When the makers of Pringles potato chips came out with new Fat-Free Pringles, they wanted the fat-free chips to taste just as good as their already successful regular Pringles. Did they succeed? In an independent effort to answer this question, USA Today hired registered dietitian Diane Wilke to give 44 people a chance to see whether they could tell the difference between the two kinds of Pringles. Each tester was given two bowls of chips—one containing Fat-Free Pringles, the other containing regular Pringles—and nobody was told which was which.

On average, if the two kinds of chips really taste the same, we’d expect such testers to have a 50% chance of correctly identifying the bowl containing the fat-free chips. However, 25 of the 44 testers (56.8%) successfully identified the bowl with the fat-free chips.

Does this result mean that Pringles failed in its attempt to make the products taste the same, or could the difference between the observed 56.8% and the theoretical 50% have happened just by chance? Actually, if the chips really taste the same and we were to repeat this type of test many times, pure chance would lead to about 1/5 of the tests yielding a sample percentage at least as high as the 56.8% observed here. Thus, this particular test would not allow us to rule out the possibility that the chips taste the same. After reading Sections 10.3 and 10.6 of this chapter, you’ll be able to verify how we reached this conclusion. For now, just trust us and read on. Thanks.

In statistics, as in life, nothing is as certain as the presence of uncertainty. However, just because we’re not 100% sure of something, that’s no reason why we can’t reach some conclusions that are highly likely to be true. For example, if a coin were to land heads 20 times in a row, we might be wrong in concluding that it’s unfair, but we’d still be wise to avoid engaging in gambling contests with its owner. In this chapter, we’ll examine the very important process of reaching conclusions based on sample information—in particular, of evaluating hypotheses based on claims like the following:

- Titus Walsh, the director of a municipal transit authority, claims that 35% of the system’s ridership consists of senior citizens. In a recent study, independent researchers find that only 23% of the riders observed are senior citizens. Should the claim of Walsh be considered false?
- Jackson T. Backus has just received a railroad car of canned beets from his grocery supplier, who claims that no more than 20% of the cans are dented. Jackson, a born skeptic, examines a random sample from the shipment and finds that 25% of the cans sampled are dented. Has Mr. Backus bought a batch of botched beets?

Each of the preceding cases raises a question of “believability” that can be examined by the techniques of this chapter. These methods represent inferential statistics, because information from a sample is used in reaching a conclusion about the population from which the sample was drawn.

### Null and Alternative Hypotheses

The first step in examining claims like the preceding is to form a null hypothesis, expressed as $H_0$ (“H sub naught”). The null hypothesis is a statement about the value of a population parameter and is put up for testing in the face of numerical evidence. The null hypothesis is either rejected or fails to be rejected.
The null hypothesis tends to be a “business as usual, nothing out of the ordinary is happening” statement that practically invites you to challenge its truthfulness. In the philosophy of hypothesis testing, the null hypothesis is assumed to be true unless we have statistically overwhelming evidence to the contrary. In other words, it gets the benefit of the doubt.

The alternative hypothesis, $H_1$ (“$H$ sub one”), is an assertion that holds if the null hypothesis is false. For a given test, the null and alternative hypotheses include all possible values of the population parameter, so either one or the other must be false.

There are three possible choices for the set of null and alternative hypotheses to be used for a given test. Described in terms of an (unknown) population mean ($\mu$), they might be listed as shown below. Notice that each null hypothesis has an equality term in its statement (i.e., “$=$”, “$\geq$”, or “$\leq$”).

<table>
<thead>
<tr>
<th>Null Hypothesis</th>
<th>Alternative Hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$: $\mu = $10</td>
<td>$H_1$: $\mu \neq $10</td>
</tr>
<tr>
<td>$H_0$: $\mu \geq $10</td>
<td>$H_1$: $\mu &lt; $10</td>
</tr>
<tr>
<td>$H_0$: $\mu \leq $10</td>
<td>$H_1$: $\mu &gt; $10</td>
</tr>
</tbody>
</table>

(\mu is $10, or it isn’t.)  
(\mu is at least $10, or it is less.)  
(\mu is no more than $10, or it is more.)

Directional and Nondirectional Testing

A directional claim or assertion holds that a population parameter is greater than ($>$), at least ($\geq$), no more than ($\leq$), or less than ($<$) some quantity. For example, Jackson’s supplier claims that no more than 20% of the beet cans are dented.

A nondirectional claim or assertion states that a parameter is equal to some quantity. For example, Titus Walsh claims that 35% of his transit riders are senior citizens.

Directional assertions lead to what are called one-tail tests, where a null hypothesis can be rejected by an extreme result in one direction only. A nondirectional assertion involves a two-tail test, in which a null hypothesis can be rejected by an extreme result occurring in either direction.

Hypothesis Testing and the Nature of the Test

When formulating the null and alternative hypotheses, the nature, or purpose, of the test must also be taken into account. To demonstrate how (1) directionality versus nondirectionality and (2) the purpose of the test can guide us toward the appropriate testing approach, we will consider the two examples at the beginning of the chapter. For each situation, we’ll examine (1) the claim or assertion leading to the test, (2) the null hypothesis to be evaluated, (3) the alternative hypothesis, (4) whether the test will be two-tail or one-tail, and (5) a visual representation of the test itself.

**Titus Walsh**

1. Titus’ assertion: “35% of the riders are senior citizens.”
2. Null hypothesis: $H_0$: $\pi = 0.35$, where $\pi$ = the population proportion. The null hypothesis is identical to his statement since he’s claimed an exact value for the population parameter.
3. Alternative hypothesis: $H_1$: $\pi \neq 0.35$. If the population proportion is not 0.35, then it must be some other value.
4. A two-tail test is used because the null hypothesis is nondirectional.
5. As part (a) of Figure 10.1 shows, $\pi = 0.35$ is at the center of the hypothesized distribution, and a sample with either a very high proportion or a very low proportion of senior citizens would lead to rejection of the null hypothesis. Accordingly, there are reject areas at both ends of the distribution.

### Jackson T. Backus

1. Supplier’s assertion: “No more than 20% of the cans are dented.”
2. Null hypothesis: $H_0: \pi \leq 0.20$, where $\pi$ = the population proportion. In this situation, the null hypothesis happens to be the same as the claim that led to the test. This is not always the case when the test involves a directional claim or assertion.
3. Alternative hypothesis: $H_1: \pi > 0.20$. Jackson’s purpose in conducting the test is to determine whether the population proportion of dented cans could really be greater than 0.20.
4. A one-tail test is used because the null hypothesis is directional.
5. As part (b) of Figure 10.1 shows, a sample with a very high proportion of dented cans would lead to the rejection of the null hypothesis. A one-tail test in which the rejection area is at the right is known as a right-tail test. Note that in part (b) of Figure 10.1, the center of the hypothesized distribution is identified as $\pi = 0.20$. This is the highest value for which the null hypothesis could be true. From Jackson’s standpoint, this may be viewed as somewhat conservative, but remember that the null hypothesis tends to get the benefit of the doubt.
In directional tests, the directionality of the null and alternative hypotheses will be in opposite directions and will depend on the purpose of the test. For example, in the case of Jackson Backus, Jackson was interested in rejecting $H_0: \mu \leq 0.20$ only if evidence suggested $\mu$ to be higher than 0.20. As we proceed with the examples in the chapter, we’ll get more practice in formulating null and alternative hypotheses for both nondirectional and directional tests. Table 10.1 offers general guidelines for proceeding from a verbal statement to typical null and alternative hypotheses.

**Errors in Hypothesis Testing**

Whenever we reject a null hypothesis, there is a chance that we have made a mistake—i.e., that we have rejected a true statement. Rejecting a true null hypothesis is referred to as a Type I error, and our probability of making such an error is represented by the Greek letter alpha ($\alpha$). This probability, which is referred to as the significance level of the test, is of primary concern in hypothesis testing.

On the other hand, we can also make the mistake of failing to reject a false null hypothesis—this is a Type II error. Our probability of making it is represented by the Greek letter beta ($\beta$). Naturally, if we either fail to reject a true null hypothesis or reject a false null hypothesis, we’ve acted correctly. The probability of rejecting a false null hypothesis is called the power of the test, and it will be discussed in Section 10.7. The four possibilities are shown in Table 10.2. In hypothesis testing, there is a necessary trade-off between Type I and Type II errors: For a given sample size, reducing the probability of a Type I error increases the probability of a Type II error, and vice versa. The only sure way to avoid accepting false claims is to never accept any claims. Likewise, the only sure way to avoid rejecting true claims is to never reject any claims. Of course, each of these extreme approaches is impractical, and we must usually compromise by accepting a reasonable risk of committing either type of error.

### Table 10.1

<table>
<thead>
<tr>
<th>Categories of verbal statements and typical null and alternative hypotheses for each.</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>A. Verbal Statement Is an Equality, “=”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example: “Average tire life $\leq 35,000$ miles.”</td>
</tr>
<tr>
<td>$H_0: \mu = 35,000$ miles</td>
</tr>
<tr>
<td>$H_1: \mu \neq 35,000$ miles</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B. Verbal Statement is “$\geq$” or “$\leq$” (not $&gt;$ or $&lt;$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example: “Average tire life $\geq 35,000$ miles.”</td>
</tr>
<tr>
<td>$H_0: \mu \geq 35,000$ miles</td>
</tr>
<tr>
<td>$H_1: \mu &lt; 35,000$ miles</td>
</tr>
</tbody>
</table>

Example: “Average tire life $\leq 35,000$ miles.”

| $H_0: \mu = 35,000$ miles |
| $H_1: \mu > 35,000$ miles |
10.1 What is the difference between a null hypothesis and an alternative hypothesis? Is the null hypothesis always the same as the verbal claim or assertion that led to the test? Why or why not?

10.2 For each of the following pairs of null and alternative hypotheses, determine whether the pair would be appropriate for a hypothesis test. If a pair is deemed inappropriate, explain why.

a. \( H_0: \mu \geq 10, \quad H_1: \mu < 10 \)
b. \( H_0: \mu = 30, \quad H_1: \mu \neq 30 \)
c. \( H_0: \mu > 90, \quad H_1: \mu \leq 90 \)
d. \( H_0: \mu \leq 75, \quad H_1: \mu > 85 \)
e. \( H_0: \bar{x} \geq 15, \quad H_1: \bar{x} < 15 \)
f. \( H_0: \bar{x} = 58, \quad H_1: \bar{x} \neq 58 \)

10.3 For each of the following pairs of null and alternative hypotheses, determine whether the pair would be appropriate for a hypothesis test. If a pair is deemed inappropriate, explain why.

a. \( H_0: \pi \geq 0.30, \quad H_1: \pi < 0.35 \)
b. \( H_0: \pi = 0.72, \quad H_1: \pi = 0.72 \)
c. \( H_0: \pi \leq 0.25, \quad H_1: \pi > 0.25 \)
d. \( H_0: \pi \geq 0.48, \quad H_1: \pi > 0.48 \)
e. \( H_0: \pi \leq 0.70, \quad H_1: \pi > 0.70 \)
f. \( H_0: p \geq 0.65, \quad H_1: p < 0.65 \)

10.4 The president of a company that manufactures central home air conditioning units has told an investigative reporter that at least 85% of its homeowner customers claim to be “completely satisfied” with the overall purchase experience. If the reporter were to subject the president’s statement to statistical scrutiny by questioning a sample of the company’s residential customers, would the test be one-tail or two-tail? What would be the appropriate null and alternative hypotheses?

10.5 On CNN and other news networks, guests often express their opinions in rather strong, persuasive, and sometimes frightening terms. For example, a scientist who strongly believes that global warming is taking place will warn us of the dire consequences (such as rising sea levels, coastal flooding, and global climate change) she foresees if we do not take her arguments seriously. If the scientist is correct, and the world does not take her seriously, would this be a Type I error or a Type II error? Briefly explain your reasoning.

10.6 Many law enforcement agencies use voice-stress analysis to help determine whether persons under interrogation are lying. If the sound frequency of a person’s voice changes when asked a question, the presumption is that the person is being untruthful. For this situation, state the null and alternative hypotheses in verbal terms, then identify what would constitute a Type I error and a Type II error in this situation.

10.7 Following a major earthquake, the city engineer must determine whether the stadium is structurally sound for an upcoming athletic event. If the null hypothesis is “the stadium is structurally sound,” and the alternative hypothesis is “the stadium is not structurally sound,” which type of error (Type I or Type II) would the engineer least like to commit?

10.8 A state representative is reported as saying that about 10% of reported auto thefts involve owners whose cars have not really been stolen, but who are trying to defraud their insurance company. What null and alternative
Chapter 10: Hypothesis Tests Involving a Sample Mean or Proportion

HYPOTHESIS TESTING: BASIC PROCEDURES

There are several basic steps in hypothesis testing. They are briefly presented here and will be further explained through examples that follow.

1. Formulate the null and alternative hypotheses. As described in the preceding section, the null hypothesis asserts that a population parameter is equal to, no more than, or no less than some exact value, and it is evaluated in the face of numerical evidence. An appropriate alternative hypothesis covers other possible values for the parameter.

2. Select the significance level. If we end up rejecting the null hypothesis, there's a chance that we're wrong in doing so—i.e., that we've made a Type I error. The significance level is the maximum probability that we'll make such a mistake. In Figure 10.1, the significance level is represented by the shaded area(s) beneath each curve. For two-tail tests, the level of significance is the sum of both tail areas. In conducting a hypothesis test, we can choose any significance level we desire. In practice, however, levels of 0.10, 0.05, and 0.01 tend to be most common—in other words, if we reject a null hypothesis, the maximum chance of our being wrong would be 10%, 5%, or 1%, respectively. This significance level will be used to later identify the critical value(s).

3. Select the test statistic and calculate its value. For the tests of this chapter, the test statistic will be either \( z \) or \( t \), corresponding to the normal and \( t \) distributions, respectively. Figure 10.2 shows how the test statistic is selected. An important consideration in tests involving a sample mean is whether the population standard deviation (\( \sigma \)) is known. As Figure 10.2 indicates, the \( z \)-test (normal distribution and test statistic, \( z \)) will be used for hypothesis tests involving a sample proportion.

hypotheses would be appropriate in evaluating the statement made by this legislator?

10.9 In response to the assertion made in Exercise 10.8, suppose an insurance company executive were to claim the percentage of fraudulent auto theft reports to be “no more than 10%.” What null and alternative hypotheses would be appropriate in evaluating the executive’s statement?

10.10 For each of the following statements, formulate appropriate null and alternative hypotheses. Indicate whether the appropriate test will be one-tail or two-tail, then sketch a diagram that shows the approximate location of the “rejection” region(s) for the test.

   a. “The average college student spends no more than \$300 per semester at the university’s bookstore.”
   b. “The average adult drinks 1.5 cups of coffee per day.”
   c. “The average SAT score for entering freshmen is at least 1200.”
   d. “The average employee put in 3.5 hours of overtime last week.”

10.11 In administering a “field sobriety” test to suspected drunks, officers may ask a person to walk in a straight line or close his eyes and touch his nose. Define the Type I and Type II errors in terms of this setting. Speculate on physiological variables (besides the drinking of alcoholic beverages) that might contribute to the chance of each type of error.

10.12 In the judicial system, the defense attorney argues for the null hypothesis that the defendant is innocent. In general, what would be the result if judges instructed juries to

   a. never make a Type I error?
   b. never make a Type II error?
   c. compromise between Type I and Type II errors?

10.13 Regarding the testing of pharmaceutical companies’ claims that their drugs are safe, a U.S. Food and Drug Administration official has said that it’s “better to turn down 1000 good drugs than to approve one that’s unsafe.” If the null hypothesis is \( H_0: “The drug is not harmful,” \) what type of error does the official appear to favor?
Part 4: Hypothesis Testing

FIGURE 10.2

An overview of the process of selecting a test statistic for single-sample hypothesis testing. Key assumptions are reviewed in the figure notes.

Hypothesis test, one population

Population mean, \( \mu \)

\( \sigma \) known

Is the population truly or approximately normally distributed?

Yes

\( z \)-test, with test statistic

\[ z = \frac{\bar{x} - \mu_0}{\sigma} \]

where

\[ \sigma = \frac{\sigma}{\sqrt{n}} \]

and \( \mu_0 \) is from \( H_0 \)

Section 10.3

Note 1

No

\( \sigma \) unknown

Is the population truly or approximately normally distributed?

Yes

No

Yes

No

Is \( n \geq 30 \)?

Yes

No

Yes

No

Is \( n \geq 5 \) and \( n(1 - \pi) \geq 5 \)?

Yes

No

Convert to underlying binomial distribution.

Use distribution-free test.

\( \bar{x} \)-test, with test statistic

\[ t = \frac{\bar{x} - \mu_0}{s} \]

where

\[ df = n - 1 \]

\[ s = \frac{s}{\sqrt{n}} \]

and \( \mu_0 \) is from \( H_0 \)

Section 10.5

Note 2

\( \bar{x} \)-test, with test statistic

\[ z = \frac{\bar{x} - \pi_0}{\sigma_p} \]

where

\[ \sigma_p = \sqrt{\frac{\pi_0(1 - \pi_0)}{n}} \]

and \( \pi_0 \) is from \( H_0 \)

Section 10.6

Note 3

1 The \( z \) distribution: If the population is not normally distributed, \( n \) should be \( \geq 30 \) for the central limit theorem to apply. The population \( \sigma \) is usually not known.

2 The \( t \) distribution: For an unknown \( \sigma \), and when the population is approximately normally distributed, the \( t \)-test is appropriate regardless of the sample size. As \( n \) increases, the normality assumption becomes less important. If \( n < 30 \) and the population is not approximately normal, nonparametric testing (e.g., the sign test for central tendency, in Chapter 14) may be applied. The \( t \)-test is "robust" in terms of not being adversely affected by slight departures from the population normality assumption.

3 When \( n \pi \geq 5 \) and \( n(1 - \pi) \geq 5 \), the normal distribution is considered to be a good approximation to the binomial distribution. If this condition is not met, the exact probabilities must be derived from the binomial distribution. Most practical business settings involving proportions satisfy this condition, and the normal approximation is used in this chapter.
4. Identify critical value(s) for the test statistic and state the decision rule. The critical value(s) will bound rejection and nonrejection regions for the null hypothesis, $H_0$. Such regions are shown in Figure 10.1. They are determined from the significance level selected in step 2. In a one-tail test, there will be one critical value since $H_0$ can be rejected by an extreme result in just one direction. Two-tail tests will require two critical values since $H_0$ can be rejected by an extreme result in either direction. If the null hypothesis were really true, there would still be some probability (the significance level, $\alpha$) that the test statistic would be so extreme as to fall into a rejection region. The rejection and nonrejection regions can be stated as a decision rule specifying the conclusion to be reached for a given outcome of the test (e.g., “Reject $H_0$ if $z > 1.645$, otherwise do not reject”).

5. Compare calculated and critical values and reach a conclusion about the null hypothesis. Depending on the calculated value of the test statistic, it will fall into either a rejection region or the nonrejection region. If the calculated value is in a rejection region, the null hypothesis will be rejected. Otherwise, the null hypothesis cannot be rejected. Failure to reject a null hypothesis does not constitute proof that it is true, but rather that we are unable to reject it at the level of significance being used for the test.

6. Make the related business decision. After rejecting or failing to reject the null hypothesis, the results are applied to the business decision situation that precipitated the test in the first place. For example, Jackson T. Backus may decide to return the entire shipment of beets to his distributor.

**EXERCISES**

10.14 A researcher wants to carry out a hypothesis test involving the mean for a sample of size $n = 18$. She does not know the true value of the population standard deviation, but is reasonably sure that the underlying population is approximately normally distributed. Should she use a $z$-test or a $t$-test in carrying out the analysis? Why?

10.15 A research firm claims that 62% of women in the 40–49 age group save in a 401(k) or individual retirement account. If we wished to test whether this percentage could be the same for women in this age group living in New York City and selected a random sample of 300 such individuals from New York, what would be the null and alternative hypotheses? Would the test be a $z$-test or a $t$-test? Why?

10.16 In hypothesis testing, what is meant by the decision rule? What role does it play in the hypothesis-testing procedure?

10.17 A manufacturer informs a customer’s design engineers that the mean tensile strength of its rivets is at least 3000 pounds. A test is set up to measure the tensile strength of a sample of rivets, with the null and alternative hypotheses, $H_0$: $\mu \geq 3000$ and $H_1$: $\mu < 3000$. For each of the following individuals, indicate whether the person would tend to prefer a numerically very high (e.g., $\alpha = 0.20$) or a numerically very low (e.g., $\alpha = 0.0001$) level of significance to be specified for the test.
   a. The marketing director for a major competitor of the rivet manufacturer.
   b. The rivet manufacturer’s advertising agency, which has already made the “at least 3000 pounds” claim in national ads.

10.18 It has been claimed that no more than 5% of the units coming off an assembly line are defective. Formulate a null hypothesis and an alternative hypothesis for this situation. Will the test be one-tail or two-tail? Why? If the test is one-tail, will it be left-tail or right-tail? Why?
Situations can occur where the population mean is unknown but past experience has provided us with a trustworthy value for the population standard deviation. Although this possibility is more likely in an industrial production setting, it can sometimes apply to employees, consumers, or other nonmechanical entities.

In addition to the assumption that \( \sigma \) is known, the procedure of this section assumes either (1) that the sample size is large \((n \geq 30)\), or (2) that, if \( n < 30 \), the underlying population is normally distributed. These assumptions are summarized in Figure 10.2. If the sample size is large, the central limit theorem assures us that the sample means will be approximately normally distributed, regardless of the shape of the underlying distribution. The larger the sample size, the better this approximation becomes. Because it is based on the normal distribution, the test is known as the \( z \)-test, and the test statistic is as follows:

\[
    z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} \quad \text{where } \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}
\]

where:
- \( \bar{x} \) = sample mean
- \( \mu_0 \) = hypothesized population mean
- \( n \) = sample size

The symbol \( \mu_0 \) is the value of \( \mu \) that is assumed for purposes of the hypothesis test.

### Two-Tail Testing of a Mean, \( \sigma \) Known

#### Example

**Two-Tail Test**

When a robot welder is in adjustment, its mean time to perform its task is 1.3250 minutes. Past experience has found the standard deviation of the cycle time to be 0.0396 minutes. An incorrect mean operating time can disrupt the efficiency of other activities along the production line. For a recent random sample of 80 jobs, the mean cycle time for the welder was 1.3229 minutes. The underlying data are in file `CX10WELD`. Does the machine appear to be in need of adjustment?

**Solution**

**Formulate the Null and Alternative Hypotheses**

\[
    H_0: \mu = 1.3250 \text{ minutes} \quad \text{The machine is in adjustment.}
\]

\[
    H_1: \mu \neq 1.3250 \text{ minutes} \quad \text{The machine is out of adjustment.}
\]

In this test, we are concerned that the machine might be running at a mean speed that is either too fast or too slow. Accordingly, the null hypothesis could be
rejected by an extreme sample result in either direction. The hypothesized value for the population mean is $\mu_0 = 1.3250$ minutes, shown at the center of the distribution in Figure 10.3.

Select the Significance Level
The significance level used will be $\alpha = 0.05$. If the machine is running properly, there is only a 0.05 probability of our making the mistake of concluding that it requires adjustment.

Select the Test Statistic and Calculate Its Value
The population standard deviation ($\sigma$) is known and the sample size is large, so the normal distribution is appropriate and the test statistic will be $z$, calculated as 

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{1.3229 - 1.3250}{0.0396/\sqrt{80}} = -0.0021/0.00443 = -0.47$$

Identify Critical Values for the Test Statistic and State the Decision Rule
For a two-tail test using the normal distribution and $\alpha = 0.05$, $z = -1.96$ and $z = +1.96$ will be the respective boundaries for lower and upper tails of 0.025 each. These are the critical values for the test, and they identify the rejection and non-rejection regions shown in Figure 10.3. The decision rule can be stated as “Reject $H_0$ if calculated $z < -1.96$ or $> +1.96$, otherwise do not reject.”

Compare Calculated and Critical Values and Reach a Conclusion for the Null Hypothesis
The calculated value, $z = -0.47$, falls within the non-rejection region of Figure 10.3. At the 0.05 level of significance, the null hypothesis cannot be rejected.

Make the Related Business Decision
Based on these results, the robot welder is not in need of adjustment. The difference between the hypothesized population mean, $\mu_0 = 1.3250$ minutes, and the observed sample mean, $\bar{x} = 1.3229$, is judged to have been merely the result of chance variation.
If we had used the sample information and the techniques of Chapter 9 to construct a 95% confidence interval for $\mu$, the interval would have been
\[
\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 1.3229 \pm 1.96 \frac{0.0396}{\sqrt{80}}, \text{ or from 1.3142 to 1.3316 minutes}
\]
Notice that the hypothesized value, $\mu_0 = 1.3250$ minutes, falls within the 95% confidence interval—that is, the confidence interval tells us that $\mu$ could be 1.3250 minutes. This is the same conclusion we get from the nondirectional hypothesis test using $\alpha = 0.05$, and it is not a coincidence. A $100(1 - \alpha)%$ confidence interval is equivalent to a nondirectional hypothesis test at the $\alpha$ level, a relationship that will be discussed further in Section 10.4.

### One-Tail Testing of a Mean, $\sigma$ Known

**Example:**

**One-Tail Test**

The lightbulbs in an industrial warehouse have been found to have a mean lifetime of 1030.0 hours, with a standard deviation of 90.0 hours. The warehouse manager has been approached by a representative of Extendabulb, a company that makes a device intended to increase bulb life. The manager is concerned that the average lifetime of Extendabulb-equipped bulbs might not be any greater than the 1030 hours historically experienced. In a subsequent test, the manager tests 40 bulbs equipped with the device and finds their mean life to be 1061.6 hours. The underlying data are in file `CX10BULB`. Does Extendabulb really work?

**Solution**

**Formulate the Null and Alternative Hypotheses**

The warehouse manager’s concern that Extendabulb-equipped bulbs might not be any better than those used in the past leads to a directional test. Accordingly, the null and alternative hypotheses are:

\[ H_0: \mu \leq 1030.0 \text{ hours} \quad \text{Extendabulb is no better than the present system.} \]
\[ H_1: \mu > 1030.0 \text{ hours} \quad \text{Extendabulb really does increase bulb life.} \]

At the center of the hypothesized distribution will be the highest possible value for which $H_0$ could be true, $\mu_0 = 1030.0$ hours.

**Select the Significance Level**

The level chosen for the test will be $\alpha = 0.05$. If Extendabulb really has no favorable effect, the maximum probability of our mistakenly concluding that it does will be 0.05.

**Select the Test Statistic and Calculate Its Value**

As in the previous test, the population standard deviation ($\sigma$) is known and the sample size is large, so the normal distribution is appropriate and the test statistic will be $z$. It is calculated as

\[
z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{1061.6 - 1030.0}{90.0/\sqrt{40}} = 2.22
\]
Select the Critical Value for the Test Statistic and State the Decision Rule
For a right-tail $z$-test in which $\alpha = 0.05$, $z = +1.645$ will be the boundary separating the nonrejection and rejection regions. This critical value for the test is included in Figure 10.4. The decision rule can be stated as “Reject $H_0$ if calculated $z > +1.645$, otherwise do not reject.”

Compare Calculated and Critical Values and Reach a Conclusion for the Null Hypothesis
The calculated value, $z = +2.22$, falls within the rejection region of the diagram in Figure 10.4. At the 0.05 level of significance, the null hypothesis is rejected.

Make the Related Business Decision
The results suggest that Extendabulb does increase the mean lifetime of the bulbs. The difference between the mean of the hypothesized distribution, $\mu_0 = 1030.0$ hours, and the observed sample mean, $\bar{x} = 1061.6$, is judged too great to have occurred by chance. The firm may wish to incorporate Extendabulb into its warehouse lighting system.

Other Levels of Significance
This test was conducted at the 0.05 level, but would the conclusion have been different if other levels of significance had been used instead? Consider the following possibilities:

- For the 0.05 level of significance at which the test was conducted. The critical $z$ is +1.645, and the calculated value, $z = 2.22$, exceeds it. The null hypothesis is rejected, and we conclude that Extendabulb does increase bulb life.
- For the 0.025 level of significance. The critical $z$ is +1.96, and the calculated value, $z = 2.22$, exceeds it. The null hypothesis is rejected, and we again conclude that Extendabulb increases bulb life.
- For the 0.005 level of significance. The critical $z$ is +2.58, and the calculated value, $z = 2.22$, does not exceed it. The null hypothesis is not rejected, and we conclude that Extendabulb does not increase bulb life.
As these possibilities suggest, using different levels of significance can lead to quite different conclusions. Although the primary purpose of this exercise was to give you a little more practice in hypothesis testing, consider these two key questions: (1) If you were the manufacturer of Extendabulb, which level of significance would you prefer to use in evaluating the test results? (2) On which level of significance might the manufacturer of a competing product wish to rely in discussing the Extendabulb test? We will now examine these questions in the context of describing the \( p \)-value method for hypothesis testing.

## The \( p \)-value Approach to Hypothesis Testing

There are two basic approaches to conducting a hypothesis test:

- **Using a predetermined level of significance, establish critical value(s), then see whether the calculated test statistic falls into a rejection region for the test.** This is similar to placing a high-jump bar at a given height, then seeing whether you can clear it.
- **Determine the exact level of significance associated with the calculated value of the test statistic.** In this case, we’re identifying the most extreme critical value that the test statistic would be capable of exceeding. This is equivalent to your jumping as high as you can with no bar in place, then having the judges tell you how high you would have cleared if there had been a crossbar.

In the two tests carried out previously, we used the first of these approaches, making the hypothesis test a “yes–no” decision. In the Extendabulb example, however, we did allude to what we’re about to do here by trying several different significance levels in our one-tail test examining the ability of Extendabulb to increase the lifetime of lightbulbs.

We saw that Extendabulb showed a significant improvement at the 0.05 and 0.025 levels, but was not shown to be effective at the 0.005 level. In our high-jumping analogy, we might say that Extendabulb “cleared the bar” at the 0.05 level, cleared it again when it was raised to the more demanding 0.025 level, but couldn’t quite make the grade when the bar was raised to the very demanding 0.005 level of significance. In summary:

- **0.05 level** Extendabulb significantly increases bulb life (e.g., “clears the high-jump bar”).
- **0.025 level** Extendabulb significantly increases bulb life (“clears the bar”).
- **\( p \)-value level** Extendabulb just barely shows significant improvement in bulb life (“clears the bar, but lightly touches it on the way over”).
- **0.005 level** Extendabulb shows no significant improvement in bulb life (“insufficient height, fails to clear”).

As suggested by the preceding, and illustrated in part (a) of Figure 10.5, there is some level of significance (the \( p \)-value) where the calculated value of the test statistic is exactly the same as the critical value. For a given set of data, the \( p \)-value is sometimes referred to as the observed level of significance. It is the lowest possible level of significance at which the null hypothesis can be rejected. (Note: The lowercase \( p \) in “\( p \)-value” is not related to the symbol for the sample proportion.) For the Extendabulb test, the calculated value of the test statistic was \( z = 2.22 \). For a critical \( z = +2.22 \), the right-tail area can be found using the normal distribution table at the back of the book.
Chapter 10: Hypothesis Tests Involving a Sample Mean or Proportion

Referring to the normal distribution table, we see that 2.22 standard error units to the right of the mean includes a cumulative area of 0.9868, leaving (1.0000 − 0.9868), or 0.0132, in the right-tail area. This identifies the most demanding level of significance that Extendabulb could have achieved. If we had originally specified a significance level of 0.0132 for our test, the critical value for \( z \) would have been exactly the same as the value calculated. Thus, the \( p \)-value for the Extendabulb test is found to be 0.0132.

The Extendabulb example was a one-tail test—accordingly, the \( p \)-value was the area in just one tail. For two-tail tests, such as the robot welder example of Figure 10.3, the \( p \)-value will be the sum of both tail areas, as shown in part (b) of Figure 10.5. The calculated test statistic was \( z = -0.47 \), resulting in a cumulative area of 0.3192 in the left tail of the distribution. Since the robot welder test was two-tail, the 0.3192 must be multiplied by 2 to get the \( p \)-value of 0.6384.

FIGURE 10.5
The \( p \)-value of a test is the level of significance where the observed value of the test statistic is exactly the same as a critical value for that level. These diagrams show the \( p \)-values, as calculated in the text, for two of the tests performed in this section. When the hypothesis test is two-tail, as in part (b), the \( p \)-value is the sum of two tail areas.
Computer-Assisted Hypothesis Tests and \( p \)-values

When the hypothesis test is computer-assisted, the output will include a \( p \)-value for your interpretation. Regardless of whether a \( p \)-value has been approximated by your own calculations and table reference, or is a more exact value included in a computer printout, it can be interpreted as follows:

Interpreting the \( p \)-value in a computer printout:

- **Yes**: Reject the null hypothesis. The sample result is more extreme than you would have been willing to attribute to chance.
- **No**: Do not reject the null hypothesis. The sample result is not more extreme than you would have been willing to attribute to chance.

Computer Solutions 10.1 shows how we can use Excel or Minitab to carry out a hypothesis test for the mean when the population standard deviation is known or assumed. In this case, we are replicating the hypothesis test in Figure 10.4, using the 40 data values in file **CX10BULB**. The printouts in Computer Solutions 10.1 show the \( p \)-value (0.0132) for the test. This \( p \)-value is essentially making the following statement: “If the population mean really is 1030 hours, there is only a 0.0132 probability of getting a sample mean this large (1061.6 hours) just by chance.” Because the \( p \)-value is less than the level of significance we are using to reach our conclusion (i.e., \( p \)-value = 0.0132 is \( < \alpha = 0.05 \)), \( H_0: \mu = 1030 \) is rejected.
Excel hypothesis test for $\mu$ based on raw data and $\sigma$ known

1. For example, for the 40 bulb lifetimes (file CX10BULB) on which Figure 10.4 is based, with the label and 40 data values in A1:A41: From the Add-Ins ribbon, click Data Analysis Plus. Click Z-Test: Mean. Click OK.

2. Enter A1:A41 into the Input Range box. Enter the hypothesized mean (1030) into the Hypothesized Mean box. Enter the known population standard deviation (90.0) into the Standard Deviation (SIGMA) box. Click Labels, since the variable name is in the first cell within the field. Enter the level of significance for the test (0.05) into the Alpha box. Click OK. The printout includes the $p$-value for this one-tail test, 0.0132.

Excel hypothesis test for $\mu$ based on summary statistics and $\sigma$ known

1. For example, with $\bar{x} = 1061.6$, $\sigma = 90.0$, and $n = 40$, as in Figure 10.4: Open the TEST STATISTICS workbook.

2. Using the arrows at the bottom left, select the z-Test_Mean worksheet. Enter the sample mean (1061.6), the known sigma (90.0), the sample size (40), the hypothesized population mean (1030), and the level of significance for the test (0.05). (Note: As an alternative, you can use Excel worksheet template TMZTEST. The steps are described within the template.)

MINITAB

Minitab hypothesis test for $\mu$ based on raw data and $\sigma$ known

1. For example, using the data (file CX10BULB) on which Figure 10.4 is based, with the 40 data values in column C1: Click Stat. Select Basic Statistics. Click 1-Sample Z.

2. Select Samples in Columns and enter C1 into the box. Enter the known population standard deviation (90.0) into the Standard deviation box. Select Perform hypothesis test and enter the hypothesized population mean (1030) into the Hypothesized mean: box.

3. Click Options. Enter the desired confidence level as a percentage (95.0) into the Confidence Level box. Within the Alternative box, select greater than. Click OK. By default, this test also provides the lower boundary of the 95% confidence interval (unless another confidence level has been specified).

Minitab hypothesis test for $\mu$ based on summary statistics and $\sigma$ known

Follow the procedure in steps 1 through 3, above, but in step 2 select Summarized data and enter 40 and 1061.6 into the Sample size and Mean boxes, respectively.

**EXERCISES**

10.19 What is the central limit theorem, and how is it applicable to hypothesis testing?

10.20 If the population standard deviation is known, but the sample size is less than 30, what assumption is necessary to use the z-statistic in carrying out a hypothesis test for the population mean?

10.21 What is a $p$-value, and how is it relevant to hypothesis testing?
10.22 The p-value for a hypothesis test has been reported as 0.03. If the test result is interpreted using the α = 0.05 level of significance as a criterion, will $H_0$ be rejected? Explain.

10.23 The p-value for a hypothesis test has been reported as 0.04. If the test result is interpreted using the α = 0.01 level of significance as a criterion, will $H_0$ be rejected? Explain.

10.24 A hypothesis test is carried out using the α = 0.01 level of significance, and $H_0$ cannot be rejected. What is the most accurate statement we can make about the p-value for this test?

10.25 For each of the following tests and z values, determine the p-value for the test:
   a. Right-tail test and $z = 1.54$
   b. Left-tail test and $z = -1.03$
   c. Two-tail test and $z = -1.83$

10.26 For each of the following tests and z values, determine the p-value for the test:
   a. Left-tail test and $z = -1.62$
   b. Right-tail test and $z = 1.43$
   c. Two-tail test and $z = 1.27$

10.27 For a sample of 35 items from a population for which the standard deviation is $\sigma = 20.5$, the sample mean is 458.0. At the 0.05 level of significance, test $H_0$: $\mu = 450$ versus $H_1$: $\mu \neq 450$. Determine and interpret the p-value for the test.

10.28 For a sample of 12 items from a normally distributed population for which the standard deviation is $\sigma = 17.0$, the sample mean is 230.8. At the 0.05 level of significance, test $H_0$: $\mu \leq 220$ versus $H_1$: $\mu > 220$. Determine and interpret the p-value for the test.

10.29 A quality-assurance inspector periodically examines the output of a machine to determine whether it is properly adjusted. When set properly, the machine produces nails having a mean length of 2.000 inches, with a standard deviation of 0.070 inches. For a sample of 34 tubes, the mean diameter to “drift” away from the desired 2.500 inches, with a standard deviation of 0.070 inches. For a recent random sample of 34 tubes, the mean diameter was 2.509 inches. At the 0.01 level of significance, does the machine appear to be in need of maintenance and calibration? Determine and interpret the p-value for the test.

10.30 In the past, patrons of a cinema complex have spent an average of $5.00 for popcorn and other snacks, with a standard deviation of $1.80. The amounts of these expenditures have been normally distributed. Following an intensive publicity campaign by a local medical society, the mean expenditure for a sample of 18 patrons is found to be $4.20. In a one-tail test at the 0.05 level of significance, does this recent experience suggest a decline in spending? Determine and interpret the p-value for the test.

10.31 Following maintenance and calibration, an extrusion machine produces aluminum tubing with a mean outside diameter of 2.500 inches, with a standard deviation of 0.027 inches. As the machine functions over an extended number of work shifts, the standard deviation remains unchanged, but the combination of accumulated deposits and mechanical wear causes the mean diameter to “drift” away from the desired 2.500 inches. For a recent random sample of 34 tubes, the mean diameter was 2.509 inches. At the 0.01 level of significance, does the machine appear to be in need of maintenance and calibration? Determine and interpret the p-value for the test.

10.32 A manufacturer of electronic kits has found that the mean time required for novices to assemble its new circuit tester is 3 hours, with a standard deviation of 0.20 hours. A consultant has developed a new instructional booklet intended to reduce the time an inexperienced kit builder will need to assemble the device. In a test of the effectiveness of the new booklet, 15 novices require a mean of 2.90 hours to complete the job. Assuming the population of times is normally distributed, and using the 0.05 level of significance, should we conclude that the new booklet is effective? Determine and interpret the p-value for the test.

10.33 According to bankrate.com, the average cost to remodel a home office is $10,526. Assuming a population standard deviation of $2000 and the sample of home office conversion prices charged for 40 recent jobs performed by builders in a region of the United States, examine whether the mean price for home office conversions for builders in this region might be different from the average for the nation as a whole. The underlying data are in file XR10033. Identify and interpret the p-value for the test. Using the 0.05 level of significance, what conclusion will be reached? Source: bankrate.com, July 23, 2009.

10.34 A machine that fills shipping containers with driveway filler mix is set to deliver a mean fill weight of 70.0 pounds. The standard deviation of fill weights delivered by the machine is known to be 1.0 pounds. For a recent sample of 35 containers, the fill weights are listed in data file XR10034. Using the mean for this sample, and assuming that the population standard deviation has remained unchanged at 1.0 pounds, examine whether the mean fill weight delivered by the machine might now be something other than 70.0 pounds. Identify and interpret the p-value for the test. Using the 0.05 level of significance, what conclusion will be reached?
In Chapter 9, we constructed confidence intervals for a population mean or proportion. In this chapter, we sometimes carry out nondirectional tests for the null hypothesis that the population mean or proportion could have a given value. Although the purposes may differ, the concepts are related.

In the previous section, we briefly mentioned this relationship in the context of the nondirectional test summarized in Figure 10.3. Consider this nondirectional test, carried out at the $\alpha = 0.05$ level:

1. Null and alternative hypotheses: \( H_0: \mu = 1.3250 \) minutes and \( H_1: \mu \neq 1.3250 \) minutes.
2. The standard error of the mean: \( \sigma_x = \sigma/\sqrt{n} = 0.0396/\sqrt{80} \), or 0.00443 minutes.
3. The critical \( z \) values for a two-tail test at the $\alpha = 0.05$ level are $z = -1.96$ and $z = +1.96$.
4. Expressing these \( z \) values in terms of the sample mean, critical values for \( \bar{x} \) would be calculated as \( 1.325 \pm 1.96(0.00443) \), or 1.3163 minutes and 1.3337 minutes.
5. The observed sample mean was \( \bar{x} = 1.3229 \) minutes. This fell within the acceptable limits and we were not able to reject \( H_0 \).

Based on the $\alpha = 0.05$ level, the nondirectional hypothesis test led us to conclude that \( H_0: \mu = 1.3250 \) minutes was believable. The observed sample mean (1.3229 minutes) was close enough to the 1.3250 hypothesized value that the difference could have happened by chance.

Now let’s approach the same situation by using a 95% confidence interval. As noted previously, the standard error of the sample mean is 0.00443 minutes. Based on the sample results, the 95% confidence interval for \( \mu \) is \( 1.3229 \pm 1.96(0.00443) \), or from 1.3142 minutes to 1.3316 minutes. In other words, we have 95% confidence that the population mean is somewhere between 1.3142 minutes and 1.3316 minutes. If someone were to suggest that the population mean were actually 1.3250 minutes, we would find this believable, since 1.3250 falls within the likely values for \( \mu \) that our confidence interval represents.

The nondirectional hypothesis test was done at the $\alpha = 0.05$ level, the confidence interval was for the 95% confidence level, and the conclusion was the same in each case. As a general rule, we can state that the conclusion from a nondirectional hypothesis test for a population mean at the $\alpha$ level of significance will be the same as the conclusion based on a confidence interval at the $100(1 - \alpha)$% confidence level.

When a hypothesis test is nondirectional, this equivalence will be true. This exact statement cannot be made about confidence intervals and directional tests—although they can also be shown to be related, such a demonstration would take us beyond the purposes of this chapter. Suffice it to say that confidence intervals and hypothesis tests are both concerned with using sample information to make a statement about the (unknown) value of a population mean or proportion. Thus, it is not surprising that their results are related.

By using Seeing Statistics Applet 12, at the end of the chapter, you can see how the confidence interval (and the hypothesis test conclusion) would change in response to various possible values for the sample mean.
10.35 Based on sample data, a confidence interval has been constructed such that we have 90% confidence that the population mean is between 120 and 180. Given this information, provide the conclusion that would be reached for each of the following hypothesis tests at the \( \alpha = 0.10 \) level:

a. \( H_0: \mu = 170 \) versus \( H_1: \mu \neq 170 \)

b. \( H_0: \mu = 110 \) versus \( H_1: \mu \neq 110 \)

c. \( H_0: \mu = 130 \) versus \( H_1: \mu \neq 130 \)

d. \( H_0: \mu = 200 \) versus \( H_1: \mu \neq 200 \)

10.36 Given the information in Exercise 10.27, construct a 95% confidence interval for the population mean, then reach a conclusion regarding whether \( \mu \) could actually be equal to the value that has been hypothesized. How does this conclusion compare to that reached in Exercise 10.27? Why?

10.37 Given the information in Exercise 10.29, construct a 99% confidence interval for the population mean, then reach a conclusion regarding whether \( \mu \) could actually be equal to the value that has been hypothesized. How does this conclusion compare to that reached in Exercise 10.29? Why?

10.38 Use an appropriate confidence interval in reaching a conclusion regarding the problem situation and null hypothesis for Exercise 10.31.

10.35 Based on sample data, a confidence interval has been constructed such that we have 90% confidence that the population mean is between 120 and 180. Given this information, provide the conclusion that would be reached for each of the following hypothesis tests at the \( \alpha = 0.10 \) level:

a. \( H_0: \mu = 170 \) versus \( H_1: \mu \neq 170 \)

b. \( H_0: \mu = 110 \) versus \( H_1: \mu \neq 110 \)

c. \( H_0: \mu = 130 \) versus \( H_1: \mu \neq 130 \)

d. \( H_0: \mu = 200 \) versus \( H_1: \mu \neq 200 \)

10.36 Given the information in Exercise 10.27, construct a 95% confidence interval for the population mean, then reach a conclusion regarding whether \( \mu \) could actually be equal to the value that has been hypothesized. How does this conclusion compare to that reached in Exercise 10.27? Why?
Two-Tail Testing of a Mean, $\sigma$ Unknown

**EXAMPLE**

Two-Tail Test

The credit manager of a large department store claims that the mean balance for the store’s charge account customers is $410. An independent auditor selects a random sample of 18 accounts and finds a mean balance of $511.33 and a standard deviation of $s = 183.75$. The sample data are in file **CX10CRED**. If the manager’s claim is not supported by these data, the auditor intends to examine all charge account balances. If the population of account balances is assumed to be approximately normally distributed, what action should the auditor take?

**SOLUTION**

Formulate the Null and Alternative Hypotheses

\[ H_0: \mu = 410 \quad \text{The mean balance is actually } 410. \]
\[ H_1: \mu \neq 410 \quad \text{The mean balance is some other value.} \]

In evaluating the manager’s claim, a two-tail test is appropriate since it is a nondirectional statement that could be rejected by an extreme result in either direction. The center of the hypothesized distribution of sample means for samples of \( n = 18 \) will be \( \mu_0 = 410 \).

Select the Significance Level

For this test, we will use the 0.05 level of significance. The sum of the two tail areas will be 0.05.

Select the Test Statistic and Calculate Its Value

The test statistic is \( t = \frac{\bar{x} - \mu_0}{s_x} \), and the \( t \) distribution will be used to describe the sampling distribution of the mean for samples of \( n = 18 \). The center of the distribution is \( \mu_0 = 410 \), which corresponds to \( t = 0.000 \). Since the population standard deviation is unknown, \( s \) is used to estimate \( \sigma \). The sampling distribution has an estimated standard error of

\[ s_x = \frac{s}{\sqrt{n}} = \frac{183.75}{\sqrt{18}} = 43.31 \]

and the calculated value of \( t \) will be

\[ t = \frac{\bar{x} - \mu_0}{s_x} = \frac{511.33 - 410.00}{43.31} = 2.340 \]

Identify Critical Values for the Test Statistic and State the Decision Rule

For this test, \( \alpha = 0.05 \), and the number of degrees of freedom will be \( df = (n - 1) \), or \( 18 - 1 = 17 \). The \( t \) distribution table at the back of the book provides one-tail areas, so we must identify the boundaries where each tail area is one-half of \( \alpha \), or 0.025. Referring to the 0.025 column and 17th row of the table, we find the critical values for the test statistic to be \( t = -2.110 \) and \( t = +2.110 \). (Although the “-2.110” is not shown in the table, we can identify this as the left-tail boundary because the distribution is symmetrical.) The rejection and nonrejection areas
are shown in Figure 10.6, and the decision rule can be stated as “Reject $H_0$ if the calculated $t$ is either $<-2.110$ or $>+2.110$, otherwise do not reject.”

**Compare the Calculated and Critical Values and Reach a Conclusion for the Null Hypothesis**

The calculated test statistic, $t = 2.340$, exceeds the upper boundary and falls into this rejection region. $H_0$ is rejected.

**Make the Related Business Decision**

The results suggest that the mean charge account balance is some value other than $410. The auditor should proceed to examine all charge account balances.

---

**One-Tail Testing of a Mean, $\sigma$ Unknown**

**EXAMPLE**

**One-Tail Test**

The Chekzar Rubber Company, in financial difficulties because of a poor reputation for product quality, has come out with an ad campaign claiming that the mean lifetime for Chekzar tires is at least 60,000 miles in highway driving. Skeptical, the editors of a consumer magazine purchase 36 of the tires and test them in highway use. The mean tire life in the sample is $\bar{x} = 58,341.69$ miles, with a sample standard deviation of $s = 3632.53$ miles. The sample data are in file CX10CHEK.
Chapter 10: Hypothesis Tests Involving a Sample Mean or Proportion

SOLUTION

Formulate the Null and Alternative Hypotheses
Because of the directional nature of the ad claim and the editors’ skepticism regarding its truthfulness, the null and alternative hypotheses are

\[ H_0: \mu \geq 60,000 \text{ miles} \quad \text{The mean tire life is at least 60,000 miles.} \]
\[ H_1: \mu < 60,000 \text{ miles} \quad \text{The mean tire life is under 60,000 miles.} \]

Select the Significance Level
For this test, the significance level will be specified as 0.01.

Select the Test Statistic and Calculate Its Value
The test statistic is \( t = \frac{\bar{x} - \mu_0}{s_\bar{x}} \) and the \( t \) distribution will be used to describe the sampling distribution of the mean for samples of \( n = 36 \). The center of the distribution is the lowest possible value for which \( H_0 \) could be true, or \( \mu_0 = 60,000 \) miles. Since the population standard deviation is unknown, \( s \) is used to estimate \( \sigma \). The sampling distribution has an estimated standard error of

\[ s_\bar{x} = \frac{s}{\sqrt{n}} = \frac{3632.53 \text{ miles}}{\sqrt{36}} = 605.42 \text{ miles} \]

and the calculated value of \( t \) will be

\[ t = \frac{\bar{x} - \mu_0}{s_\bar{x}} = \frac{58,341.69 - 60,000.00}{605.42} = -2.739 \]

Identify the Critical Value for the Test Statistic and State the Decision Rule
For this test, \( \alpha \) has been specified as 0.01. The number of degrees of freedom is \( df = (n - 1) \), or (36 - 1) = 35. The \( t \) distribution table is now used in finding the value of \( t \) that corresponds to a one-tail area of 0.01 and \( df = 35 \) degrees of freedom. Referring to the 0.01 column and 35th row of the table, we find this critical value to be \( t = -2.438 \). (Although the value listed is positive, remember that the distribution is symmetrical, and we are looking for the left-tail boundary.) The rejection and nonrejection regions are shown in Figure 10.7, and the

![](Fig_10.7.png)

The Chekzar Rubber Company has claimed that, in highway use, the mean lifetime of its tires is at least 60,000 miles. At the 0.01 level in this left-tail test, the claim is not supported.
decision rule can be stated as “Reject $H_0$ if the calculated $t$ is less than $-2.438$,
otherwise do not reject.”

**Compare the Calculated and Critical Values and Reach a Conclusion
for the Null Hypothesis**

The calculated test statistic, $t = -2.739$, is less than the critical value, $t = -2.438$,
and falls into the rejection region of the test. The null hypothesis, $H_0: \mu \geq 60,000$
miles, must be rejected.

**Make the Related Business Decision**

The test results support the editors’ doubts regarding Chekzar’s ad claim. The magazine may wish to exert either readership or legal pressure on Chekzar to modify its claim.

Compared to the $t$-test, the $z$-test is a little easier to apply if the analysis is carried out by pocket calculator and references to a statistical table. (There are lesser “gaps” between areas listed in the normal distribution table compared to values provided in the $t$ table.) Also, courtesy of the central limit theorem, results can be fairly satisfactory when $n$ is large and $s$ is a close estimate of $\sigma$.

Nevertheless, the $t$-test remains the appropriate procedure whenever $\sigma$ is unknown and is being estimated by $s$. In addition, this is the method you will either use or come into contact with when dealing with computer statistical packages handling the kinds of analyses in this section. For example, with Excel, Minitab, SPSS, SAS, and others, we can routinely (and correctly) apply the $t$-test whenever $s$ has been used to estimate $\sigma$.

An important note when using statistical tables to determine $p$-values: For $t$-tests, the $p$-value can’t be determined as exactly as with the $z$-test, because the $t$ table areas include greater “gaps” (e.g., the 0.005, 0.01, 0.025 columns, and so on). However, we can narrow down the $t$-test $p$-value to a range, such as “between 0.01 and 0.025.”

For example, in the Chekzar Rubber Company $t$-test of Figure 10.7, the calculated $t$ statistic was $t = -2.739$. We were able to reject the null hypothesis at the 0.01 level (critical value, $t = -2.438$), and would also have been able to reject $H_0$ at the 0.005 level (critical value, $t = -2.724$). Based on the $t$ table, the most accurate conclusion we can reach is that the $p$-value for the Chekzar test is less than 0.005. Had we used the computer in performing this test, we would have found the actual $p$-value to be 0.0048.

Computer Solutions 10.2 shows how we can use Excel or Minitab to carry out a hypothesis test for the mean when the population standard deviation is unknown. In this case, we are replicating the hypothesis test shown in Figure 10.6, using the 18 data values in file **CX10CRED**. The printouts in Computer Solutions 10.2 show the $p$-value (0.032) for the test. This $p$-value represents the following statement: “If the population mean really is $410$, there is only a 0.032 probability of getting a sample mean this far away from $410$ just by chance.” Because the $p$-value is less than the level of significance we are using to reach a conclusion (i.e., $p$-value = 0.032 < $\alpha$ = 0.05), $H_0$: $\mu = 410$ is rejected.

In the Minitab portion of Computer Solutions 10.2, the 95% confidence interval is shown as $420.0$ to $602.7$. The hypothesized population mean ($410$) does not fall within the 95% confidence interval; thus, at this confidence level, the results suggest that the population mean is some value other than $410$. This same conclusion was reached in our two-tail test at the 0.05 level of significance.
Hypothesis Test for Population Mean, \( \sigma \) Unknown

These procedures show how to carry out a hypothesis test for the population mean when the population standard deviation is unknown.

**EXCEL**

Excel hypothesis test for \( \mu \) based on raw data and \( \sigma \) unknown

1. For example, for the credit balances (file CX10CRED) on which Figure 10.6 is based, with the label and 18 data values in A1:A19: From the Add-Ins ribbon, click Data Analysis Plus. Click t-Test: Mean. Click OK.
2. Enter A1:A19 into the Input Range box. Enter the hypothesized mean (410) into the Hypothesized Mean box. Click Labels. Enter the level of significance for the test (0.05) into the Alpha box. Click OK. The printout shows the p-value for this two-tail test, 0.0318.

Excel hypothesis test for \( \mu \) based on summary statistics and \( \sigma \) unknown

1. For example, with \( \bar{x} = 511.33 \), \( s = 183.75 \), and \( n = 18 \), as in Figure 10.6: Open the TEST STATISTICS workbook.
2. Using the arrows at the bottom left, select the t-Test_Mean worksheet. Enter the sample mean (511.33), the sample standard deviation (183.75), the sample size (18), the hypothesized population mean (410), and the level of significance for the test (0.05).

(Note: As an alternative, you can use Excel worksheet template TMTTEST. The steps are described within the template.)

**MINITAB**

Minitab hypothesis test for \( \mu \) based on raw data and \( \sigma \) unknown

1. For example, using the data (file CX10CRED) on which Figure 10.6 is based, with the 18 data values in column C1: Click Stat. Select Basic Statistics. Click 1-Sample t.
2. Select Samples in Columns and enter C1 into the box. Select Perform hypothesis test and enter the hypothesized population mean (410) into the Hypothesized mean: box.

(continued)
10.39 Under what circumstances should the t-statistic be used in carrying out a hypothesis test for the population mean?

10.40 For a simple random sample of 40 items, \( \bar{x} = 25.9 \) and \( s = 4.2 \). At the 0.01 level of significance, test \( H_0: \mu = 24.0 \) versus \( H_1: \mu \neq 24.0 \).

10.41 For a simple random sample of 15 items from a population that is approximately normally distributed, \( \bar{x} = 82.0 \) and \( s = 20.5 \). At the 0.05 level of significance, test \( H_0: \mu \geq 90.0 \) versus \( H_1: \mu < 90.0 \).

10.42 The average age of passenger cars in use in the United States is 9.0 years. For a simple random sample of 34 vehicles observed in the employee parking area of a large manufacturing plant, the average age is 10.4 years, with a standard deviation of 3.1 years. At the 0.01 level of significance, can we conclude that the average age of cars driven to work by the plant’s employees is greater than the national average? Source: polk.com, August 9, 2006.

10.43 The average length of a flight by regional airlines in the United States has been reported as 464 miles. If a simple random sample of 30 flights by regional airlines were to have \( \bar{x} = 479.6 \) miles and \( s = 42.8 \) miles, would this tend to cast doubt on the reported average of 464 miles? Use a two-tail test and the 0.05 level of significance in arriving at your answer. Source: Bureau of the Census, Statistical Abstract of the United States 2009, p. 664.

10.44 The International Coffee Association has reported the mean daily coffee consumption for U.S. residents as 1.65 cups. Assume that a sample of 38 people from a North Carolina city consumed a mean of 1.84 cups of coffee per day, with a standard deviation of 0.85 cups. In a two-tail test at the 0.05 level, could the residents of this city be said to be significantly different from their counterparts across the nation? Source: coffeeresearch.org, August 8, 2006.

10.45 Taxco, a firm specializing in the preparation of income tax returns, claims the mean refund for customers who received refunds last year was $150. For a random sample of 12 customers who received refunds last year, the mean amount was found to be $125, with a standard deviation of $43. Assuming that the population is approximately normally distributed, and using the 0.10 level in a two-tail test, do these results suggest that Taxco’s assertion may be accurate?

10.46 The new director of a local YMCA has been told by his predecessors that the average member has belonged for 8.7 years. Examining a random sample of 15 membership files, he finds the mean length of membership to be 7.2 years, with a standard deviation of 2.5 years. Assuming the population is approximately normally distributed, and using the 0.05 level, does this result suggest that the actual mean length of membership may be some value other than 8.7 years?

10.47 A scrap metal dealer claims that the mean of his cash sales is “no more than $80,” but an Internal Revenue Service agent believes the dealer is untruthful. Observing a sample of 20 cash customers, the agent finds the mean purchase to be $91, with a standard deviation of $21. Assuming the population is approximately normally distributed, and using the 0.05 level of significance, is the agent’s suspicion confirmed?

10.48 During 2008, college work-study students earned a mean of $1478. Assume that a sample consisting of 45 of the work-study students at a large university was found to have earned a mean of $1503 during that year, with a standard deviation of $210. Would a one-tail test at the 0.05 level suggest the average earnings of this university’s work-study students were significantly higher than the national mean? Source: Bureau of the Census, Statistical Abstract of the United States 2009, p. 178.

10.49 According to the Federal Reserve Board, the mean net worth of U.S. households headed by persons 75 years or older is $640,000. Suppose a simple random sample of 50 households in this age group is obtained from a certain
Chapter 10: Hypothesis Tests Involving a Sample Mean or Proportion

10.50 Using the sample results in Exercise 10.49, construct and interpret the 95% confidence interval for the population mean. Is the hypothesized population mean ($640,000) within the interval? Given the presence or absence of the $640,000 value within the interval, is this consistent with the findings of the hypothesis test conducted in Exercise 10.49?

10.51 It has been reported that the average life for halogen lightbulbs is 4000 hours. Learning of this figure, a plant manager would like to find out whether the vibration and temperature conditions that the facility’s bulbs encounter might be having an adverse effect on the service life of bulbs in her plant. In a test involving 15 halogen bulbs installed in various locations around the plant, she finds the average life for bulbs in the sample is 3882 hours, with a standard deviation of 200 hours. Assuming the population of halogen bulb lifetimes to be approximately normally distributed, and using the 0.025 level of significance, do the test results tend to support the manager’s suspicion that adverse conditions might be detrimental to the operating lifespan of halogen lightbulbs used in her plant? Source: Cindy Hall and Gary Visgaitis, “Bulbs Lasting Longer,” USA Today, March 9, 2000, p. 10.

10.52 In response to an inquiry from its national office, the manager of a local bank has stated that her bank’s average service time for a drive-through customer is 93 seconds. A student intern working at the bank happens to be taking a statistics course and is curious as to whether the true average might be some value other than 93 seconds. The intern observes a simple random sample of 50 drive-through customers whose average service time is 89.5 seconds, with a standard deviation of 11.3 seconds. From these sample results, and using the 0.05 level of significance, what conclusion would the student reach with regard to the bank manager’s claim?

10.53 Using the sample results in Exercise 10.52, construct and interpret the 95% confidence interval for the population mean. Is the hypothesized population mean (93 seconds) within the interval? Given the presence or absence of the 93 seconds value within the interval, is this consistent with the findings of the hypothesis test conducted in Exercise 10.52?

10.54 The U.S. Census Bureau says the 52-question “long form” received by 1 in 6 households during the 2000 census takes a mean of 38 minutes to complete. Suppose a simple random sample of 35 persons is given the form, and their mean time to complete it is 36.8 minutes, with a standard deviation of 4.0 minutes. From these sample results, and using the 0.10 level of significance, what conclusion do you reach?

10.55 Using the sample results in Exercise 10.54, construct and interpret the 90% confidence interval for the population mean. Is the hypothesized population mean (38 minutes) within the interval? Given the presence or absence of the 38 minutes value within the interval, is this consistent with the findings of the hypothesis test conducted in Exercise 10.54?

(DATA SET) Note: Exercises 10.56–10.58 require a computer and statistical software.

10.56 The International Council of Shopping Centers reports that the average teenager spends $57 during a shopping trip to the mall. The promotions director of a local mall has used a variety of strategies to attract area teens to his mall, including live bands and “teen-appreciation days” that feature special bargains for this age group. He believes teen shoppers at his mall respond to his promotional efforts by shopping there more often and spending more when they do. Mall management decides to evaluate the promotions director’s success by surveying a simple random sample of 45 local teens and finding out how much they spent on their most recent shopping visit to the mall. The results are listed in data file XR10056. Use a suitable hypothesis test in examining whether the mean mall shopping expenditure for teens in this area might be higher than for U.S. teens as a whole. Identify and interpret the p-value for the test. Using the 0.025 level of significance, what conclusion do you reach? Source: icsc.org, July 23, 2009.

10.57 According to the Insurance Information Institute, the mean annual expenditure for automobile insurance for U.S. motorists is $817. Suppose that a government official in North Carolina has surveyed a simple random sample of 80 residents of her state, and that their auto insurance expenditures for the most recent year are in data file XR10057. Based on these data, examine whether the mean annual auto insurance expenditure for motorists in North Carolina might be different from the $817 for the country as a whole. Identify and interpret the p-value for the test. Using the 0.05 level of significance, what conclusion do you reach? Source: ii.org, July 23, 2009.

10.58 Using the sample data in Exercise 10.57, construct and interpret the 95% confidence interval for the population mean. Is the hypothesized population mean ($817) within the interval? Given the presence or absence of the $817 value within the interval, is this consistent with the findings of the hypothesis test conducted in Exercise 10.57?
Occasions may arise when we wish to compare a sample proportion, \( p \), with a value that has been hypothesized for the population proportion, \( \pi \). As we noted in Figure 10.2, the theoretically correct distribution for dealing with proportions is the binomial distribution. However, the normal distribution is a good approximation when \( n\pi \geq 5 \) and \( n(1 - \pi) \geq 5 \). The larger the sample size, the better this approximation becomes, and for most practical settings, this condition is satisfied. When using the normal distribution for hypothesis tests of a sample proportion, the test statistic is as follows:

\[
z = \frac{p - \pi_0}{\sigma_p}
\]

where
- \( p \) = sample proportion
- \( \pi_0 \) = hypothesized population proportion
- \( n \) = sample size
- \( \sigma_p \) = standard error of the distribution of the sample proportion

\[
\sigma_p = \sqrt{\frac{\pi_0(1 - \pi_0)}{n}}
\]

### Two-Tail Testing of a Proportion

Whenever the null hypothesis involves a proportion and is nondirectional, this technique is appropriate. To demonstrate how it works, consider the following situation.

#### EXAMPLE

**Two-Tail Test**

The career services director of Hobart University has said that 70% of the school’s seniors enter the job market in a position directly related to their undergraduate field of study. In a sample consisting of 200 of the graduates from last year’s class, 66% have entered jobs related to their field of study. The underlying data are in file `CX10GRAD`, with values coded as 1 = no job in field, 2 = job in field.

**SOLUTION**

**Formulate the Null and Alternative Hypotheses**

The director’s statement is nondirectional and leads to null and alternative hypotheses of

- \( H_0: \pi = 0.70 \)  The proportion of graduates entering jobs in their field is 0.70.
- \( H_1: \pi \neq 0.70 \)  The proportion is some value other than 0.70.

**Select the Significance Level**

For this test, the 0.05 level will be used. The sum of the two tail areas will be 0.05.
Select the Test Statistic and Calculate Its Value
The test statistic will be \( z \), the number of standard error units from the hypothesized population proportion, \( \pi_0 = 0.70 \), to the sample proportion, \( p = 0.66 \). The standard error of the sample proportion is

\[
\sigma_p = \sqrt{\frac{\pi_0(1 - \pi_0)}{n}} = \sqrt{\frac{0.70(1 - 0.70)}{200}} = 0.0324
\]

and the calculated value of \( z \) will be

\[
z = \frac{p - \pi_0}{\sigma_p} = \frac{0.66 - 0.70}{0.0324} = -1.23
\]

Identify Critical Values for the Test Statistic and State the Decision Rule
Since the test is two-tail and the selected level of significance is 0.05, the critical values will be \( z = -1.96 \) and \( z = +1.96 \). The decision rule can be stated as “Reject \( H_0 \) if calculated \( z \) is either < -1.96 or > +1.96, otherwise do not reject.”

Compare the Calculated and Critical Values and Reach a Conclusion for the Null Hypothesis
The calculated value of the test statistic, \( z = -1.23 \), falls between the two critical values, placing it in the nonrejection region of the distribution shown in Figure 10.8. The null hypothesis is not rejected.

Make the Related Decision
Failure to reject the null hypothesis leads us to conclude that the proportion of graduates who enter the job market in careers related to their field of study could indeed be equal to the claimed value of 0.70. If the career services director has been making this claim to her students or their parents, this analysis would suggest that her assertion not be challenged.
One-Tail Testing of a Proportion

Directional tests for a proportion are similar to the preceding example, but have only one tail area in which the null hypothesis can be rejected. Consider the following actual case.

EXAMPLE

One-Tail Test

In an administrative decision, the U.S. Veterans Administration (VA) closed the cardiac surgery units of several VA hospitals that either performed fewer than 150 operations per year or had mortality rates higher than 5.0%.

In one of the closed surgery units, 100 operations had been performed during the preceding year, with a mortality rate of 7.0%. The underlying data are in file CX10HOSP, with values coded as 1 = nonfatality, 2 = fatality. At the 0.01 level of significance, was the mortality rate of this hospital significantly greater than the 5.0% cutoff point? Consider the hospital’s performance as representing a sample from the population of possible operations it might have performed if the patients had been available.

SOLUTION

Formulate the Null and Alternative Hypotheses

The null hypothesis makes the assumption that the “population” mortality rate for the hospital cardiac surgery unit is really no greater than 0.05, and that the observed proportion, \( p = 0.07 \), was simply due to chance variation.

\[
H_0: \pi \leq 0.05 \quad \text{The true mortality rate for the unit is no more than 0.05.}
\]

\[
H_1: \pi > 0.05 \quad \text{The true mortality rate is greater than 0.05.}
\]

The center of the hypothesized distribution, \( \pi_0 = 0.05 \), is the highest possible value for which the null hypothesis could be true.

Select the Significance Level

The significance level has been specified as \( \alpha = 0.01 \). If the null hypothesis were really true, there would be no more than a 0.01 probability of incorrectly rejecting it.

Select the Test Statistic and Calculate Its Value

The test statistic will be \( z \), calculated as \( z = (p - \pi_0)/\sigma_p \). The standard error of the sample proportion and the calculated value of the test statistic are

\[
\sigma_p = \sqrt{\frac{\pi_0(1 - \pi_0)}{n}} = \sqrt{\frac{0.05(1 - 0.05)}{100}} = 0.02179
\]

and

\[
z = \frac{p - \pi_0}{\sigma_p} = \frac{0.07 - 0.05}{0.02179} = 0.92
\]

Identify the Critical Value for the Test Statistic and State the Decision Rule

For the 0.01 level, the critical value of \( z \) is \( z = +2.33 \). The decision rule can be stated as “Reject \( H_0 \) if calculated \( z > +2.33 \), otherwise do not reject.”

Compare the Calculated and Critical Values and Reach a Conclusion for the Null Hypothesis

Since the calculated value, \( z = 0.92 \), is less than the critical value, it falls into the nonrejection region of Figure 10.9 and the null hypothesis, \( H_0: \pi \leq 0.05 \), cannot be rejected.

Make the Related Business Decision

The cardiac surgery mortality rate for this hospital could have been as high as 0.07 merely by chance, and closing it could not be justified strictly on the basis of a “significantly greater than 0.05” guideline. [Notes: (1) The VA may have been striving for some lower population proportion not mentioned in the article, and (2) because the cardiac unit did not meet the minimum requirement of 150 operations per year, the VA would have closed it anyway.]

Statistics in Action 10.1 addresses a serious issue but can provide no conclusions. Use the Internet, find the latest data, and you can do your own testing.

Computer Solutions 10.3 shows how we can use Excel or Minitab to carry out a hypothesis test for a proportion. In this case, we are replicating the hypothesis test shown in Figure 10.8, using summary information. The printouts in Computer Solutions 10.3 show the \( p \)-value (0.217) for the test. This \( p \)-value represents the following statement: “If the population proportion really is 0.70, there is a 0.217 probability of getting a sample proportion this far away from 0.70 just by chance.” The \( p \)-value is not less than the level of significance we are using to reach a conclusion (i.e., \( p \)-value = 0.217 is not less than \( \alpha = 0.05 \)), and \( H_0: \pi = 0.70 \) cannot be rejected.

In the Minitab portion of Computer Solutions 10.3, the 95% confidence interval is shown as (0.594349 to 0.725651). The hypothesized population proportion (0.70) falls within the 95% confidence interval and, at this confidence level, it appears that the population proportion could be 0.70. This relationship between confidence intervals and two-tail hypothesis tests was discussed in Section 10.4.
STATISTICS IN ACTION

10.1

9/11 First Responders: Still Paying the Price?

As most everyone knows, September 11, 2001, was a sad day for both America and the world. The World Trade Center Twin Towers in New York City were struck by fuel-filled, hijacked aircraft and crumbled to the ground, resulting in both catastrophic fatalities and an environment filled with toxic smoke and dust. The police, firefighters, and other emergency-response heroes on the scene were forced to breathe a mixture of air and toxic dusts containing who-knows-what. The same goes for those who worked at the site during search, rescue, and clean-up operations.

In August 2009, an article in the Journal of Occupational and Environmental Medicine reported that a relatively unusual number of first-responders, especially those in the under-45 age group, are now suffering from a form of cancer that has historically been considered a disease of the elderly. Of the 28,252 responders studied, eight cases of multiple myeloma were diagnosed between September 11, 2001, and September 10, 2007, and half of these were in persons now under the age of 45.

According to the study authors, the expected number of cases of this disease among the under-45 responders would be 1.2, but the observed number was 4.0. Was this relatively high number of diagnoses among the under-45 due to chance variation, or did that toxic cloud inflict what medical professionals might call a major bodily insult in the form of a ticking time bomb that will eventually cause even greater physical damage to those who were first-responders on that fateful day?

In many Statistics in Action presentations, we are able to present data, findings, and conclusions. However, this one is an exception that calls for your own research and consideration. Before we could do any hypothesis testing, we would need more information in terms of how many of the first-responders are now younger than 45 and the percentage of the general population in this age group who are victims of this disease. We could then use either the normal or (preferably) the binomial distribution to determine the probability that the number of cases could have been this large simply by chance. Depending on what data your future Internet searching reveals on this serious matter, you may gain insights by using either the binomial distribution or the normal distribution approximation.

As a business statistics student, and as a concerned and caring citizen, you may find it relevant to use the Internet and other sources to keep abreast of these developments and be aware of the after-effects being experienced by the 9/11 first-responders. Perhaps more than any example we might provide within these pages, this one exemplifies that statistics are more than mere numbers.


COMPUTER 10.3 SOLUTIONS

Hypothesis Test for a Population Proportion

These procedures show how to carry out a hypothesis test for the population proportion.

EXCEL

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>z-Test of a Proportion</td>
<td>0.66</td>
<td>z Stat</td>
<td>-1.23</td>
</tr>
<tr>
<td>2</td>
<td>Sample proportion</td>
<td>0.66</td>
<td>z Stat</td>
<td>-1.23</td>
</tr>
<tr>
<td>3</td>
<td>Sample size</td>
<td>200</td>
<td>P[Z&lt;=z] one-tail</td>
<td>0.1085</td>
</tr>
<tr>
<td>4</td>
<td>Hypothesized proportion</td>
<td>0.70</td>
<td>z Critical one-tail</td>
<td>1.645</td>
</tr>
<tr>
<td>5</td>
<td>Alpha</td>
<td>0.05</td>
<td>P[Z&lt;=z] two-tail</td>
<td>0.2170</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td>z Critical two-tail</td>
<td>1.960</td>
</tr>
</tbody>
</table>
Excel hypothesis test for \( \pi \) based on summary statistics

1. For example, with \( n = 200 \) and \( p = 0.66 \), as in Figure 10.8: Open the TEST STATISTICS workbook.
2. Using the arrows at the bottom left, select the \textit{z-Test\_Proportion} worksheet. Enter the sample proportion (0.66), the sample size (200), the hypothesized population proportion (0.70), and the level of significance for the test (0.05). The \( p \)-value for this two-tail test is shown as 0.2170.

(Note: As an alternative, you can use Excel worksheet template TMPTEST. The steps are described within the template.)

Excel hypothesis test for \( \pi \) based on raw data

1. For example, using data file \textit{CX10GRAD}, with the label and 200 data values in A1:A201 and data coded as 1 = no job in field, 2 = job in field: From the Add-Ins ribbon, click Data Analysis Plus. Click \textit{Z-Test: Proportion}. Click OK.
2. Enter A1:A201 into the Input Range box. Enter 2 into the Code for Success box. Enter 0.70 into the Hypothesized Proportion box. Click Labels. Enter the level of significance for the test (0.05) into the Alpha box. Click OK.

MINITAB

Minitab hypothesis test for \( \pi \) based on summary statistics

![Minitab Output](image.jpg)

1. This interval is based on the summary statistics for Figure 10.8: Click Stat. Select Basic Statistics. Click 1 Proportion. Select Summarized Data. Enter the sample size (200) into the Number of Trials box. Multiply the sample proportion (0.66) times the sample size (200) to get the number of “events” or “successes” (0.66)(200) = 132. (Had this not been an integer, it would have been necessary to round to the nearest integer.) Enter the number of “successes” (132) into the Number of events box. Select Perform hypothesis test and enter the hypothesized population proportion (0.70) into the Hypothesized proportion box.
2. Click Options. Enter the desired confidence level as a percentage (95.0) into the Confidence Level box. Within the Alternative box, select not equal. Click to select Use test and interval based on normal distribution. Click OK. Click OK.

Minitab hypothesis test for \( \pi \) based on raw data

1. For example, using file \textit{CX10GRAD}, with column C1 containing the 200 assumed data values (coded as 1 = no job in field, 2 = job in field): Click Stat. Select Basic Statistics. Click 1 Proportion. Select Samples in columns and enter C1 into the dialog box. Select Perform hypothesis test and enter the hypothesized proportion (0.70) into the Hypothesized proportion box.
2. Follow step 2 in the summary-information procedure, above. Note: Minitab will select the larger of the two codes (i.e., 2 = job in field) as the “success” and provide the sample proportion and the confidence interval for the population proportion of graduates having jobs in their fields. To obtain the results for those not having jobs in their fields, just re-code the data so graduates without jobs in their fields will have the higher code number: Click Data. Select Code. Click Numeric to Numeric. Enter C1 into both the Code data from columns box and the Store coded data in columns box. Enter 1 into the Original values box. Enter 3 into the New box. Click OK. The new codes will be 3 = no job in field, 2 = job in field.
Had we used the computer to perform the test summarized in Figure 10.9, we would have found the \( p \)-value for this one-tail test to be 0.179. The \( p \)-value for the test says, “If the population proportion really is 0.05, there is a 0.179 probability of getting a sample proportion this large (0.07) just by chance.” The \( p \)-value (0.179) is not less than the level of significance we are using as a criterion (\( \alpha = 0.01 \)), so we are unable to reject \( H_0: \pi \leq 0.05 \).

### Exercises

10.59 When carrying out a hypothesis test for a population proportion, under what conditions is it appropriate to use the normal distribution as an approximation to the (theoretically correct) binomial distribution?

10.60 For a simple random sample, \( n = 200 \) and \( p = 0.34 \). At the 0.01 level, test \( H_0: \pi = 0.40 \) versus \( H_1: \pi \neq 0.40 \).

10.61 For a simple random sample, \( n = 1000 \) and \( p = 0.47 \). At the 0.05 level, test \( H_0: \pi \geq 0.50 \) versus \( H_1: \pi < 0.50 \).

10.62 For a simple random sample, \( n = 700 \) and \( p = 0.63 \). At the 0.025 level, test \( H_0: \pi \leq 0.60 \) versus \( H_1: \pi > 0.60 \).

10.63 A simple random sample of 300 items is selected from a large shipment, and testing reveals that 4% of the sampled items are defective. The supplier claims that no more than 2% of the items in the shipment are defective. Carry out an appropriate hypothesis test and comment on the credibility of the supplier’s claim.

10.64 The director of admissions at a large university says that 15% of high school juniors to whom she sends university literature eventually apply for admission. In a sample of 300 persons to whom materials were sent, 30 students applied for admission. In a two-tail test at the 0.05 level of significance, should we reject the director’s claim?

10.65 According to the human resources director of a plant, no more than 5% of employees hired in the past year have violated their preemployment agreement not to use any of five illegal drugs. The agreement specified that random urine checks could be carried out to ascertain compliance. In a random sample of 400 employees, screening detected at least one of these drugs in the systems of 8% of those tested. At the 0.025 level, is the human resources director’s claim credible? Determine and interpret the \( p \)-value for the test.

10.66 It has been claimed that 65% of homeowners would prefer to heat with electricity instead of gas. A study finds that 60% of 200 homeowners prefer electric heating to gas. In a two-tail test at the 0.05 level of significance, can we conclude that the percentage who prefer electric heating may differ from 65%? Determine and interpret the \( p \)-value for the test.

10.67 In the past, 44% of those taking a public accounting qualifying exam have passed the exam on their first try. Lately, the availability of exam preparation books and tutoring sessions may have improved the likelihood of an individual’s passing on his or her first try. In a sample of 250 recent applicants, 130 passed on their first attempt. At the 0.05 level of significance, can we conclude that the proportion passing on the first try has increased? Determine and interpret the \( p \)-value for the test.

10.68 Heritage Union has said that 66% of U.S. adults have purchased life insurance. Suppose that for a random sample of 50 adults from a given U.S. city, a researcher finds that only 56% of them have purchased life insurance. At the 0.05 level in a one-tail test, examine whether the percentage of sample homes having a fireplace could have differed from 55% simply by chance. Determine and interpret the \( p \)-value for the test.

10.69 According to the National Association of Home Builders, 55% of new single-family homes built during 2005 had a fireplace. Suppose a nationwide homebuilder has claimed that its homes are “a cross section of America,” but a simple random sample of 600 of its single-family homes built during that year included only 50.0% that had a fireplace. Using the 0.05 level of significance in a two-tail test, examine whether the percentage of sample homes having a fireplace could have differed from 55% simply by chance. Determine and interpret the \( p \)-value for the test.


10.70 Based on the sample results in Exercise 10.69, construct and interpret the 95% confidence interval for the population proportion. Is the hypothesized proportion (0.55) within the interval? Given the presence or absence of the 0.55 value within the interval, is this consistent with the findings of the hypothesis test conducted in Exercise 10.69?

10.71 According to the U.S. Bureau of Labor Statistics, 7.0% of female hourly workers who are 16 to 24 years old are being paid minimum wage or less. (Note that some workers in some industries are exempt from the minimum wage requirement of the Fair Labor Standards Act and, thus, could be legally earning less than the “minimum” wage.) A prominent politician is interested in how young working women within her county compare to this national percentage, and selects a simple random sample of 500 female hourly workers who are 16 to 24 years old. Of the women in the sample, 42 are being paid minimum wage or less. From these sample results, and using the 0.10 level of significance, could the politician conclude that the percentage of young female hourly workers who are low-paid in her county might be the same as the percentage of young women who are low-paid in the nation as a whole? Determine and interpret the \( p \)-value for the test. 


10.72 Using the sample results in Exercise 10.71, construct and interpret the 90% confidence interval for the population proportion. Is the hypothesized population proportion (0.07) within the interval? Given the presence or absence of the 0.07 value within the interval, is this consistent with the findings of the hypothesis test conducted in Exercise 10.71?

10.73 Brad Davenport, a consumer reporter for a national cable TV channel, is working on a story evaluating generic food products and comparing them to their brand-name counterparts. According to Brad, consumers claim to like the brand-name products better than the generics, but they can’t even tell which is which. To test his theory, Brad gives each of 200 consumers two potato chips—one generic, the other a brand name—and asks them which one is the brand-name chip. Fifty-five percent of the subjects correctly identify the brand-name chip. At the 0.025 level, is this significantly greater than the 50% that could be expected simply by chance? Determine and interpret the \( p \)-value for the test.

10.74 It has been reported that 80% of taxpayers who are audited by the Internal Revenue Service end up paying more money in taxes. Assume that auditors are randomly assigned to cases, and that one of the ways the IRS oversees its auditors is to monitor the percentage of cases that result in the taxpayer paying more taxes. If a sample of 400 cases handled by an individual auditor has 77.0% of those she audited paying more taxes, is there reason to believe her overall “pay more” percentage might be some value other than 80%? Use the 0.10 level of significance in reaching a conclusion. Determine and interpret the \( p \)-value for the test. 


10.75 Based on the sample results in Exercise 10.74, construct and interpret the 90% confidence interval for the population proportion. Is the hypothesized proportion (0.80) within the interval? Given the presence or absence of the 0.80 value within the interval, is this consistent with the findings of the hypothesis test conducted in Exercise 10.74?

10.76 According to the National Collegiate Athletic Association (NCAA), 41% of male basketball players graduate within 6 years of enrolling in their college or university, compared to 56% for the student body as a whole. Assume that data file XR10076 shows the current status for a sample of 200 male basketball players who enrolled in New England colleges and universities 6 years ago. The data codes are 1 = left school, 2 = still in school, 3 = graduated. Using these data and the 0.10 level of significance, does the graduation rate for male basketball players from schools in this region differ significantly from the 41% for male basketball players across the nation? Identify and interpret the \( p \)-value for the test. 

Source: “NCAA Basketball ‘Reforms’ Come Up Short,” USA Today, April 1, 2000, p. 17A.

10.77 Using the sample results in Exercise 10.76, construct and interpret the 90% confidence interval for the population proportion. Is the hypothesized proportion (0.41) within the interval? Given the presence or absence of the 0.41 value within the interval, is this consistent with the findings of the hypothesis test conducted in Exercise 10.76?

10.78 Website administrators sometimes use analysis tools or service providers to “track” the movements of visitors to the various portions of their site. Overall, the administrator of a political action website has found that 35% of the visitors who visit the “Environmental Issues” page go on to visit the “Here’s What You Can Do” page. In an effort to increase this rate, the administrator places a photograph of an oil-covered sea otter on the Issues page. Of the next 300 visitors to the Issues page, 40% also visit the Can Do page. The data are in file XR10078, coded as 1 = did not go on to visit Can Do page and 2 = went on to visit Can Do page. At the 0.05 level, is the 40% rate significantly greater than the 35% that had been occurring in the past, or might this higher rate be simply due to chance variation? Identify and interpret the \( p \)-value for the test.
Hypothesis Testing Errors and the Power of a Test

As discussed previously in the chapter, incorrect conclusions can result from hypothesis testing. As a quick review, the mistakes are of two kinds:

- **Type I error**, rejecting a true hypothesis:
  \[ \alpha = \text{probability of rejecting } H_0 \text{ when } H_0 \text{ is true} \]
  or
  \[ \alpha = P(\text{reject } H_0 | H_0 \text{ true}) \]
  \[ \alpha = \text{the level of significance of a test} \]

- **Type II error**, failing to reject a false hypothesis:
  \[ \beta = \text{probability of failing to reject } H_0 \text{ when } H_0 \text{ is false} \]
  or
  \[ \beta = P(\text{fail to reject } H_0 | H_0 \text{ false}) \]
  \[ 1 - \beta = \text{probability of rejecting } H_0 \text{ when } H_0 \text{ is false} \]
  \[ 1 - \beta = \text{the power of a test} \]

In this section, our focus will be on \((1 - \beta)\), the power of a test. As mentioned previously, there is a trade-off between \(\alpha\) and \(\beta\): For a given sample size, reducing \(\alpha\) tends to increase \(\beta\), and vice versa; with larger sample sizes, however, *both* \(\alpha\) and \(\beta\) can be decreased for a given test.

In wishing people luck, we sometimes tell them, “Don’t take any wooden nickels.” As an analogy, the power of a hypothesis test is the probability that the test will correctly reject the “wooden nickel” represented by a false null hypothesis. In other words, \((1 - \beta)\), the *power of a test*, is the probability that the test will respond correctly by rejecting a false null hypothesis.

The Power of a Test: An Example

As an example, consider the Extendabulb test, presented in Section 10.3 and illustrated in Figure 10.4. The test can be summarized as follows:

- **Null and alternative hypotheses:**
  \[ H_0: \mu \leq 1030 \text{ hours} \quad \text{Extendabulb is no better than the previous system.} \]
  \[ H_1: \mu > 1030 \text{ hours} \quad \text{Extendabulb does increase bulb life.} \]

- **Significance level selected:** 0.05
- **Calculated value of test statistic:**
  \[ z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{1061.6 - 1030.0}{90/\sqrt{40}} = 2.22 \]
Chapter 10: Hypothesis Tests Involving a Sample Mean or Proportion

- Critical value for test statistic:  \( z = +1.645 \)
- Decision rule:  Reject \( H_0 \) if calculated \( z > +1.645 \), otherwise do not reject.

For purposes of determining the power of the test, we will first convert the critical value, \( z = +1.645 \), into the equivalent mean bulb life for a sample of this size. This will be 1.645 standard error units to the right of the mean of the hypothesized distribution (1030 hours). The standard error for the distribution of sample means is \( \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{90}{\sqrt{40}} = 14.23 \) hours. The critical \( z \) value can now be converted into a critical sample mean:

\[
\text{Sample mean, } \bar{x} \quad \text{corresponding to } \quad z = +1.645
\]

\[
\bar{x} = 1030.00 + 1.645(14.23) = 1053.41 \text{ hours}
\]

and the decision rule, “Reject \( H_0 \) if calculated test statistic is greater than \( z = +1.645 \)” can be restated as “Reject \( H_0 \) if sample mean is greater than 1053.41 hours.”

The power of a test to correctly reject a false hypothesis depends on the true value of the population mean, a quantity that we do not know. At this point, we will assume that the true mean has a value that would cause the null hypothesis to be false, then the decision rule of the test will be applied to see whether this “wooden nickel” is rejected, as it should be.

As an arbitrary choice, the true mean life of Extendabulb-equipped bulbs will be assumed to be \( \mu = 1040 \) hours. The next step is to see how the decision rule, “Reject \( H_0 \) if the sample mean is greater than 1053.41 hours,” is likely to react. In particular, interest is focused on the probability that the decision rule will correctly reject the false null hypothesis that the mean is no more than 1030 hours.

As part (a) of Figure 10.10 shows, the distribution of sample means is centered on \( \mu = 1040 \) hours, the true value assumed for bulb life. The standard error of the distribution of sample means remains the same, so the spread of the sampling distribution is unchanged compared to that in Figure 10.4. In part (a) of Figure 10.10, however, the entire distribution has been “shifted” 10 hours to the right.

If the true mean is 1040 hours, the shaded portion of the curve in part (a) of Figure 10.10 represents the power of the hypothesis test—that is, the probability that it will correctly reject the false null hypothesis. Using the standard error of the sample mean, \( \sigma_{\bar{x}} = 14.23 \) hours, we can calculate the number of standard error units from 1040 to 1053.41 hours as

\[
z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{1053.41 - 1040.00}{14.23} = \frac{13.41}{14.23} = 0.94 \text{ standard error units to the right of the population mean}
\]

From the normal distribution table, we find the cumulative area to \( z = +0.94 \) is 0.8264. Since the total area beneath the curve is 1.0000, we can calculate the shaded area as 1.0000 – 0.8264, or 0.1736. Thus, if the true mean life of Extendabulb-equipped bulbs is 1040 hours, there is a 0.1736 probability that a sample of 40 bulbs will have a mean in the “reject \( H_0 \)” region of our test and that we will correctly reject the false null hypothesis that \( \mu \) is no more than 1030 hours. For a true mean of 1040 hours, the power of the test is 0.1736.
The power of the test \((1 - \beta)\) is the probability that the decision rule will correctly reject a false null hypothesis. For example, if the population mean were really 1040 hours (part a), there would be a 0.1736 probability that the decision rule would correctly reject the null hypothesis that \(H_0: \mu \leq 1030\) hours. Similarly, if the actual mean were 1060 hours (part b), the probability of correctly rejecting the null hypothesis would be 0.6772. If the actual mean were 1080 hours (part c), the probability of rejecting the null hypothesis would be 0.9693.
The Power Curve for a Hypothesis Test

One-Tail Test

In the preceding example, we arbitrarily selected one value ($\mu = 1040$ hours) that would make the null hypothesis false, then found the probability that the decision rule of the test would correctly reject the false null hypothesis. In other words, we calculated the power of the test ($1 - \beta$) for just one possible value of the actual population mean. If we were to select many such values (e.g., $\mu = 1060$, $\mu = 1080$, $\mu = 1100$, and so on) for which $H_0$ is false, we could calculate a corresponding value of $(1 - \beta)$ for each of them.

For example, part (b) of Figure 10.10 illustrates the power of the test whenever the Extendabulb-equipped bulbs are assumed to have a true mean life of 1060 hours. In part (b), the power of the test is 0.6772. This is obtained by the same approach used when the true mean life was assumed to be 1040 hours, but we are now using $\mu = 1060$ hours instead of 1040.

The diagram in part (c) of Figure 10.10 repeats this process for an assumed value of 1080 hours for the true population mean. Notice how the shape of the distribution is the same in diagrams (a), (b), and (c), but that the distribution itself shifts from one diagram to the next, reflecting the new true value being assumed for $\mu$.

Calculating the power of the test $(1 - \beta)$ for several more possible values for the population mean, we arrive at the power curve shown by the lower line in Figure 10.11. (The upper line in the figure will be discussed shortly.) As Figure 10.11 shows, the power of the test becomes stronger as the true population mean exceeds 1030 by a greater margin. For example, our test is almost
certain (probability = 0.9949) to reject the null hypothesis whenever the true population mean is 1090 hours. In Figure 10.11, the power of the test drops to zero whenever the true population mean is 1030 hours. This would also be true for all values lower than 1030 as well, because such actual values for the mean would result in the null hypothesis actually being true—in such cases, it would not be possible to reject a false null hypothesis, since the null hypothesis would not be false.

A complement to the power curve is known as the operating characteristic (OC) curve. Its horizontal axis would be the same as that in Figure 10.11, but the vertical axis would be identified as \( \beta \) instead of \( 1 - \beta \). In other words, the operating characteristic curve plots the probability that the hypothesis test will not reject the null hypothesis for each of the selected values for the population mean.

Two-Tail Test

In two-tail tests, the power curve will have a zero value when the assumed population mean equals the hypothesized value, then will increase toward 1.0 in both directions from that assumed value for the mean. In appearance, it will somewhat resemble an upside-down normal curve. The basic principle for power curve construction will be the same as for the one-tail test: Assume different population mean values for which the null hypothesis would be false, then determine the probability that an observed sample mean would fall into a rejection region originally specified by the decision rule of the test.

The Effect of Increased Sample Size on Type I and Type II Errors

For a given sample size, we can change the decision rule so as to decrease \( \beta \), the probability of making a Type II error. However, this will increase \( \alpha \), the probability of making a Type I error. Likewise, for a given sample size, changing the decision rule so as to decrease \( \alpha \) will increase \( \beta \). In either of these cases, we are involved in a trade-off between \( \alpha \) and \( \beta \). On the other hand, we can decrease both \( \alpha \) and \( \beta \) by using a larger sample size. With the larger sample size, (1) the sampling distribution of the mean or the proportion will be narrower, and (2) the resulting decision rule will be more likely to lead us to the correct conclusion regarding the null hypothesis.

If a test is carried out at a specified significance level (e.g., \( \alpha = 0.05 \)), using a larger sample size will change the decision rule but will not change \( \alpha \). This is because \( \alpha \) has been decided upon in advance. However, in this situation the larger sample size will reduce the value of \( \beta \), the probability of making a Type II error. As an example, suppose that the Extendabulb test of Figure 10.4 had involved a sample consisting of \( n = 60 \) bulbs instead of just 40. With the greater sample size, the test would now appear as follows:

- The test is unchanged with regard to the following:
  - Null hypothesis: \( H_0: \mu \leq 1030 \) hours
  - Alternative hypothesis: \( H_1: \mu > 1030 \) hours
  - Population standard deviation: \( \sigma = 90 \) hours
  - Level of significance specified: \( \alpha = 0.05 \)
The following are changed as the result of $n = 60$ instead of $n = 40$:

The standard error of the sample mean, $\sigma_{\bar{x}}$, is now

$$\frac{\sigma}{\sqrt{n}} = \frac{90}{\sqrt{60}} = 11.62 \text{ hours}$$

The critical $z$ of +1.645 now corresponds to a sample mean of

$$1030.00 + 1.645(11.62) = 1049.11 \text{ hours}$$

The decision rule becomes, “Reject $H_0$ if $\bar{x} > 1049.11$ hours.”

With the larger sample size and this new decision rule, if we were to repeat the process that led to Figure 10.10, we would find the following values for the power of the test. In the accompanying table, they are compared with those reported in Figure 10.10, with each test using its own decision rule for the 0.05 level of significance.

<table>
<thead>
<tr>
<th>True Value of Population</th>
<th>With $n = 60$ prob($\bar{x} &gt; 1049.11$)</th>
<th>With $n = 40$ prob($\bar{x} &gt; 1053.41$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1040</td>
<td>0.2177</td>
<td>0.1736</td>
</tr>
<tr>
<td>1060</td>
<td>0.8264</td>
<td>0.6772</td>
</tr>
<tr>
<td>1080</td>
<td>0.9961</td>
<td>0.9693</td>
</tr>
</tbody>
</table>

For example, for $n = 60$ and the decision rule shown,

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{1049.11 - 1080.00}{11.62} = -2.66$$

with the normal curve area to the right of $z = -2.66$ equal to 0.9961.

As these figures indicate, the same test with $n = 60$ (and the corresponding decision rule) would have a greater probability of correctly rejecting the false null hypothesis for each of the possible population means in Figure 10.10. If the Extendabulb example had remained the same, but only the sample size were changed, the $n = 60$ sample size would result in much higher $(1 - \beta)$ values than the ones previously calculated for $n = 40$. This curve is shown by the upper line in Figure 10.11. Combining two or more power curves into a display similar to Figure 10.11 can reveal the effect of various sample sizes on the susceptibility of a test to Type II error. Seeing Statistics Applet 13, at the end of the chapter, allows you to do “what-if” analyses involving a nondirectional test and its power curve values.

Computer Solutions 10.4 shows how we can use Excel or Minitab to determine the power of the test and generate a power curve like the ones in Figure 10.11. With Excel, we specify the assumed population means one at a time, while Minitab offers the capability of finding power values for an entire set of assumed population means at the same time. In either case, the procedure is much more convenient than using a pocket calculator and statistical tables, as was done in generating the values shown in Figure 10.10.
The Power Curve for a Hypothesis Test

These procedures show how to construct the power curve for a hypothesis test. The example used is the Extendabulb test described in Figures 10.4, 10.10, and 10.11.

**EXCEL**

**Excel procedure for obtaining the power curve for a hypothesis test**

1. Open the BETA-MEAN workbook. Using the arrows at the bottom left, select the Right-tail Test worksheet.
2. Enter 1030 as the hypothesized population mean, $H_0: \mu$. Enter 90 as SIGMA, 40 as the Sample size, and 0.05 as ALPHA. Enter 1040 as the value for $H_1: \mu$, the assumed population mean for which we want the power of the test. Excel automatically calculates the critical value for the sample mean as 1053.41 hours, and the power of the test appears in row 5 as 0.1731. This is comparable to part (a) of Figure 10.10, but is slightly more accurate because its calculation does not rely upon printed tables and the inherent gaps within them. We now have one point on the power curve for $n = 40$: Power = 0.1731 on the horizontal axis for an Assumed Actual $\mu$ of 1040 hours on the horizontal axis. This result is shown above, in the upper portion of the Excel printout.
3. Repeat step 2 to obtain the power of the test for each of several other assumed actual values for $\mu$—for example, 1030, 1050, 1060, 1070, 1080, and 1090. The resulting points can then be plotted as shown in the lower portion of the Excel printout.
MINITAB

Minitab procedure for obtaining the power curve for a hypothesis test

1. Click **Stat.** Select **Power and Sample Size.** Click **1-Sample Z.** Enter 40 into the **Sample Sizes** box. Enter 10 20 30 40 50 60 (separated by spaces) into the **Differences** box. These are the differences between the hypothesized population mean (1030) and the selected values for the assumed actual mean (i.e., 1040, 1050, 1060, 1070, 1080, and 1090). Enter 90 into the **Standard deviation** box.

2. Click **Options.** Within the **Alternative Hypothesis** box, select **Greater than.** Enter 0.05 into the **Significance level** box. Click **OK.** Click **OK.** The printout shows the power of the test for sample sizes of 40 and for assumed actual population means of 1040 (difference from 1030 is 10), 1050, 1060, 1070, 1080, and 1090. It corresponds to the values plotted in Figure 10.11. Minitab will also generate a power curve graph (not shown here).

---

**EXERCISES**

**10.79** What is a power curve and how is it applicable to hypothesis testing?

**10.80** What is an operating characteristic curve and how is it related to the power curve for a test?

**10.81** A hypothesis test has been set up and is to be conducted at the $\alpha = 0.05$ level of significance. If the sample size is doubled, what will be the effect on $\alpha$? On $\beta$?

**10.82** For the test described in Exercise 10.31, if the true population mean is really 2.520 inches, what is the probability that the inspector will correctly reject the false null hypothesis that $\mu = 2.500$ inches?

**10.83** For the test described in Exercise 10.32, assume that the true population mean for the new booklet is $\mu = 2.80$ hours. Under this assumption, what is the probability that the false null hypothesis, $H_0: \mu \geq 3.00$ hours, will be rejected?
10.84 Using assumed true population means of 2.485, 2.490, 2.495, 2.500, 2.505, 2.510, and 2.515 inches, plot the power curve for the test in Exercise 10.31.

10.85 Using assumed true population means of 2.80, 2.85, 2.90, 2.95, and 3.00 hours, plot the power curve for the test in Exercise 10.32.

10.86 Using assumed true population percentages of 2%, 3%, 4%, 5%, 6%, and 7%, plot the power curve for the test in Exercise 10.63.

10.87 When a thread-cutting machine is operating properly, only 2% of the units produced are defective. Since the machine was bumped by a forklift truck, however, the quality-control manager is concerned that it may require expensive downtime and a readjustment. The manager would like to set up a right-tail test at the 0.01 level of significance, then identify the proportion of defectives in a sample of 400 units produced. His null hypothesis is $H_0: \pi \leq 0.02$, with $H_1: \pi > 0.02$.

a. What critical value of $z$ will be associated with this test?

b. For the critical $z$ determined in part (a), identify the sample proportion that this $z$ value represents, then use this value of $p$ in stating a decision rule for the test.

c. What is the probability that the quality-control manager will fail to reject a false $H_0$ if the actual population proportion of defectives is $\pi = 0.02$? If $\pi = 0.03$? If $\pi = 0.04$? If $\pi = 0.05$? If $\pi = 0.06$?

d. Making use of the probabilities calculated in part (c), plot the power and operating characteristic curves for the test.

10.88 Plot the operating characteristic curve that corresponds to the power curve constructed for Exercise 10.84.

10.89 Plot the operating characteristic curve that corresponds to the power curve constructed for Exercise 10.86.

SUMMARY

- **Hypothesis testing: null and alternative hypotheses**
  Hypothesis testing is a method of using sample information to evaluate a claim that has been converted into a set of statements called the null and alternative hypotheses. The null hypothesis can be either rejected or not rejected; if it is not rejected, the alternative hypothesis, which asserts the null hypothesis to be wrong, must be rejected. The null hypothesis is given the benefit of the doubt and is assumed to be true unless we are faced with statistically overwhelming evidence to the contrary.

- **Directional and nondirectional hypothesis tests**
  A verbal statement from which a hypothesis test originates may be either directional or nondirectional. Accordingly, the appropriate hypothesis test will also be either directional (one-tail test) or nondirectional (two-tail test). The object of the hypothesis test may be a single population mean or proportion, as discussed in this chapter, or the difference between two population means or proportions (covered in the next chapter).

- **Type I and Type II errors**
  When we reject or fail to reject the null hypothesis, there is the possibility that our conclusion may be mistaken. Rejecting a true null hypothesis is known as a Type I error, while failing to reject a false null hypothesis results in a Type II error. Alpha ($\alpha$), the probability of incorrectly rejecting a true null hypothesis, is also known as the level of significance of the test. Beta ($\beta$) is the probability of incorrectly failing to reject a false null hypothesis. With sample size and type of test (one-tail versus two-tail) being equal, reducing $\alpha$ tends to increase $\beta$, and vice versa, though random samples of larger size can simultaneously reduce both $\alpha$ and $\beta$. 

Chapter 10: Hypothesis Tests Involving a Sample Mean or Proportion

• The hypothesis testing procedure
Hypothesis testing involves (1) formulating the null and alternative hypotheses, (2) selecting the significance level ($\alpha$) to be used in the test, (3) selecting the test statistic and calculating its value, (4) identifying the critical value(s) for the test statistic and stating the decision rule, (5) comparing calculated and critical value(s) for the test statistic and reaching a conclusion about the null hypothesis, and (6) making the related business decision for the situation that precipitated the hypothesis test in the first place. In a two-tail test, there will be two critical values for the test statistic; in a one-tail test, a single critical value.

• Tests involving a sample mean or proportion
When testing a sample mean, an important consideration in the selection of the test statistic ($z$ or $t$) is whether the population standard deviation ($\sigma$) is known. If other assumptions are met, the $t$-test is appropriate whenever $\sigma$ is unknown. Figure 10.2 summarizes the test statistic selection process and related assumptions.

If the calculated test statistic falls into a rejection region defined by the critical value(s), the difference between the sample mean or proportion and its hypothesized population value is judged to be too great to have occurred by chance. In such a case, the null hypothesis is rejected.

• The $p$-value for a hypothesis test
The $p$-value for a hypothesis test is the level of significance where the calculated value of the test statistic is exactly the same as a critical value, and is another way of expressing the level of significance ($\alpha$) of the test. When the hypothesis test is computer assisted, the computer output will include a $p$-value for use in interpreting the results. The $p$-value is the lowest level of significance at which the null hypothesis can be rejected.

• Confidence intervals and nondirectional tests
Although confidence intervals and hypothesis tests have different purposes, the concepts are actually related. In the case of a nondirectional test for a population mean, the conclusion at the $\alpha$ level of significance will be the same as that reached on the basis of a confidence interval constructed for the $100(1 - \alpha)\%$ confidence level.

• Power and operating characteristic curves for a test
Another measure of the performance of a hypothesis test is the power of the test. This is the probability ($1 - \beta$) that the test will avoid a Type II error by correctly rejecting a false null hypothesis. When the power of the test (vertical axis) is plotted against a range of population parameter values for which the null hypothesis would be false (horizontal axis), the result is known as the power curve for the test. A complement to the power curve is the operating characteristic (OC) curve, which has the same horizontal axis, but uses $\beta$ instead of $(1 - \beta)$ for the vertical axis. The OC curve expresses the probability that the test will make a Type II error by incorrectly failing to reject the null hypothesis for a range of population parameter values for which the null hypothesis would be false.
Testing a Mean, Population Standard Deviation Known

\[ z = \frac{\bar{x} - \mu_0}{\sigma_x} \]

where

- \( \sigma_x = \) standard error for the sample mean, \( = \frac{\sigma}{\sqrt{n}} \)
- \( \bar{x} = \) sample mean
- \( \mu_0 = \) hypothesized population mean
- \( n = \) sample size

If the population is not normally distributed, \( n \) should be \( \geq 30 \) for the central limit theorem to apply. The population \( \sigma \) is usually not known.

Testing a Mean, Population Standard Deviation Unknown

\[ t = \frac{\bar{x} - \mu_0}{s_x} \]

where

- \( s_x = \) estimated standard error for the sample mean, \( = \frac{s}{\sqrt{n}} \)
- \( \bar{x} = \) sample mean
- \( \mu_0 = \) hypothesized population mean
- \( n = \) sample size and \( df = n - 1 \)

For an unknown \( \sigma \), and when the population is approximately normally distributed, the \( t \)-test is appropriate regardless of the sample size. As \( n \) increases, the normality assumption becomes less important. If \( n < 30 \) and the population is not approximately normal, non-parametric testing (e.g., the sign test for central tendency, in Chapter 14) may be applied. The \( t \)-test is “robust” in terms of not being adversely affected by slight departures from the population normality assumption.

Testing a Proportion

\[ z = \frac{p - \pi_0}{\sigma_p} \]

where

- \( p = \) sample proportion
- \( \pi_0 = \) hypothesized population proportion
- \( n = \) sample size
- \( \sigma_p = \) standard error of the distribution of the sample proportion

\[ \sigma_p = \sqrt{\frac{\pi_0(1 - \pi_0)}{n}} \]

When \( n\pi \geq 5 \) and \( n(1 - \pi) \geq 5 \), the normal distribution is considered to be a good approximation to the binomial distribution. If this condition is not met, the exact probabilities must be derived from the binomial table. Most practical business settings involving proportions satisfy this condition, and the normal approximation is used in the chapter.

Confidence Intervals and Hypothesis Testing

Confidence intervals and hypothesis testing are related. For example, in the case of a non-directional test for a population mean, the conclusion at the \( \alpha \) level of significance will be the same as that reached on the basis of a confidence interval constructed for the 100(1 - \( \alpha \))% confidence level.
10.90 For each of the following situations, determine whether a one-tail test or a two-tail test would be appropriate. Describe the test, including the null and alternative hypotheses, then explain your reasoning in selecting it.
   a. A machine that has not been serviced for several months is producing output in which 5% of the items are defective. The machine has just been serviced and quality should now be improved.
   b. In a speech during her campaign for re-election, a Republican candidate claims that 55% of registered Democrats in her county intend to vote for her.
   c. Of those who have bought a new car in the past, a dealer has found that 70% experience three or more mechanical problems in the first four months of ownership. Unhappy with this percentage, the dealer has heavily revised the procedure by which predelivery mechanical checks are carried out.

10.91 Strands of human hair absorb elements from the bloodstream and provide a historical record of both health and the use or nonuse of chemical substances. Hair grows at the rate of about one-half inch per month, and a person with long hair might be accused or absolved on the basis of a segment of hair that sprouted many months or years ago. By separately analyzing sections of the hair strand, scientists can even approximate the periods of time during which drug use was heavy, moderate, light, or altogether absent. If a transit employee provides a strand of hair for drug testing, state the null and alternative hypotheses in verbal terms, then identify what would constitute a Type I error and a Type II error in this situation. Source: omegalabs.com, July 24, 2009.

10.92 Before accepting a large shipment of bolts, the director of an elevator construction project checks the tensile strength of a simple random sample consisting of 20 bolts. She is concerned that the bolts may be counterfeits, which bear the proper markings for this grade of bolt, but are made from inferior materials. For this application, the genuine bolts are known to have a tensile strength that is normally distributed with a mean of 1400 pounds and a standard deviation of 30 pounds. The mean tensile strength for the bolts tested is 1385 pounds. Formulate and carry out a hypothesis test to examine the possibility that the bolts in the shipment might not be genuine.

10.93 Before the hiring of an efficiency expert, the mean productivity of a firm’s employees was 45.4 units per hour, with a standard deviation of 4.5 units per hour. After incorporating the changes recommended by the expert, it was found that a sample of 30 workers produced a mean of 47.5 units per hour. Using the 0.01 level of significance, can we conclude that the mean productivity has increased?

10.94 The average U.S. family includes 3.13 persons. To determine whether families in her city tend to be representative in size compared to those across the United States, a city council member selects a simple random sample of 40 families. She finds the average number of persons in a family to be 3.40, with a standard deviation of 1.10. Source: Bureau of the Census, Statistical Abstract of the United States 2009, p. 52.
   a. Using a nondirectional hypothesis test and a level of significance of your choice, what conclusion would be reached?
   b. For the significance level used in part (a), construct the appropriate confidence interval and verify the conclusion reached in part (a).

10.95 The average U.S. household has $23,560 in life insurance. A local insurance agent would like to see how households in his city compare to the national average, and selects a simple random sample of 30 households from the city. For households in the sample, the average amount of life insurance is $245,800, with $ = $25,500. Source: reuters.com, September 23, 2008.
   a. Using a nondirectional hypothesis test and a level of significance of your choice, what conclusion would be reached?
   b. For the significance level used in part (a), construct the appropriate confidence interval and verify the conclusion reached in part (a).

10.96 Technical data provided with a room dehumidifier explain that units tested from the production line consume a mean of 800 watts of power, with a standard deviation of 12 watts. The superintendent of a large office complex, upon evaluating 30 of the units recently purchased and finding their mean power consumption to be 805 watts, claims that her shipment does not meet company specifications. At the 0.05 level of significance, is the superintendent’s complaint justified?

10.97 A regional office of the Internal Revenue Service randomly distributes returns to be audited to the pool of auditors. Over the thousands of returns audited last year, the average amount of extra taxes collected was $356 per audited return. One of the auditors, Jeffrey Jones, is suspected of being too lenient with persons whose returns are being audited. For a simple random sample of 30 of the returns audited by Mr. Jones last year, an average of $322 in extra taxes was collected, with a standard deviation of $90. Based on this information and a hypothesis test of...
your choice, do the suspicions regarding Mr. Jones appear to be justified?

10.98 A state transportation official claims that the mean waiting time at exit booths from a toll road near the capitol is no more than 0.40 minutes. For a sample of 35 motorists exiting the toll road, it was found that the mean waiting time was 0.46 minutes, with a standard deviation of 0.16 minutes. At the 0.05 level of significance, can we reject the official’s claim?

10.99 During 2006, 3.0% of all U.S. households were burglary victims. For a simple random sample of 300 households from a certain region, suppose that 18 households were victimized by burglary during that year. Apply an appropriate hypothesis test and the 0.05 level of significance in determining whether the region should be considered as having a burglary problem greater than that for the nation as a whole. Source: Bureau of the Census, Statistical Abstract of the United States 2009, p. 195.

10.100 A 1931 issue of *Time* magazine contained an advertisement for *Liberty* magazine. According to a study cited by *Liberty*, it was found that 15% of *Liberty* families had a “mechanical refrigerator,” compared to just 8% for all U.S. families. Assuming that the study included a sample of 120 *Liberty* families, and using the 0.01 level of significance, could the percentage of *Liberty* families owning this modern (for 1931) convenience be higher than for the nation overall? Source: *Time*, August 10, 1931, p. 49.

10.101 The administrator of a local hospital has told the governing board that 30% of its emergency room patients are not really in need of emergency treatment (i.e., the problems could just as easily be handled by an appointment with their family physician). In checking a random sample of 400 emergency room patients, a board member finds that 35% of those treated were not true emergency cases. Using an appropriate hypothesis test and the 0.05 level, evaluate the administrator’s statement.

10.102 In a taste comparison test, it was found that 58 of 100 persons preferred the chunky version of a peanut butter over the creamy type. An interested observer would like to determine whether this proportion (0.58) is significantly greater than the (0.50) proportion that would tend to result from chance. Using the 0.05 level of significance, what conclusion will be reached?

10.103 An exterminator claims that no more than 10% of the homes he treats have termite problems within 1 year after treatment. In a sample of 100 homes, local officials find that 14 had termites less than 1 year after being treated. At the 0.05 level of significance, evaluate the credibility of the exterminator’s statement.

10.104 A national chain of health clubs says the mean amount of weight lost by members during the past month was at least 5 pounds. Skeptical of this claim, a consumer advocate believes the chain’s assertion is an exaggeration. She interviews a random sample of 40 members, finding their mean weight loss to be 4.6 pounds, with a standard deviation of 1.5 pounds. At the 0.01 level of significance, evaluate the health club’s contention.

10.105 In an interview with a local newspaper, a respected trial lawyer claims that he wins at least 75% of his court cases. Bert, a skeptical statistics student, sets up a one-tail test at the 0.05 level of significance to evaluate the attorney’s claim. The student plans to examine a random sample of 40 cases tried by the attorney and determine the proportion of these cases that were won. The null and alternative hypotheses are $H_0: \pi \geq 0.75$ and $H_1: \pi < 0.75$. Using the techniques of this chapter, Bert sets up a hypothesis test in which the decision rule is “Reject $H_0$ if $z < -1.645$, otherwise do not reject.” What is the probability that Bert will make a Type II error (fail to reject a false null hypothesis) if the attorney’s true population proportion of wins is actually

a. $\pi = 0.75$? b. $\pi = 0.70$? c. $\pi = 0.65$?

d. $\pi = 0.60$? e. $\pi = 0.55$?

f. Making use of the probabilities calculated in parts (a) through (e), describe and plot the power curve for Bert’s hypothesis test.

10.106 A consumer agency suspects that a pet food company may be underfilling packages for one of its brands. The package label states “1600 grams net weight,” and the president of the company claims the average weight is at least 1600 grams. For a simple random sample of 35 boxes collected by the consumer agency, the mean and standard deviation were 1591.7 grams and 18.5 grams, respectively.

a. At the 0.05 level of significance, what conclusion should be reached by the consumer agency? Would the president of the company prefer to use a different level of significance in reaching a conclusion? Explain.

b. Use the decision rule associated with part (a) and a range of selected assumed values for $\mu$ in constructing the power curve for the test.

10.107 Historically, Shop-Mart has gotten an average of 2000 hours of use from its G&E fluorescent lightbulbs. Because its fixtures are attached to the ceiling, the bulbs are rather cumbersome to replace, and Shop-Mart is looking into the possibility of switching to Phipps bulbs, which cost the same. A sample of Phipps bulbs lasted an average of 2080 hours, and the $p$-value in a right-tail test ($H_0: \mu \leq 2000$ versus $H_1: \mu > 2000$) is 0.012. Shop-Mart has shared these results with both G&E and Phipps.

a. Give Shop-Mart a recommendation. In doing so, use “0.012” in a sentence that would be understood by a Shop-Mart executive who has had no statistical training but is comfortable with probabilities as used by weather forecasters.
b. In interpreting these results, what level of significance might G&E like to use in reaching a conclusion? Explain.

c. In interpreting these results, what level of significance might Phipps like to use in reaching a conclusion? Explain.

d. If the test had been two-tail (i.e., \( H_0: \mu = 2000 \) versus \( H_1: \mu \neq 2000 \)) instead of right-tail, would the \( p \)-value still be 0.012? If not, what would the \( p \)-value be? Explain.

( DATA SET ) Note: Exercises 10.108–10.113 require a computer and statistical software.

10.108 Before installing a high-tech laser exhibit near the exit area, a museum of American technology found the average contribution by patrons was $2.75 per person. For a sample of 30 patrons following installation of the new exhibit, the contributions are as listed in data file XR10108. Based on the sample data, and using the 0.025 level of significance, can we conclude that the new exhibit has increased the average contribution of exhibit patrons? Identify and interpret the \( p \)-value for the test.

10.109 According to the U.S. Department of Justice, 25% of violent crimes involve the use of a weapon. We will assume that data file XR10109 contains a sample of the crime information for a given city, with data for the 400 crimes coded as 1 = crime involved a weapon and 2 = crime did not involve a weapon. Using the 0.05 level in a two-tail test, was the percentage of violent crimes involving a weapon in this city significantly different from the 25% in the nation as a whole? Identify and interpret the \( p \)-value for the test. Source: Bureau of the Census, Statistical Abstract of the United States 2009, p. 194.

10.110 Based on the sample results in Exercise 10.109, construct and interpret the 95% confidence interval for the population proportion. Is the hypothesized proportion (0.25) within the interval? Given the presence or absence of the 0.25 value within the interval, is this consistent with the findings of the hypothesis test conducted in Exercise 10.109?

10.111 In making aluminum castings into alternator housings, an average of 3.5 ounces per casting must be trimmed off and recycled as a raw material. A new manufacturing procedure has been proposed to reduce the amount of aluminum that must be recycled in this way. For a sample of 35 castings made with the new process, data file XR10111 lists the weights of aluminum trimmed and recycled. Based on the sample data, and using the 0.01 level, has the new procedure significantly reduced the average amount of aluminum trimmed and recycled? Identify and interpret the \( p \)-value for the test.

10.112 Use the decision rule associated with Exercise 10.111 and a range of selected assumed values for \( \mu \) to construct the power curve for the test.

10.113 In the past, the mean lifetime of diesel engine injection pumps has been 12,000 operating hours. A new injection pump is available that is promoted as lasting longer than the old version. In a test of 50 of the new pumps, the lifetimes are as listed in data file XR10113. Based on the sample data, and using the 0.025 level of significance, examine the possibility that the new pumps might have a mean life that is no more than that of the old design. Identify and interpret the \( p \)-value for the test.

INTEGRATED CASES

Thorndike Sports Equipment

The first item on Ted Thorndike’s agenda for today is a meeting with Martha Scott, a technical sales representative with Cromwell Industries. Cromwell produces a racquet-stringing machine that is promoted as being the fastest in the industry. In their recent telephone conversation, Martha offered Ted the chance to try the machine for a full week, then return it to Cromwell if he decides it doesn’t offer a worthwhile improvement.

Currently, the Thorndike racquetball and tennis racquets are strung using a type of machine that has been around for about 10 years, and the Thorndikes have been very pleased with the results. When efficiency experts visited the plant last year, they found it took an average of 3.25 minutes to string a racquetball racquet and an average of 4.13 minutes for a tennis racquet. Both distributions were found to be approximately normal.

Ms. Scott comes by, explains the operation of the machine, and assists in its installation. Because its controls are similar in function and layout to the older model, no operator training is necessary. During the 1-week trial, the new machine will be working side-by-side with older models in the racquet-stringing department.

After Ms. Scott leaves, Luke and Ted Thorndike discuss the machine further. Luke indicates that the oldest of the
current machines is about due for replacement and, if the Cromwell model is really faster, he will buy the Cromwell to replace it. It’s possible that Cromwell models would also be purchased as other stringing machines approach the end of their operating lifetimes and are retired. Luke cautions Ted that the Cromwell must be purchased only if it is indeed faster than the current models—although Ms. Scott seems to have an honest face, he’s never trusted anybody whose briefcase contains brochures. The Cromwell must be quicker in stringing both racquetball racquets and tennis racquets. Otherwise, the firm will continue purchasing and using the current model.

Evaluating the Cromwell model, Ted measures the exact time required for each racquet in separate samples consisting of 40 racquetball racquets and 40 tennis racquets. The times are given here and are also in data file THORN10.

Since Luke seems quite adamant against buying something that isn’t any faster than the current models, Ted must be very confident that any increased speeds measured are not simply due to chance. Using a computer statistical package to analyze the sample times, do the respective tests appear to warrant purchase of the Cromwell machine, or should it be returned to Cromwell at the end of the week?

40 Racquetball Racquets:

<table>
<thead>
<tr>
<th>Time (minutes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.97</td>
</tr>
<tr>
<td>3.43</td>
</tr>
<tr>
<td>3.10</td>
</tr>
<tr>
<td>2.79</td>
</tr>
<tr>
<td>3.41</td>
</tr>
<tr>
<td>3.22</td>
</tr>
<tr>
<td>3.36</td>
</tr>
<tr>
<td>3.46</td>
</tr>
<tr>
<td>3.11</td>
</tr>
<tr>
<td>3.26</td>
</tr>
<tr>
<td>3.36</td>
</tr>
<tr>
<td>3.48</td>
</tr>
<tr>
<td>3.09</td>
</tr>
<tr>
<td>3.06</td>
</tr>
<tr>
<td>2.67</td>
</tr>
<tr>
<td>3.13</td>
</tr>
<tr>
<td>2.97</td>
</tr>
<tr>
<td>3.09</td>
</tr>
<tr>
<td>3.18</td>
</tr>
<tr>
<td>2.71</td>
</tr>
<tr>
<td>3.41</td>
</tr>
<tr>
<td>2.83</td>
</tr>
<tr>
<td>3.20</td>
</tr>
<tr>
<td>3.32</td>
</tr>
<tr>
<td>2.70</td>
</tr>
<tr>
<td>3.17</td>
</tr>
<tr>
<td>3.16</td>
</tr>
<tr>
<td>3.34</td>
</tr>
<tr>
<td>2.91</td>
</tr>
<tr>
<td>3.34</td>
</tr>
<tr>
<td>2.97</td>
</tr>
<tr>
<td>3.14</td>
</tr>
<tr>
<td>3.36</td>
</tr>
<tr>
<td>3.89</td>
</tr>
<tr>
<td>3.21</td>
</tr>
<tr>
<td>3.13</td>
</tr>
<tr>
<td>2.92</td>
</tr>
<tr>
<td>2.94</td>
</tr>
<tr>
<td>3.18</td>
</tr>
</tbody>
</table>

40 Tennis Racquets:

<table>
<thead>
<tr>
<th>Time (minutes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.77</td>
</tr>
<tr>
<td>3.97</td>
</tr>
<tr>
<td>4.08</td>
</tr>
<tr>
<td>4.36</td>
</tr>
<tr>
<td>3.85</td>
</tr>
<tr>
<td>4.65</td>
</tr>
<tr>
<td>3.89</td>
</tr>
<tr>
<td>3.99</td>
</tr>
<tr>
<td>3.51</td>
</tr>
<tr>
<td>3.43</td>
</tr>
<tr>
<td>4.31</td>
</tr>
<tr>
<td>3.84</td>
</tr>
<tr>
<td>3.86</td>
</tr>
<tr>
<td>3.50</td>
</tr>
<tr>
<td>4.23</td>
</tr>
<tr>
<td>3.88</td>
</tr>
<tr>
<td>3.91</td>
</tr>
<tr>
<td>4.47</td>
</tr>
<tr>
<td>4.27</td>
</tr>
<tr>
<td>3.70</td>
</tr>
<tr>
<td>3.80</td>
</tr>
<tr>
<td>4.54</td>
</tr>
<tr>
<td>4.46</td>
</tr>
<tr>
<td>3.97</td>
</tr>
<tr>
<td>4.49</td>
</tr>
<tr>
<td>3.74</td>
</tr>
<tr>
<td>4.17</td>
</tr>
<tr>
<td>3.68</td>
</tr>
<tr>
<td>4.09</td>
</tr>
<tr>
<td>4.27</td>
</tr>
<tr>
<td>3.84</td>
</tr>
<tr>
<td>3.51</td>
</tr>
<tr>
<td>3.71</td>
</tr>
<tr>
<td>3.77</td>
</tr>
<tr>
<td>3.90</td>
</tr>
<tr>
<td>4.14</td>
</tr>
<tr>
<td>4.03</td>
</tr>
<tr>
<td>4.61</td>
</tr>
<tr>
<td>3.63</td>
</tr>
<tr>
<td>4.62</td>
</tr>
</tbody>
</table>

**Springdale Shopping Survey**

The Springdale computer database exercise of Chapter 2 listed 30 shopping-related questions asked of 150 respondents in the community of Springdale. The data are in file SHOPPING. The coding key for these responses was also provided in the Chapter 2 exercise. In this section, hypothesis testing will be used to examine the significance of some of the survey results.

From a managerial perspective, the information gained from this database exercise would be useful in determining whether the mean attitude toward a given area differed significantly from the “neutral” level, in observing and comparing the mean attitude scores exhibited toward the three shopping areas, and in examining the strengths and weaknesses of a manager’s shopping area relative to the other two areas. If such a study were repeated from time to time, management could observe the extent to which the overall attitude toward their shopping area was becoming stronger or weaker in the eyes of the consumer, thus helping them select business strategies that could retain and take advantage of their area’s strengths and either correct or minimize the effect of its weaknesses.

1. Item C in the description of the data collection instrument lists variables 7, 8, and 9, which represent the respondent’s general attitude toward each of the three shopping areas. Each of these variables has numerically equal distances between the possible responses, and for purposes of analysis they may be considered to be of the interval scale of measurement.
   a. Calculate the mean attitude score toward the Springdale Mall shopping area and compute the sample standard deviation. Does this area seem to be well regarded by the respondents?
   b. In a two-tail test at the 0.10 level, compare the sample mean calculated in part (a) to the 3.0 “neutral” value. Use the null hypothesis $H_0: \mu = 3.0$ and the alternative hypothesis $H_1: \mu \neq 3.0$.
   c. Generate the 90% confidence interval for the population mean and use this interval in verifying the conclusion obtained in part (b).
   d. What is the $p$-value associated with the hypothesis test?
   e. Repeat parts (a) through (d) for the Downtown shopping area.
   f. Repeat parts (a) through (d) for the West Mall shopping area.

2. If Springdale Mall offered exactly the same benefits as the other two shopping areas, we would expect exactly one-third of those who expressed an opinion to select it as the area best fitting the description for variable number 10 ("Easy to return/exchange goods"). In testing whether Springdale Mall differs significantly from this expected proportion, the null and alternative hypotheses will be $H_0: \pi = 0.333$ and $H_1: \pi \neq 0.333$. Carry out the following analyses for Springdale Mall:
   a. Analyzing the data for variable 10, use the preceding null and alternative hypotheses in testing the null hypothesis at the 0.05 level.
   b. Determine the $p$-value for the test conducted in part (a).
   c. Repeat parts (a) and (b) for variables 11–17.
   d. Based on the preceding analyses, identify the principal strengths and weaknesses exhibited by Springdale Mall.
Pronto Pizza is a family-owned pizza restaurant in Vinemont, a small town of 20,000 people in upstate New York. Antonio Scapelli started the business 30 years ago as Antonio’s Restaurant with just a few thousand dollars. Antonio, his wife, and their children, most of whom are now grown, operate the business. Several years ago, one of Antonio’s sons, Tony Jr., graduated from NYU with an undergraduate degree in business administration. After graduation, he came back to manage the family business. Pronto Pizza was one of the earliest pizza restaurants to offer pizza delivery to homes. Fortunately, Tony had the foresight to make this business decision a few years ago. At the same time, he changed the restaurant’s name from Antonio’s to Pronto Pizza to emphasize the pizza delivery service. The restaurant has thrived since then and has become one of the leading businesses in the area. While many of their customers still “dine in” at the restaurant, nearly 90% of Pronto’s current business is derived from the pizza delivery service.

Recently, one of the national chains of fast-food pizza delivery services found its way to Vinemont, New York. In order to attract business, this new competitor guarantees delivery of its pizzas within 30 minutes after the order is placed. If the delivery is not made within 30 minutes, the customer receives the order without charge. Before long, there were signs that this new pizza restaurant was taking business away from Pronto Pizza. Tony realized that Pronto Pizza would have to offer a similar guarantee in order to remain competitive.

After a careful cost analysis, Tony determined that to offer a guarantee of 29 minutes or less, Pronto’s average delivery time would have to be 25 minutes or less. Tony thought that this would limit the percentage of “free pizzas” under the guarantee to about 5% of all deliveries, which he had figured to be the break-even point for such a promotion. To be sure of Pronto’s ability to deliver on a promise of 29 minutes or less, Tony knew that he needed to collect data on Pronto’s pizza deliveries.

Pronto Pizza’s delivery service operates from 4:00 p.m. to midnight every day of the week. After an order for a pizza is phoned in, one of the two cooks is given the order for preparation. When the crust is prepared and the ingredients have been added, the pizza is placed on the belt of the conveyor oven. The speed of the conveyor is set so that pizzas come out perfectly, time after time. Once the pizza is ready and one of Pronto’s drivers is available to make the delivery, the pizza is taken in a heat-insulated bag to the customer. Pronto uses approximately five to six drivers each night for deliveries. Most of the drivers hired by Pronto Pizza are juniors and seniors at the local high school.

Given the large number of deliveries made each evening, Tony knew that he could not possibly monitor every single delivery. He had thought of the possibility of having someone else collect the data, but given the importance of accurate data, he decided to make all of the measurements himself. This, of course, meant taking a random sample of, rather than all, deliveries over some time period. Tony decided to monitor deliveries over the course of a full month. During each hour of delivery service operation, he randomly selected a phoned-in order. He then carefully measured the time required to prepare the order and the amount of time that the order had to wait for a delivery person to become available. Tony would then go with the delivery person to accurately measure the delivery time. After returning, Tony randomly selected an order placed during the next hour and repeated the process. At the end of the month, Tony had collected data on 240 deliveries.

Once the data were available, Tony knew there were several issues that should be addressed. He was committed to going with the 29-minute delivery guarantee unless the data strongly indicated that the true average delivery time was greater than 25 minutes. How would he make this decision? Tony also realized that there were three components that could affect pizza delivery times: the preparation time, the waiting time for an available driver, and the travel time to deliver the pizza to the customer. Tony hoped that he had collected sufficient data to allow him to determine how he might improve the delivery operation by reducing the overall delivery time.

Assignment
Tony has asked you for some assistance in interpreting the data he has collected. In particular, he needs to know whether the true average delivery time for Pronto Pizza might be more than 25 minutes. Use the data in the file PRONTO in answering his question. A description of this data set is given in the accompanying Data Description section. Based on your examination of the data, provide Tony with some suggestions that would help him in making his decision about the 29-minute delivery guarantee and in improving his pizza delivery service. The case questions will assist you in your analysis of the data. Use important details from your analysis to support your recommendations.
**Data Description**

The PRONTO file contains the data collected by Tony Scapelli over the past month on pizza deliveries. Data recorded for each delivery order are as shown in the table below.

The variables are defined as follows:

<table>
<thead>
<tr>
<th>DAY</th>
<th>HOUR</th>
<th>PREP_TIME</th>
<th>WAIT_TIME</th>
<th>TRAVEL_TIME</th>
<th>DISTANCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>4</td>
<td>14.86</td>
<td>3.08</td>
<td>6.02</td>
<td>2.5</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>14.84</td>
<td>13.81</td>
<td>5.47</td>
<td>3.3</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>15.41</td>
<td>9.91</td>
<td>8.99</td>
<td>4.9</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>16.34</td>
<td>2.08</td>
<td>7.98</td>
<td>3.8</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>15.19</td>
<td>2.69</td>
<td>9.01</td>
<td>4.9</td>
</tr>
</tbody>
</table>

Note that the confidence interval limits shown in the graph may sometimes differ slightly from those we would calculate using the pocket calculator and our standard normal distribution table. This is because the applet is using more exact values for z than we are able to show within printed tables like the one in the text.

**Applet Exercises**

12.1 When the applet initially loads, identify the sample mean and describe the 95% confidence interval for \( \mu \). Based on this confidence interval, would it seem believable that the true population mean might be 1.325 minutes?

12.2 Using the slider, increase the sample mean to approximately 1.330 minutes. What is the 95% confidence interval for \( \mu \)? Based on this confidence interval, would it seem believable that the true population mean might be 1.325 minutes?

12.3 Using the slider, decrease the sample mean to approximately 1.310 minutes. Based on this confidence interval, would it seem believable that the true population mean might be 1.325 minutes?

Try out the Applets: http://www.cengage.com/bstatistics/weiers
3. Based on the data Tony has already collected, approximately what percentage of the time will Pronto Pizza fail to meet its guaranteed time of 29 minutes or less? Explain how you arrived at your answer and discuss any assumptions that you are making. Does it appear that they will meet their requirement of failing to meet the guarantee 5% of the time or less?

4. Compare the average delivery times for different hours of the day. Does the time of day seem to have an effect on the average time a customer will have to wait for his or her pizza? Attach any supporting output for your answer.

5. Based on your analysis of the data, what action (or actions) would you recommend to the owners of Pronto Pizza to improve their operation? Attach any supporting output that led to your conclusion.

This applet allows "what-if" analyses of the power of a nondirectional test with null and alternative hypotheses $H_0: \mu = 10$ versus $H_1: \mu \neq 10$. There are three sliders, allowing us to choose the $\alpha$ level for the test (left slider), the sample size (right slider), and the actual value of $\mu$ (identified as "Alt Hyp" and controlled by the bottom slider). As we move any of the sliders, we can immediately see the power of the test in the display at the top of the graph. Recall that the power of the test is $(1 - \beta)$, and that it is the probability of rejecting a false null hypothesis.

The distribution in the upper portion of the chart shows the hypothesized population mean ($H_0: \mu = 10$) as well as the lower and upper critical values for the sample mean in a nondirectional test for the combination of $\alpha$ and $n$ being used. The sum of the blue areas in the lower distribution show, for a given $\alpha$, actual $\mu$, and $n$, the probability that $\bar{x}$ will fall outside the critical values and cause us to reject the null hypothesis.

**Applet Exercises**

13.1 Using the left and right sliders, set up the test so that $\alpha = 0.10$ and $n = 20$. Now move the bottom slider so that the actual $\mu$ is as close as possible to 10 without being equal to 10 (e.g., 10.01). What is the value for the power of the test? Is this the value you would expect?

13.2 Set the left, right, and bottom sliders so that $\alpha = 0.05$, $n = 15$, and $\mu$ = the same value you selected in Applet Exercise 13.1. Now move the left slider upward and downward to change the $\alpha$ level for the test. In what way(s) do these changes in $\alpha$ affect the power of the test?

13.3 Set the left and bottom sliders so that $\alpha = 0.05$ and $\mu = 11.2$. Now move the right slider upward and downward to change the sample size for the test. Observe the power of the test for each of the following values for $n$: 2, 10, 20, 40, 60, 80, and 100.

13.4 Set the right and bottom sliders so that $n = 20$ and $\mu = 11.2$. Now move the left slider upward and downward to change the $\alpha$ level for the test. Observe the power of the test for each of the following values for $\alpha$: 0.01, 0.02, 0.05, 0.10, 0.20, 0.30, 0.40, and 0.50.

Try out the Applets: http://www.cengage.com/bstatistics/weiers