9 Fundamentals of Hypothesis Testing: One-Sample Tests

Learning Objectives

In this chapter, you learn:

- The basic principles of hypothesis testing
- How to use hypothesis testing to test a mean or proportion
- The assumptions of each hypothesis-testing procedure, how to evaluate them, and the consequences if they are seriously violated
- How to avoid the pitfalls involved in hypothesis testing
- Ethical issues involved in hypothesis testing

USING STATISTICS

@ Oxford Cereals, Part II

As in Chapter 7, you again find yourself as plant operations manager for Oxford Cereals. You are responsible for monitoring the amount in each cereal box filled. Company specifications require a mean weight of 368 grams per box. It is your responsibility to adjust the process when the mean fill weight in the population of boxes differs from 368 grams. How can you make the decision about whether to adjust the process when you are unable to weigh every single box as it is being filled? You begin by selecting and weighing a random sample of 25 cereal boxes. After computing the sample mean, how do you proceed?

In Chapter 7, you learned methods to determine whether the value of a sample mean is consistent with a known population mean. In this Oxford Cereals scenario, you want to use a sample mean to validate a claim about the population mean, a somewhat different problem. For this type of problem, you use an inferential method called **hypothesis testing**. Hypothesis testing requires that you state a claim unambiguously. In this scenario, the claim is that the population mean is 368 grams. You examine a sample statistic to see if it better supports the stated claim, called the *null hypothesis*, or the mutually exclusive alternative hypothesis (for this scenario, that the population mean is not 368 grams).

In this chapter, you will learn several applications of hypothesis testing. You will learn how to make inferences about a population parameter by *analyzing differences* between the results observed, the sample statistic, and the results you would expect to get if an underlying hypothesis were actually true. For the Oxford Cereals scenario, hypothesis testing allows you to infer one of the following:

- The mean weight of the cereal boxes in the sample is a value consistent with what you would expect if the mean of the entire population of cereal boxes is 368 grams.
- The population mean is not equal to 368 grams because the sample mean is significantly different from 368 grams.

9.1 Fundamentals of Hypothesis-Testing Methodology

Hypothesis testing typically begins with a theory, a claim, or an assertion about a particular parameter of a population. For example, your initial hypothesis in the cereal example is that the process is working properly, so the mean fill is 368 grams, and no corrective action is needed.

The Null and Alternative Hypotheses

The hypothesis that the population parameter is equal to the company specification is referred to as the null hypothesis. A **null hypothesis** is often one of status quo and is identified by the symbol $\boldsymbol{H_0}$. Here the null hypothesis is that the filling process is working properly, and therefore the mean fill is the 368-gram specification provided by Oxford Cereals. This is stated as

$$H_0: \mu = 368$$

Even though information is available only from the sample, the null hypothesis is stated in terms of the population parameter because your focus is on the population of all cereal boxes. You use the sample statistic to make inferences about the entire filling process. One inference may be that the results observed from the sample data indicate that the null hypothesis is false. If the null hypothesis is considered false, something else must be true.

Whenever a null hypothesis is specified, an alternative hypothesis is also specified, and it must be true if the null hypothesis is false. The **alternative hypothesis, $\boldsymbol{H_1}$**, is the opposite of the null hypothesis, H_0. This is stated in the cereal example as

$$H_1: \mu \neq 368$$

The alternative hypothesis represents the conclusion reached by rejecting the null hypothesis. The null hypothesis is rejected when there is sufficient evidence from the sample data that the null hypothesis is false. In the cereal example, if the weights of the sampled boxes are sufficiently above or below the expected 368-gram mean specified by Oxford Cereals, you reject the null hypothesis in favor of the alternative hypothesis that the mean fill is different from 368 grams. You stop production and take whatever action is necessary to correct the problem. If the null hypothesis is not rejected, you should continue to believe that the process is working correctly and therefore no corrective action is necessary. In this second circumstance, you have not proven that the process is working correctly. Rather, you have failed to prove that it is working incorrectly, and therefore you continue your belief (although unproven) in the null hypothesis.

In hypothesis testing, you reject the null hypothesis when the sample evidence suggests that it is far more likely that the alternative hypothesis is true. However, failure to reject the null hypothesis is not proof that it is true. You can never prove that the null hypothesis is correct because the decision is based only on the sample information, not on the entire population. Therefore, if you fail to reject the null hypothesis, you can only conclude that there is insufficient evidence to warrant its rejection. The following key points summarize the null and alternative hypotheses:

- The null hypothesis, H_0, represents the current belief in a situation.
- The alternative hypothesis, H_1, is the opposite of the null hypothesis and represents a research claim or specific inference you would like to prove.
- If you reject the null hypothesis, you have statistical proof that the alternative hypothesis is correct.
- If you do not reject the null hypothesis, you have failed to prove the alternative hypothesis. The failure to prove the alternative hypothesis, however, does not mean that you have proven the null hypothesis.
- The null hypothesis, H_0, always refers to a specified value of the population parameter (such as μ), not a sample statistic (such as $\bar{X}$).
- The statement of the null hypothesis always contains an equal sign regarding the specified value of the population parameter (e.g., $H_0: \mu = 368$ grams).
- The statement of the alternative hypothesis never contains an equal sign regarding the specified value of the population parameter (e.g., $H_1: \mu \neq 368$ grams).

EXAMPLE 9.1

The Null and Alternative Hypotheses

You are the manager of a fast-food restaurant. You want to determine whether the waiting time to place an order has changed in the past month from its previous population mean value of 4.5 minutes. State the null and alternative hypotheses.

SOLUTION The null hypothesis is that the population mean has not changed from its previous value of 4.5 minutes. This is stated as

$$H_0: \mu = 4.5$$

The alternative hypothesis is the opposite of the null hypothesis. Because the null hypothesis is that the population mean is 4.5 minutes, the alternative hypothesis is that the population mean is not 4.5 minutes. This is stated as

$$H_1: \mu \neq 4.5$$

The Critical Value of the Test Statistic

The logic of hypothesis testing involves determining how likely the null hypothesis is to be true by considering the data collected in a sample. In the Oxford Cereal Company scenario, the null hypothesis is that the mean amount of cereal per box in the entire filling process is 368 grams (the population parameter specified by the company). You select a sample of boxes from the filling process, weigh each box, and compute the sample mean. This statistic is an estimate of the corresponding parameter (the population mean, μ). Even if the null hypothesis is true, the statistic (the sample mean, $\bar{X}$) is likely to differ from the value of the parameter (the population mean, μ) because of variation due to sampling. However, you expect the sample statistic to be close to the population parameter if the null hypothesis is true. If the sample statistic is close to the population parameter, you have insufficient evidence to reject the null hypothesis. For example, if the sample mean is 367.9, you might conclude that the population mean has not changed (i.e., $\mu = 368$) because a sample mean of 367.9 is very close to the hypothesized value of 368. Intuitively, you think that it is likely that you could get a sample mean of 367.9 from a population whose mean is 368.

However, if there is a large difference between the value of the statistic and the hypothesized value of the population parameter, you might conclude that the null hypothesis is false. For example, if the sample mean is 320, you might conclude that the population mean is not 368 (i.e., $\mu \neq 368$) because the sample mean is very far from the hypothesized value of 368.

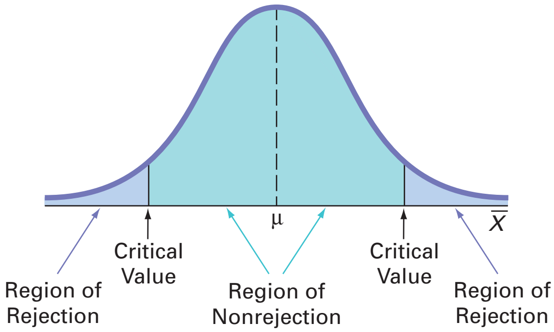
In such a case, you conclude that it is very unlikely to get a sample mean of 320 if the population mean is really 368. Therefore, it is more logical to conclude that the population mean is not equal to 368. Here you reject the null hypothesis.

However, the decision-making process is not always so clear-cut. Determining what is "very close" and what is "very different" is arbitrary without clear definitions. Hypothesis-testing methodology provides clear definitions for evaluating differences. Furthermore, it enables you to quantify the decision-making process by computing the probability of getting a certain sample result if the null hypothesis is true. You calculate this probability by determining the sampling distribution for the sample statistic of interest (e.g., the sample mean) and then computing the particular **test statistic** based on the given sample result. Because the sampling distribution for the test statistic often follows a well-known statistical distribution, such as the standardized normal distribution or *t* distribution, you can use these distributions to help determine whether the null hypothesis is true.

Regions of Rejection and Nonrejection

The sampling distribution of the test statistic is divided into two regions, a **region of rejection** (sometimes called the critical region) and a **region of nonrejection** (see Figure 9.1). If the test statistic falls into the region of nonrejection, you do not reject the null hypothesis. In the Oxford Cereals scenario, you conclude that there is insufficient evidence that the population mean fill is different from 368 grams. If the test statistic falls into the rejection region, you reject the null hypothesis. In this case, you conclude that the population mean is not 368 grams.

FIGURE 9.1
Regions of rejection and nonrejection in hypothesis testing



The region of rejection consists of the values of the test statistic that are unlikely to occur if the null hypothesis is true. These values are much more likely to occur if the null hypothesis is false. Therefore, if a value of the test statistic falls into this rejection region, you reject the null hypothesis because that value is unlikely if the null hypothesis is true.

To make a decision concerning the null hypothesis, you first determine the **critical value** of the test statistic. The critical value divides the nonrejection region from the rejection region. Determining the critical value depends on the size of the rejection region. The size of the rejection region is directly related to the risks involved in using only sample evidence to make decisions about a population parameter.

Risks in Decision Making Using Hypothesis Testing

Using hypothesis testing involves the risk of reaching an incorrect conclusion. You might wrongly reject a true null hypothesis, H_0, or, conversely, you might wrongly *not* reject a false null hypothesis, H_0. These types of risk are called Type I and Type II errors.

TYPE I AND TYPE II ERRORS

A **Type I error** occurs if you reject the null hypothesis, H_0, when it is true and should not be rejected. A Type I error is a "false alarm." The probability of a Type I error occurring is α.

A **Type II error** occurs if you do not reject the null hypothesis, H_0, when it is false and should be rejected. A Type II error represents a "missed opportunity" to take some corrective action. The probability of a Type II error occurring is β.

In the Oxford Cereals scenario, you would make a Type I error if you concluded that the population mean fill is *not* 368 when it *is* 368. This error causes you to needlessly adjust the filling process (the "false alarm") even though the process is working properly. In the same scenario, you would make a Type II error if you concluded that the population mean fill *is* 368 when it is *not* 368. In this case, you would allow the process to continue without adjustment, even though an adjustment is needed (the "missed opportunity").

Traditionally, you control the Type I error by determining the risk level, α (the lowercase Greek letter *alpha*) that you are willing to have of rejecting the null hypothesis when it is true. This risk, or probability, of committing a Type I error is called the *level of significance* (α). Because you specify the level of significance before you perform the hypothesis test, you directly control the risk of committing a Type I error. Traditionally, you select a level of 0.01, 0.05, or 0.10. The choice of a particular risk level for making a Type I error depends on the cost of making a Type I error. After you specify the value for α, you can then determine the critical values that divide the rejection and nonrejection regions. You know the size of the rejection region because α is the probability of rejection when the null hypothesis is true. From this, you can then determine the critical value or values that divide the rejection and nonrejection regions.

The probability of committing a Type II error is called the β *risk*. Unlike a Type I error, which you control through the selection of α, the probability of making a Type II error depends on the difference between the hypothesized and actual values of the population parameter. Because large differences are easier to find than small ones, if the difference between the hypothesized and actual value of the population parameter is large, β is small. For example, if the population mean is 330 grams, there is a small chance (β) that you will conclude that the mean has not changed from 368. However, if the difference between the hypothesized and actual value of the parameter is small, β is large. For example, if the population mean is actually 367 grams, there is a large chance (β) that you will conclude that the mean is still 368 grams.

PROBABILITY OF TYPE I AND TYPE II ERRORS

The **level of significance (α)** of a statistical test is the probability of committing a Type I error.

The **β risk** is the probability of committing a Type II error.

The complement of the probability of a Type I error, $(1 - \alpha)$, is called the *confidence coefficient*. The confidence coefficient is the probability that you will not reject the null hypothesis, H_0, when it is true and should not be rejected. In the Oxford Cereals scenario, the confidence coefficient measures the probability of concluding that the population mean fill is 368 grams when it is actually 368 grams.

The complement of the probability of a Type II error, $(1 - \beta)$, is called the *power of a statistical test*. The power of a statistical test is the probability that you will reject the null hypothesis when it is false and should be rejected. In the Oxford Cereals scenario, the power of the test is the probability that you will correctly conclude that the mean fill amount is not 368 grams when it actually is not 368 grams. For an extended discussion of the power of a statistical test, read the **Section 9.6** online topic file that is available in on this book's companion website. (See Appendix C to learn how to access the online topic files.)

COMPLEMENTS OF TYPE I AND TYPE II ERRORS

The **confidence coefficient**, $(1 - \alpha)$, is the probability that you will not reject the null hypothesis, H_0, when it is true and should not be rejected.

The **power of a statistical test**, $(1 - \beta)$, is the probability that you will reject the null hypothesis when it is false and should be rejected.

Risks in Decision Making: A Delicate Balance Table 9.1 illustrates the results of the two possible decisions (do not reject H_0 or reject H_0) that you can make in any hypothesis test. You can make a correct decision or make one of two types of errors.

TABLE 9.1

Hypothesis Testing and Decision Making

	Actual Situation	
Statistical Decision	**H_0 True**	**H_0 False**
Do not reject H_0	Correct decision Confidence $= (1 - \alpha)$	Type II error $P(\text{Type II error}) = \beta$
Reject H_0	Type I error $P(\text{Type I error}) = \alpha$	Correct decision Power $= (1 - \beta)$

One way to reduce the probability of making a Type II error is by increasing the sample size. Large samples generally permit you to detect even very small differences between the hypothesized values and the actual population parameters. For a given level of α, increasing the sample size decreases β and therefore increases the power of the statistical test to detect that the null hypothesis, H_0, is false.

However, there is always a limit to your resources, and this affects the decision of how large a sample you can select. For any given sample size, you must consider the trade-offs between the two possible types of errors. Because you can directly control the risk of Type I error, you can reduce this risk by selecting a smaller value for α. For example, if the negative consequences associated with making a Type I error are substantial, you could select $\alpha = 0.01$ instead of 0.05. However, when you decrease α, you increase β, so reducing the risk of a Type I error results in an increased risk of a Type II error. However, to reduce β, you could select a larger value for α. Therefore, if it is important to try to avoid a Type II error, you can select α of 0.05 or 0.10 instead of 0.01.

In the Oxford Cereals scenario, the risk of a Type I error occurring involves concluding that the mean fill amount has changed from the hypothesized 368 grams when it actually has not changed. The risk of a Type II error occurring involves concluding that the mean fill amount has not changed from the hypothesized 368 grams when it actually has changed. The choice of reasonable values for α and β depends on the costs inherent in each type of error. For example, if it is very costly to change the cereal-filling process, you would want to be very confident that a change is needed before making any changes. In this case, the risk of a Type I error occurring is more important, and you would choose a small α. However, if you want to be very certain of detecting changes from a mean of 368 grams, the risk of a Type II error occurring is more important, and you would choose a higher level of α.

Now that you have been introduced to hypothesis testing, recall that in the Using Statistics scenario on page 325, the business problem facing Oxford Cereals is to determine whether the cereal-filling process is working properly (i.e., whether the mean fill throughout the entire packaging process remains at the specified 368 grams, and no corrective action is needed). To evaluate the 368-gram requirement, you select a random sample of 25 boxes, weigh each box, compute the sample mean, $\overline{X}$, and then evaluate the difference between this sample statistic and the hypothesized population parameter by comparing the sample mean weight (in grams) to the expected population mean of 368 grams specified by the company. The null and alternative hypotheses are

$$H_0: \mu = 368$$
$$H_1: \mu \neq 368$$

When the standard deviation, σ, is known (which rarely occurs), you use the ***Z* test for the mean** if the population is normally distributed. If the population is not normally distributed, you can still use the Z test if the sample size is large enough for the Central Limit Theorem to take effect (see Section 7.4). Equation (9.1) defines the Z_{STAT} test statistic for determining the difference between the sample mean, $\overline{X}$, and the population mean, μ, when the standard deviation, σ, is known.

Z TEST FOR THE MEAN (σ KNOWN)

$$Z_{STAT} = \frac{\overline{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}} \tag{9.1}$$

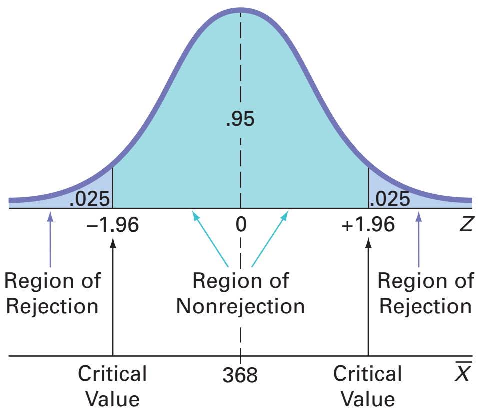
In Equation (9.1), the numerator measures the difference between the observed sample mean, $\bar{X}$, and the hypothesized mean, μ. The denominator is the standard error of the mean, so Z_{STAT} represents the difference between $\bar{X}$ and μ in standard error units.

Hypothesis Testing Using the Critical Value Approach

The critical value approach compares the computed Z_{STAT} test statistic value from Equation (9.1) to critical values that divide the normal distribution into regions of rejection and nonrejection. The critical values are expressed as standardized Z values that are determined by the level of significance.

For example, if you use a level of significance of 0.05, the size of the rejection region is 0.05. Because the rejection region is divided into the two tails of the distribution, you divide the 0.05 into two equal parts of 0.025 each. For this **two-tail test**, a rejection region of 0.025 in each tail of the normal distribution results in a cumulative area of 0.025 below the lower critical value and a cumulative area of 0.975 $(1 - 0.025)$ below the upper critical value (which leaves an area of 0.025 in the upper tail). According to the cumulative standardized normal distribution table (Table E.2), the critical values that divide the rejection and nonrejection regions are -1.96 and $+1.96$. Figure 9.2 illustrates that if the mean is actually 368 grams, as H_0 claims, the values of the Z_{STAT} test statistic have a standardized normal distribution centered at $Z = 0$ (which corresponds to an $\bar{X}$ value of 368 grams). Values of Z_{STAT} greater than $+1.96$ or less than -1.96 indicate that $\bar{X}$ is sufficiently different from the hypothesized $\mu = 368$ that it is unlikely that such an $\bar{X}$ value would occur if H_0 were true.

FIGURE 9.2
Testing a hypothesis about the mean (σ known) at the 0.05 level of significance



Therefore, the decision rule is

$$\text{Reject } H_0 \text{ if } Z_{STAT} > +1.96$$
$$\text{or if } Z_{STAT} < -1.96;$$
$$\text{otherwise, do not reject } H_0.$$

Suppose that the sample of 25 cereal boxes indicates a sample mean, $\bar{X}$ of 372.5 grams, and the population standard deviation, σ, is 15 grams. Using Equation (9.1) on page 330,

$$Z_{STAT} = \frac{\bar{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}} = \frac{372.5 - 368}{\dfrac{15}{\sqrt{25}}} = +1.50$$

Because $Z_{STAT} = +1.50$ is between -1.96 and $+1.96$, you do not reject H_0 (see Figure 9.3). You continue to believe that the mean fill amount is 368 grams. To take into account the possibility of a Type II error, you state the conclusion as "there is insufficient evidence that the mean fill is different from 368 grams."

Exhibit 9.1 summarizes the critical value approach to hypothesis testing. Steps 1 though 4 correspond to the Define task, step 5 combines the Collect and Organize tasks, and step 6 corresponds to the Visualize and Analyze tasks of the business problem-solving methodology first introduced in Chapter 2.

FIGURE 9.3

Testing a hypothesis about the mean cereal weight (σ known) at the 0.05 level of significance

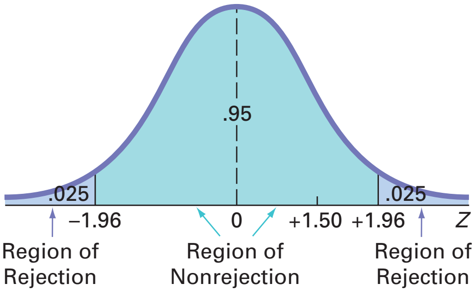


EXHIBIT 9.1 THE CRITICAL VALUE APPROACH TO HYPOTHESIS TESTING

1. State the null hypothesis, H_0, and the alternative hypothesis, H_1.
2. Choose the level of significance, α, and the sample size, n. The level of significance is based on the relative importance of the risks of committing Type I and Type II errors in the problem.
3. Determine the appropriate test statistic and sampling distribution.
4. Determine the critical values that divide the rejection and nonrejection regions.
5. Collect the sample data, organize the results, and compute the value of the test statistic.
6. Make the statistical decision and state the managerial conclusion. If the test statistic falls into the nonrejection region, you do not reject the null hypothesis. If the test statistic falls into the rejection region, you reject the null hypothesis. The managerial conclusion is written in the context of the real-world problem.

EXAMPLE 9.2

Applying the Critical Value Approach to Hypothesis Testing at Oxford Cereals

State the critical value approach to hypothesis testing at Oxford Cereals.

SOLUTION

Step 1: State the null and alternative hypotheses. The null hypothesis, H_0, is always stated as a mathematical expression, using population parameters. In testing whether the mean fill is 368 grams, the null hypothesis states that μ equals 368. The alternative hypothesis, H_1, is also stated as a mathematical expression, using population parameters. Therefore, the alternative hypothesis states that μ is not equal to 368 grams.

Step 2: Choose the level of significance and the sample size. You choose the level of significance, α, according to the relative importance of the risks of committing Type I and Type II errors in the problem. The smaller the value of α, the less risk there is of making a Type I error. In this example, making a Type I error means that you conclude that the population mean is not 368 grams when it is 368 grams. Thus, you will take corrective action on the filling process even though the process is working properly. Here, $\alpha = 0.05$ is selected. The sample size, n, is 25.

Step 3: Select the appropriate test statistic. Because σ is known from information about the filling process, you use the normal distribution and the Z_{STAT} test statistic.

Step 4: Determine the rejection region. Critical values for the appropriate test statistic are selected so that the rejection region contains a total area of α when H_0 is true and the nonrejection region contains a total area of $1 - \alpha$ when H_0 is true. Because $\alpha = 0.05$ in the cereal example, the critical values of the Z_{STAT} test statistic are -1.96 and $+1.96$. The rejection region is therefore $Z_{STAT} < -1.96$ or $Z_{STAT} > +1.96$. The nonrejection region is $-1.96 \le Z_{STAT} \le +1.96$.

Step 5: Collect the sample data and compute the value of the test statistic. In the cereal example, $\bar{X} = 372.5$, and the value of the test statistic is $Z_{STAT} = +1.50$.

Step 6: State the statistical decision and the managerial conclusion. First, determine whether the test statistic has fallen into the rejection region or the nonrejection region. For the cereal example, $Z_{STAT} = +1.50$ is in the region of nonrejection because

$-1.96 \le Z_{STAT} = +1.50 \le +1.96$. Because the test statistic falls into the nonrejection region, the statistical decision is to not reject the null hypothesis, H_0. The managerial conclusion is that insufficient evidence exists to prove that the mean fill is different from 368 grams. No corrective action on the filling process is needed.

EXAMPLE 9.3

Testing and Rejecting a Null Hypothesis

You are the manager of a fast-food restaurant. The business problem is to determine whether the population mean waiting time to place an order has changed in the past month from its previous population mean value of 4.5 minutes. From past experience, you can assume that the population is normally distributed, with a population standard deviation of 1.2 minutes. You select a sample of 25 orders during a one-hour period. The sample mean is 5.1 minutes. Use the six-step approach listed in Exhibit 9.1 on page 332 to determine whether there is evidence at the 0.05 level of significance that the population mean waiting time to place an order has changed in the past month from its previous population mean value of 4.5 minutes.

SOLUTION

Step 1: The null hypothesis is that the population mean has not changed from its previous value of 4.5 minutes:

$$H_0: \mu = 4.5$$

The alternative hypothesis is the opposite of the null hypothesis. Because the null hypothesis is that the population mean is 4.5 minutes, the alternative hypothesis is that the population mean is not 4.5 minutes:

$$H_1: \mu \neq 4.5$$

Step 2: You have selected a sample of $n = 25$. The level of significance is 0.05 (i.e., $\alpha = 0.05$).

Step 3: Because σ is assumed known, you use the normal distribution and the Z_{STAT} test statistic.

Step 4: Because $\alpha = 0.05$, the critical values of the Z_{STAT} test statistic are -1.96 and $+1.96$. The rejection region is $Z_{STAT} < -1.96$ or $Z_{STAT} > +1.96$. The nonrejection region is $-1.96 \le Z_{STAT} \le +1.96$

Step 5: You collect the sample data and compute $\bar{X} = 5.1$. Using Equation (9.1) on page 330, you compute the test statistic:

$$Z_{STAT} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{5.1 - 4.5}{\frac{1.2}{\sqrt{25}}} = +2.50$$

Step 6: Because $Z_{STAT} = +2.50 > +1.96$, you reject the null hypothesis. You conclude that there is evidence that the population mean waiting time to place an order has changed from its previous value of 4.5 minutes. The mean waiting time for customers is longer now than it was last month. As the manager, you would now want to determine how waiting time could be reduced to improve service.

Hypothesis Testing Using the *p*-Value Approach

Using the *p*-value to determine rejection and nonrejection is another approach to hypothesis testing.

> *p*-VALUE
>
> The ***p*-value** is the probability of getting a test statistic equal to or more extreme than the sample result, given that the null hypothesis, H_0, is true. The *p*-value is also known as the *observed level of significance*.

The decision rules for rejecting H_0 in the *p*-value approach are

- If the *p*-value is greater than or equal to α, do not reject the null hypothesis.
- If the *p*-value is less than α, reject the null hypothesis.

Many people confuse these rules, mistakenly believing that a high *p*-value is reason for rejection. You can avoid this confusion by remembering the following:

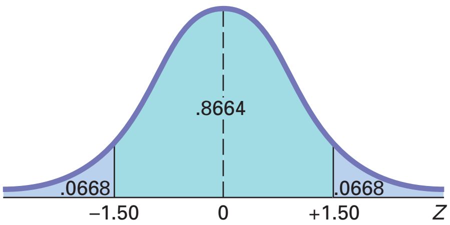
> If the *p*-value is low, then H_0 must go.

To understand the *p*-value approach, consider the Oxford Cereals scenario. You tested whether the mean fill was equal to 368 grams. The test statistic resulted in a Z_{STAT} value of $+1.50$, and you did not reject the null hypothesis because $+1.50$ was less than the upper critical value of $+1.96$ and greater than the lower critical value of -1.96.

To use the *p*-value approach for the *two-tail test*, you find the probability of getting a test statistic Z_{STAT} that is equal to or *more extreme than* 1.50 standard error units from the center of a standardized normal distribution. In other words, you need to compute the probability of a Z_{STAT} value greater than $+1.50$, along with the probability of a Z_{STAT} value less than -1.50. Table E.2 shows that the probability of a Z_{STAT} value below -1.50 is 0.0668. The probability of a value below $+1.50$ is 0.9332, and the probability of a value above $+1.50$ is $1 - 0.9332 = 0.0668$. Therefore, the *p*-value for this two-tail test is $0.0668 + 0.0668 = 0.1336$ (see Figure 9.4). Thus, the probability of a test statistic equal to or more extreme than the sample result is 0.1336. Because 0.1336 is greater than $\alpha = 0.05$, you do not reject the null hypothesis.

FIGURE 9.4

Finding a *p*-value for a two-tail test



In this example, the observed sample mean is 372.5 grams, 4.5 grams above the hypothesized value, and the *p*-value is 0.1336. Thus, if the population mean is 368 grams, there is a 13.36% chance that the sample mean differs from 368 grams by at least 4.5 grams (i.e., is ≥ 372.5 grams or ≤ 363.5 grams). Therefore, even though 372.5 is above the hypothesized value of 368, a result as extreme as or more extreme than 372.5 is not highly unlikely when the population mean is 368.

Unless you are dealing with a test statistic that follows the normal distribution, you will only be able to approximate the *p*-value from the tables of the distribution. However, Excel and Minitab can compute the *p*-value for any hypothesis test, and this allows you to substitute the *p*-value approach for the critical value approach when you do hypothesis testing.

Figure 9.5 shows the results for the cereal-filling example discussed in this section, as computed by Excel (left results) and Minitab (right results). These results include the Z_{STAT} test statistic and critical values.

FIGURE 9.5

Excel and Minitab results for the *Z* test for the mean (σ known) for the cereal-filling example

	A	B	
1	Z Test for the Mean		
2			
3	Data		
4	Null Hypothesis μ=	368	
5	Level of Significance	0.05	
6	Population Standard Deviation	15	
7	Sample Size	25	
8	Sample Mean	372.5	
9			
10	Intermediate Calculations		
11	Standard Error of the Mean	3	=B6/SQRT(B7)
12	Z Test Statistic	1.5	=(B8 - B4)/B11
13			
14	Two-Tail Test		
15	Lower Critical Value	-1.9600	=NORMSINV(B5/2)
16	Upper Critical Value	1.9600	=NORMSINV(1 - B5/2)
17	*p* Value	0.1336	=2 * (1 - NORMSDIST(ABS(B12)))
18	Do not reject the null hypothesis		=IF(B17 < B5, "Reject the null hypothesis", "Do not reject the null hypothesis")

One-Sample Z

```
Test of mu = 368 vs not = 368
The assumed standard deviation = 15

 N    Mean  SE Mean          95% CI              Z      P
25  372.50     3.00  (366.62, 378.38)  1.50  0.134
```

Exhibit 9.2 summarizes the p-value approach to hypothesis testing.

EXHIBIT 9.2 THE p-VALUE APPROACH TO HYPOTHESIS TESTING

1. State the null hypothesis, H_0, and the alternative hypothesis, H_1.
2. Choose the level of significance, α, and the sample size, n. The level of significance is based on the relative importance of the risks of committing Type I and Type II errors in the problem.
3. Determine the appropriate test statistic and the sampling distribution.
4. Collect the sample data, compute the value of the test statistic, and compute the p-value.
5. Make the statistical decision and state the managerial conclusion. If the p-value is greater than or equal to α, do not reject the null hypothesis. If the p-value is less than α, reject the null hypothesis. The managerial conclusion is written in the context of the real-world problem.

EXAMPLE 9.4

Testing and Rejecting a Null Hypothesis Using the p-Value Approach

You are the manager of a fast-food restaurant. The business problem is to determine whether the population mean waiting time to place an order has changed in the past month from its previous value of 4.5 minutes. From past experience, you can assume that the population standard deviation is 1.2 minutes and the population waiting time is normally distributed. You select a sample of 25 orders during a one-hour period. The sample mean is 5.1 minutes. Use the five-step p-value approach of Exhibit 9.2 to determine whether there is evidence that the population mean waiting time to place an order has changed in the past month from its previous population mean value of 4.5 minutes.

SOLUTION

Step 1: The null hypothesis is that the population mean has not changed from its previous value of 4.5 minutes:

$$H_0: \mu = 4.5$$

The alternative hypothesis is the opposite of the null hypothesis. Because the null hypothesis is that the population mean is 4.5 minutes, the alternative hypothesis is that the population mean is not 4.5 minutes:

$$H_1: \mu \neq 4.5$$

Step 2: You have selected a sample of $n = 25$, and you have chosen a 0.05 level of significance (i.e., $\alpha = 0.05$).

Step 3: Select the appropriate test statistic. Because σ is assumed known, you use the normal distribution and the Z_{STAT} test statistic.

Step 4: You collect the sample data and compute $\bar{X} = 5.1$. Using Equation (9.1) on page 330, you compute the test statistic as follows:

$$Z_{STAT} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{5.1 - 4.5}{\frac{1.2}{\sqrt{25}}} = +2.50$$

To find the probability of getting a Z_{STAT} test statistic that is equal to or more extreme than 2.50 standard error units from the center of a standardized normal distribution, you compute the probability of a Z_{STAT} value greater than $+2.50$ along with the probability of a Z_{STAT} value less than -2.50. From Table E.2, the probability of a Z_{STAT} value below -2.50 is 0.0062. The probability of a value below $+2.50$ is 0.9938. Therefore, the probability of a value above $+2.50$ is $1 - 0.9938 = 0.0062$. Thus, the p-value for this two-tail test is $0.0062 + 0.0062 = 0.0124$.

Step 5: Because the p-value $= 0.0124 < \alpha = 0.05$, you reject the null hypothesis. You conclude that there is evidence that the population mean waiting time to place an order has changed from its previous population mean value of 4.5 minutes. The mean waiting time for customers is longer now than it was last month.

A Connection Between Confidence Interval Estimation and Hypothesis Testing

This chapter and Chapter 8 discuss confidence interval estimation and hypothesis testing, the two major elements of statistical inference. Although confidence interval estimation and hypothesis testing share the same conceptual foundation, they are used for different purposes. In Chapter 8, confidence intervals estimated parameters. In this chapter, hypothesis testing makes decisions about specified values of population parameters. Hypothesis tests are used when trying to determine whether a parameter is less than, more than, or not equal to a specified value. Proper interpretation of a confidence interval, however, can also indicate whether a parameter is less than, more than, or not equal to a specified value. For example, in this section, you tested whether the population mean fill amount was different from 368 grams by using Equation (9.1) on page 330:

$$Z_{STAT} = \frac{\overline{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$$

Instead of testing the null hypothesis that $\mu = 368$ grams, you can reach the same conclusion by constructing a confidence interval estimate of μ. If the hypothesized value of $\mu = 368$ is contained within the interval, you do not reject the null hypothesis because 368 would not be considered an unusual value. However, if the hypothesized value does not fall into the interval, you reject the null hypothesis because $\mu = 368$ grams is then considered an unusual value. Using Equation (8.1) on page 283 and the following data:

$$n = 25, \overline{X} = 372.5 \text{ grams}, \sigma = 15 \text{ grams}$$

for a confidence level of 95% (i.e., $\alpha = 0.05$),

$$\overline{X} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$372.5 \pm (1.96)\frac{15}{\sqrt{25}}$$

$$372.5 \pm 5.88$$

so that

$$366.62 \le \mu \le 378.38$$

Because the interval includes the hypothesized value of 368 grams, you do not reject the null hypothesis. There is insufficient evidence that the mean fill amount over the entire filling process is not 368 grams. You reached the same decision by using two-tail hypothesis testing.

Can You Ever Know the Population Standard Deviation?

The end of Section 8.1 on page 285 discussed how learning a confidence interval estimation method that required knowing σ, the population standard deviation, served as an effective introduction to the concept of a confidence interval. That passage then revealed that you would be unlikely to use that procedure for most practical applications for several reasons.

Likewise, for most practical applications, you are unlikely to use a hypothesis-testing method that requires knowing σ. If you knew the population standard deviation, you would also know the population mean and would not need to form a hypothesis about the mean and then test that hypothesis. So why study a hypothesis testing of the mean that requires that σ is known? Using such a test makes it much easier to explain the fundamentals of hypothesis testing. With a known population standard deviation, you can use the normal distribution and compute p-values using the tables of the normal distribution.

Because it is important that you understand the concept of hypothesis testing when reading the rest of this book, review this section carefully—even if you anticipate never having a practical reason to use the test represented by Equation (9.1).

Problems for Section 9.1

LEARNING THE BASICS

9.1 If you use a 0.05 level of significance in a (two-tail) hypothesis test, what will you decide if $Z_{STAT} = -0.76$?

9.2 If you use a 0.05 level of significance in a (two-tail) hypothesis test, what will you decide if $Z_{STAT} = +2.21$?

9.3 If you use a 0.10 level of significance in a (two-tail) hypothesis test, what is your decision rule for rejecting a null hypothesis that the population mean is 500 if you use the *Z* test?

9.4 If you use a 0.01 level of significance in a (two-tail) hypothesis test, what is your decision rule for rejecting $H_0: \mu = 12.5$ if you use the *Z* test?

9.5 What is your decision in Problem 9.4 if $Z_{STAT} = -2.61$?

9.6 What is the *p*-value if, in a two-tail hypothesis test, $Z_{STAT} = +2.00$?

9.7 In Problem 9.6, what is your statistical decision if you test the null hypothesis at the 0.10 level of significance?

9.8 What is the *p*-value if, in a two-tail hypothesis test, $Z_{STAT} = -1.38$?

APPLYING THE CONCEPTS

9.9 In the U.S. legal system, a defendant is presumed innocent until proven guilty. Consider a null hypothesis, H_0, that the defendant is innocent, and an alternative hypothesis, H_1, that the defendant is guilty. A jury has two possible decisions: Convict the defendant (i.e., reject the null hypothesis) or do not convict the defendant (i.e., do not reject the null hypothesis). Explain the meaning of the risks of committing either a Type I or Type II error in this example.

9.10 Suppose the defendant in Problem 9.9 is presumed guilty until proven innocent, as in some other judicial systems. How do the null and alternative hypotheses differ from those in Problem 9.9? What are the meanings of the risks of committing either a Type I or Type II error here?

9.11 Many consumer groups feel that the U.S. Food and Drug Administration (FDA) drug approval process is too easy and, as a result, too many drugs are approved that are later found to be unsafe. On the other hand, a number of industry lobbyists have pushed for a more lenient approval process so that pharmaceutical companies can get new drugs approved more easily and quickly. Consider a null hypothesis that a new, unapproved drug is unsafe and an alternative hypothesis that a new, unapproved drug is safe.

a. Explain the risks of committing a Type I or Type II error.
b. Which type of error are the consumer groups trying to avoid? Explain.
c. Which type of error are the industry lobbyists trying to avoid? Explain.
d. How would it be possible to lower the chances of both Type I and Type II errors?

9.12 As a result of complaints from both students and faculty about lateness, the registrar at a large university wants to determine whether the scheduled break between classes should be changed and, therefore, is ready to undertake a study. Until now, the registrar has believed that there should be 20 minutes between scheduled classes. State the null hypothesis, H_0, and the alternative hypothesis, H_1.

9.13 Do students at your school study more than, less than, or about the same as students at other business schools? *BusinessWeek* reported that at the top 50 business schools, students studied an average of 14.6 hours per week. (Data extracted from "Cracking the Books," Special Report/Online Extra, **www.businessweek.com**, March 19, 2007.) Set up a hypothesis test to try to prove that the mean number of hours studied at your school is different from the 14.6-hour-per-week benchmark reported by *BusinessWeek*.

a. State the null and alternative hypotheses.
b. What is a Type I error for your test?
c. What is a Type II error for your test?

SELF Test **9.14** The quality-control manager at a light bulb factory needs to determine whether the mean life of a large shipment of light bulbs is equal to 375 hours. The population standard deviation is 100 hours. A random sample of 64 light bulbs indicates a sample mean life of 350 hours.

a. At the 0.05 level of significance, is there evidence that the mean life is different from 375 hours?
b. Compute the *p*-value and interpret its meaning.
c. Construct a 95% confidence interval estimate of the population mean life of the light bulbs.
d. Compare the results of (a) and (c). What conclusions do you reach?

9.15 Suppose that in Problem 9.14, the standard deviation is 120 hours.

a. Repeat (a) through (d) of Problem 9.14, assuming a standard deviation of 120 hours.
b. Compare the results of (a) to those of Problem 9.14.

9.16 The manager of a paint supply store wants to determine whether the mean amount of paint contained in 1-gallon cans purchased from a nationally known manufacturer is actually 1 gallon. You know from the manufacturer's specifications that the standard deviation of the amount of paint is 0.02 gallon. You select a random sample of 50 cans, and the mean amount of paint per 1-gallon can is 0.995 gallon.

a. Is there evidence that the mean amount is different from 1.0 gallon? (Use $\alpha = 0.01$.)
b. Compute the *p*-value and interpret its meaning.

c. Construct a 99% confidence interval estimate of the population mean amount of paint.
d. Compare the results of (a) and (c). What conclusions do you reach?

9.17 Suppose that in Problem 9.16, the standard deviation is 0.012 gallon.

a. Repeat (a) through (d) of Problem 9.16, assuming a standard deviation of 0.012 gallon.
b. Compare the results of (a) to those of Problem 9.16.

9.2 *t* Test of Hypothesis for the Mean (σ Unknown)

In virtually all hypothesis-testing situations concerning the population mean, μ, you do not know the population standard deviation, σ. Instead, you use the sample standard deviation, S. If you assume that the population is normally distributed, the sampling distribution of the mean follows a t distribution with $n - 1$ degrees of freedom, and you use the ***t* test for the mean**. If the population is not normally distributed, you can still use the t test if the sample size is large enough for the Central Limit Theorem to take effect (see Section 7.4). Equation (9.2) defines the test statistic for determining the difference between the sample mean, $\bar{X}$, and the population mean, μ, when using the sample standard deviation, S.

t TEST FOR THE MEAN (σ UNKNOWN)

$$t_{STAT} = \frac{\bar{X} - \mu}{\dfrac{S}{\sqrt{n}}} \tag{9.2}$$

where the t_{STAT} test statistic follows a t distribution having $n - 1$ degrees of freedom.

To illustrate the use of the t test for the mean, return to the Chapter 8 Saxon Home Improvement scenario on page 279. The business objective is to determine whether the mean amount per sales invoice is unchanged from the $120 of the past five years. As an accountant for the company, you need to determine whether this amount changes. In other words, the hypothesis test is used to try to determine whether the mean amount per sales invoice is increasing or decreasing.

The Critical Value Approach

To perform this two-tail hypothesis test, you use the six-step method listed in Exhibit 9.1 on page 332.

Step 1 You define the following hypotheses:

$$H_0: \mu = \$120$$
$$H_1: \mu \neq \$120$$

The alternative hypothesis contains the statement you are trying to prove. If the null hypothesis is rejected, then there is statistical evidence that the population mean amount per sales invoice is no longer $120. If the statistical conclusion is "do not reject H_0," then you will conclude that there is insufficient evidence to prove that the mean amount differs from the long-term mean of $120.

Step 2 You collect the data from a sample of $n = 12$ sales invoices. You decide to use $\alpha = 0.05$.

Step 3 Because σ is unknown, you use the t distribution and the t_{STAT} test statistic. You must assume that the population of sales invoices is normally distributed because the sample size of 12 is too small for the Central Limit Theorem to take effect. This assumption is discussed on page 340.

Step 4 For a given sample size, n, the test statistic t_{STAT} follows a t distribution with $n - 1$ degrees of freedom. The critical values of the t distribution with $12 - 1 = 11$ degrees of freedom are found in Table E.3, as illustrated in Table 9.2 and Figure 9.6. The alternative hypothesis, $H_1: \mu \neq \$120$, has two tails. The area in the rejection region of the

t distribution's left (lower) tail is 0.025, and the area in the rejection region of the t distribution's right (upper) tail is also 0.025.

From the t table as given in Table E.3, a portion of which is shown in Table 9.2, the critical values are ± 2.2010. The decision rule is

$$\text{Reject } H_0 \text{ if } t_{STAT} < -2.2010$$
$$\text{or if } t_{STAT} > +2.2010;$$
$$\text{otherwise, do not reject } H_0.$$

TABLE 9.2

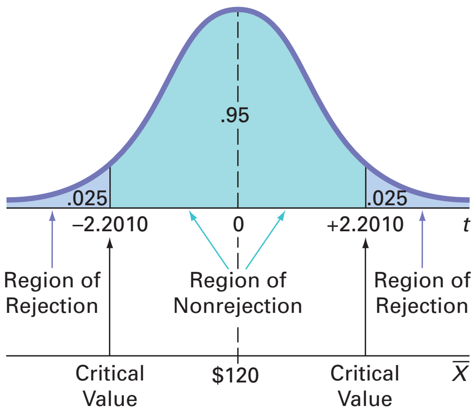
Determining the Critical Value from the t Table for an Area of 0.025 in Each Tail, with 11 Degrees of Freedom

	Cumulative Probabilities					
	.75	.90	.95	$\boxed{.975}$	.99	.995
	Upper-Tail Areas					
Degrees of Freedom	.25	.10	.05	$\boxed{.025}$	.01	.005
1	1.0000	3.0777	6.3138	12.7062	31.8207	63.6574
2	0.8165	1.8856	2.9200	4.3027	6.9646	9.9248
3	0.7649	1.6377	2.3534	3.1824	4.5407	5.8409
4	0.7407	1.5332	2.1318	2.7764	3.7469	4.6041
5	0.7267	1.4759	2.0150	2.5706	3.3649	4.0322
6	0.7176	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.7111	1.4149	1.8946	2.3646	2.9980	3.4995
8	0.7064	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.7027	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693
$\boxed{11}$	0.6974	1.3634	1.7959	$\boxed{2.2010}$	2.7181	3.1058

Source: Extracted from Table E.3.

FIGURE 9.6

Testing a hypothesis about the mean (σ unknown) at the 0.05 level of significance with 11 degrees of freedom



Step 5 You organize and store the data from a random sample of 12 sales invoices in **Invoices**:

108.98 152.22 111.45 110.59 127.46 107.26
93.32 91.97 111.56 75.71 128.58 135.11

Using Equations (3.1) and (3.7) on pages 97 and 103,

$$\bar{X} = \$112.85 \text{ and } S = \$20.80$$

From Equation (9.2) on page 338,

$$t_{STAT} = \frac{\bar{X} - \mu}{\dfrac{S}{\sqrt{n}}} = \frac{112.85 - 120}{\dfrac{20.80}{\sqrt{12}}} = -1.1908$$

Figure 9.7 shows the results for this test of hypothesis, as computed by Excel and Minitab.

FIGURE 9.7

Excel and Minitab results for the *t* test of sales invoices

	A	B	
1	t Test for the Hypothesis of the Mean		
2			
3	Data		
4	Null Hypothesis μ=	120	
5	Level of Significance	0.05	
6	Sample Size	12	
7	Sample Mean	112.85	
8	Sample Standard Deviation	20.8	
9			
10	Intermediate Calculations		
11	Standard Error of the Mean	6.0044	=B8/SQRT(B6)
12	Degrees of Freedom	11	=B6 - 1
13	*t* Test Statistic	-1.1908	=(B7 - B4)/B11
14			
15	Two-Tail Test		
16	Lower Critical Value	-2.2010	=-TINV(B5, B12)
17	Upper Critical Value	2.2010	=TINV(B5, B12)
18	*p*-Value	0.2588	=TDIST(ABS(B13), B12, 2)
19	Do not reject the null hypothesis		=IF(B18 < B5, "Reject the null hypothesis", "Do not reject the null hypothesis")

One-Sample T

```
Test of mu = 120 vs not = 120

 N    Mean  StDev  SE Mean        95% CI           T      P
12  112.85  20.80     6.00  (99.63, 126.07)    -1.19  0.259
```

Step 6 Because $-2.2010 < t_{STAT} = -1.1908 < 2.2010$, you do not reject H_0. You have insufficient evidence to conclude that the mean amount per sales invoice differs from \$120. The audit suggests that the mean amount per invoice has not changed.

The *p*-Value Approach

To perform this two-tail hypothesis test, you use the five-step method listed in Exhibit 9.2 on page 335.

Step 1–3 These steps are the same as in the critical value approach.

Step 4 From the Figure 9.7 results, the $t_{STAT} = -1.19$ and p-value $= 0.2588$.

Step 5 Because the p-value of 0.2588 is greater than $\alpha = 0.05$, you do not reject H_0. The data provide insufficient evidence to conclude that the mean amount per sales invoice differs from \$120. The audit suggests that the mean amount per invoice has not changed. The p-value indicates that if the null hypothesis is true, the probability that a sample of 12 invoices could have a sample mean that differs by \$7.15 or more from the stated \$120 is 0.2588. In other words, if the mean amount per sales invoice is truly \$120, then there is a 25.88% chance of observing a sample mean below \$112.85 or above \$127.15.

In the preceding example, it is incorrect to state that there is a 25.88% chance that the null hypothesis is true. Remember that the p-value is a conditional probability, calculated by *assuming* that the null hypothesis is true. In general, it is proper to state the following:

> If the null hypothesis is true, there is a (p-value) × 100% chance of observing a test statistic at least as contradictory to the null hypothesis as the sample result.

Checking the Normality Assumption

You use the t test when the population standard deviation, σ, is not known and is estimated using the sample standard deviation, S. To use the t test, you assume that the data represent a random sample from a population that is normally distributed. In practice, as long as the sample size is not very small and the population is not very skewed, the t distribution provides a good approximation of the sampling distribution of the mean when σ is unknown.

There are several ways to evaluate the normality assumption necessary for using the t test. You can examine how closely the sample statistics match the normal distribution's theoretical properties. You can also construct a histogram, stem-and-leaf display, boxplot, or normal probability plot to visualize the distribution of the sales invoice amounts. For details on evaluating normality, see Section 6.3 on pages 230–234.

Figures 9.8 through 9.10 show the descriptive statistics, boxplot, and normal probability plot for the sales invoice data.

FIGURE 9.8

Excel and Minitab descriptive statistics for the sales invoice data

	A	B
1	*Invoice Amount*	
2		
3	Mean	112.8508
4	Standard Error	6.0039
5	Median	111.02
6	Mode	#N/A
7	Standard Deviation	20.7980
8	Sample Variance	432.5565
9	Kurtosis	0.1727
10	Skewness	0.1336
11	Range	76.51
12	Minimum	75.71
13	Maximum	152.22
14	Sum	1354.21
15	Count	12

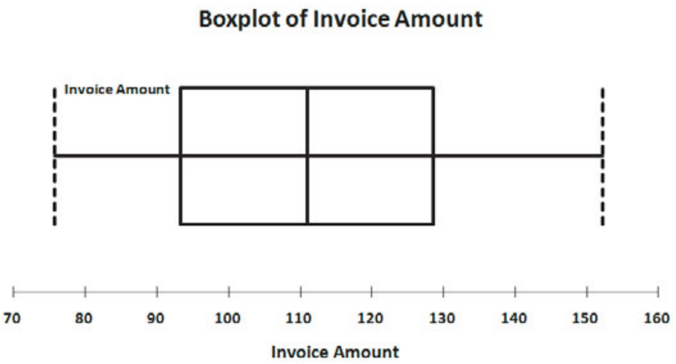
Descriptive Statistics: Invoice Amount

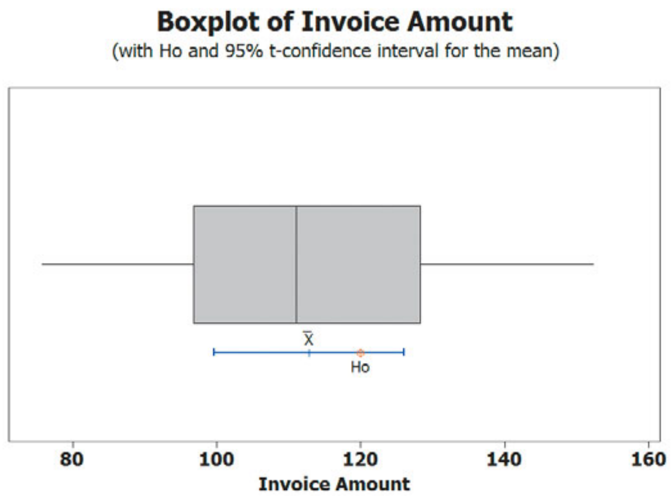
```
                Total
Variable        Count    Mean  StDev  Variance  CoefVar  Minimum     Q1  Median
Invoice Amount     12  112.85  20.80    432.56    18.43    75.71  96.80  111.02

Variable            Q3  Maximum  Range    IQR  Skewness  Kurtosis
Invoice Amount  128.30   152.22  76.51  31.50      0.13      0.17
```

FIGURE 9.9

Excel and Minitab boxplots for the sales invoice data



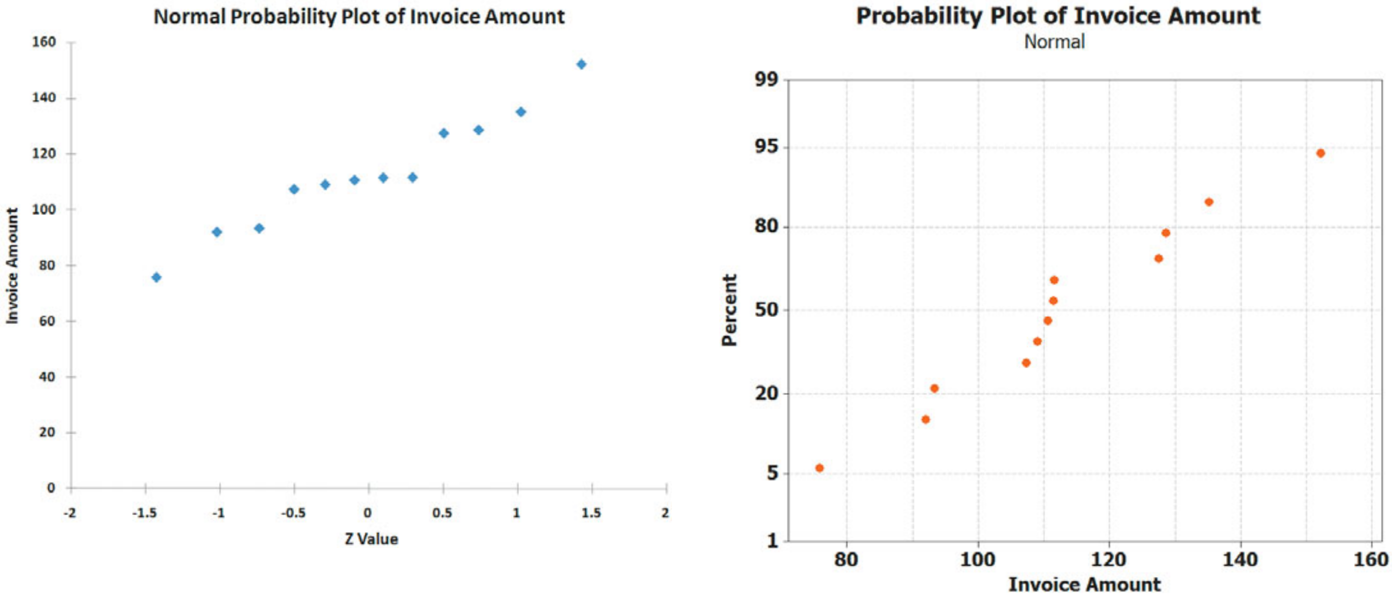


The mean is very close to the median, and the points on the normal probability plots on page 342 appear to be increasing approximately in a straight line. The boxplots appear to be approximately symmetrical. Thus, you can assume that the population of sales invoices is approximately normally distributed. The normality assumption is valid, and therefore the auditor's results are valid.

The t test is a **robust** test. A robust test does not lose power if the shape of the population departs somewhat from a normal distribution, particularly when the sample size is large enough to enable the test statistic t to be influenced by the Central Limit Theorem (see Section 7.4). However, you can reach erroneous conclusions and can lose statistical power if you use the t test incorrectly. If the sample size, n, is small (i.e., less than 30) and you cannot easily make the assumption that the underlying population is at least approximately normally distributed, then *nonparametric* testing procedures are more appropriate (see references 1 and 2).

FIGURE 9.10

Excel and Minitab normal probability plots for the sales invoice data



Problems for Section 9.2

LEARNING THE BASICS

9.18 If, in a sample of $n = 16$ selected from a normal population, $\bar{X} = 56$ and $S = 12$, what is the value of t_{STAT} if you are testing the null hypothesis $H_0: \mu = 50$?

9.19 In Problem 9.18, how many degrees of freedom does the t test have?

9.20 In Problems 9.18 and 9.19, what are the critical values of t if the level of significance, α, is 0.05 and the alternative hypothesis, H_1, is $\mu \neq 50$?

9.21 In Problems 9.18, 9.19, and 9.20, what is your statistical decision if the alternative hypothesis, H_1, is $\mu \neq 50$?

9.22 If, in a sample of $n = 16$ selected from a left-skewed population, $\bar{X} = 65$, and $S = 21$, would you use the t test to test the null hypothesis $H_0: \mu = 60$? Discuss.

9.23 If, in a sample of $n = 160$ selected from a left-skewed population, $\bar{X} = 65$, and $S = 21$, would you use the t test to test the null hypothesis $H_0: \mu = 60$? Discuss.

APPLYING THE CONCEPTS

SELF Test **9.24** You are the manager of a restaurant for a fast-food franchise. Last month, the mean waiting time at the drive-through window for branches in your geographical region, as measured from the time a customer places an order until the time the customer receives the order, was 3.7 minutes. You select a random sample of 64 orders. The sample mean waiting time is 3.57 minutes, with a sample standard deviation of 0.8 minute.

a. At the 0.05 level of significance, is there evidence that the population mean waiting time is different from 3.7 minutes?

b. Because the sample size is 64, do you need to be concerned about the shape of the population distribution when conducting the t test in (a)? Explain.

9.25 A manufacturer of chocolate candies uses machines to package candies as they move along a filling line. Although the packages are labeled as 8 ounces, the company wants the packages to contain a mean of 8.17 ounces so that virtually none of the packages contain less than 8 ounces. A sample of 50 packages is selected periodically, and the packaging process is stopped if there is evidence that the mean amount packaged is different from 8.17 ounces. Suppose that in a particular sample of 50 packages, the mean amount dispensed is 8.159 ounces, with a sample standard deviation of 0.051 ounce.

a. Is there evidence that the population mean amount is different from 8.17 ounces? (Use a 0.05 level of significance.)

b. Determine the p-value and interpret its meaning.

9.26 A stationery store wants to estimate the mean retail value of greeting cards that it has in its inventory. A random sample of 100 greeting cards indicates a mean value of \$2.55 and a standard deviation of \$0.44.

a. Is there evidence that the population mean retail value of the greeting cards is different from \$2.50? (Use a 0.05 level of significance.)

b. Determine the p-value and interpret its meaning.

9.27 The U.S. Department of Transportation requires tire manufacturers to provide performance information on tire sidewalls to help prospective buyers make their purchasing decisions. One very important piece of information is the tread wear index, which indicates the tire's resistance to tread wear. A tire with a grade of 200 should last twice as long, on average, as a tire with a grade of 100.

A consumer organization wants to test the actual tread wear index of a brand name of tires that claims "graded 200" on the sidewall of the tire. A random sample of $n = 18$ indicates a sample mean tread wear index of 195.3 and a sample standard deviation of 21.4.

a. Is there evidence that the population mean tread wear index is different from 200? (Use a 0.05 level of significance.)
b. Determine the p-value and interpret its meaning.

9.28 The file **FastFood** contains the amount that a sample of nine customers spent for lunch ($) at a fast-food restaurant:

4.20	5.03	5.86	6.45	7.38	7.54	8.46	8.47	9.87

a. At the 0.05 level of significance, is there evidence that the mean amount spent for lunch is different from $6.50?
b. Determine the p-value in (a) and interpret its meaning.
c. What assumption must you make about the population distribution in order to conduct the t test in (a) and (b)?
d. Because the sample size is 9, do you need to be concerned about the shape of the population distribution when conducting the t test in (a)? Explain.

9.29 In New York State, savings banks are permitted to sell a form of life insurance called savings bank life insurance (SBLI). The approval process consists of underwriting, which includes a review of the application, a medical information bureau check, possible requests for additional medical information and medical exams, and a policy compilation stage in which the policy pages are generated and sent to the bank for delivery. The ability to deliver approved policies to customers in a timely manner is critical to the profitability of this service. During a period of one month, a random sample of 27 approved policies is selected, and the total processing time, in days, is recorded (and stored in **Insurance**):

73	19	16	64	28	28	31	90	60	56	31	56	22	18
45	48	17	17	17	91	92	63	50	51	69	16	17	

a. In the past, the mean processing time was 45 days. At the 0.05 level of significance, is there evidence that the mean processing time has changed from 45 days?
b. What assumption about the population distribution is needed in order to conduct the t test in (a)?
c. Construct a boxplot or a normal probability plot to evaluate the assumption made in (b).
d. Do you think that the assumption needed in order to conduct the t test in (a) is valid? Explain.

9.30 The following data (in **Drink**) represent the amount of soft-drink filled in a sample of 50 consecutive 2-liter bottles. The results, listed horizontally in the order of being filled, were

2.109	2.086	2.066	2.075	2.065	2.057	2.052	2.044	2.036	2.038
2.031	2.029	2.025	2.029	2.023	2.020	2.015	2.014	2.013	2.014
2.012	2.012	2.012	2.010	2.005	2.003	1.999	1.996	1.997	1.992
1.994	1.986	1.984	1.981	1.973	1.975	1.971	1.969	1.966	1.967
1.963	1.957	1.951	1.951	1.947	1.941	1.941	1.938	1.908	1.894

a. At the 0.05 level of significance, is there evidence that the mean amount of soft drink filled is different from 2.0 liters?
b. Determine the p-value in (a) and interpret its meaning.
c. In (a), you assumed that the distribution of the amount of soft drink filled was normally distributed. Evaluate this assumption by constructing a boxplot or a normal probability plot.
d. Do you think that the assumption needed in order to conduct the t test in (a) is valid? Explain.
e. Examine the values of the 50 bottles in their sequential order, as given in the problem. Does there appear to be a pattern to the results? If so, what impact might this pattern have on the validity of the results in (a)?

9.31 One of the major measures of the quality of service provided by any organization is the speed with which it responds to customer complaints. A large family-held department store selling furniture and flooring, including carpet, had undergone a major expansion in the past several years. In particular, the flooring department had expanded from 2 installation crews to an installation supervisor, a measurer, and 15 installation crews. The store had the business objective of improving its response to complaints. The variable of interest was defined as the number of days between when the complaint was made and when it was resolved. Data were collected from 50 complaints that were made in the past year. The data, stored in **Furniture**, are as follows:

54	5	35	137	31	27	152	2	123	81	74	27
11	19	126	110	110	29	61	35	94	31	26	5
12	4	165	32	29	28	29	26	25	1	14	13
13	10	5	27	4	52	30	22	36	26	20	23
33	68										

a. The installation supervisor claims that the mean number of days between the receipt of a complaint and the resolution of the complaint is 20 days. At the 0.05 level of significance, is there evidence that the claim is not true (i.e., that the mean number of days is different from 20)?
b. What assumption about the population distribution is needed in order to conduct the t test in (a)?
c. Construct a boxplot or a normal probability plot to evaluate the assumption made in (b).
d. Do you think that the assumption needed in order to conduct the t test in (a) is valid? Explain.

9.32 A manufacturing company produces steel housings for electrical equipment. The main component part of the housing is a steel trough that is made out of a 14-gauge steel coil. It is produced using a 250-ton progressive punch press with a wipe-down operation that puts two 90-degree forms in the flat steel to make the trough. The distance from one side of the form to the other is critical because of weatherproofing in outdoor applications. The company requires that the width of the trough be between 8.31 inches and 8.61 inches. The file **Trough** contains the widths of the troughs, in inches, for a sample of $n = 49$:

8.312	8.343	8.317	8.383	8.348	8.410	8.351	8.373	8.481	8.422
8.476	8.382	8.484	8.403	8.414	8.419	8.385	8.465	8.498	8.447
8.436	8.413	8.489	8.414	8.481	8.415	8.479	8.429	8.458	8.462
8.460	8.444	8.429	8.460	8.412	8.420	8.410	8.405	8.323	8.420
8.396	8.447	8.405	8.439	8.411	8.427	8.420	8.498	8.409	

a. At the 0.05 level of significance, is there evidence that the mean width of the troughs is different from 8.46 inches?
b. What assumption about the population distribution is needed in order to conduct the *t* test in (a)?
c. Evaluate the assumption made in (b).
d. Do you think that the assumption needed in order to conduct the *t* test in (a) is valid? Explain.

9.33 One operation of a steel mill is to cut pieces of steel into parts that are used in the frame for front seats in an automobile. The steel is cut with a diamond saw and requires the resulting parts must be cut to be within $\pm$ 0.005 inch of the length specified by the automobile company. The file **Steel** contains a sample of 100 steel parts. The measurement reported is the difference, in inches, between the actual length of the steel part, as measured by a laser measurement device, and the specified length of the steel part. For example, a value of –0.002 represents a steel part that is 0.002 inch shorter than the specified length.
a. At the 0.05 level of significance, is there evidence that the mean difference is not equal to 0.0 inches?
b. Construct a 95% confidence interval estimate of the population mean. Interpret this interval.
c. Compare the conclusions reached in (a) and (b).
d. Because $n = 100$, do you have to be concerned about the normality assumption needed for the *t* test and *t* interval?

9.34 In Problem 3.67 on page 135, you were introduced to a tea-bag-filling operation. An important quality characteristic of interest for this process is the weight of the tea in the individual bags. The file **Teabags** contains an ordered array of the weight, in grams, of a sample of 50 tea bags produced during an eight-hour shift.
a. Is there evidence that the mean amount of tea per bag is different from 5.5 grams? (Use $\alpha = 0.01$.)
b. Construct a 99% confidence interval estimate of the population mean amount of tea per bag. Interpret this interval.
c. Compare the conclusions reached in (a) and (b).

9.35 Although many people think they can put a meal on the table in a short period of time, an article reported that they end up spending about 40 minutes doing so. (Data extracted from N. Hellmich, "Americans Go for the Quick Fix for Dinner," *USA Today*, February 14, 2006.) Suppose another study is conducted to test the validity of this statement. A sample of 25 people is selected, and the length of time to prepare and cook dinner (in minutes) is recorded, with the following results (in **Dinner**):

44.0	51.9	49.7	40.0	55.5	33.0	43.4	41.3	45.2	40.7	41.1	49.1	30.9
45.2	55.3	52.1	55.1	38.8	43.1	39.2	58.6	49.8	43.2	47.9	46.6	

a. Is there evidence that the population mean time to prepare and cook dinner is different from 40 minutes? Use the *p*-value approach and a level of significance of 0.05.
b. What assumption about the population distribution is needed in order to conduct the *t* test in (a)?
c. Make a list of the various ways you could evaluate the assumption noted in (b).
d. Evaluate the assumption noted in (b) and determine whether the *t* test in (a) is valid.

9.3 One-Tail Tests

In Section 9.1, hypothesis testing was used to examine the question of whether the population mean amount of cereal filled is 368 grams. The alternative hypothesis $(H_1 : \mu \neq 368)$ contains two possibilities: Either the mean is less than 368 grams or the mean is more than 368 grams. For this reason, the rejection region is divided into the two tails of the sampling distribution of the mean. In Section 9.2, a two-tail test was used to determine whether the mean amount per invoice had changed from \$120.

In contrast to these two examples, many situations require an alternative hypothesis that focuses on a *particular direction*. For example, the population mean is *less than* a specified value. One such situation involves the business problem concerning the service time at the drive-through window of a fast-food restaurant. The speed with which customers are served is of critical importance to the success of the service (see **www.qsrmagazine.com/reports/drive-thru_time_study**). In one past study, McDonald's had a mean service time of 174.22 seconds, which was only ninth best in the industry. Suppose that McDonald's began a quality improvement effort to reduce the service time by deploying an improved drive-through service process in a sample of 25 stores. Because McDonald's would want to institute the new process in all of its stores only if the test sample saw a *decreased* drive-through time, the entire rejection region is located in the lower tail of the distribution.

The Critical Value Approach

You wish to determine whether the new drive-through process has a mean that is less than 174.22 seconds. To perform this one-tail hypothesis test, you use the six-step method listed in Exhibit 9.1 on page 332.

Step 1 You define the null and alternative hypotheses:

$$H_0: \mu \geq 174.22$$
$$H_1: \mu < 174.22$$

The alternative hypothesis contains the statement for which you are trying to find evidence. If the conclusion of the test is "reject H_0," there is statistical evidence that the mean drive-through time is less than the drive-through time in the old process. This would be reason to change the drive-through process for the entire population of stores. If the conclusion of the test is "do not reject H_0," then there is insufficient evidence that the mean drive-through time in the new process is significantly less than the drive-through time in the old process. If this occurs, there would be insufficient reason to institute the new drive-through process in the population of stores.

Step 2 You collect the data by selecting a drive-through time sample of $n = 25$ stores. You decide to use $\alpha = 0.05$.

Step 3 Because σ is unknown, you use the t distribution and the t_{STAT} test statistic. You need to assume that the service time is normally distributed because only a sample of 25 drive-through times is selected.

Step 4 The rejection region is entirely contained in the lower tail of the sampling distribution of the mean because you want to reject H_0 only when the sample mean is significantly less than 174.22 seconds. When the entire rejection region is contained in one tail of the sampling distribution of the test statistic, the test is called a **one-tail test**, or **directional test**. If the alternative hypothesis includes the *less than* sign, the critical value of t is negative. As shown in Table 9.3 and Figure 9.11, because the entire rejection region is in the lower tail of the t distribution and contains an area of 0.05, due to the symmetry of the t distribution, the critical value of the t test statistic with $25 - 1 = 24$ degrees of freedom is -1.7109. The decision rule is

$$\text{Reject } H_0 \text{ if } t_{STAT} < -1.7109;$$
$$\text{otherwise, do not reject } H_0.$$

TABLE 9.3

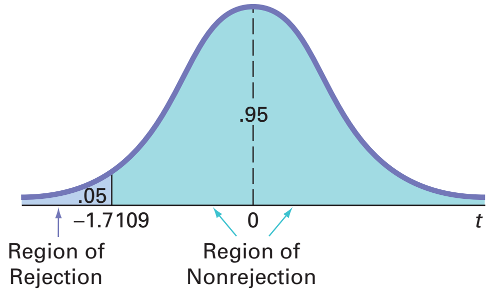
Determining the Critical Value from the t Table for an Area of 0.05 in the Lower Tail, with 24 Degrees of Freedom

	Cumulative Probabilities					
	.75	**.90**	**$\boxed{.95}$**	**.975**	**.99**	**.995**
	Upper-Tail Areas					
Degrees of Freedom	**.25**	**.10**	**$\boxed{.05}$**	**.025**	**.01**	**.005**
1	1.0000	3.0777	6.3138	12.7062	31.8207	63.6574
2	0.8165	1.8856	2.9200	4.3027	6.9646	9.9248
3	0.7649	1.6377	2.3534	3.1824	4.5407	5.8409
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
23	0.6853	1.3195	1.7139	2.0687	2.4999	2.8073
$\boxed{24}$	0.6848	1.3178	$\boxed{1.7109}$	2.0639	2.4922	2.7969
25	0.6844	1.3163	1.7081	2.0595	2.4851	2.7874

Source: Extracted from Table E.3.

FIGURE 9.11

One-tail test of hypothesis for a mean (σ unknown) at the 0.05 level of significance



Step 5 From the sample of 25 stores you selected, you find that the sample mean service time at the drive-through equals 162.96 seconds and the sample standard deviation equals 20.2 seconds. Using $n = 25$, $\bar{X} = 162.96$, $S = 20.2$, and Equation (9.2) on page 338,

$$t_{STAT} = \frac{\bar{X} - \mu}{\dfrac{S}{\sqrt{n}}} = \frac{162.96 - 174.22}{\dfrac{20.2}{\sqrt{25}}} = -2.7871$$

Step 6 Because $t_{STAT} = -2.7871 < -1.7109$, you reject the null hypothesis (see Figure 9.11). You conclude that the mean service time at the drive-through is less than 174.22 seconds. There is sufficient evidence to change the drive-through process for the entire population of stores.

The *p*-Value Approach

Use the five steps listed in Exhibit 9.2 on page 335 to illustrate the t test for the drive-through time study using the p-value approach.

Step 1–3 These steps are the same as in the critical value approach on page 332.

Step 4 $t_{STAT} = -2.7871$ (see step 5 of the critical value approach). Because the alternative hypothesis indicates a rejection region entirely in the lower tail of the sampling distribution, to compute the p-value, you need to find the probability that the t_{STAT} test statistic will be less than –2.7871. Figure 9.12 shows that the p-value is 0.0051.

FIGURE 9.12

Excel and Minitab t test results for the drive-through time study

	A		B	
1	t Test for the Hypothesis of the Mean			
2				
3	Data			
4	Null Hypothesis	μ=	174.22	
5	Level of Significance		0.05	
6	Sample Size		25	
7	Sample Mean		162.96	
8	Sample Standard Deviation		20.2	
9				
10	Intermediate Calculations			
11	Standard Error of the Mean		4.0400	=B8/SQRT(B6)
12	Degrees of Freedom		24	=B6 - 1
13	*t* Test Statistic		-2.7871	=(B7 - B4)/B11
14				
15	Lower-Tail Test			
16	Lower Critical Value		-1.7109	=-TINV(2 * B5, B12)
17	*p*-Value		0.0051	=IF(B13 < 0, E22, E23)
18	Reject the null hypothesis			=IF(B17 < B5, "Reject the null hypothesis", "Do not reject the null hypothesis")

Not shown
Cell E22: =TDIST(ABS(B13), B12, 1)
Cell E23: =1 - E22

One-Sample T

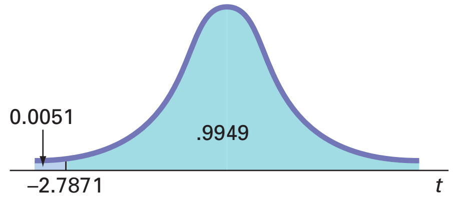
```
Test of mu = 174.22 vs < 174.22

                                    95% Upper
 N    Mean   StDev   SE Mean        Bound       T       P
25  162.96   20.20      4.04       169.87   -2.79   0.005
```

Step 5 The p-value of 0.0051 is less than $\alpha = 0.05$ (see Figure 9.13). You reject H_0 and conclude that the mean service time at the drive-through is less than 174.22 seconds. There is sufficient evidence to change the drive-through process for the entire population of stores.

FIGURE 9.13

Determining the p-value for a one-tail test



EXAMPLE 9.5

A One-Tail Test for the Mean

A company that manufactures chocolate bars is particularly concerned that the mean weight of a chocolate bar is not greater than 6.03 ounces. A sample of 50 chocolate bars is selected; the sample mean is 6.034 ounces, and the sample standard deviation is 0.02 ounces. Using the $\alpha = 0.01$ level of significance, is there evidence that the population mean weight of the chocolate bars is greater than 6.03 ounces?

SOLUTION Using the critical value approach, listed in Exhibit 9.1 on page 332,

Step 1 First, you define your hypotheses:

$$H_0 : \mu \le 6.03$$
$$H_1 : \mu > 6.03$$

Step 2 You collect the data from a sample of $n = 50$. You decide to use $\alpha = 0.01$.

Step 3 Because σ is unknown, you use the t distribution and the t_{STAT} test statistic.

Step 4 The rejection region is entirely contained in the upper tail of the sampling distribution of the mean because you want to reject H_0 only when the sample mean is significantly greater than 6.03 ounces. Because the entire rejection region is in the upper tail of the t distribution and contains an area of 0.01, the critical value of the t distribution with $50 - 1 = 49$ degrees of freedom is 2.4049 (see Table E.3).

The decision rule is

$$\text{Reject } H_0 \text{ if } t_{STAT} > 2.4049;$$
$$\text{otherwise, do not reject } H_0.$$

Step 5 From your sample of 50 chocolate bars, you find that the sample mean weight is 6.034 ounces, and the sample standard deviation is 0.02 ounces. Using $n = 50$, $\bar{X} = 6.034$, $S = 0.02$, and Equation (9.2) on page 338,

$$t_{STAT} = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{6.034 - 6.03}{\frac{0.02}{\sqrt{50}}} = 1.414$$

Step 6 Because $t_{STAT} = 1.414 < 2.4049$, or using Microsoft Excel or Minitab, the p-value is $0.0818 > 0.01$, you do not reject the null hypothesis. There is insufficient evidence to conclude that the population mean weight is greater than 6.03 ounces.

To perform one-tail tests of hypotheses, you must properly formulate H_0 and H_1. A summary of the null and alternative hypotheses for one-tail tests is as follows:

- The null hypothesis, H_0, represents the status quo or the current belief in a situation.
- The alternative hypothesis, H_1, is the opposite of the null hypothesis and represents a research claim or specific inference you would like to prove.
- If you reject the null hypothesis, you have statistical proof that the alternative hypothesis is correct.
- If you do not reject the null hypothesis, you have failed to prove the alternative hypothesis. The failure to prove the alternative hypothesis, however, does not mean that you have proven the null hypothesis.
- The null hypothesis always refers to a specified value of the *population parameter* (such as μ), not to a *sample statistic* (such as $\bar{X}$).

- The statement of the null hypothesis *always* contains an equal sign regarding the specified value of the parameter (e.g., $H_0: \mu \geq 174.22$).
- The statement of the alternative hypothesis *never* contains an equal sign regarding the specified value of the parameter (e.g., $H_1: \mu < 174.22$).

Problems for Section 9.3

LEARNING THE BASICS

9.36 In a one-tail hypothesis test where you reject H_0 only in the *upper* tail, what is the *p*-value if $Z_{STAT} = +2.00$?

9.37 In Problem 9.36, what is your statistical decision if you test the null hypothesis at the 0.05 level of significance?

9.38 In a one-tail hypothesis test where you reject H_0 only in the *lower* tail, what is the *p*-value if $Z_{STAT} = -1.38$?

9.39 In Problem 9.38, what is your statistical decision if you test the null hypothesis at the 0.01 level of significance?

9.40 In a one-tail hypothesis test where you reject H_0 only in the *lower* tail, what is the *p*-value if $Z_{STAT} = +1.38$?

9.41 In Problem 9.40, what is the statistical decision if you test the null hypothesis at the 0.01 level of significance?

9.42 In a one-tail hypothesis test where you reject H_0 only in the *upper* tail, what is the critical value of the *t*-test statistic with 10 degrees of freedom at the 0.01 level of significance?

9.43 In Problem 9.42, what is your statistical decision if $t_{STAT} = +2.39$?

9.44 In a one-tail hypothesis test where you reject H_0 only in the *lower* tail, what is the critical value of the t_{STAT} test statistic with 20 degrees of freedom at the 0.01 level of significance?

9.45 In Problem 9.44, what is your statistical decision if $t_{STAT} = -1.15$?

APPLYING THE CONCEPTS

9.46 In a recent year, the Federal Communications Commission reported that the mean wait for repairs for Verizon customers was 36.5 hours. In an effort to improve this service, suppose that a new repair service process was developed. This new process, used for a sample of 100 repairs, resulted in a sample mean of 34.5 hours and a sample standard deviation of 11.7 hours.

a. Is there evidence that the population mean amount is less than 36.5 hours? (Use a 0.05 level of significance.)
b. Determine the *p*-value and interpret its meaning.

9.47 In a recent year, the Federal Communications Commission reported that the mean wait for repairs for AT&T customers was 25.3 hours. In an effort to improve this service, suppose that a new repair service process was developed. This new process, used for a sample of 100 repairs, resulted in a sample mean of 22.3 hours and a sample standard deviation of 8.3 hours.

a. Is there evidence that the population mean amount is less than 25.3 hours? (Use a 0.05 level of significance.)
b. Determine the *p*-value and interpret its meaning.

SELF Test **9.48** Southside Hospital in Bay Shore, New York, commonly conducts stress tests to study the heart muscle after a person has a heart attack. Members of the diagnostic imaging department conducted a quality improvement project with the objective of reducing the turnaround time for stress tests. Turnaround time is defined as the time from when a test is ordered to when the radiologist signs off on the test results. Initially, the mean turnaround time for a stress test was 68 hours. After incorporating changes into the stress-test process, the quality improvement team collected a sample of 50 turnaround times. In this sample, the mean turnaround time was 32 hours, with a standard deviation of 9 hours. (Data extracted from E. Godin, D. Raven, C. Sweetapple, and F. R. Del Guidice, "Faster Test Results," *Quality Progress*, January 2004, 37(1), pp. 33–39.)

a. If you test the null hypothesis at the 0.01 level of significance, is their evidence that the new process has reduced turnaround time?
b. Interpret the meaning of the *p*-value in this problem.

9.49 You are the manager of a restaurant that delivers pizza to college dormitory rooms. You have just changed your delivery process in an effort to reduce the mean time between the order and completion of delivery from the current 25 minutes. A sample of 36 orders using the new delivery process yields a sample mean of 22.4 minutes and a sample standard deviation of 6 minutes.

a. Using the six-step critical value approach, at the 0.05 level of significance, is there evidence that the population mean delivery time has been reduced below the previous population mean value of 25 minutes?
b. At the 0.05 level of significance, use the five-step *p*-value approach.
c. Interpret the meaning of the *p*-value in (b).
d. Compare your conclusions in (a) and (b).

9.50 The per-store daily customer count (i.e., the mean number of customers in a store in one day) for a nationwide convenience store chain that operates nearly 10,000 stores has been steady, at 900, for some time. To increase the customer count, the chain is considering cutting prices for coffee beverages by approximately half. The small size will now be \$0.59 instead of \$0.99, and the medium size will be \$0.69 instead of \$1.19. Even with this reduction in price, the chain will have a 40% gross margin on coffee. To test the new initiative, the chain has reduced coffee prices in a sample of 34 stores, where customer counts have been running almost exactly at the national average of 900. After four weeks, the sample stores stabilize at a mean customer count of 974 and a standard deviation of 96. This increase seems like a substantial amount to you, but it also seems like a pretty small sample. Do you think reducing coffee prices is a good strategy for increasing the mean customer count?

a. State the null and alternative hypotheses.

b. Explain the meaning of the Type I and Type II errors in the context of this scenario.

c. At the 0.01 level of significance, is there evidence that reducing coffee prices is a good strategy for increasing the mean customer count?

d. Interpret the meaning of the p-value in (c).

9.51 The population mean waiting time to check out of a supermarket has been 10.73 minutes. Recently, in an effort to reduce the waiting time, the supermarket has experimented with a system in which there is a single waiting line with multiple checkout servers. A sample of 100 customers was selected, and their mean waiting time to check out was 9.52 minutes, with a sample standard deviation of 5.8 minutes.

a. At the 0.05 level of significance, using the critical value approach to hypothesis testing, is there evidence that the population mean waiting time to check out is less than 10.73 minutes?

b. At the 0.05 level of significance, using the p-value approach to hypothesis testing, is there evidence that the population mean waiting time to check out is less than 10.73 minutes?

c. Interpret the meaning of the p-value in this problem.

d. Compare your conclusions in (a) and (b).

9.4 *Z* Test of Hypothesis for the Proportion

In some situations, you want to test a hypothesis about the proportion of events of interest in the population, π, rather than test the population mean. To begin, you select a random sample and compute the **sample proportion**, $p = X/n$. You then compare the value of this statistic to the hypothesized value of the parameter, π, in order to decide whether to reject the null hypothesis. If the number of events of interest (X) and the number of events that are not of interest $(n - X)$ are each at least five, the sampling distribution of a proportion approximately follows a normal distribution. You use the ***Z* test for the proportion** given in Equation (9.3) to perform the hypothesis test for the difference between the sample proportion, p, and the hypothesized population proportion, π.

Z TEST FOR THE PROPORTION

$$Z_{STAT} = \frac{p - \pi}{\sqrt{\dfrac{\pi(1 - \pi)}{n}}} \tag{9.3}$$

where

$$p = \text{Sample proportion} = \frac{X}{n} = \frac{\text{Number of events of interest in the sample}}{\text{Sample size}}$$

$$\pi = \text{Hypothesized proportion of events of interest in the population}$$

The Z_{STAT} test statistic approximately follows a standardized normal distribution when X and $(n - X)$ are each at least 5.

Alternatively, by multiplying the numerator and denominator by n, you can write the Z_{STAT} test statistic in terms of the number of events of interest, X, as shown in Equation (9.4).

Z TEST FOR THE PROPORTION IN TERMS OF THE NUMBER OF EVENTS OF INTEREST

$$Z_{STAT} = \frac{X - n\pi}{\sqrt{n\pi(1 - \pi)}} \tag{9.4}$$

The Critical Value Approach

To illustrate the Z test for a proportion, consider a survey conducted for American Express that sought to determine the reasons adults wanted Internet access while on vacation. (Data extracted from "Wired Vacationers," *USA Today*, June 4, 2010, p. 1A.) Of 2,000 adults, 1,540 said that they wanted Internet access so they could check personal e-mail while on vacation. A survey conducted in the previous year indicated that 75% of adults wanted Internet access so they could check personal e-mail while on vacation. Is there evidence that the percentage of adults who wanted Internet access to check personal e-mail while on vacation has changed from the previous year? To investigate this question, the null and alternative hypotheses are follows:

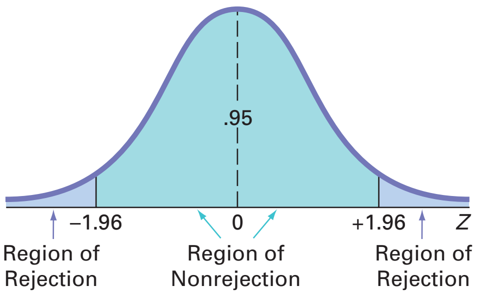
$H_0: \pi = 0.75$ (i.e., the proportion of adults who want Internet access to check personal email while on vacation has not changed from the previous year)

$H_1: \pi \neq 0.75$ (i.e., the proportion of adults who want Internet access to check personal email while on vacation has changed from the previous year)

Because you are interested in determining whether the population proportion of adults who want Internet access to check personal email while on vacation has changed from 0.75 in the previous year, you use a two-tail test. If you select the $\alpha = 0.05$ level of significance, the rejection and nonrejection regions are set up as in Figure 9.14, and the decision rule is

$$\text{Reject } H_0 \text{ if } Z_{STAT} < -1.96 \text{ or if } Z_{STAT} > +1.96;$$
$$\text{otherwise, do not reject } H_0.$$

FIGURE 9.14
Two-tail test of hypothesis for the proportion at the 0.05 level of significance



Because 1,540 of the 2,000 adults stated that they wanted Internet access to check personal email while on vacation,

$$p = \frac{1{,}540}{2{,}000} = 0.77$$

Since $X = 1{,}540$ and $n - X = 460$, each > 5, using Equation (9.3),

$$Z_{STAT} = \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} = \frac{0.77 - 0.75}{\sqrt{\frac{0.75(1-0.75)}{2{,}000}}} = \frac{0.02}{0.0097} = 2.0656$$

or, using Equation (9.4),

$$Z_{STAT} = \frac{X - n\pi}{\sqrt{n\pi(1-\pi)}} = \frac{1{,}540 - (2{,}000)(0.75)}{\sqrt{2{,}000(0.75)(0.25)}} = \frac{40}{19.3649} = 2.0656$$

Because $Z_{STAT} = 2.0656 > 1.96$, you reject H_0. There is evidence that the population proportion of all adults who want Internet access to check personal e-mail while on vacation has changed from 0.75 in the previous year. Figure 9.15 presents results for these data, as computed by Excel and Minitab.

FIGURE 9.15

Excel and Minitab results for the Z test for whether the proportion of adults who want Internet access to check personal email while on vacation has changed from the previous year

	A	B	
1	Z Test of Hypothesis for the Proportion		
2			
3	Data		
4	Null Hypothesis *p* =	0.75	
5	Level of Significance	0.05	
6	Number of Items of Interest	1540	
7	Sample Size	2000	
8			
9	Intermediate Calculations		
10	Sample Proportion	0.7700	=B6/B7
11	Standard Error	0.0097	=SQRT(B4*(1 - B4)/B7)
12	Z Test Statistic	2.0656	=(B10 - B4)/B11
13			
14	Two-Tail Test		
15	Lower Critical Value	-1.9600	=NORMSINV(B5/2)
16	Upper Critical Value	1.9600	=NORMSINV(1 - B5/2)
17	*p*-Value	0.0389	=2 * (1 - NORMSDIST(ABS(B12)))
18	Reject the null hypothesis		=IF(B17 < B5, "Reject the null hypothesis", "Do not reject the null hypothesis")

Test and CI for One Proportion

```
Test of p = 0.75 vs p not = 0.75

Sample     X     N  Sample p        95% CI              Z-Value  P-Value
1       1540  2000  0.770000  (0.751557, 0.788443)        2.07    0.039

Using the normal approximation.
```

The *p*-Value Approach

As an alternative to the critical value approach, you can compute the *p*-value. For this two-tail test in which the rejection region is located in the lower tail and the upper tail, you need to find the area below a Z value of –2.0656 and above a Z value of +2.0656. Figure 9.15 reports a *p*-value of 0.0389. Because this value is less than the selected level of significance ($\alpha = 0.05$), you reject the null hypothesis.

EXAMPLE 9.6

Testing a Hypothesis for a Proportion

A fast-food chain has developed a new process to ensure that orders at the drive-through are filled correctly. The business problem is defined as determining whether the new process can increase the percentage of orders processed correctly. The previous process filled orders correctly 85% of the time. Data are collected from a sample of 100 orders using the new process. The results indicate that 94 orders were filled correctly. At the 0.01 level of significance, can you conclude that the new process has increased the proportion of orders filled correctly?

SOLUTION The null and alternative hypotheses are

$H_0: \pi \leq 0.85$ (i.e., the population proportion of orders filled correctly using the new process is less than or equal to 0.85)
$H_1: \pi > 0.85$ (i.e., the population proportion of orders filled correctly using the new process is greater than 0.85)

Since $X = 94$ and $n - X = 6$, both > 5, using Equation (9.3) on page 349,

$$p = \frac{X}{n} = \frac{94}{100} = 0.94$$

$$Z_{STAT} = \frac{p - \pi}{\sqrt{\dfrac{\pi(1 - \pi)}{n}}} = \frac{0.94 - 0.85}{\sqrt{\dfrac{0.85(1 - 0.85)}{100}}} = \frac{0.09}{0.0357} = 2.52$$

The *p*-value for $Z_{STAT} > 2.52$ is 0.0059.

Using the critical value approach, you reject H_0 if $Z_{STAT} > 2.33$. Using the *p*-value approach, you reject H_0 if *p*-value < 0.01. Because $Z_{STAT} = 2.52 > 2.33$ or the *p-value* $= 0.0059 < 0.01$, you reject H_0. You have evidence that the new process has increased the proportion of correct orders above 0.85.

Problems for Section 9.4

LEARNING THE BASICS

9.52 If, in a random sample of 400 items, 88 are defective, what is the sample proportion of defective items?

9.53 In Problem 9.52, if the null hypothesis is that 20% of the items in the population are defective, what is the value of Z_{STAT}?

9.54 In Problems 9.52 and 9.53, suppose you are testing the null hypothesis $H_0: \pi = 0.20$ against the two-tail alternative hypothesis $H_1: \pi \neq 0.20$ and you choose the level of significance $\alpha = 0.05$. What is your statistical decision?

APPLYING THE CONCEPTS

9.55 The U.S. Department of Education reports that 46% of full-time college students are employed while attending college. (Data extracted from "The Condition of Education 2009," *National Center for Education Statistics*, **nces.ed.gov**.) A recent survey of 60 full-time students at Miami University found that 29 were employed.

a. Use the five-step p-value approach to hypothesis testing and a 0.05 level of significance to determine whether the proportion of full-time students at Miami University is different from the national norm of 0.46.

b. Assume that the study found that 36 of the 60 full-time students were employed and repeat (a). Are the conclusions the same?

9.56 Online magazines make it easy for readers to link to an advertiser's website directly from an advertisement placed in the digital magazine. A recent survey indicated that 56% of online magazine readers have clicked on an advertisement and linked directly to the advertiser's website. The survey was based on a sample size of $n = 6{,}403$. (Data extracted from "Metrics," *EContent*, January/February, 2007, p. 20.)

a. Use the five-step p-value approach to try to determine whether there is evidence that more than half of all the readers of online magazines have linked to an advertiser's website. (Use the 0.05 level of significance.)

b. Suppose that the sample size was only $n = 100$, and, as before, 56% of the online magazine readers indicated that they had clicked on an advertisement to link directly to the advertiser's website. Use the five-step p-value approach to try to determine whether there is evidence that more than half of all the readers of online magazines have linked to an advertiser's website. (Use the 0.05 level of significance.)

c. Discuss the effect that sample size has on hypothesis testing.

d. What do you think are your chances of rejecting any null hypothesis concerning a population proportion if a sample size of $n = 20$ is used?

9.57 One of the issues facing organizations is increasing diversity throughout the organization. One of the ways to evaluate an organization's success at increasing diversity is to compare the percentage of employees in the organization in a particular position with a specific background to the percentage in a particular position with that specific background in the general workforce. Recently, a large academic medical center determined that 9 of 17 employees in a particular position were female, whereas 55% of the employees for this position in the general workforce were female. At the 0.05 level of significance, is there evidence that the proportion of females in this position at this medical center is different from what would be expected in the general workforce?

SELF Test **9.58** Of 1,000 respondents aged 24 to 35, 65% reported that they preferred to "look for a job in a place where I would like to live" rather than "look for the best job I can find, the place where I live is secondary." (Data extracted from L. Belkin, "What Do Young Jobseekers Want? (Something Other Than a Job)," *The New York Times*, September 6, 2007, p. G2.) At the 0.05 level of significance, is there evidence that the proportion of all young jobseekers aged 24 to 35 who preferred to "look for a job in a place where I would like to live" rather than "look for the best job I can find, the place where I live is secondary" is different from 60%?

9.59 The telephone company wants to investigate the desirability of beginning a marketing campaign that would offer customers the right to purchase an additional telephone line at a substantially reduced installation cost. The campaign will be initiated if there is evidence that more than 20% of the customers would consider purchasing an additional telephone line if it were made available at a substantially reduced installation cost. A random sample of 500 households is selected. The results indicate that 135 of the households would purchase the additional telephone line at a reduced installation cost.

a. At the 0.05 level of significance, is there evidence that more than 20% of the customers would purchase the additional telephone line?

b. How would the manager in charge of promotional programs concerning residential customers use the results in (a)?

9.60 A study by the Pew Internet and American Life Project (**pewinternet.org**) found that Americans had a complex and ambivalent attitude toward technology. (Data extracted from M. Himowitz, "How to Tell What Kind of Tech User You Are," *Newsday*, May 27, 2007, p. F6.) The study reported that 8% of the respondents were "Omnivores" who are gadget lovers, text messengers, and online gamers (often with their own blogs or web pages), video makers, and YouTube posters. You believe that the percentage of students at your school who are Omnivores is greater than 8%, and you plan to carry out a study to prove that this is so.

a. State the null and alternative hypotheses.

b. You select a sample of 200 students at your school and find that 30 students can be classified as Omnivores. Use either the six-step critical value hypothesis-testing approach or the five-step p-value approach to determine at the 0.05 level of significance whether there is evidence that the percentage of Omnivores at your school is greater than 8%.

9.5 Potential Hypothesis-Testing Pitfalls and Ethical Issues

To this point, you have studied the fundamental concepts of hypothesis testing. You have used hypothesis testing to analyze differences between sample statistics and hypothesized population parameters in order to make business decisions concerning the underlying population characteristics. You have also learned how to evaluate the risks involved in making these decisions.

When planning to carry out a hypothesis test based on a survey, research study, or designed experiment, you must ask several questions to ensure that you use proper methodology. You need to raise and answer questions such as the following in the planning stage:

- What is the goal of the survey, study, or experiment? How can you translate the goal into a null hypothesis and an alternative hypothesis?
- Is the hypothesis test a two-tail test or one-tail test?
- Can you select a random sample from the underlying population of interest?
- What types of data will you collect in the sample? Are the variables numerical or categorical?
- At what level of significance should you conduct the hypothesis test?
- Is the intended sample size large enough to achieve the desired power of the test for the level of significance chosen?
- What statistical test procedure should you use and why?
- What conclusions and interpretations can you reach from the results of the hypothesis test?

Failing to consider these questions early in the planning process can lead to biased or incomplete results. Proper planning can help ensure that the statistical study will provide objective information needed to make good business decisions.

Statistical Significance Versus Practical Significance

You need to make a distinction between the existence of a statistically significant result and its practical significance in a field of application. Sometimes, due to a very large sample size, you may get a result that is statistically significant but has little practical significance. For example, suppose that prior to a national marketing campaign focusing on a series of expensive television commercials, you believe that the proportion of people who recognize your brand is 0.30. At the completion of the campaign, a survey of 20,000 people indicates that 6,168 recognized your brand. A one-tail test trying to prove that the proportion is now greater than 0.30 results in a p-value of 0.0047, and the correct statistical conclusion is that the proportion of consumers recognizing your brand name has now increased. Was the campaign successful? The result of the hypothesis test indicates a statistically significant increase in brand awareness, but is this increase practically important? The population proportion is now estimated at $6{,}168/20{,}000 = 0.3084$, or 30.84%. This increase is less than 1% more than the hypothesized value of 30%. Did the large expenses associated with the marketing campaign produce a result with a meaningful increase in brand awareness? Because of the minimal real-world impact that an increase of less than 1% has on the overall marketing strategy and the huge expenses associated with the marketing campaign, you should conclude that the campaign was not successful. On the other hand, if the campaign increased brand awareness from 30% to 50%, you could conclude that the campaign was successful.

Reporting of Findings

In conducting research, you should document both good and bad results. You should not just report the results of hypothesis tests that show statistical significance but omit those for which there is insufficient evidence in the findings. In instances in which there is insufficient evidence to reject H_0, you must make it clear that this does not prove that the null hypothesis is true. What the result does indicate is that with the sample size used, there is not enough information to *disprove* the null hypothesis.

Ethical Issues

You need to distinguish between poor research methodology and unethical behavior. Ethical considerations arise when the hypothesis-testing process is manipulated. Some of the areas where ethical issues can arise include the use of human subjects in experiments, the data collection method, the type of test (one-tail or two-tail test), the choice of the level of significance, the cleansing and discarding of data, and the failure to report pertinent findings.

9.6 *Online Topic:* The Power of a Test

Section 9.1 defines Type I and Type II errors and the power of a test. To examine the power of a test in greater depth, read the **Section 9.6** online topic file that is available on this book's companion website. (See Appendix C to learn how to access the online topic files.)

USING STATISTICS @ Oxford Cereals, Part II Revisited

As the plant operations manager for Oxford Cereals, you were responsible for the cereal-filling process. It was your responsibility to adjust the process when the mean fill weight in the population of boxes deviated from the company specification of 368 grams. Because weighing all the cereal boxes would be too time-consuming and impractical, you needed to select and weigh a sample of boxes and conduct a hypothesis test.

You determined that the null hypothesis should be that the population mean fill was 368 grams. If the mean weight of the sampled boxes were sufficiently above or below the expected 368-gram mean specified by Oxford Cereals, you would reject the null hypothesis in favor of the alternative hypothesis that the mean fill was different from 368 grams. If this happened, you would stop production and take whatever action is necessary to correct the problem. If the null hypothesis was not rejected, you would continue to believe in the status quo—that the process was working correctly—and therefore take no corrective action.

Before proceeding, you considered the risks involved with hypothesis tests. If you rejected a true null hypothesis, you would make a Type I error and conclude that the population mean fill was not 368 when it actually was 368. This error would result in adjusting the filling process even though the process was working properly. If you did not reject a false null hypothesis, you would make a Type II error and conclude that the population mean fill was 368 when it actually was not 368. Here, you would allow the process to continue without adjustment even though the process was not working properly.

After collecting a random sample of 25 cereal boxes, you used the six-step critical value approach to hypothesis testing. Because the test statistic fell into the nonrejection region, you did not reject the null hypothesis. You concluded that there was insufficient evidence to prove that the mean fill differed from 368 grams. No corrective action on the filling process was needed.

SUMMARY

This chapter presented the foundation of hypothesis testing. You learned how to perform tests on the population mean and on the population proportion. The chapter developed both the critical value approach and the p-value approach to hypothesis testing.

In deciding which test to use, you should ask the following question: Does the test involve a numerical variable or a categorical variable? If the test involves a numerical variable, use the t test for the mean. If the test involves a categorical variable, use the Z test for the proportion. Table 9.4 provides a list of hypothesis tests covered in the chapter.

TABLE 9.4

Summary of Topics in Chapter 9

	Type of Data	
Type of Analysis	**Numerical**	**Categorical**
Hypothesis test concerning a single parameter	t test of hypothesis for the mean (Section 9.2)	Z test of hypothesis for the proportion (Section 9.4)

KEY EQUATIONS

***Z* Test for the Mean (*σ* Known)**

$$Z_{STAT} = \frac{\bar{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}} \tag{9.1}$$

***t* Test for the Mean (*σ* Unknown)**

$$t_{STAT} = \frac{\bar{X} - \mu}{\dfrac{S}{\sqrt{n}}} \tag{9.2}$$

***Z* Test for the Proportion**

$$Z_{STAT} = \frac{p - \pi}{\sqrt{\dfrac{\pi(1 - \pi)}{n}}} \tag{9.3}$$

***Z* Test for the Proportion in Terms of the Number of Events of Interest**

$$Z_{STAT} = \frac{X - n\pi}{\sqrt{n\pi(1 - \pi)}} \tag{9.4}$$

KEY TERMS

alternative hypothesis (H_1) 326
β risk 329
confidence coefficient 329
critical value 328
directional test 345
hypothesis testing 326
level of significance (α) 329
null hypothesis (H_0) 326
one-tail test 345
p-value 333
power of a statistical test 329
region of nonrejection 328
region of rejection 328
robust 341
sample proportion 349
t test for the mean 338
test statistic 328
two-tail test 331
Type I error 328
Type II error 328
Z test for the mean 330
Z test for the proportion 349

CHAPTER REVIEW PROBLEMS

CHECKING YOUR UNDERSTANDING

9.61 What is the difference between a null hypothesis, H_0, and an alternative hypothesis, H_1?

9.62 What is the difference between a Type I error and a Type II error?

9.63 What is meant by the power of a test?

9.64 What is the difference between a one-tail test and a two-tail test?

9.65 What is meant by a p-value?

9.66 How can a confidence interval estimate for the population mean provide conclusions for the corresponding two-tail hypothesis test for the population mean?

9.67 What is the six-step critical value approach to hypothesis testing?

9.68 What is the five-step p-value approach to hypothesis testing?

APPLYING THE CONCEPTS

9.69 An article in *Marketing News* (T. T. Semon, "Consider a Statistical Insignificance Test," *Marketing News*, February 1, 1999) argued that the level of significance used when comparing two products is often too low—that is, sometimes you should be using an α value greater than 0.05. Specifically, the article recounted testing the proportion of potential customers with a preference for product 1 over product 2. The null hypothesis was that the population proportion of potential customers preferring product 1 was 0.50, and the alternative hypothesis was that it was not equal to 0.50. The p-value for the test was 0.22. The article suggested that, in some cases, this should be enough evidence to reject the null hypothesis.

a. State, in statistical terms, the null and alternative hypotheses for this example.
b. Explain the risks associated with Type I and Type II errors in this case.
c. What would be the consequences if you rejected the null hypothesis for a p-value of 0.22?
d. Why do you think the article suggested raising the value of α?
e. What would you do in this situation?
f. What is your answer in (e) if the p-value equals 0.12? What if it equals 0.06?

9.70 La Quinta Motor Inns developed a computer model to help predict the profitability of sites that are being considered as locations for new hotels. If the computer model predicts large profits, La Quinta buys the proposed site and builds a new hotel. If the computer model predicts small or moderate profits, La Quinta chooses not to proceed with that site. (Data extracted from S. E. Kimes and J. A. Fitzsimmons, "Selecting Profitable Hotel Sites at La Quinta Motor Inns," *Interfaces*, Vol. 20, March–April 1990, pp. 12–20.) This decision-making procedure can be expressed in the hypothesis-testing framework. The null hypothesis is that the site is not a profitable location. The alternative hypothesis is that the site is a profitable location.

a. Explain the risks associated with committing a Type I error in this case.
b. Explain the risks associated with committing a Type II error in this case.
c. Which type of error do you think the executives at La Quinta Motor Inns want to avoid? Explain.
d. How do changes in the rejection criterion affect the probabilities of committing Type I and Type II errors?

9.71 Webcredible, a UK-based consulting firm specializing in websites, intranets, mobile devices, and applications, conducted a survey of 1,132 mobile phone users between February and April 2009. The survey found that 52% of mobile phone users are now using the mobile Internet. (Data extracted from "Email and Social Networking Most Popular Mobile Internet Activities," **www.webcredible.co.uk**, May 13, 2009.) The authors of the article imply that the survey proves that more than half of all mobile phone users are now using the mobile Internet.

a. Use the five-step p-value approach to hypothesis testing and a 0.05 level of significance to try to prove that more than half of all mobile phone users are now using the mobile Internet.
b. Based on your result in (a), is the claim implied by the authors valid?
c. Suppose the survey found that 53% of mobile phone users are now using the mobile Internet. Repeat parts (a) and (b).
d. Compare the results of (b) and (c).

9.72 The owner of a gasoline station wants to study gasoline purchasing habits of motorists at his station. He selects a random sample of 60 motorists during a certain week, with the following results:

- The amount purchased was $\bar{X} = 11.3$ gallons, $S = 3.1$ gallons.
- Eleven motorists purchased premium-grade gasoline.

a. At the 0.05 level of significance, is there evidence that the population mean purchase was different from 10 gallons?
b. Determine the p-value in (a).
c. At the 0.05 level of significance, is there evidence that less than 20% of all the motorists at the station purchased premium-grade gasoline?
d. What is your answer to (a) if the sample mean equals 10.3 gallons?
e. What is your answer to (c) if 7 motorists purchased premium-grade gasoline?

9.73 An auditor for a government agency is assigned the task of evaluating reimbursement for office visits to physicians paid by Medicare. The audit was conducted on a sample of 75 of the reimbursements, with the following results:

- In 12 of the office visits, there was an incorrect amount of reimbursement.
- The amount of reimbursement was $\bar{X} = \$93.70$, $S = \$34.55$.

a. At the 0.05 level of significance, is there evidence that the population mean reimbursement was less than $100?
b. At the 0.05 level of significance, is there evidence that the proportion of incorrect reimbursements in the population was greater than 0.10?
c. Discuss the underlying assumptions of the test used in (a).
d. What is your answer to (a) if the sample mean equals $90?
e. What is your answer to (b) if 15 office visits had incorrect reimbursements?

9.74 A bank branch located in a commercial district of a city had the business objective of improving the process for serving customers during the noon-to-1:00 P.M. lunch period. The waiting time (defined as the time the customer enters the line until he or she reaches the teller window) of all customers during this hour is recorded over a period of a week. Data were collected from a random sample of 15 customers, and the results are organized (and stored in **Bank1**) as follows:

4.21	5.55	3.02	5.13	4.77	2.34	3.54	3.20
4.50	6.10	0.38	5.12	6.46	6.19	3.79	

a. At the 0.05 level of significance, is there evidence that the population mean waiting time is less than 5 minutes?
b. What assumption about the population distribution is needed in order to conduct the t test in (a)?
c. Construct a boxplot or a normal probability plot to evaluate the assumption made in (b).
d. Do you think that the assumption needed in order to conduct the t test in (a) is valid? Explain.
e. As a customer walks into the branch office during the lunch hour, she asks the branch manager how long she can expect to wait. The branch manager replies, "Almost certainly not longer than 5 minutes." On the basis of the results of (a), evaluate this statement.

9.75 A manufacturing company produces electrical insulators. If the insulators break when in use, a short circuit is likely to occur. To test the strength of the insulators, destructive testing is carried out to determine how much force is required to break the insulators. Force is measured by observing the number of pounds of force applied to the insulator before it breaks. The following data (stored in **Force**) are from 30 insulators subjected to this testing:

1,870	1,728	1,656	1,610	1,634	1,784	1,522	1,696	1,592	1,662
1,866	1,764	1,734	1,662	1,734	1,774	1,550	1,756	1,762	1,866
1,820	1,744	1,788	1,688	1,810	1,752	1,680	1,810	1,652	1,736

a. At the 0.05 level of significance, is there evidence that the population mean force is greater than 1,500 pounds?
b. What assumption about the population distribution is needed in order to conduct the t test in (a)?
c. Construct a histogram, boxplot, or normal probability plot to evaluate the assumption made in (b).
d. Do you think that the assumption needed in order to conduct the t test in (a) is valid? Explain.

9.76 An important quality characteristic used by the manufacturer of Boston and Vermont asphalt shingles is the amount of moisture the shingles contain when they are packaged. Customers may feel that they have purchased a product lacking in quality if they find moisture and wet shingles inside the packaging. In some cases, excessive moisture can cause the granules attached to the shingle for texture and coloring purposes to fall off the shingle, resulting in appearance problems. To monitor the amount of moisture present, the company conducts moisture tests. A shingle is weighed and then dried. The shingle is then reweighed, and, based on the amount of moisture taken out of the product, the pounds of moisture per 100 square feet are calculated. The company would like to show that the mean moisture content is less than 0.35 pound per 100 square feet. The file **Moisture** includes 36 measurements (in pounds per 100 square feet) for Boston shingles and 31 for Vermont shingles.
a. For the Boston shingles, is there evidence at the 0.05 level of significance that the population mean moisture content is less than 0.35 pound per 100 square feet?
b. Interpret the meaning of the p-value in (a).
c. For the Vermont shingles, is there evidence at the 0.05 level of significance that the population mean moisture content is less than 0.35 pound per 100 square feet?
d. Interpret the meaning of the p-value in (c).
e. What assumption about the population distribution is needed in order to conduct the t tests in (a) and (c)?
f. Construct histograms, boxplots, or normal probability plots to evaluate the assumption made in (a) and (c).
g. Do you think that the assumption needed in order to conduct the t tests in (a) and (c) is valid? Explain.

9.77 Studies conducted by the manufacturer of Boston and Vermont asphalt shingles have shown product weight to be a major factor in the customer's perception of quality. Moreover, the weight represents the amount of raw materials being used and is therefore very important to the company from a cost standpoint. The last stage of the assembly line packages the shingles before the packages are placed on wooden pallets. Once a pallet is full (a pallet for most brands holds 16 squares of shingles), it is weighed, and the measurement is recorded. The file **Pallet** contains the weight (in pounds) from a sample of 368 pallets of Boston shingles and 330 pallets of Vermont shingles.
a. For the Boston shingles, is there evidence that the population mean weight is different from 3,150 pounds?
b. Interpret the meaning of the p-value in (a).
c. For the Vermont shingles, is there evidence that the population mean weight is different from 3,700 pounds?
d. Interpret the meaning of the p-value in (c).
e. In (a) through (d), do you have to worry about the normality assumption? Explain.

9.78 The manufacturer of Boston and Vermont asphalt shingles provides its customers with a 20-year warranty on most of its products. To determine whether a shingle will last through the warranty period, accelerated-life testing is conducted at the manufacturing plant. Accelerated-life testing exposes the shingle to the stresses it would be subject to in a lifetime of normal use in a laboratory setting via an experiment that takes only a few minutes to conduct. In this test, a shingle is repeatedly scraped with a brush for a short period of time, and the shingle granules removed by the brushing are weighed (in grams). Shingles that experience low amounts of granule loss are expected to last longer in normal use than shingles that experience high amounts of granule loss. The file **Granule** contains a sample of 170 measurements made on the company's Boston shingles and 140 measurements made on Vermont shingles.

a. For the Boston shingles, is there evidence that the population mean granule loss is different from 0.50 grams?
b. Interpret the meaning of the *p*-value in (a).
c. For the Vermont shingles, is there evidence that the population mean granule loss is different from 0.50 grams?
d. Interpret the meaning of the *p*-value in (c).
e. In (a) through (d), do you have to worry about the normality assumption? Explain.

REPORT WRITING EXERCISE

9.79 Referring to the results of Problems 9.76 through 9.78 concerning Boston and Vermont shingles, write a report that evaluates the moisture level, weight, and granule loss of the two types of shingles.

MANAGING ASHLAND MULTICOMM SERVICES

Continuing its monitoring of the upload speed first described in the Chapter 6 Managing Ashland MultiComm Services case on page 244, the technical operations department wants to ensure that the mean target upload speed for all Internet service subscribers is at least 0.97 on a standard scale in which the target value is 1.0. Each day, upload speed was measured 50 times, with the following results (stored in **AMS9**).

0.854	1.023	1.005	1.030	1.219	0.977	1.044	0.778	1.122	1.114
1.091	1.086	1.141	0.931	0.723	0.934	1.060	1.047	0.800	0.889
1.012	0.695	0.869	0.734	1.131	0.993	0.762	0.814	1.108	0.805
1.223	1.024	0.884	0.799	0.870	0.898	0.621	0.818	1.113	1.286
1.052	0.678	1.162	0.808	1.012	0.859	0.951	1.112	1.003	0.972

Calculate the sample statistics and determine whether there is evidence that the population mean upload speed is less than 0.97. Write a memo to management that summarizes your conclusions.

DIGITAL CASE

Apply your knowledge about hypothesis testing in this Digital Case, which continues the cereal-fill-packaging dispute first discussed in the Digital Case from Chapter 7.

In response to the negative statements made by the Concerned Consumers About Cereal Cheaters (CCACC) in the Chapter 7 Digital Case, Oxford Cereals recently conducted an experiment concerning cereal packaging. The company claims that the results of the experiment refute the CCACC allegations that Oxford Cereals has been cheating consumers by packaging cereals at less than labeled weights.

Open **OxfordCurrentNews.pdf**, a portfolio of current news releases from Oxford Cereals. Review the relevant press releases and supporting documents. Then answer the following questions:

1. Are the results of the experiment valid? Why or why not? If you were conducting the experiment, is there anything you would change?
2. Do the results support the claim that Oxford Cereals is not cheating its customers?
3. Is the claim of the Oxford Cereals CEO that many cereal boxes contain *more* than 368 grams surprising? Is it true?
4. Could there ever be a circumstance in which the results of the Oxford Cereals experiment *and* the CCACC's results are both correct? Explain

REFERENCES

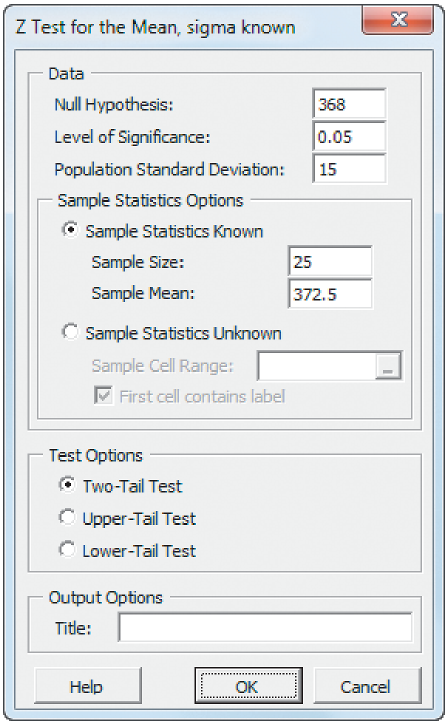
1. Bradley, J. V., *Distribution-Free Statistical Tests* (Upper Saddle River, NJ: Prentice Hall, 1968).
2. Daniel, W., *Applied Nonparametric Statistics*, 2nd ed. (Boston: Houghton Mifflin, 1990).
3. *Microsoft Excel 2010* (Redmond, WA: Microsoft Corp., 2007).
4. *Minitab Release 16* (State College, PA: Minitab Inc., 2010).

CHAPTER 9 EXCEL GUIDE

EG9.1 FUNDAMENTALS of HYPOTHESIS-TESTING METHODOLOGY

PHStat2 Use **Z Test for the Mean, sigma known** to perform the Z test for the mean when σ is known. For example, to perform the Z test for the Figure 9.5 cereal-filling example on page 334, select **PHStat ➔ One-Sample Tests ➔ Z Test for the Mean, sigma known**. In the procedure's dialog box (shown below):

1. Enter **368** as the **Null Hypothesis**.
2. Enter **0.05** as the **Level of Significance**.
3. Enter **15** as the **Population Standard Deviation**.
4. Click **Sample Statistics Known** and enter **25** as the **Sample Size** and **372.5** as the **Sample Mean**.
5. Click **Two-Tail Test**.
6. Enter a **Title** and click **OK**.



For problems that use unsummarized data, click **Sample Statistics Unknown** in step 4 and enter the cell range of the unsummarized data as the **Sample Cell Range**.

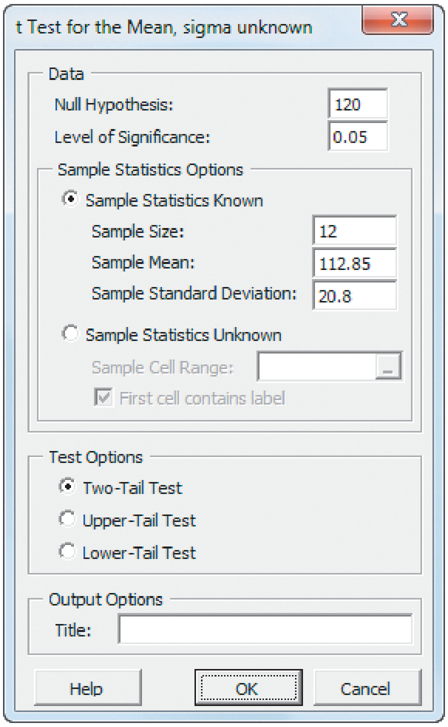
In-Depth Excel Use the **COMPUTE worksheet** of the **Z Mean workbook**, shown in Figure 9.5 on page 334, as a template for performing the two-tail Z test. The worksheet contains the data for the Section 9.1 cereal-filling example. For other problems, change the values in cells B4 through B8 as necessary.

In cells B15 and B16, **NORMSINV(*level of significance* / 2)** and **NORMSINV(1 - *level of significance* / 2)** computes the lower and upper critical values. The expression **2 * (1 − NORMSDIST (*absolute value of the Z test statistic*))** computes the p-value for the two-tail test in cell B17. In cell A18, **IF(*p-value* < *level of significance*, *display reject message*, *display do not reject message*)** determines which message to display in the cell.

EG9.2 *t* TEST of HYPOTHESIS for the MEAN (σ UNKNOWN)

PHStat2 Use ***t* Test for the Mean, sigma unknown** to perform the t test for the mean when σ is unknown. For example, to perform the t test for the Figure 9.7 sales invoice example on page 340, select **PHStat ➔ One-Sample Tests ➔ t Test for the Mean, sigma unknown**. In the procedure's dialog box (shown on the top of page 360):

1. Enter **120** as the **Null Hypothesis**.
2. Enter **0.05** as the **Level of Significance**.
3. Click **Sample Statistics Known** and enter **12** as the **Sample Size**, **112.85** as the **Sample Mean**, and **20.8** as the **Sample Standard Deviation**.
4. Click **Two-Tail Test**.
5. Enter a **Title** and click **OK**.



For problems that use unsummarized data, click **Sample Statistics Unknown** in step 3 and enter the cell range of the unsummarized data as the **Sample Cell Range**.

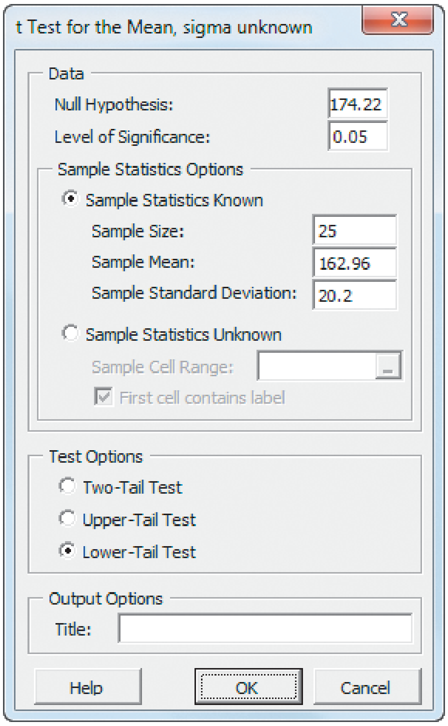
In-Depth Excel Use the **COMPUTE worksheet** of the **T mean workbook**, shown in Figure 9.7 on page 340, as a template for performing the two-tail *t* test. The worksheet contains the data for the Section 9.2 sales invoice example. For other problems, change the values in cells B4 through B8 as necessary.

In cells B16 and B17, the worksheet uses the expressions **-TINV(*level of significance*, *degrees of freedom*)** and **TINV(*level of significance*, *degrees of freedom*)** to compute the lower and upper critical values, respectively. In cell B18, the worksheet uses **TDIST(*absolute value of the t test statistic*, *degrees of freedom*, 2)** to compute the *p*-value. The worksheet also uses an **IF** function to determine which message to display in cell A19.

EG9.3 ONE-TAIL TESTS

PHStat2 Click either **Lower-Tail Test** or **Upper-Tail Test** in the procedure dialog boxes discussed in Sections EG9.1 and EG9.2 to perform a one-tail test. For example, to perform the Figure 9.12 one-tail test for the drive-through time study example on page 346, select **PHStat → One-Sample Tests → t Test for the Mean, sigma unknown**. In the procedure's dialog box (shown below):

1. Enter **174.22** as the **Null Hypothesis**.
2. Enter **0.05** as the **Level of Significance**.
3. Click **Sample Statistics Known** and enter **25** as the **Sample Size**, **162.96** as the **Sample Mean**, and **20.2** as the **Sample Standard Deviation**.
4. Click **Lower-Tail Test**.
5. Enter a **Title** and click **OK**.



In-Depth Excel Modify the functions discussed in Section EG9.1 and EG9.2 to perform one-tail tests. For the Section EG9.1 *Z* test, enter **NORMSINV(*level of significance*)** or **NORMSINV(1 − *level of significance*)** to compute the lower-tail or upper-tail critical value. Enter **NORMSDIST(*Z test statistic*)** or **1 − NORMSDIST(*absolute value of the Z test statistic*)** to compute the lower-tail or upper-tail *p*-value. For the Section EG9.2 *t* test, enter **-TINV(2 * *level of significance*, *degrees of freedom*)** or **TINV(2 * *level of significance*, *degrees of freedom*)** to compute the lower-tail or upper-tail critical values.

Computing *p*-values is more complex. If the *t* test statistic is less than zero, the lower-tail *p*-value is equal to **TDIST(*absolute value of the t test statistic*, *degrees of**

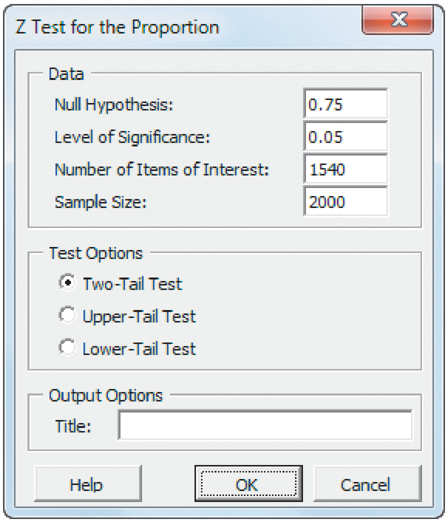
***freedom*, 1)**, and the upper-tail p-value is equal to **1 − TDIST(*absolute value of the t test statistic*, *degrees of freedom*, 1)**. If the t test statistic is greater than or equal to zero, the values are reversed.

Use the **COMPUTE_LOWER worksheet** or the **COMPUTE_UPPER worksheet** of the **Z Mean workbook** or the **T mean workbook** as a template for performing one-tail t tests. Open the **COMPUTE_ALL_FORMULAS worksheet** of either workbook to examine all formulas.

EG9.4 *Z* TEST of HYPOTHESIS for the PROPORTION

PHStat2 Use **Z Test for the Proportion** to perform the Z test of hypothesis for the proportion. For example, to perform the Z test for the Figure 9.15 vacation Internet access study example on page 351, select **PHStat ➔ One-Sample Tests ➔ Z Test for the Proportion**. In the procedure's dialog box (shown in the right column):

1. Enter **0.75** as the **Null Hypothesis**.
2. Enter **0.05** as the **Level of Significance**.
3. Enter **1540** as the **Number of Items of Interest**.
4. Enter **2000** as the **Sample Size**.
5. Click **Two-Tail Test**.
6. Enter a **Title** and click **OK**.



In-Depth Excel Use the **COMPUTE worksheet** of the **Z Proportion workbook**, shown in Figure 9.15 on page 351, as a template for performing the two-tail Z test. The worksheet contains the data for the Section 9.4 vacation Internet access study example. For other problems, change the values in cells B4 through B7 as necessary.

The worksheet uses **NORMSINV(*level of significance* / 2)** and **NORMSINV(1 - *level of significance* / 2)** to compute the lower and upper critical values in cells B15 and B16. In cell B17, the worksheet uses the expression **2 * (1 − NORMSDIST(*absolute value of the Z test statistic*)** to compute the p-value. The worksheet also uses an **IF** function to determine which message to display in cell A18.

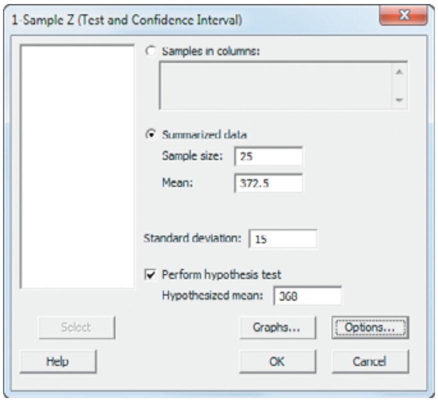
Use the **COMPUTE_LOWER worksheet** or **COMPUTE_UPPER worksheet** as a template for performing one-tail tests. Open to the **COMPUTE_ALL_FORMULAS worksheet** to examine all formulas in the one-tail test worksheets.

CHAPTER 9 MINITAB GUIDE

MG9.1 FUNDAMENTALS of HYPOTHESIS-TESTING METHODOLOGY

Use **1-Sample Z** to perform the Z test for the mean when σ is known. For example, to perform the Z test for the Figure 9.5 cereal-filling example on page 334, select **Stat ➔ Basic Statistics ➔ 1-Sample Z**. In the "1-Sample Z (Test and Confidence Interval)" dialog box (shown below):

1. Click **Summarized data**.
2. Enter **25** in the **Sample size** box and **372.5** in the **Mean** box.
3. Enter **15** in the **Standard deviation** box.
4. Check **Perform hypothesis test** and enter **368** in the **Hypothesized mean** box.
5. Click **Options**.



In the 1-Sample Z - Options dialog box:

6. Enter **95.0** in the **Confidence level** box.
7. Select **not equal** from the **Alternative** drop-down list.
8. Click **OK**.
9. Back in the original dialog box, click **OK**.

For problems that use unsummarized data, open the worksheet that contains the data and replace steps 1 and 2 with these steps:

1. Click **Samples in columns**.
2. Enter the name of the column containing the unsummarized data in the **Samples in column** box.

MG9.2 *t* TEST of HYPOTHESIS for the MEAN (σ UNKNOWN)

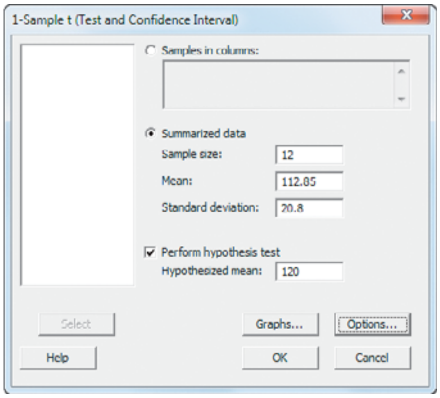
Use ***t* Test for the Mean, sigma unknown** to perform the t test for the mean when σ is unknown. For example, to perform the t test for the Figure 9.7 sales invoice example on page 340, select **Stat ➔ Basic Statistics ➔ 1-Sample t**.

In the 1-Sample t (Test and Confidence Interval) dialog box (shown below):

1. Click **Summarized data**.
2. Enter **12** in the **Sample size** box, **112.85** in the **Mean** box, and **20.8** in the **Standard deviation** box.
3. Check **Perform hypothesis test** and enter **120** in the **Hypothesized mean** box.
4. Click **Options**.

In the 1-Sample t - Options dialog box:

5. Enter **95.0** in the **Confidence level** box.
6. Select **not equal** from the **Alternative** drop-down list.
7. Click **OK**.
8. Back in the original dialog box, click **OK**.



For problems that use unsummarized data, open the worksheet that contains the data and replace steps 1 and 2 with these steps:

1. Click **Samples in columns**.
2. Enter the name of the column containing the unsummarized in the **Samples in column** box.

To create a boxplot of the unsummarized data, replace step 8 with the following steps 8 through 10:

8. Back in the original dialog box, click **Graphs**.
9. In the 1-Sample t - Graphs dialog box, check **Boxplot of data** and then click **OK**.
10. Back in the original dialog box, click **OK.**

MG9.3 ONE-TAIL TESTS

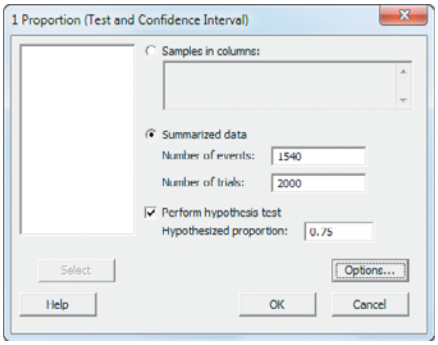
To perform a one-tail test for **1-Sample Z**, select **less than** or **greater than** from the drop-down list in step 7 of the Section MG9.1 instructions.

To perform a one-tail test for **1-Sample t**, select **less than** or **greater than** from the drop-down list in step 6 of the Section MG9.2 instructions.

MG9.4 Z TEST of HYPOTHESIS for the PROPORTION

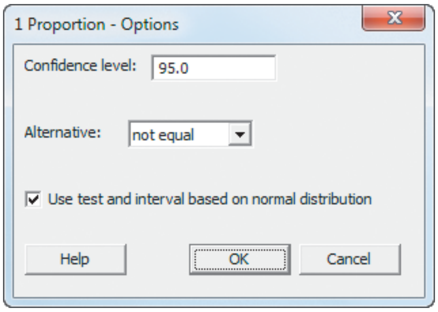
Use **1 Proportion** to perform the *Z* test of hypothesis for the proportion. For example, to perform the *Z* test for the Figure 9.15 vacation Internet access study example on page 351, select **Stat ➔ Basic Statistics ➔ 1 Proportion**. In the 1 Proportion (Test and Confidence Interval) dialog box (shown below):

1. Click **Summarized data**.
2. Enter **1540** in the **Number of events** box and **2000** in the **Number of trials** box.
3. Check **Perform hypothesis test** and enter **0.75** in the **Hypothesized proportion** box.
4. Click **Options**.



In the 1-Proportion - Options dialog box (shown in right column):

5. Enter **95.0** in the **Confidence level** box.
6. Select **not equal** from the **Alternative** drop-down list.
7. Check **Use test and interval based on normal distribution**.
8. Click **OK**.



9. Back in the original dialog box, click **OK**.

To perform a one-tail test, select **less than** or **greater than** from the drop-down list in step 6. For problems that use unsummarized data, open the worksheet that contains the data and replace steps 1 and 2 with these steps:

1. Click **Samples in columns**.
2. Enter the name of the column containing the unsummarized in the **Samples in column** box.