perform Hypothesis Test**
Enter 7 in the **Hypothesized mean** box
Select **Options**
Step 5. When the 1-Sample t-options dialog box appears:
Enter 95 in the **Confidence level** box
Select **greater than** in the **Alternative** box
Click **OK**
Step 6. Click **OK**

The Heathrow Airport rating study involved a greater than alternative hypothesis. The preceding steps can be easily modified for other hypothesis tests by selecting the less than or not equal options in the **Alternative** box in step 5.

Population Proportion

CD file WomenGolf

We illustrate using the Pine Creek golf course example in Section 9.5. The data with responses Female and Male are in column C1 of a Minitab worksheet. Minitab uses an alphabetical ordering of the responses and selects the *second response* for the population

*Minitab provides both hypothesis testing and interval estimation results simultaneously. The user may select any confidence level for the interval estimate of the population mean: 95% confidence is suggested here.

proportion of interest. In this example, Minitab uses the alphabetical ordering Female-Male to provide results for the population proportion of Male responses. Because Female is the response of interest, we change Minitab's ordering as follows: Select any cell in the column and use the sequence: Editor > Column > Value Order. Then choose the option of entering a user-specified order. Enter Male-Female in the **Define-an-order** box and click OK. Minitab's 1 Proportion routine will then provide the hypothesis test results for the population proportion of female golfers. We proceed as follows:

Step 1. Select the **Stat** menu
Step 2. Choose **Basic Statistics**
Step 3. Choose **1 Proportion**
Step 4. When the 1 Proportion dialog box appears:
  Enter C1 in the **Samples in Columns** box
  Select **Perform Hypothesis Test**
  Enter .20 in the **Hypothesized proportion** box
  Select **Options**
Step 5. When the 1 Proportion-Options dialog box appears:
  Enter 95 in the **Confidence level** box
  Select greater than in the **Alternative** box
  Select **Use test and interval based on normal distribution**
  Click **OK**
Step 6. Click **OK**

Appendix 9.2 Hypothesis Testing with Excel

Excel does not provide built-in routines for the hypothesis tests presented in this chapter. To handle these situations, we present Excel worksheets that we designed to use as templates for testing hypotheses about a population mean and a population proportion. The worksheets are easy to use and can be modified to handle any sample data. The worksheets are available on the CD that accompanies this book.

Population Mean: σ Known

We illustrate using the MaxFlight golf ball distance example in Section 9.3. The data are in column A of an Excel worksheet. The population standard deviation $\sigma = 12$ is assumed known and the level of significance is $\alpha = .05$. The following steps can be used to test the hypothesis H_0: $\mu = 295$ versus H_a: $\mu \neq 295$.

CD file
Hyp Sigma Known

Refer to Figure 9.8 as we describe the procedure. The worksheet in the background shows the cell formulas used to compute the results shown in the foreground worksheet. The data are entered into cells A2:A51. The following steps are necessary to use the template for this data set.

Step 1. Enter the data range A2:A51 into the =COUNT cell formula in cell D4
Step 2. Enter the data range A2:A51 into the =AVERAGE cell formula in cell D5
Step 3. Enter the population standard deviation $\sigma = 12$ into cell D6
Step 4. Enter the hypothesized value for the population mean 295 into cell D8

The remaining cell formulas automatically provide the standard error, the value of the test statistic z, and three p-values. Because the alternative hypothesis ($\mu_0 \neq 295$) indicates a two-tailed test, the p-value (Two Tail) in cell D15 is used to make the rejection decision. With p-value $= .1255 > \alpha = .05$, the null hypothesis cannot be rejected. The p-values in cells D13 or D14 would be used if the hypotheses involved a one-tailed test.

FIGURE 9.8 EXCEL WORKSHEET FOR HYPOTHESIS TESTS ABOUT A POPULATION MEAN WITH σ KNOWN

	A	B	C	D	E
1	**Yards**		**Hypothesis Test About a Population Mean**		
2	303		**With σ Known**		
3	282				
4	289		**Sample Size**	=COUNT(A2:A51)	
5	298		**Sample Mean**	=AVERAGE(A2:A51)	
6	283		**Population Std. Deviation**	12	
7	317				
8	297		**Hypothesized Value**	295	
9	308				
10	317		**Standard Error**	=D6/SQRT(D4)	
11	293		**Test Statistic** z	=(D5-D8)/D10	
12	284				
13	290		p**-value (Lower Tail)**	=NORMSDIST(D11)	
14	304		p**-value (Upper Tail)**	=1-D13	
15	290		p**-value (Two Tail)**	=2*MIN(D13,D14)	
16	311				
17	305				
49	303				
50	301				
51	292				
52					

	A	B	C	D	E
1	**Yards**		**Hypothesis Test About a Population Mean**		
2	303		**With σ Known**		
3	282				
4	289		**Sample Size**	50	
5	298		**Sample Mean**	297.6	
6	283		**Population Std. Deviation**	12	
7	317				
8	297		**Hypothesized Value**	295	
9	308				
10	317		**Standard Error**	1.70	
11	293		**Test Statistic** z	1.53	
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				
17	305				
49	303				
50	301				
51	292				
52					

Note: Rows 18 to 48 are hidden.

This template can be used to make hypothesis testing computations for other applications. For instance, to conduct a hypothesis test for a new data set, enter the new sample data into column A of the worksheet. Modify the formulas in cells D4 and D5 to correspond to the new data range. Enter the population standard deviation into cell D6 and the hypothesized value for the population mean into cell D8 to obtain the results. If the new sample data have already been summarized, the new sample data do not have to be entered into the worksheet. In this case, enter the sample size into cell D4, the sample mean into cell D5, the population standard deviation into cell D6, and the hypothesized value for the population mean into cell D8 to obtain the results. The worksheet in Figure 9.8 is available in the file Hyp Sigma Known on the CD that accompanies this book.

Population Mean: σ Unknown

We illustrate using the Heathrow Airport rating example in Section 9.4. The data are in column A of an Excel worksheet. The population standard deviation σ is unknown and will be estimated by the sample standard deviation s. The level of significance is $\alpha = .05$. The following steps can be used to test the hypothesis H_0: $\mu \leq 7$ versus H_a: $\mu > 7$.

CD file
Hyp Sigma Unknown

Refer to Figure 9.9 as we describe the procedure. The background worksheet shows the cell formulas used to compute the results shown in the foreground version of the worksheet. The data are entered into cells A2:A61. The following steps are necessary to use the template for this data set.

Step 1. Enter the data range A2:A61 into the =COUNT cell formula in cell D4
Step 2. Enter the data range A2:A61 into the =AVERAGE cell formula in cell D5
Step 3. Enter the data range A2:A61 into the =STDEV cell formula in cell D6
Step 4. Enter the hypothesized value for the population mean 7 into cell D8

The remaining cell formulas automatically provide the standard error, the value of the test statistic t, the number of degrees of freedom, and three p-values. Because the alternative hypothesis ($\mu > 7$) indicates an upper tail test, the p-value (Upper Tail) in cell D15 is used to make the decision. With p-value $= .0353 < \alpha = .05$, the null hypothesis is rejected. The p-values in cells D14 or D16 would be used if the hypotheses involved a lower tail test or a two-tailed test.

This template can be used to make hypothesis testing computations for other applications. For instance, to conduct a hypothesis test for a new data set, enter the new sample data into column A of the worksheet and modify the formulas in cells D4, D5, and D6 to correspond to the new data range. Enter the hypothesized value for the population mean into cell D8 to obtain the results. If the new sample data have already been summarized, the new sample data do not have to be entered into the worksheet. In this case, enter the sample size into cell D4, the sample mean into cell D5, the sample standard deviation into cell D6, and the hypothesized value for the population mean into cell D8 to obtain the results. The worksheet in Figure 9.9 is available in the file Hyp Sigma Unknown on the CD that accompanies this book.

Population Proportion

CD file
Hypothesis p

We illustrate using the Pine Creek golf course survey data presented in Section 9.5. The data of Male or Female golfer are in column A of an Excel worksheet. Refer to Figure 9.10 as we describe the procedure. The background worksheet shows the cell formulas used to compute the results shown in the foreground worksheet. The data are entered into cells A2:A401. The following steps can be used to test the hypothesis H_0: $p \leq .20$ versus H_a: $p > .20$.

Step 1. Enter the data range A2:A401 into the =COUNTA cell formula in cell D3
Step 2. Enter Female as the response of interest in cell D4
Step 3. Enter the data range A2:A401 into the =COUNTIF cell formula in cell D5
Step 4. Enter the hypothesized value for the population proportion .20 into cell D8

The remaining cell formulas automatically provide the standard error, the value of the test statistic z, and three p-values. Because the alternative hypothesis ($p_0 > .20$)

FIGURE 9.9 EXCEL WORKSHEET FOR HYPOTHESIS TESTS ABOUT A POPULATION MEAN WITH σ UNKNOWN

	A	B	C	D	E
1	Rating		Hypothesis Test About a Population Mean		
2	5		With σ Unknown		
3	7				
4	8		Sample Size	=COUNT(A2:A61)	
5	7		Sample Mean	=AVERAGE(A2:A61)	
6	8		Sample Std. Deviation	=STDEV(A2:A61)	
7	8				
8	8		Hypothesized Value	7	
9	7				
10	8		Standard Error	=D6/SQRT(D4)	
11	10		Test Statistic *t*	=(D5-D8)/D10	
12	6		Degrees of Freedom	=D4-1	
13	7				
14	8		*p*-value (Lower Tail)	=IF(D11<0,TDIST(-D11,D12,1),1-TDIST(D11,D12,1))	
15	8		*p*-value (Upper Tail)	=1-D14	
16	9		*p*-value (Two Tail)	=2*MIN(D14,D15)	
17	7				
59	7				
60	7				
61	8				
62					

	A	B	C	D	E
1	Rating		Hypothesis Test About a Population Mean		
2	5		With σ Unknown		
3	7				
4	8		Sample Size	60	
5	7		Sample Mean	7.25	
6	8		Sample Std. Deviation	1.05	
7	8				
8	8		Hypothesized Value	7	
9	7				
10	8		Standard Error	0.136	
11	10		Test Statistic *t*	1.841	
12	6		Degrees of Freedom	59	
13	7				
14	8		*p*-value (Lower Tail)	0.9647	
15	8		*p*-value (Upper Tail)	0.0353	
16	9		*p*-value (Two Tail)	0.0706	
17	7				
59	7				
60	7				
61	8				
62					

Note: Rows 18 to 58 are hidden.

indicates an upper tail test, the p-value (Upper Tail) in cell D14 is used to make the decision. With p-value $= .0062 < \alpha = .05$, the null hypothesis is rejected. The p-values in cells D13 or D15 would be used if the hypothesis involved a lower tail test or a two-tailed test.

This template can be used to make hypothesis testing computations for other applications. For instance, to conduct a hypothesis test for a new data set, enter the new sample data into column A of the worksheet. Modify the formulas in cells D3 and D5 to correspond to the new data range. Enter the response of interest into cell D4 and the hypothesized value for the population proportion into cell D8 to obtain the results. If the new sample data have

FIGURE 9.10 EXCEL WORKSHEET FOR HYPOTHESIS TESTS ABOUT A POPULATION PROPORTION

	A	B	C	D	E
1	**Golfer**		**Hypothesis Test About a Population Proportion**		
2	Female				
3	Male		**Sample Size**	=COUNTA(A2:A401)	
4	Female		**Response of Interest**	Female	
5	Male		**Count for Response**	=COUNTIF(A2:A401,D4)	
6	Male		**Sample Proportion**	=D5/D3	
7	Female				
8	Male		**Hypothesized Value**	0.20	
9	Male				
10	Female		**Standard Error**	=SQRT(D8*(1-D8)/D3)	
11	Male		**Test Statistic *z***	=(D6-D8)/D10	
12	Male				
13	Male		***p*-value (Lower Tail)**	=NORMSDIST(D11)	
14	Male		***p*-value (Upper Tail)**	=1-D13	
15	Male		***p*-value (Two Tail)**	=2*MIN(D13,D14)	
16	Female				
400	Male				
401	Male				
402					

	A	B	C	D	E
1	**Golfer**		**Hypothesis Test About a Population Proportion**		
2	Female				
3	Male		**Sample Size**	400	
4	Female		**Response of Interest**	Female	
5	Male		**Count for Response**	100	
6	Male		**Sample Proportion**	0.2500	
7	Female				
8	Male		**Hypothesized Value**	0.20	
9	Male				
10	Female		**Standard Error**	0.0200	
11	Male		**Test Statistic *z***	2.50	
12	Male				
13	Male		***p*-value (Lower Tail)**	0.9938	
14	Male		***p*-value (Upper Tail)**	0.0062	
15	Male		***p*-value (Two Tail)**	0.0124	
16	Female				
400	Male				
401	Male				
402					

Note: Rows 17 to 399 are hidden.

already been summarized, the new sample data do not have to be entered into the worksheet. In this case, enter the sample size into cell D3, the sample proportion into cell D6, and the hypothesized value for the population proportion into cell D8 to obtain the results. The worksheet in Figure 9.10 is available in the file Hypothesis p on the CD that accompanies this book.

CHAPTER 10

Comparisons Involving Means, Experimental Design, and Analysis of Variance

CONTENTS