Most American college students make use of the Internet for both academic and social purposes. The authors of the paper “U.S. College Students’ Internet Use: Race, Gender and Digital Divides” (Journal of Computer-Mediated Communication [2009]: 244–264) describe the results of a survey of 7421 students at 40 colleges and universities. The sample was selected in a way that the authors believed would result in a sample that reflected general demographics of college students in the U.S. The authors wanted to use the sample data to estimate the proportion of college students who spend more than 3 hours a day on the Internet. The methods introduced in this chapter will be used to produce the desired estimate. Because the estimate will be based only on a sample rather than on a census of all U.S. college students it is important that this estimate be constructed in a way that also conveys information about the anticipated accuracy.

The objective of inferential statistics is to use sample data to decrease our uncertainty about some characteristic of the corresponding population, such as a population mean $\mu$ or a population proportion $p$. One way to accomplish this uses the
sample data to arrive at a single number that represents a plausible value for the characteristic of interest. Alternatively, an entire range of plausible values for the characteristic can be reported. These two estimation techniques, point estimation and interval estimation, are introduced in this chapter.

### 9.1 Point Estimation

The simplest approach to estimating a population characteristic involves using sample data to compute a single number that can be regarded as a plausible value of the characteristic. For example, sample data might suggest that 1000 hours is a plausible value for \( \mu \), the true mean lifetime for lightbulbs of a particular brand. In a different setting, a sample survey of students at a particular university might lead to the statement that \( .41 \) is a plausible value for \( p \), the proportion of all students who favor a fee for recreational facilities.

In the examples just given, 1000 is a point estimate of \( \mu \) and \( .41 \) is a point estimate of \( p \). The adjective point reflects the fact that the estimate corresponds to a single point on the number line.

A point estimate is obtained by first selecting an appropriate statistic. The estimate is then the value of the statistic for the given sample. For example, the computed value of the sample mean is one point estimate of a population mean \( \mu \), and the sample proportion is a point estimate of a population proportion, \( p \).

#### Definition

A **point estimate** of a population characteristic is a single number that is based on sample data and represents a plausible value of the characteristic.

In the examples just given, 1000 is a point estimate of \( \mu \) and \( .41 \) is a point estimate of \( p \). The adjective point reflects the fact that the estimate corresponds to a single point on the number line.

A point estimate is obtained by first selecting an appropriate statistic. The estimate is then the value of the statistic for the given sample. For example, the computed value of the sample mean is one point estimate of a population mean \( \mu \), and the sample proportion is a point estimate of a population proportion, \( p \).

#### Example 9.1 Internet Use by College Students

One of the purposes of the survey described in the chapter introduction was to estimate the proportion of college students who spend more than 3 hours a day on the Internet. Based on information given in the paper, 2998 of the 7421 students surveyed reported Internet use of more than 3 hours per day. We can use this information to estimate \( p \), where \( p \) is the proportion of all U.S. college students who use the Internet more than 3 hours a day. With a success identified as a student who uses the Internet more than 3 hours a day, \( p \) is then the population proportion of successes. The statistic

\[
\hat{p} = \frac{\text{number of successes in the sample}}{n}
\]

which is the sample proportion of successes, is an obvious choice for obtaining a point estimate of \( p \). Based on the reported information, the point estimate of \( p \) is

\[
\hat{p} = \frac{2998}{7421} = .404
\]

That is, based on this random sample, we estimate that 40.4% of college students in the United States spend more than 3 hours a day on the Internet.

For purposes of estimating a population proportion \( p \), there is no obvious alternative to the statistic \( \hat{p} \). In other situations, such as the one illustrated in Example 9.2, there may be several statistics that can be used to obtain an estimate.
The paper “The Impact of Internet and Television Use on the Reading Habits and Practices of College Students” (Journal of Adolescent and Adult Literacy [2009]: 609–619) investigates the reading habits of college students. The authors distinguished between recreational reading and academic reading and asked students to keep track of time spent reading. The following observations represent the number of hours spent on academic reading in 1 week by 20 college students (these data are compatible with summary values given in the paper and have been arranged in order from smallest to largest):

1.7 3.8 4.7 9.6 11.7 12.3 12.3 12.4 12.6 13.4
14.1 14.2 15.8 15.9 18.7 19.4 21.2 21.9 23.3 28.2

A dotplot of the data is shown here:

From the dotplot, we can see that the distribution of academic reading time is approximately symmetric.

If a point estimate of \( \mu \), the mean academic reading time per week for all college students, is desired, an obvious choice of a statistic for estimating \( \mu \) is the sample mean, \( \bar{x} \). However, there are other possibilities. We might consider using a trimmed mean or even the sample median, because the data set exhibits some symmetry. (If the corresponding population distribution is symmetric, the population mean \( \mu \) and the population median are equal).

The three statistics and the resulting estimates of \( \mu \) calculated from the data are

Sample mean = \( \bar{x} = \frac{\sum x}{n} = \frac{287.2}{20} = 14.36 \)

Sample median = \( \frac{13.4 + 14.1}{2} = 13.75 \)

10% trimmed mean = \( \frac{\text{average of middle 16 observations}}{16} = \frac{230.2}{16} = 14.39 \)

The estimates of the mean academic reading time per week for college students differ somewhat from one another. The choice from among them should depend on which statistic tends, on average, to produce an estimate closest to the true value of \( \mu \). The following subsection discusses criteria for choosing among competing statistics.

### Choosing a Statistic for Computing an Estimate

As illustrated in Example 9.2, more than one statistic may be reasonable to use to obtain a point estimate of a specified population characteristic. We would like to use a statistic that tends to produce an accurate estimate—that is, an estimate close to the value of the population characteristic. Information about the accuracy of estimation for a particular statistic is provided by the statistic’s sampling distribution.
Figure 9.1 displays the sampling distributions of three different statistics. The value of the population characteristic, which is denoted by *true value* in the figure, is marked on the measurement axis. The distribution in Figure 9.1(a) is that of a statistic unlikely to yield an estimate close to the true value. The distribution is centered to the right of the true value, making it very likely that an estimate (a value of the statistic for a particular sample) will be larger than the true value. If this statistic is used to compute an estimate based on a first sample, then another estimate based on a second sample, and another estimate based on a third sample, and so on, the long-run average value of these estimates will be greater than the true value.

The sampling distribution of Figure 9.1(b) is centered at the true value. Thus, although one estimate may be smaller than the true value and another may be larger, when this statistic is used many times over with different samples, there will be no long-run tendency to over- or underestimate the true value. Note that even though the sampling distribution is correctly centered, it spreads out quite a bit about the true value. Because of this, some estimates resulting from the use of this statistic will be far above or far below the true value, even though there is no systematic tendency to underestimate or overestimate the true value.

In contrast, the mean value of the statistic with the distribution shown in Figure 9.1(c) is equal to the true value of the population characteristic (implying no systematic error in estimation), and the statistic’s standard deviation is relatively small. Estimates based on this third statistic will almost always be quite close to the true value—certainly more often than estimates resulting from the statistic with the sampling distribution shown in Figure 9.1(b).

**DEFINITION**

A statistic whose mean value is equal to the value of the population characteristic being estimated is said to be an **unbiased statistic**. A statistic that is not unbiased is said to be **biased**.

As an example of a statistic that is biased, consider using the sample range as an estimate of the population range. Because the range of a population is defined as the difference between the largest value in the population and the smallest value, the range for a sample tends to underestimate the population range. This is because the largest value in a sample must be less than or equal to the largest value in the population and the smallest sample value must be greater than or equal to the smallest value in the population. The sample range equals the population range *only* if the sample includes both the largest and the smallest values in the population; in all other in-
stances, the sample range is smaller than the population range. Thus, $\mu_{\text{sample range}}$ is less than the population range, implying bias.

Let $x_1, x_2, \ldots, x_n$ represent the values in a random sample. One of the general results concerning the sampling distribution of $\bar{x}$, the sample mean, is that $\mu_{\bar{x}} = \mu$. This result says that the $\bar{x}$ values from all possible random samples of size $n$ center around $\mu$, the population mean. For example, if $\mu = 100$, the $\bar{x}$ distribution is centered at 100, whereas if $\mu = 5200$, then the $\bar{x}$ distribution is centered at 5200. Therefore, $\bar{x}$ is an unbiased statistic for estimating $\mu$. Similarly, because the sampling distribution of $\hat{p}$ is centered at $p$, it follows that $\hat{p}$ is an unbiased statistic for estimating a population proportion.

Using an unbiased statistic that also has a small standard deviation ensures that there will be no systematic tendency to under- or overestimate the value of the population characteristic and that estimates will almost always be relatively close to the value of the population characteristic.

Consider the problem of estimating a population mean, $\mu$. The obvious choice of statistic for obtaining a point estimate of $\mu$ is the sample mean, $\bar{x}$, an unbiased statistic for this purpose. However, when the population distribution is symmetric, $\bar{x}$ is not the only choice. Other unbiased statistics for estimating $\mu$ in this case include the sample median and any trimmed mean (with the same number of observations trimmed from each end of the ordered sample). Which statistic should be used? The following facts are helpful in making a choice.

1. If the population distribution is normal, then $\bar{x}$ has a smaller standard deviation than any other unbiased statistic for estimating $\mu$. However, in this case, a trimmed mean with a small trimming percentage (such as 10%) performs almost as well as $\bar{x}$.
2. When the population distribution is symmetric with heavy tails compared to the normal curve, a trimmed mean is a better statistic than $\bar{x}$ for estimating $\mu$.

When the population distribution is unquestionably normal, the choice is clear: Use $\bar{x}$ to estimate $\mu$. However, with a heavy-tailed distribution, a trimmed mean gives protection against one or two outliers in the sample that might otherwise have a large effect on the value of the estimate.

Now consider estimating another population characteristic, the population variance, $\sigma^2$. The sample variance

$$s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}$$

is a good choice for obtaining a point estimate of the population variance, $\sigma^2$. It can be shown that $s^2$ is an unbiased statistic for estimating $\sigma^2$; that is, whatever the value of $\sigma^2$, the sampling distribution of $s^2$ is centered at that value. It is precisely for this reason—to obtain an unbiased statistic—that the divisor $(n - 1)$ is used. An alternative statistic is the average squared deviation

$$\frac{\sum (x - \bar{x})^2}{n}$$
which one might think has a more natural divisor than \( s^2 \). However, the average squared deviation is biased, with its values tending to be smaller, on average, than the value of \( \sigma^2 \).

**EXAMPLE 9.3  Airborne Times for Flights from San Francisco to Washington, D.C.**

The Bureau of Transportation Statistics provides data on U.S. airline flights. The airborne times (in minutes) for nonstop flights from San Francisco to Washington Dulles airport for 10 randomly selected flights in June 2009 are:

270  256  267  285  274  275  266  258  271  281

For these data \( \sum x = 2703, \sum x^2 = 731,373, n = 10, \) and

\[
\sum (x - \bar{x})^2 = \sum x^2 - \frac{(\sum x)^2}{n}
\]

\[
= 731,373 - \frac{(2703)^2}{10}
\]

\[
= 752.1
\]

Let \( \sigma^2 \) denote the true variance in airborne time for June, 2009 nonstop flights from San Francisco to Washington Dulles airport. Using the sample variance \( s^2 \) to provide a point estimate of \( \sigma^2 \) yields

\[
s^2 = \frac{\sum(x - \bar{x})^2}{n - 1} = \frac{752.1}{9} = 83.57
\]

Using the average squared deviation (with divisor \( n = 10 \)), the resulting point estimate is

\[
\frac{\sum(x - \bar{x})^2}{n} = \frac{752.1}{10} = 75.21
\]

Because \( s^2 \) is an unbiased statistic for estimating \( \sigma^2 \), most statisticians would recommend using the point estimate 83.57.

An obvious choice of a statistic for estimating the population standard deviation \( \sigma \) is the sample standard deviation \( s \). For the data given in Example 9.3,

\[
s = \sqrt{83.57} = 9.14
\]

Unfortunately, the fact that \( s^2 \) is an unbiased statistic for estimating \( \sigma^2 \) does not imply that \( s \) is an unbiased statistic for estimating \( \sigma \). The sample standard deviation tends to underestimate slightly the true value of \( \sigma \). However, unbiasedness is not the only criterion by which a statistic can be judged, and there are other good reasons for using \( s \) to estimate \( \sigma \). In what follows, whenever we need to estimate \( \sigma \) based on a single random sample, we will use the statistic \( s \) to obtain a point estimate.
9.1 Three different statistics are being considered for estimating a population characteristic. The sampling distributions of the three statistics are shown in the following illustration:

Which statistic would you recommend? Explain your choice.

9.2 Why is an unbiased statistic generally preferred over a biased statistic for estimating a population characteristic? Does unbiasedness alone guarantee that the estimate will be close to the true value? Explain. Under what circumstances might you choose a biased statistic over an unbiased statistic if two statistics are available for estimating a population characteristic?

9.3 Consumption of fast food is a topic of interest to researchers in the field of nutrition. The article “Effects of Fast-Food Consumption on Energy Intake and Diet Quality Among Children” (Pediatrics [2004]: 112–118) reported that 1720 of those in a random sample of 6212 U.S. children indicated that on a typical day, they ate fast food. Estimate \( p \), the proportion of children in the United States who eat fast food on a typical day.

9.4 Data consistent with summary quantities in the article referenced in Exercise 9.3 on total calorie consumption on a particular day are given for a sample of children who did not eat fast food on that day and for a sample of children who did eat fast food on that day. Assume that it is reasonable to regard these samples as representative of the population of children in the United States.

No Fast Food
2331 1918 1009 1730 1469 2053 2143 1981 1852 1777 1765 1827 1648 1506 2669

9.5 Each person in a random sample of 20 students at a particular university was asked whether he or she is registered to vote. The responses (R = registered, N = not registered) are given here:

R R N R N N R R N R R N R R R R R N R R R N

Use these data to estimate \( p \), the proportion of all students at the university who are registered to vote.

9.6 Suppose that each of 935 smokers received a nicotine patch, which delivers nicotine to the bloodstream but at a much slower rate than cigarettes do. Dosage was decreased to 0 over a 12-week period. Suppose that 245 of the subjects were still not smoking 6 months after treatment. Assuming it is reasonable to regard this sample as representative of all smokers, estimate the percentage of all smokers who, when given this treatment, would refrain from smoking for at least 6 months.

9.7 Given below are the sodium contents (in mg) for seven brands of hot dogs rated as “very good” by Consumer Reports (www.consumerreports.org):

<table>
<thead>
<tr>
<th>Sodium Content (mg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>420</td>
</tr>
<tr>
<td>470</td>
</tr>
<tr>
<td>350</td>
</tr>
<tr>
<td>360</td>
</tr>
<tr>
<td>270</td>
</tr>
<tr>
<td>550</td>
</tr>
<tr>
<td>530</td>
</tr>
</tbody>
</table>

a. Use the given data to produce a point estimate of \( \mu \), the true mean sodium content for hot dogs.

b. Use the given data to produce a point estimate of \( \sigma^2 \), the variance of sodium content for hot dogs.

c. Use the given data to produce an estimate of \( \sigma \), the standard deviation of sodium content. Is the statistic you used to produce your estimate unbiased?
9.8  A random sample of \( n = 12 \) four-year-old red pine trees was selected, and the diameter (in inches) of each tree’s main stem was measured. The resulting observations are as follows:

\[
11.3 \quad 10.7 \quad 12.4 \quad 15.2 \quad 10.1 \quad 12.1 \quad 16.2 \quad 10.5 \\
11.4 \quad 11.0 \quad 10.7 \quad 12.0
\]

a. Compute a point estimate of \( \sigma \), the population standard deviation of main stem diameter. What statistic did you use to obtain your estimate?
b. Making no assumptions about the shape of the population distribution of diameters, give a point estimate for the population median diameter. What statistic did you use to obtain the estimate?
c. Suppose that the population distribution of diameter is symmetric but with heavier tails than the normal distribution. Give a point estimate of the population mean diameter based on a statistic that gives some protection against the presence of outliers in the sample. What statistic did you use?
d. Suppose that the diameter distribution is normal. Then the 90th percentile of the diameter distribution is \( \mu + 1.28\sigma \) (so 90% of all trees have diameters less than this value). Compute a point estimate for this percentile. (Hint: First compute an estimate of \( \mu \) in this case; then use it along with your estimate of \( \sigma \) from Part (a).)

9.9  A random sample of 10 houses heated with natural gas in a particular area, is selected, and the amount of gas (in therms) used during the month of January is determined for each house. The resulting observations are as follows:

\[
103 \quad 156 \quad 118 \quad 89 \quad 125 \quad 147 \quad 122 \quad 109 \quad 138 \quad 99
\]

a. Let \( \mu_J \) denote the average gas usage during January by all houses in this area. Compute a point estimate of \( \mu_J \).
b. Suppose that 10,000 houses in this area use natural gas for heating. Let \( \tau \) denote the total amount of gas used by all of these houses during January. Estimate \( \tau \) using the given data. What statistic did you use in computing your estimate?
c. Use the data in Part (a) to estimate \( p \), the proportion of all houses that used at least 100 therms.
d. Give a point estimate of the population median usage based on the sample of Part (a). Which statistic did you use?

9.2  **Large-Sample Confidence Interval for a Population Proportion**

In Section 9.1, we saw how to use a statistic to produce a point estimate of a population characteristic. The value of a point estimate depends on which sample, out of all the possible samples, happens to be selected. Different samples usually produce different estimates as a result of chance differences from one sample to another. Because of sampling variability, rarely is the point estimate from a sample exactly equal to the actual value of the population characteristic. We hope that the chosen statistic produces an estimate that is close, on average, to the true value.

Although a point estimate may represent our best single-number guess for the value of the population characteristic, it is not the only plausible value. As an alternative to a point estimate, we can use the sample data to report an interval of plausible values for the population characteristic. For example, we might be confident that for all text messages sent from cell phones, the proportion \( p \) of messages that are longer than 50 characters is in the interval from .53 to .57. The narrowness of this interval implies that we have rather precise information about the value of \( p \). If, with the same high degree of confidence, we could only state that \( p \) was between .32 and .74, it would be clear that we had relatively imprecise knowledge of the value of \( p \).
A *confidence interval* (CI) for a population characteristic is an interval of plausible values for the characteristic. It is constructed so that, with a chosen degree of confidence, the actual value of the population characteristic will be between the lower and upper endpoints of the interval.

Associated with each confidence interval is a *confidence level*. The confidence level provides information on how much “confidence” we can have in the *method* used to construct the interval estimate (not our confidence in any one particular interval). Usual choices for confidence levels are 90%, 95%, and 99%, although other confidence levels are also possible. If we were to construct a 95% confidence interval using the technique to be described shortly, we would be using a method that is “successful” 95% of the time. That is, if this method was used to generate an interval estimate over and over again with different samples, in the long run 95% of the resulting intervals would include the actual value of the characteristic being estimated. Similarly, a 99% confidence interval is one that is constructed using a method that is, in the long run, successful in capturing the actual value of the population characteristic 99% of the time.

One goal of many statistical studies is to estimate the proportion of individuals or objects in a population that possess a particular property of interest. For example, a university administrator might be interested in the proportion of students who prefer a new registration system to the previous registration method. In a different setting, a quality control engineer might be concerned about the proportion of defective parts manufactured using a particular process.

Recall that \( \hat{p} \) denotes the proportion of the population that possess the property of interest. Previously, we used the sample proportion

\[
\hat{p} = \frac{\text{number in the sample that possess the property of interest}}{n}
\]

to calculate a point estimate of \( p \). We can also use \( \hat{p} \) to construct a confidence interval for \( p \).

Although a small-sample confidence interval for \( p \) can be obtained, our focus is on the large-sample case. The construction of the large-sample interval is based on properties of the sampling distribution of the statistic \( \hat{p} \):

1. The sampling distribution of \( \hat{p} \) is centered at \( p \); that is, \( \mu_\hat{p} = p \). Therefore, \( \hat{p} \) is an unbiased statistic for estimating \( p \).
2. As long as the sample size is less than 10% of the population size, the standard deviation of \( \hat{p} \) is well approximated by \( \sigma_\hat{p} = \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \)
3. As long as \( n \) is large (\( np \geq 10 \) and \( n(1 - \hat{p}) \geq 10 \)) the sampling distribution of \( \hat{p} \) is well approximated by a normal curve.

The accompanying box summarizes these properties.
The development of a confidence interval for \( p \) is easier to follow if we select a particular confidence level. For a confidence level of 95%, Appendix Table 2, the table of standard normal (\( z \)) curve areas, can be used to determine a value \( z^* \) such that a central area of .95 falls between \(-z^*\) and \( z^*\). In this case, the remaining area of .05 is divided equally between the two tails, as shown in Figure 9.2. The total area to the left of the desired \( z^* \) is .975 (.95 central area + .025 area below \(-z^*)\). By locating .9750 in the body of Appendix Table 2, we find that the corresponding \( z \) critical value is \( z^* = 1.96 \).

Generalizing this result to normal distributions other than the standard normal distribution tells us that for any normal distribution, about 95% of the values are within 1.96 standard deviations of the mean. For large random samples, the sampling distribution of \( \hat{p} \) is approximately normal with mean \( \mu_{\hat{p}} = p \) and standard deviation

\[
\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}},
\]

and we get the following result.

When \( n \) is large, approximately 95% of all samples of size \( n \) will result in a value of \( \hat{p} \) that is within \( 1.96 \sigma_{\hat{p}} = 1.96 \sqrt{\frac{p(1-p)}{n}} \) of the value of the population proportion \( p \).

If \( \hat{p} \) is within \( 1.96 \sqrt{\frac{p(1-p)}{n}} \) of \( p \), this means the interval

\[
\hat{p} - 1.96 \sqrt{\frac{p(1-p)}{n}} \text{ to } \hat{p} + 1.96 \sqrt{\frac{p(1-p)}{n}}
\]

will capture \( p \) (and this will happen for 95% of all possible samples). However, if \( \hat{p} \) is farther away from \( p \) than \( 1.96 \sqrt{\frac{p(1-p)}{n}} \) (which will happen for about 5% of all possible samples), the interval will not include the true value of \( p \). This is shown in Figure 9.3.

Because \( \hat{p} \) is within \( 1.96 \sigma_{\hat{p}} \) of \( p \) 95% of the time, this implies that in repeated sampling, 95% of the time the interval

\[
\hat{p} - 1.96 \sqrt{\frac{p(1-p)}{n}} \text{ to } \hat{p} + 1.96 \sqrt{\frac{p(1-p)}{n}}
\]

will contain \( p \).
Since \( p \) is unknown, \( \sqrt{\frac{p(1-p)}{n}} \) must be estimated. As long as the sample size is large, the value of \( \sqrt{\frac{p(1-p)}{n}} \) can be used in place of \( \sqrt{\frac{p(1-p)}{n}} \).

When \( n \) is large, a 95% confidence interval for \( p \) is

\[
\left( \hat{p} - 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)
\]

An abbreviated formula for the interval is

\[
\hat{p} \pm 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}
\]

where \( \hat{p} + 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \) gives the upper endpoint of the interval and \( \hat{p} - 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \) gives the lower endpoint of the interval.

The interval can be used as long as

1. \( n\hat{p} \geq 10 \) and \( n(1-\hat{p}) \geq 10 \),
2. the sample size is less than 10% of the population size if sampling is without replacement, and
3. the sample can be regarded as a random sample from the population of interest.

**EXAMPLE 9.4 College Education Essential for Success?**

The article "How Well Are U.S. Colleges Run?" (USA Today, February 17, 2010) describes a survey of 1031 adult Americans. The survey was carried out by the National Center for Public Policy and the sample was selected in a way that makes it
reasonable to regard the sample as representative of adult Americans. Of those surveyed, 567 indicated that they believed a college education is essential for success. With \( p \) denoting the proportion of all adult Americans who believe that a college education is essential for success, a point estimate of \( p \) is

\[
\hat{p} = \frac{567}{1031} = .55
\]

Before computing a confidence interval to estimate \( p \), we should check to make sure that the three necessary conditions are met:

1. \( n\hat{p} = 1031(.55) = 567 \) and \( n(1 - \hat{p}) = 1031(1 - .55) = 1031(.45) = 364 \) are both greater than or equal to 10, so the sample size is large enough to proceed.
2. The sample size of \( n = 1031 \) is much smaller than 10% of the population size (the number of adult Americans).
3. The sample was selected in a way designed to produce a representative sample. So, it is reasonable to regard the sample as a random sample from the population.

Because all three conditions are met, it is appropriate to use the sample data to construct a 95% confidence interval for \( p \).

A 95% confidence interval for \( p \) is

\[
\hat{p} \pm 1.96 \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} = .55 \pm 1.96 \sqrt{\frac{.55(1 - .55)}{1031}}
\]

\[
= .55 \pm (1.96)(.015)
\]

\[
= .55 \pm .029
\]

\[
= (.521, .579)
\]

Based on this sample, we can be 95% confident that \( p \), the proportion of adult Americans who believe a college education is essential for success, is between .521 and .579. We used a method to construct this estimate that in the long run will successfully capture the actual value of \( p \) 95% of the time.

The 95% confidence interval for \( p \) calculated in Example 9.4 is (.521, .579). It is tempting to say that there is a “probability” of .95 that \( p \) is between .521 and .579. Do not yield to this temptation! The 95% refers to the percentage of all possible samples resulting in an interval that includes \( p \). In other words, if we take sample after sample from the population and use each one separately to compute a 95% confidence interval, in the long run roughly 95% of these intervals will capture \( p \). Figure 9.4 illustrates this concept for intervals generated from 100 different random samples. In this particular set of 100 intervals, 93 include \( p \), whereas 7 do not. Any specific interval, and our interval (.521, .579) in particular, either includes \( p \) or it does not (remember, the value of \( p \) is fixed but not known to us). We cannot make a chance (probability) statement concerning this particular interval. The confidence level 95% refers to the method used to construct the interval rather than to any particular interval, such as the one we obtained.

The formula given for a 95% confidence interval can easily be adapted for other confidence levels. The choice of a 95% confidence level led to the use of the \( z \) value 1.96 (chosen to capture a central area of .95 under the standard normal curve) in the formula. Any other confidence level can be obtained by using an appropriate \( z \) critical value in place of 1.96. For example, suppose that we wanted to achieve a confidence level of 99%. To obtain a central area of .99, the appropriate \( z \) critical value would
have a cumulative area (area to the left) of .995, as illustrated in Figure 9.5. From Appendix Table 2, we find that the corresponding $z$ critical value is $z = 2.58$. A 99% confidence interval for $p$ is then obtained by using 2.58 in place of 1.96 in the formula for the 95% confidence interval.

Why settle for 95% confidence when 99% confidence is possible? Because the higher confidence level comes with a price tag. The resulting interval is wider than the 95% interval. The width of the 95% interval is $2 \left( 1.96 \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \right)$, whereas the 99% interval has width $2 \left( 2.58 \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \right)$. The higher reliability of the 99% interval (where “reliability” is specified by the confidence level) entails a loss in precision (as indicated by the wider interval). In the opinion of many investigators, a 95% confidence interval produces a reasonable compromise between reliability and precision.

FIGURE 9.5
Finding the $z$ critical value for a 99% confidence level.

FIGURE 9.4
One hundred 95% confidence intervals for $\hat{p}$ computed from 100 different random samples (asterisks identify intervals that do not include $\hat{p}$).
The Large-Sample Confidence Interval for $p$

The general formula for a confidence interval for a population proportion $p$ when

1. $\hat{p}$ is the sample proportion from a simple random sample,
2. the sample size $n$ is large ($n\hat{p} \geq 10$ and $n(1 - \hat{p}) \geq 10$), and
3. if the sample is selected without replacement, the sample size is small relative to the population size ($n$ is at most 10% of the population size)*

is

$$\hat{p} \pm (z \text{ critical value})\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

The desired confidence level determines which $z$ critical value is used. The three most commonly used confidence levels, 90%, 95%, and 99%, use $z$ critical values 1.645, 1.96, and 2.58, respectively.

Note: This interval is not appropriate for small samples. It is possible to construct a confidence interval in the small-sample case, but this is beyond the scope of this textbook.

*In Chapter 7, we saw a different situation where a similar condition is introduced, but where the requirement was that at most 5% of the population is included in the sample. Be careful not to confuse these two rules.

**EXAMPLE 9.5 Dangerous Driving**

The article “Nine Out of Ten Drivers Admit in Survey to Having Done Something Dangerous” (Knight Ridder Newspapers, July 8, 2005) reported the results of a survey of 1100 drivers. Of those surveyed, 990 admitted to careless or aggressive driving during the previous 6 months. Assuming that it is reasonable to regard this sample of 1100 as representative of the population of drivers, we can use this information to construct an estimate of $p$, the proportion of all drivers who have engaged in careless or aggressive driving in the past 6 months.

For this sample

$$\hat{p} = \frac{990}{1100} = .900$$

Because the sample size is less than 10% of the population size and $n\hat{p} = 990$ and $n(1 - \hat{p}) = 110$ are both greater than or equal to 10, the conditions necessary for appropriate use of the formula for a large-sample confidence interval are met. A 90% confidence interval for $p$ is then

$$\hat{p} \pm (z \text{ critical value})\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} = .900 \pm 1.645\sqrt{\frac{(0.900)(0.100)}{1100}}$$

$$= .900 \pm (1.645)(.009)$$

$$= .900 \pm .015$$

$$= (.885, .915)$$

Based on these sample data, we can be 90% confident that the proportion of all drivers who have engaged in careless or aggressive driving in the past 6 months is between .885 and .915. We have used a method to construct this interval estimate that has a 10% error rate.

The confidence level for the $z$ confidence interval for a population proportion is only approximate. That is, when we report a 95% confidence interval for a popula-
Large-Sample Confidence Interval for a Population Proportion

The 95% confidence level implies that we have used a method that produces an interval that includes the actual value of the population proportion 95% of the time in repeated sampling. In fact, because the normal distribution is only an approximation to the sampling distribution of \( \hat{p} \), the true confidence level may differ somewhat from the reported value. If the conditions (1) \( n\hat{p} \geq 10 \) and \( n(1 - \hat{p}) \geq 10 \) and (2) \( n \) is at most 10% of the population size if sampling without replacement are met, the normal approximation is reasonable and the actual confidence level is usually not too different from the reported level; this is why it is important to check these conditions before computing and reporting a \( z \) confidence interval for a population proportion.

What should you do if these conditions are not met? If the sample size is too small to satisfy the \( n\hat{p} \) and \( n(1 - \hat{p}) \) greater than or equal to 10 condition, an alternative procedure can be used. Consult a statistician or a more advanced textbook in this case. If the condition that the sample size is less than 10% of the population size when sampling without replacement is not satisfied, the \( z \) confidence interval tends to be conservative (that is, it tends to be wider than is necessary to achieve the desired confidence level). In this case, a finite population correction factor can be used to obtain a more precise interval. Again, it would be wise to consult a statistician or a more advanced textbook.

An Alternative to the Large-Sample \( z \) Interval

Investigators have shown that in some instances, even when the sample size conditions of the large-sample \( z \) confidence interval for a population proportion are met, the actual confidence level associated with the method may be noticeably different from the reported confidence level. A modified interval that has an actual confidence level that is closer to the reported confidence level is based on a modified sample proportion, \( \hat{p}_{mod} \), the proportion of successes after adding two successes and two failures to the sample. Then \( \hat{p}_{mod} \) is

\[
\hat{p}_{mod} = \frac{\text{number of successes} + 2}{n + 4}
\]

\( \hat{p}_{mod} \) is used in place of \( \hat{p} \) in the usual confidence interval formula. Properties of this modified confidence interval are investigated in Activity 9.2 at the end of the chapter.

General Form of a Confidence Interval

Many confidence intervals have the same general form as the large-sample \( z \) interval for \( \hat{p} \) just considered. We started with a statistic \( \hat{p} \), from which a point estimate for \( p \) was obtained. The standard deviation of this statistic is \( \sqrt{p(1 - p)/n} \). This resulted in a confidence interval of the form

\[
\left( \text{point estimate using a specified statistic} \right) \pm (\text{critical value}) \left( \text{standard deviation of the statistic} \right)
\]

Because \( p \) was unknown, we estimated the standard deviation of the statistic by \( \sqrt{\hat{p}(1 - \hat{p})/n} \), which yielded the interval

\[
\left( \text{point estimate using a specified statistic} \right) \pm (\text{critical value}) \left( \text{estimated standard deviation of the statistic} \right)
\]

For a population characteristic other than \( \hat{p} \), a statistic for estimating the characteristic is selected. Then (drawing on statistical theory) a formula for the standard deviation of the statistic is given. In practice, it is almost always necessary to estimate
this standard deviation (using something analogous to $\sqrt{\hat{p}(1 - \hat{p})/n}$ rather than $\sqrt{p(1 - p)/n}$, for example), so that the interval

$$\left( \text{point estimate using a specified statistic} \right) \pm \left( \text{critical value} \right) \left( \text{estimated standard deviation of the statistic} \right)$$

is the prototype confidence interval. It is common practice to refer to both the standard deviation of a statistic and the estimated standard deviation of a statistic as the standard error. In this textbook, when we use the term standard error, we mean the estimated standard deviation of a statistic.

**Definition**

The **standard error** of a statistic is the estimated standard deviation of the statistic.

The 95% confidence interval for $p$ is based on the fact that, for approximately 95% of all random samples, $\hat{p}$ is within $1.96\sqrt{\frac{p(1-p)}{n}}$ of $p$. The quantity $1.96\sqrt{\frac{p(1-p)}{n}}$ is sometimes called the **bound on the error of estimation** associated with a 95% confidence level—we have 95% confidence that the point estimate $\hat{p}$ is no farther than this quantity from $p$.

**Definition**

If the sampling distribution of a statistic is (at least approximately) normal, the **bound on error of estimation**, $B$, associated with a 95% confidence interval is $(1.96) \cdot (\text{standard error of the statistic})$.

### Choosing the Sample Size

Before collecting any data, an investigator may wish to determine a sample size for which a particular value of the bound on the error is achieved. For example, with $p$ representing the actual proportion of students at a university who purchase textbooks over the Internet, the objective of an investigation may be to estimate $p$ to within .05 with 95% confidence. The value of $n$ necessary to achieve this is obtained by equating .05 to $1.96\sqrt{\frac{p(1-p)}{n}}$ and solving for $n$.

In general, suppose that we wish to estimate $p$ to within an amount $B$ (the specified bound on the error of estimation) with 95% confidence. Finding the necessary sample size requires solving the equation

$$B = 1.96\sqrt{\frac{p(1-p)}{n}}$$
Solving this equation for \( n \) results in
\[
n = p(1 - p) \left( \frac{1.96}{B} \right)^2
\]

Unfortunately, the use of this formula requires the value of \( p \), which is unknown. One possible way to proceed is to carry out a preliminary study and use the resulting data to get a rough estimate of \( p \). In other cases, prior knowledge may suggest a reasonable estimate of \( p \). If there is no reasonable basis for estimating \( p \) and a preliminary study is not feasible, a conservative solution follows from the observation that \( p(1 - p) \) is never larger than .25 (its value when \( p = .5 \)). Replacing \( p(1 - p) \) with .25, the maximum value, yields
\[
n = .25 \left( \frac{1.96}{B} \right)^2
\]

Using this formula to obtain \( n \) gives us a sample size for which we can be 95% confident that \( \hat{p} \) will be within \( B \) of \( p \), no matter what the value of \( p \).

---

**EXAMPLE 9.6 Sniffing Out Cancer**

Researchers have found biochemical markers of cancer in the exhaled breath of cancer patients, but chemical analysis of breath specimens has not yet proven effective in clinical diagnosis. The authors of the paper "Diagnostic Accuracy of Canine Scent Detection in Early- and Late-Stage Lung and Breast Cancers" (Integrative Cancer Therapies [2006]: 1-10) describe a study to investigate whether dogs can be trained to identify the presence or absence of cancer by sniffing breath specimens. Suppose we want to collect data that would allow us to estimate the long-run proportion of accurate identifications for a particular dog that has completed training. The dog has been trained to lie down when presented with a breath specimen from a cancer patient and to remain standing when presented with a specimen from a person who does not have cancer. How many different breath specimens should be used if we want to estimate the long-run proportion of correct identifications for this dog to within .05 with 95% confidence?

Using a conservative value of \( p = .5 \) in the formula for required sample size gives
\[
n = p(1 - p) \left( \frac{1.96}{B} \right)^2 = (.5)(.5) \left( \frac{1.96}{.05} \right)^2 = 384.16
\]

Thus, a sample of at least 385 breath specimens should be used. Note that in sample size calculations, we always round up.
EXERCISES 9.10 - 9.33

9.10 For each of the following choices, explain which would result in a wider large-sample confidence interval for \( p \):
   a. 90% confidence level or 95% confidence level
   b. \( n = 100 \) or \( n = 400 \)

9.11 The formula used to compute a large-sample confidence interval for \( p \) is
   \[ \hat{p} \pm (z \text{ critical value}) \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \]
   What is the appropriate \( z \) critical value for each of the following confidence levels?
   a. 95% b. 90% c. 99% d. 80% e. 85%

9.12 The use of the interval
   \[ \hat{p} \pm (z \text{ critical value}) \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \]
   requires a large sample. For each of the following combinations of \( n \) and \( \hat{p} \), indicate whether the sample size is large enough for use of this interval to be appropriate.
   a. \( n = 50 \) and \( \hat{p} = .30 \)
   b. \( n = 50 \) and \( \hat{p} = .05 \)
   c. \( n = 15 \) and \( \hat{p} = .45 \)
   d. \( n = 100 \) and \( \hat{p} = .01 \)
   e. \( n = 100 \) and \( \hat{p} = .70 \)
   f. \( n = 40 \) and \( \hat{p} = .25 \)
   g. \( n = 60 \) and \( \hat{p} = .25 \)
   h. \( n = 80 \) and \( \hat{p} = .10 \)

9.13 Discuss how each of the following factors affects the width of the confidence interval for \( p \):
   a. The confidence level
   b. The sample size
   c. The value of \( \hat{p} \)

9.14 The article “Career Expert Provides DOs and DON'Ts for Job Seekers on Social Networking” (CareerBuilder.com, August 19, 2009) included data from a survey of 2667 hiring managers and human resource professionals. The article noted that many employers are using social networks to screen job applicants and that this practice is becoming more common. Of the 2667 people who participated in the survey, 1200 indicated that they use social networking sites (such as Facebook, MySpace, and LinkedIn) to research job applicants. For the purposes of this exercise, assume that the sample is representative of hiring managers and human resource professionals. Construct and interpret a 95% confidence interval for the proportion of hiring managers and human resource professionals who use social networking sites to research job applicants.

9.15 The article “Nine Out of Ten Drivers Admit in Survey to Having Done Something Dangerous” (Knight Ridder Newspapers, July 8, 2005) reported the results of a survey of 1100 drivers. Of those surveyed, 990 admitted to careless or aggressive driving during the previous 6 months. Assuming that it is reasonable to regard this sample of 1100 as representative of the population of drivers, use this information to construct a 99% confidence interval to estimate \( p \), the proportion of all drivers who have engaged in careless or aggressive driving in the previous 6 months.

9.16 In a survey on supernatural experiences, 722 of 4013 adult Americans surveyed reported that they had seen or been with a ghost (“What Supernatural Experiences We’ve Had,” USA Today, February 8, 2010). Use the following data set to construct a confidence interval to estimate the proportion of all adult Americans who have seen or been with a ghost.

   a. What assumption must be made in order for it to be appropriate to use the formula of this section to construct a confidence interval to estimate the proportion of all adult Americans who have seen or been with a ghost?
   b. Construct and interpret a 90% confidence interval for the proportion of all adult Americans who have seen or been with a ghost.
   c. Would a 99% confidence interval be narrower or wider than the interval computed in Part (b)? Justify your answer.

9.17 If a hurricane was headed your way, would you evacuate? The headline of a press release issued January 21, 2009 by the survey research company International Communications Research (icrsurvey.com) states, “Thirty-one Percent of People on High-Risk Coast Will Refuse Evacuation Order, Survey of Hurricane Preparedness Finds.” This headline was based on a survey of 5046 adults who live within 20 miles of the coast in high hurricane risk counties of eight southern states. In selecting the sample, care was taken to ensure that the sample would be representative of the population of coastal residents in these states. Use this information to estimate the proportion of coastal residents who would evacuate using a 98% confidence interval. Write a few sentences inter-
interpreting the interval and the confidence level associated with the interval.

9.18 The study “Digital Footprints” (Pew Internet & American Life Project, www.pewinternet.org, 2007) reported that 47% of Internet users have searched for information about themselves online. The 47% figure was based on a random sample of Internet users. For purposes of this exercise, suppose that the sample size was \( n = 300 \) (the actual sample size was much larger). Construct and interpret a 90% confidence interval for the proportion of Internet users who have searched online for information about themselves.

9.19 The article “Kids Digital Day: Almost 8 Hours” (USA Today, January 20, 2010) summarized results from a national survey of 2002 Americans age 8 to 18. The sample was selected in a way that was expected to result in a sample representative of Americans in this age group.

a. Of those surveyed, 1321 reported owning a cell phone. Use this information to construct and interpret a 90% confidence interval estimate of the proportion of all Americans age 8 to 18 who own a cell phone.

b. Of those surveyed, 1522 reported owning an MP3 music player. Use this information to construct and interpret a 90% confidence interval estimate of the proportion of all Americans age 8 to 18 who own an MP3 music player.

c. Explain why the confidence interval from Part (b) is narrower than the confidence interval from Part (a) even though the confidence level and the sample size used to compute the two intervals was the same.

9.20 The article “Students Increasingly Turn to Credit Cards” (San Luis Obispo Tribune, July 21, 2006) reported that 37% of college freshmen and 48% of college seniors carry a credit card balance from month to month. Suppose that the reported percentages were based on random samples of 1000 college freshmen and 1000 college seniors.

a. Construct a 90% confidence interval for the proportion of college freshmen who carry a credit card balance from month to month.

b. Construct a 90% confidence interval for the proportion of college seniors who carry a credit card balance from month to month.

c. Explain why the two 90% confidence intervals from Parts (a) and (b) are not the same width.

9.21 The article “CSI Effect Has Juries Wanting More Evidence” (USA Today, August 5, 2004) examines how the popularity of crime-scene investigation television shows is influencing jurors’ expectations of what evidence should be produced at a trial. In a survey of 500 potential jurors, one study found that 350 were regular watchers of at least one crime-scene forensics television series.

a. Assuming that it is reasonable to regard this sample of 500 potential jurors as representative of potential jurors in the United States, use the given information to construct and interpret a 95% confidence interval for the proportion of all potential jurors who regularly watch at least one crime-scene investigation series.

b. Would a 99% confidence interval be wider or narrower than the 95% confidence interval from Part (a)?

9.22 In a survey of 1000 randomly selected adults in the United States, participants were asked what their most favorite and what their least favorite subject was when they were in school (Associated Press, August 17, 2005). In what might seem like a contradiction, math was chosen more often than any other subject in both categories! Math was chosen by 230 of the 1000 as the favorite subject, and it was also chosen by 370 of the 1000 as the least favorite subject.

a. Construct a 95% confidence interval for the proportion of U.S. adults for whom math was the favorite subject.

b. Construct a 95% confidence interval for the proportion of U.S. adults for whom math was the least favorite subject.

c. Explain why the confidence interval from Part (b) is narrower than the 95% confidence interval from Part (a).

9.23 The report “2005 Electronic Monitoring & Surveillance Survey: Many Companies Monitoring, Recording, Videotaping—and Firing—Employees” (American Management Association, 2005) summarized the results of a survey of 526 U.S. businesses. The report stated that 137 of the 526 businesses had fired workers for misuse of the Internet and 131 had fired workers for e-mail misuse. For purposes of this exercise, assume that it is reasonable to regard this sample as representative of businesses in the United States.

a. Construct and interpret a 95% confidence interval for the proportion of U.S. businesses that have fired workers for misuse of the Internet.

b. What are two reasons why a 90% confidence interval for the proportion of U.S. businesses that have...
fired workers for misuse of e-mail would be narrower than the 95% confidence interval computed in Part (a)?

9.24 In an AP-AOL sports poll (Associated Press, December 18, 2005), 394 of 1000 randomly selected U.S. adults indicated that they considered themselves to be baseball fans. Of the 394 baseball fans, 272 stated that they thought the designated hitter rule should either be expanded to both baseball leagues or eliminated.

a. Construct a 95% confidence interval for the proportion of U.S. adults who consider themselves to be baseball fans.

b. Construct a 95% confidence interval for the proportion of those who consider themselves to be baseball fans who think the designated hitter rule should be expanded to both leagues or eliminated.

c. Explain why the confidence intervals of Parts (a) and (b) are not the same width even though they both have a confidence level of 95%.

9.25 The article “Viewers Speak Out Against Reality TV” (Associated Press, September 12, 2005) included the following statement: “Few people believe there’s much reality in reality TV: a total of 82% said the shows are either ‘totally made up’ or ‘mostly distorted.’” This statement was based on a survey of 1002 randomly selected adults. Compute and interpret a bound on the error of estimation for the reported percentage.

9.26 One thousand randomly selected adult Americans participated in a survey conducted by the Associated Press (June 2006). When asked “Do you think it is sometimes justified to lie or do you think lying is never justified?” 52% responded that lying was never justified. When asked about lying to avoid hurting someone’s feelings, 650 responded that this was often or sometimes okay.

a. Construct a 90% confidence interval for the proportion of adult Americans who think lying is never justified.

b. Construct a 90% confidence interval for the proportion of adult American who think that it is often or sometimes okay to lie to avoid hurting someone’s feelings.

c. Comment on the apparent inconsistency in the responses given by the individuals in this sample.

9.27 USA Today (October 14, 2002) reported that 36% of adult drivers admit that they often or sometimes talk on a cell phone when driving. This estimate was based on data from a sample of 1004 adult drivers, and a bound on the error of estimation of 3.1% was reported. Assuming a 95% confidence level, do you agree with the reported bound on the error? Explain.

9.28 The Gallup Organization conducts an annual survey on crime. It was reported that 25% of all households experienced some sort of crime during the past year. This estimate was based on a sample of 1002 randomly selected households. The report states, “One can say with 95% confidence that the margin of sampling error is \( \pm 3 \) percentage points.” Explain how this statement can be justified.

9.29 The article “Hospitals Dispute Medtronic Data on Wires” (The Wall Street Journal, February 4, 2010) describes several studies of the failure rate of defibrillators used in the treatment of heart problems. In one study conducted by the Mayo Clinic, it was reported that failures were experienced within the first 2 years by 18 of 89 patients under 50 years old and 13 of 362 patients age 50 and older who received a particular type of defibrillator. Assume it is reasonable to regard these two samples as representative of patients in the two age groups who receive this type of defibrillator.

a. Construct and interpret a 95% confidence interval for the proportion of patients under 50 years old who experience a failure within the first 2 years after receiving this type of defibrillator.

b. Construct and interpret a 99% confidence interval for the proportion of patients age 50 and older who experience a failure within the first 2 years after receiving this type of defibrillator.

c. Suppose that the researchers wanted to estimate the proportion of patients under 50 years old who experience a failure within the first 2 years after receiving this type of defibrillator to within .03 with 95% confidence. How large a sample should be used? Use the results of the study as a preliminary estimate of the population proportion.

9.30 Based on a representative sample of 511 U.S. teenagers age 12 to 17, International Communications Research estimated that the proportion of teens who support keeping the legal drinking age at 21 is \( \hat{p} = 0.64 \) (64%). The press release titled “Majority of Teens (Still) Favor the Legal Drinking Age” (www.icrsurvey.com, January 21, 2009) also reported a margin of error of 0.04 (4%) for this estimate. Show how the reported value for the margin of error was computed.
9.31 A discussion of digital ethics appears in the article “Academic Cheating, Aided by Cell Phones or Web, Shown to be Common” (Los Angeles Times, June 17, 2009). One question posed in the article is: What proportion of college students have used cell phones to cheat on an exam? Suppose you have been asked to estimate this proportion for students enrolled at a large university. How many students should you include in your sample if you want to estimate this proportion to within .02 with 95% confidence?

9.32 In spite of the potential safety hazards, some people would like to have an Internet connection in their car. A preliminary survey of adult Americans has estimated this proportion to be somewhere around .30 (USA Today, May 1, 2009).

9.33 A consumer group is interested in estimating the proportion of packages of ground beef sold at a particular store that have an actual fat content exceeding the fat content stated on the label. How many packages of ground beef should be tested to estimate this proportion to within .05 with 95% confidence?

9.34 In this section, we consider how to use information from a random sample to construct a confidence interval estimate of a population mean, \( \mu \). We begin by considering the case in which (1) \( \sigma \), the population standard deviation, is known (not realistic, but we will see shortly how to handle the more realistic situation where \( \sigma \) is unknown) and (2) the sample size \( n \) is large enough for the Central Limit Theorem to apply. In this case, the following three properties about the sampling distribution of \( \bar{x} \) hold:

1. The sampling distribution of \( \bar{x} \) is centered at \( \mu \), so \( \bar{x} \) is an unbiased statistic for estimating \( \mu \) (\( \mu_{\bar{x}} = \mu \)).

2. The standard deviation of \( \bar{x} \) is \( \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \).

3. As long as \( n \) is large (generally \( n \geq 30 \)), the sampling distribution of \( \bar{x} \) is approximately normal, even when the population distribution itself is not normal.

The same reasoning that was used to develop the large-sample confidence interval for a population proportion \( p \) can be used to obtain a confidence interval estimate for \( \mu \).

### The One-Sample z Confidence Interval for \( \mu \)

The general formula for a confidence interval for a population mean \( \mu \) when

1. \( \bar{x} \) is the sample mean from a simple random sample,
2. the sample size \( n \) is large (generally \( n \geq 30 \)), and
3. \( \sigma \), the population standard deviation, is known

is

\[
\bar{x} \pm (z \text{ critical value}) \left( \frac{\sigma}{\sqrt{n}} \right)
\]
 Cosmic radiation levels rise with increasing altitude, prompting researchers to consider how pilots and flight crews might be affected by increased exposure to cosmic radiation. The paper “Estimated Cosmic Radiation Doses for Flight Personnel” (Space Medicine and Medical Engineering [2002]: 265–269) reported a mean annual cosmic radiation dose of 219 mrems for a sample of flight personnel of Xinjiang Airlines. Suppose that this mean was based on a random sample of 100 flight crew members.

Let $\mu$ denote the mean annual cosmic radiation exposure for all Xinjiang Airlines flight crew members. Although $\sigma$, the true population standard deviation, is not usually known, suppose for illustrative purposes that $\sigma = 35$ mrem is known. Because the sample size is large and $\sigma$ is known, a 95% confidence interval for $\mu$ is

$$
\bar{x} \pm (z \text{ critical value}) \left( \frac{\sigma}{\sqrt{n}} \right) = 219 \pm (1.96) \left( \frac{35}{\sqrt{100}} \right)
$$

$$
= 219 \pm 6.86
$$

$$
= (212.14, 225.86)
$$

Based on this sample, plausible values of $\mu$, the actual mean annual cosmic radiation exposure for Xinjiang Airlines flight crew members, are between 212.14 and 225.86 mrem. A 95% confidence level is associated with the method used to produce this interval estimate.

The confidence interval just introduced is appropriate when $\sigma$ is known and $n$ is large, and it can be used regardless of the shape of the population distribution. This is because this confidence interval is based on the Central Limit Theorem, which says that when $n$ is sufficiently large, the sampling distribution of $\bar{x}$ is approximately normal for any population distribution. When $n$ is small, the Central Limit Theorem cannot be used to justify the normality of the $\bar{x}$ sampling distribution, so the $z$ confidence interval can not be used. One way to proceed in the small-sample case is to make a specific assumption about the shape of the population distribution and then to use a method that is valid under this assumption.

One instance where this is easy to do is when it is reasonable to believe that the population distribution is normal in shape. Recall that for a normal population distribution the sampling distribution of $\bar{x}$ is normal even for small sample sizes. So, if $n$ is small but the population distribution is normal, the same confidence interval formula just introduced can still be used.

If it is reasonable to believe that the distribution of values in the population is normal, a confidence interval for $\mu$ (when $\sigma$ is known) is

$$
\bar{x} \pm (z \text{ critical value}) \left( \frac{\sigma}{\sqrt{n}} \right)
$$

This interval is appropriate even when $n$ is small, as long as it is reasonable to think that the population distribution is normal in shape.

There are several ways that sample data can be used to assess the plausibility of normality. Two common ways are to look at a normal probability plot of the sample
data (looking for a plot that is reasonably straight) or to construct a boxplot of the
data (looking for approximate symmetry and no outliers).

**Confidence Interval for \( \mu \) When \( \sigma \) Is Unknown**

The confidence interval just developed has an obvious drawback: To compute the
interval endpoints, \( \sigma \) must be known. Unfortunately, this is rarely the case in prac-
tice. We now turn our attention to the situation when \( \sigma \) is unknown. The develop-
ment of the confidence interval in this instance depends on the assumption that the
population distribution is normal. This assumption is not critical if the sample size is
large, but it is important when the sample size is small.

To understand the derivation of this confidence interval, it is instructive to begin
by taking another look at the previous 95% confidence interval. We know that
\( \mu_x = \mu \) and \( \sigma_x = \frac{\sigma}{\sqrt{n}} \). Also, when the population distribution is normal, the \( \bar{x} \)
distribution is normal. These facts imply that the standardized variable

\[
    z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}
\]

has approximately a standard normal distribution. Because the interval from \(-1.96\)
to \(1.96\) captures an area of .95 under the \( z \) curve, approximately 95% of all samples
result in an \( \bar{x} \) value that satisfies

\[
-1.96 < \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} < 1.96
\]

Manipulating these inequalities to isolate \( \mu \) in the middle results in the equivalent
inequalities:

\[
\bar{x} - 1.96\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \bar{x} + 1.96\left(\frac{\sigma}{\sqrt{n}}\right)
\]

The term \( \bar{x} - 1.96\left(\frac{\sigma}{\sqrt{n}}\right) \) is the lower endpoint of the 95% large-sample confidence
interval for \( \mu \), and \( \bar{x} + 1.96\left(\frac{\sigma}{\sqrt{n}}\right) \) is the upper endpoint.

If \( \sigma \) is unknown, we must use the sample data to estimate \( \sigma \). If we use the sample
standard deviation as our estimate, the result is a different standardized variable de-
noted by \( t \):

\[
t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}
\]

The value of \( s \) may not be all that close to \( \sigma \), especially when \( n \) is small. As a
consequence, the use of \( s \) in place of \( \sigma \) introduces extra variability. The value of \( z \)
varies from sample to sample, because different samples generally result in different
\( \bar{x} \) values. There is even more variability in \( t \), because different samples may result in
different values of both \( \bar{x} \) and \( s \). Because of this, the distribution of \( t \) is more spread
out than the standard normal (\( z \)) distribution.

To develop an appropriate confidence interval, we must investigate the probabil-
ity distribution of the standardized variable \( t \) for a sample from a normal population.
This requires that we first learn about probability distributions called \( t \) distributions.
**t Distributions**

Just as there are many different normal distributions, there are also many different *t* distributions. While normal distributions are distinguished from one another by their mean $\mu$ and standard deviation $\sigma$, *t* distributions are distinguished by a positive whole number called the number of *degrees of freedom* (df). There is a *t* distribution with 1 df, another with 2 df, and so on.

### Important Properties of *t* Distributions

1. The *t* distribution corresponding to any particular number of degrees of freedom is bell shaped and centered at zero (just like the standard normal ($z$) distribution).
2. Each *t* distribution is more spread out than the standard normal ($z$) distribution.
3. As the number of degrees of freedom increases, the spread of the corresponding *t* distribution decreases.
4. As the number of degrees of freedom increases, the corresponding sequence of *t* distributions approaches the standard normal ($z$) distribution.

The properties discussed in the preceding box are illustrated in Figure 9.6, which shows two *t* curves along with the $z$ curve.

![Figure 9.6](image)

Comparison of the $z$ curve and *t* curves for 12 df and 4 df.

Appendix Table 3 gives selected critical values for various *t* distributions. The central areas for which values are tabulated are .80, .90, .95, .98, .99, .998, and .999. To find a particular critical value, go down the left margin of the table to the row labeled with the desired number of degrees of freedom. Then move over in that row to the column headed by the desired central area. For example, the value in the 12-df row under the column corresponding to central area .95 is 2.18, so 95% of the area under the *t* curve with 12 df lies between $-2.18$ and $2.18$. Moving over two columns, we find the critical value for central area .99 (still with 12 df) to be 3.06 (see Figure 9.7). Moving down the .99 column to the 20-df row, we see the critical value is 2.85, so the area between $-2.85$ and $2.85$ under the *t* curve with 20 df is .99.

![Figure 9.7](image)

$t$ critical values illustrated.
Notice that the critical values increase from left to right in each row of Appendix Table 3. This makes sense because as we move to the right, we capture larger central areas. In each column, the critical values decrease as we move downward, reflecting decreasing spread for \( t \) distributions with larger degrees of freedom.

The larger the number of degrees of freedom, the more closely the \( t \) curve resembles the \( z \) curve. To emphasize this, we have included the \( z \) critical values as the last row of the \( t \) table. Furthermore, once the number of degrees of freedom exceeds 30, the critical values change little as the number of degrees of freedom increases. For this reason, Appendix Table 3 jumps from 30 df to 40 df, then to 60 df, then to 120 df, and finally to the row of \( z \) critical values. If we need a critical value for a number of degrees of freedom between those tabulated, we just use the critical value for the closest df. For df > 120, we use the \( z \) critical values. Many graphing calculators calculate \( t \) critical values for any number of degrees of freedom, so if you are using such a calculator, it is not necessary to approximate the \( t \) critical values as described.

**One-Sample \( t \) Confidence Interval**

The fact that the sampling distribution of \( \frac{\bar{x} - \mu}{(s/\sqrt{n})} \) is approximately the \( z \) (standard normal) distribution when \( n \) is large led to the \( z \) confidence interval when \( \sigma \) is known. In the same way, the following proposition provides the key to obtaining a confidence interval when the population distribution is normal but \( \sigma \) is unknown.

\[
\text{If } \bar{x} \text{ and } s \text{ are the mean and standard deviation of a random sample from a normal population distribution, then the probability distribution of the standardized variable}
\]

\[
t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}
\]

\[
is the \ t \ distribution with df = n - 1.
\]

To see how this result leads to the desired confidence interval, consider the case \( n = 25 \). We use the \( t \) distribution with df = \( n - 1 = 24 \). From Appendix Table 3, the interval between -2.06 and 2.06 captures a central area of .95 under the \( t \) curve with 24 df. This means that 95\% of all samples (with \( n = 25 \)) from a normal population result in values of \( \bar{x} \) and \( s \) for which

\[
-2.06 < \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} < 2.06
\]

Algebraically manipulating these inequalities to isolate \( \mu \) yields

\[
\bar{x} - 2.06\left(\frac{s}{\sqrt{25}}\right) < \mu < \bar{x} + 2.06\left(\frac{s}{\sqrt{25}}\right)
\]

The 95\% confidence interval for \( \mu \) in this situation extends from the lower endpoint \( \bar{x} - 2.06\left(\frac{s}{\sqrt{25}}\right) \) to the upper endpoint \( \bar{x} + 2.06\left(\frac{s}{\sqrt{25}}\right) \). This interval can also be written

\[
\bar{x} \pm 2.06\left(\frac{s}{\sqrt{25}}\right)
\]
The differences between this interval and the interval when \( \sigma \) is known are the use of the \( t \) critical value 2.06 rather than the \( z \) critical value 1.96 and the use of the sample standard deviation as an estimate of \( \sigma \). The extra uncertainty that results from estimating \( \sigma \) causes the \( t \) interval to be wider than the \( z \) interval.

If the sample size is something other than 25 or if the desired confidence level is something other than 95\%, a different \( t \) critical value (obtained from Appendix Table 3) is used in place of 2.06.

### The One-Sample \( t \) Confidence Interval for \( \mu \)

The general formula for a confidence interval for a population mean \( \mu \) based on a sample of size \( n \) when

1. \( \bar{x} \) is the sample mean from a simple random sample,
2. the population distribution is normal, or the sample size \( n \) is large (generally \( n \geq 30 \)), and
3. \( \sigma \), the population standard deviation, is unknown

is

\[
\bar{x} \pm (t \text{ critical value}) \left( \frac{s}{\sqrt{n}} \right)
\]

where the \( t \) critical value is based on \( \text{df} = n - 1 \). Appendix Table 3 gives critical values appropriate for each of the confidence levels 90\%, 95\%, and 99\%, as well as several other less frequently used confidence levels.

If \( n \) is large (generally \( n \geq 30 \)), the normality of the population distribution is not critical. However, this confidence interval is appropriate for small \( n \) only when the population distribution is (at least approximately) normal. If this is not the case, as might be suggested by a normal probability plot or boxplot, another estimation method should be used.

### EXAMPLE 9.8 Drive-Through Medicine

During a flu outbreak, many people visit emergency rooms, where they often must wait in crowded waiting rooms where other patients may be exposed. The paper “Drive-Through Medicine: A Novel Proposal for Rapid Evaluation of Patients during an Influenza Pandemic” (Annals of Emergency Medicine [2010]: 268–273) describes an interesting study of the feasibility of a drive-through model where flu patients are evaluated while they remain in their cars. One of the interesting observations from this study was that not only were patients kept relatively isolated and away from other patients, but the time to process a patient was shorter because delays related to turning over examination rooms were eliminated.

In the experiment, 38 volunteers were each given a scenario from a randomly selected set of flu cases seen in the emergency room. The scenarios provided the volunteer with a medical history and a description of symptoms that would allow the volunteer to respond to questions from the examining physician. These volunteer patients were then processed using a drive-through procedure that was implemented in the parking structure of Stanford University Hospital and the time to process each case from admission to discharge was recorded.
Data read from a graph that appears in the paper was used to compute the following summary statistics for admission-to-discharge processing times (in minutes):

\[ n = 38 \quad \bar{x} = 26 \quad s = 1.57 \]

A boxplot of the 38 processing times did show a couple of outliers on the high end, corresponding to unusually long processing times, suggesting that it is probably not reasonable to think of the population distribution of drive-through processing times as being approximately normal. However, because the sample size is greater than 30 and the distribution of sample processing times was not extremely skewed, it is appropriate to consider using the \( t \) confidence interval to estimate the mean admission-to-discharge processing time for flu patients using the drive-through procedure. So, because the 38 flu scenarios were thought to be representative of the population of flu patients seen in emergency rooms and the sample size is large, we can use the formula for the \( t \) confidence interval to compute a 95% confidence interval.

Because \( n = 38 \), \( df = 37 \), and the appropriate \( t \) critical value is 2.02 (from the 40-df row of Appendix Table 3). The confidence interval is then

\[ \bar{x} \pm (t \text{ critical value}) \left( \frac{s}{\sqrt{n}} \right) = 26 \pm (2.02) \left( \frac{1.57}{\sqrt{38}} \right) \]

\[ = 26 \pm 0.514 \]

\[ = (25.486, 26.514) \]

Based on the sample data, we believe that the actual mean admission-to-discharge processing time for flu patients processed using the drive-through procedure is between 25.486 minutes and 26.514 minutes. We used a method that has a 5% error rate to construct this interval. The authors of the paper indicated that the average processing time for flu patients seen in the emergency room was about 90 minutes, so it appears that the drive-through procedure has promise both in terms of keeping flu patients isolated and also in reducing processing time.

---

**EXAMPLE 9.9 Waiting for Surgery**

The Cardiac Care Network in Ontario, Canada, collected information on the time between the date a patient was recommended for heart surgery and the surgery date for cardiac patients in Ontario ("Wait Times Data Guide," Ministry of Health and Long-Term Care, Ontario, Canada, 2006). The reported mean wait times (in days) for samples of patients for two cardiac procedures are given in the accompanying table. (The standard deviations in the table were estimated from information on wait-time variability included in the report.)

<table>
<thead>
<tr>
<th>Surgical Procedure</th>
<th>Sample Size</th>
<th>Mean Wait Time</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bypass</td>
<td>539</td>
<td>19</td>
<td>10</td>
</tr>
<tr>
<td>Angiography</td>
<td>847</td>
<td>18</td>
<td>9</td>
</tr>
</tbody>
</table>

If we had access to the raw data (the 539 + 847 = 1386 individual wait-time observations), we might begin by looking at boxplots. Data consistent with the given summary quantities were used to generate the boxplots of Figure 9.8. The boxplots for the two surgical procedures are similar. There are outliers in both data sets, which
might cause us to question the normality of the two wait-time distributions, but because the sample sizes are large, it is still appropriate to use the \( t \) confidence interval.

As a next step, we can use the confidence interval of this section to estimate the actual mean wait time for each of the two procedures. Let’s first focus on the sample of bypass patients. For this group,

sample size \( n = 539 \)
sample mean wait time \( \bar{x} = 19 \)
sample standard deviation \( s = 10 \)

The report referenced here indicated that it is reasonable to regard these data as representative of the Ontario population. So, with \( \mu \) denoting the mean wait time for bypass surgery in Ontario, we can estimate \( \mu \) using a 90% confidence interval.

From Appendix Table 3, we use \( t \) critical value \( = 1.645 \) (from the \( z \) critical value row because df \( = n - 1 = 538 \gg 120 \)), the largest number of degrees of freedom in the table). The 90% confidence interval for \( \mu \) is

\[
\bar{x} \pm (t \text{ critical value}) \left( \frac{s}{\sqrt{n}} \right) = 19 \pm (1.645) \left( \frac{10}{\sqrt{539}} \right) \\
= 19 \pm 0.709 \\
= (18.291, 19.709)
\]

Based on this sample, we are 90% confident that \( \mu \) is between 18.291 days and 19.709 days. This interval is fairly narrow, indicating that our information about the value of \( \mu \) is relatively precise.

A graphing calculator or any of the commercially available statistical computing packages can produce \( t \) confidence intervals. Confidence interval output from Minitab for the angiography data is shown here.

<table>
<thead>
<tr>
<th>One-Sample T</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>847</td>
</tr>
</tbody>
</table>

The 90% confidence interval for mean wait time for angiography extends from 17.4908 days to 18.5092 days. This interval is narrower than the 90% interval for bypass surgery wait time for two reasons: the sample size is larger (847 rather than 539) and the sample standard deviation is smaller (9 rather than 10).
The article “Chimps Aren’t Charitable” (Newsday, November 2, 2005) summarized the results of a research study published in the journal *Nature*. In this study, chimpanzees learned to use an apparatus that dispensed food when either of two ropes was pulled. When one of the ropes was pulled, only the chimp controlling the apparatus received food. When the other rope was pulled, food was dispensed both to the chimp controlling the apparatus and also to a chimp in the adjoining cage. The accompanying data (approximated from a graph in the paper) represent the number of times out of 36 trials that each of seven chimps chose the option that would provide food to both chimps (the “charitable” response).

23 22 21 24 19 20 20

Figure 9.9 is a normal probability plot of these data. The plot is reasonably straight, so it seems plausible that the population distribution of number of charitable responses is approximately normal.

For purposes of this example, let’s suppose it is reasonable to regard this sample of seven chimps as representative of the population of all chimpanzees. Calculation of a confidence interval for the mean number of charitable responses for the population of all chimps requires $\bar{x}$ and $s$. From the given data, we compute

$$\bar{x} = 21.29 \quad s = 1.80$$

The $t$ critical value for a 99% confidence interval based on 6 df is 3.71. The interval is

$$\bar{x} \pm (t \text{ critical value}) \left( \frac{s}{\sqrt{n}} \right) = 21.29 \pm (3.71) \left( \frac{1.80}{\sqrt{7}} \right)$$

$$= 21.29 \pm 2.52$$

$$= (18.77, 23.81)$$

A statistical software package could also have been used to compute the 99% confidence interval. The following is output from SPSS. The slight discrepancy between
the hand-calculated interval and the one reported by SPSS occurs because SPSS uses more decimal accuracy in $\bar{x}$, $s$, and $t$ critical values.

### One-Sample Statistics

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CharitableResponses</strong></td>
<td>7</td>
<td>21.2857</td>
<td>1.79947</td>
<td>.68014</td>
</tr>
</tbody>
</table>

### 99% Confidence Interval

<table>
<thead>
<tr>
<th></th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CharitableResponses</strong></td>
<td>18.7642</td>
<td>23.8073</td>
</tr>
</tbody>
</table>

With 99% confidence, we estimate the population mean number of charitable responses (out of 36 trials) to be between 18.77 and 23.81. Remember that the 99% confidence level implies that if the same formula is used to calculate intervals for sample after sample randomly selected from the population of chimps, in the long run 99% of these intervals will capture $\mu$ between the lower and upper confidence limits.

Notice that based on this interval, we would conclude that, on average, chimps choose the charitable option more than half the time (more than 18 out of 36 trials). The *Newsday* headline “Chimps Aren’t Charitable” was based on additional data from the study indicating that chimps’ charitable behavior was no different when there was another chimp in the adjacent cage than when the adjacent cage was empty. We will revisit this study in Chapter 11 to investigate this further.

## Choosing the Sample Size

When estimating $\mu$ using a large sample or using a small sample from a normal population, the bound $B$ on the error of estimation associated with a 95% confidence interval is

$$B = 1.96\left(\frac{\sigma}{\sqrt{n}}\right)$$

Before collecting any data, an investigator may wish to determine a sample size for which a particular value of the bound is achieved. For example, with $\mu$ representing the average fuel efficiency (in miles per gallon, mpg) for all cars of a certain type, the objective of an investigation may be to estimate $\mu$ to within 1 mpg with 95% confidence. The value of $n$ necessary to achieve this is obtained by setting $B = 1$ and then solving $1 = 1.96\left(\frac{\sigma}{\sqrt{n}}\right)$ for $n$.

In general, suppose that we wish to estimate $\mu$ to within an amount $B$ (the specified bound on the error of estimation) with 95% confidence. Finding the necessary sample size requires solving the equation $B = 1.96\left(\frac{\sigma}{\sqrt{n}}\right)$ for $n$. The result is

$$n = \left(\frac{1.96\sigma}{B}\right)^2$$

Notice that the greater the variability in the population (larger $\sigma$), the greater the required sample size will be. And, of course, the smaller the desired bound on error is, the larger the required sample size will be.

Use of the sample-size formula requires that $\sigma$ be known, but this is rarely the case in practice. One possible strategy for estimating $\sigma$ is to carry out a preliminary study and use the resulting sample standard deviation (or a somewhat larger value, to
be conservative) to determine $n$ for the main part of the study. Another possibility is simply to make an educated guess about the value of $\sigma$ and to use that value to calculate $n$. For a population distribution that is not too skewed, dividing the anticipated range (the difference between the largest and the smallest values) by 4 often gives a rough idea of the value of the standard deviation.

The sample size required to estimate a population mean $\mu$ to within an amount $B$ with 95% confidence is

$$n = \left( \frac{1.96\sigma}{B} \right)^2$$

If $\sigma$ is unknown, it may be estimated based on previous information or, for a population that is not too skewed, by using (range)/4.

If the desired confidence level is something other than 95%, 1.96 is replaced by the appropriate $z$ critical value (for example, 2.58 for 99% confidence).

**EXAMPLE 9.11 Cost of Textbooks**

The financial aid office wishes to estimate the mean cost of textbooks per quarter for students at a particular university. For the estimate to be useful, it should be within $20 of the true population mean. How large a sample should be used to be 95% confident of achieving this level of accuracy?

To determine the required sample size, we must have a value for $\sigma$. The financial aid office is pretty sure that the amount spent on books varies widely, with most values between $150 and $550. A reasonable estimate of $\sigma$ is then

$$\frac{\text{range}}{4} = \frac{550 - 150}{4} = \frac{400}{4} = 100$$

The required sample size is

$$n = \left( \frac{1.96\sigma}{B} \right)^2 = \left( \frac{1.96(100)}{20} \right)^2 = (9.8)^2 = 96.04$$

Rounding up, a sample size of 97 or larger is recommended.

**EXERCISES 9.34 - 9.52**

9.34 Given a variable that has a $t$ distribution with the specified degrees of freedom, what percentage of the time will its value fall in the indicated region?

- a. 10 df, between $-1.81$ and $1.81$
- b. 10 df, between $-2.23$ and $2.23$
- c. 24 df, between $-2.06$ and $2.06$
- d. 24 df, between $-2.80$ and $2.80$
- e. 24 df, outside the interval from $-2.80$ to $2.80$

9.35 The formula used to compute a confidence interval for the mean of a normal population when $n$ is small is

$$\bar{x} \pm (t \text{ critical value}) \frac{s}{\sqrt{n}}$$

What is the appropriate $t$ critical value for each of the following confidence levels and sample sizes?
a. 95% confidence, n = 17
b. 90% confidence, n = 12
c. 99% confidence, n = 24
d. 90% confidence, n = 25
e. 90% confidence, n = 13
f. 95% confidence, n = 10

9.36 The two intervals (114.4, 115.6) and (114.1, 115.9) are confidence intervals (computed using the same sample data) for \( \mu = \) true average resonance frequency (in hertz) for all tennis rackets of a certain type.

- a. What is the value of the sample mean resonance frequency?
- b. The confidence level for one of these intervals is 90% and for the other it is 99%. Which is which, and how can you tell?

9.37 Samples of two different models of cars were selected, and the actual speed for each car was determined when the speedometer registered 50 mph. The resulting 95% confidence intervals for mean actual speed were (51.3, 52.7) and (49.4, 50.6). Assuming that the two sample standard deviations are equal, which confidence interval is based on the larger sample size? Explain your reasoning.

9.38 The authors of the paper “Deception and Design: The Impact of Communication Technology on Lying Behavior” (Proceedings of Computer Human Interaction [2004]) asked 30 students in an upper division communications course at a large university to keep a journal for 7 days, recording each social interaction and whether or not they told any lies during that interaction. A lie was defined as “any time you intentionally try to mislead someone.” The paper reported that the mean number of lies per day for the 30 students was 1.58 and the standard deviation of number of lies per day was 1.02.

- a. What assumption must be made in order for the \( t \) confidence interval of this section to be an appropriate method for estimating \( \mu \), the mean number of lies per day for all students at this university?
- b. Would you recommend using the \( t \) confidence interval to construct an estimate of \( \mu \) as defined in Part (a)? Explain why or why not.

9.39 In a study of academic procrastination, the authors of the paper “Correlates and Consequences of Behavioral Procrastination” (Procrastination, Current Issues and New Directions [2000]) reported that for a sample of 411 undergraduate students at a midsize public university preparing for a final exam in an introductory psychology course, the mean time spent studying for the exam was 7.74 hours and the standard deviation of study times was 3.40 hours. For purposes of this exercise, assume that it is reasonable to regard this sample as representative of students taking introductory psychology at this university.

- a. Construct a 95% confidence interval to estimate \( \mu \), the mean time spent studying for the final exam for students taking introductory psychology at this university.
- b. The paper also gave the following sample statistics for the percentage of study time that occurred in the 24 hours prior to the exam:

\[
\begin{align*}
&\text{Sample statistics} \\
&n = 411, \quad \bar{x} = 43.18, \quad s = 21.46
\end{align*}
\]

Construct and interpret a 90% confidence interval for the mean percentage of study time that occurs in the 24 hours prior to the exam.

9.40 How much money do people spend on graduation gifts? In 2007, the National Retail Federation (www.nrf.com) surveyed 2815 consumers who reported that they bought one or more graduation gifts that year. The sample was selected in a way designed to produce a sample representative of adult Americans who purchased graduation gifts in 2007. For this sample, the mean amount spent per gift was $55.05. Suppose that the sample standard deviation was $20. Construct and interpret a 98% confidence interval for the mean amount of money spent per graduation gift in 2007.

9.41 In June 2009, Harris Interactive conducted its Great Schools Survey. In this survey, the sample consisted of 1086 adults who were parents of school-aged children. The sample was selected in a way that makes it reasonable to regard it as representative of the population of parents of school-aged children. One question on the survey asked respondents how much time (in hours) per month they spent volunteering at their children’s school during the previous school year. The following summary statistics for time volunteered per month were given:

\[
\begin{align*}
&n = 1086, \quad \bar{x} = 5.6, \quad \text{median} = 1
\end{align*}
\]

- a. What does the fact that the mean is so much larger than the median tell you about the distribution of time spent volunteering per month?
- b. Based on your answer to Part (a), explain why it is not reasonable to assume that the population distribution of time spent volunteering is approximately normal.
c. Explain why it is appropriate to use the *t* confidence interval to estimate the mean time spent volunteering for the population of parents of school-aged children even though the population distribution is not approximately normal.

d. Suppose that the sample standard deviation was $s = 5.2$. Compute and interpret a 98% confidence interval for $\mu$, the mean time spent volunteering for the population of parents of school-aged children.

9.42 The authors of the paper “Driven to Distraction” (*Psychological Science* [2001]: 462–466) describe an experiment to evaluate the effect of using a cell phone on reaction time. Subjects were asked to perform a simulated driving task while talking on a cell phone. While performing this task, occasional red and green lights flashed on the computer screen. If a green light flashed, subjects were to continue driving, but if a red light flashed, subjects were to brake as quickly as possible and the reaction time (in msec) was recorded. The following summary statistics are based on a graph that appeared in the paper:

\[ n = 48 \quad \bar{x} = 530 \quad s = 70 \]

a. Construct and interpret a 95% confidence interval for $\mu$, the mean time to react to a red light while talking on a cell phone. What assumption must be made in order to generalize this confidence interval to the population of all drivers?

b. Suppose that the researchers wanted to estimate the mean reaction time to within 5 msec with 95% confidence. Using the sample standard deviation from the study described as a preliminary estimate of the standard deviation of reaction times, compute the required sample size.

c. Consider the following statement: If the process of selecting a sample of size 50 and then computing the corresponding 95% confidence interval is repeated 100 times, 95 of the resulting intervals will include $\mu$. Is this statement correct? Why or why not?

9.44 Acrylic bone cement is sometimes used in hip and knee replacements to fix an artificial joint in place. The force required to break an acrylic bone cement bond was measured for six specimens under specified conditions, and the resulting mean and standard deviation were 306.09 Newtons and 41.97 Newtons, respectively. Assuming that it is reasonable to believe that breaking force under these conditions has a distribution that is approximately normal, estimate the mean breaking force for acrylic bone cement under the specified conditions using a 95% confidence interval.

9.45 The article “The Association Between Television Viewing and Irregular Sleep Schedules Among Children Less Than 3 Years of Age” (*Pediatrics* [2005]: 851–856) reported the accompanying 95% confidence intervals for average TV viewing time (in hours per day) for three different age groups.

<table>
<thead>
<tr>
<th>Age Group</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 12 months</td>
<td>(0.8, 1.0)</td>
</tr>
<tr>
<td>12 to 23 months</td>
<td>(1.4, 1.8)</td>
</tr>
<tr>
<td>24 to 35 months</td>
<td>(2.1, 2.5)</td>
</tr>
</tbody>
</table>

a. Suppose that the sample sizes for each of the three age group samples were equal. Based on the given confidence intervals, which of the age group samples had the greatest variability in TV viewing time? Explain your choice.

b. Now suppose that the sample standard deviations for the three age group samples were equal, but that the three sample sizes might have been different. Which of the three age-group samples had the largest sample size? Explain your choice.

c. The interval (.768, 1.032) is either a 90% confidence interval or a 99% confidence interval for the mean TV viewing time computed using the sample data for children less than 12 months old. Is the confidence level for this interval 90% or 99%? Explain your choice.

9.46 The article “Most Canadians Plan to Buy Treats, Many Will Buy Pumpkins, Decorations and/or Costumes” (*Ipsos-Reid*, October 24, 2005) summarized results from a survey of 1000 randomly selected Canadian residents. Each individual in the sample was asked how much he or she anticipated spending on Halloween during 2005. The resulting sample mean and standard deviation were $46.65 and $83.70, respectively.
a. Explain how it could be possible for the standard deviation of the anticipated Halloween expense to be larger than the mean anticipated expense.

b. Is it reasonable to think that the distribution of the variable anticipated Halloween expense is approximately normal? Explain why or why not.

c. Is it appropriate to use the $t$ confidence interval to estimate the mean anticipated Halloween expense for Canadian residents? Explain why or why not.

d. If appropriate, construct and interpret a 99% confidence interval for the mean anticipated Halloween expense for Canadian residents.

9.47 Because of safety considerations, in May 2003 the Federal Aviation Administration (FAA) changed its guidelines for how small commuter airlines must estimate passenger weights. Under the old rule, airlines used 180 pounds as a typical passenger weight (including carry-on luggage) in warm months and 185 pounds as a typical weight in cold months. The *Alaska Journal of Commerce* (May 25, 2003) reported that Frontier Airlines conducted a study to estimate average passenger plus carry-on weights. They found an average summer weight of 183 pounds and a winter average of 190 pounds. Suppose that each of these estimates was based on a random sample of 100 passengers and that the sample standard deviations were 20 pounds for the summer weights and 23 pounds for the winter weights.

a. Construct and interpret a 95% confidence interval for the mean summer weight (including carry-on luggage) of Frontier Airlines passengers.

b. Construct and interpret a 95% confidence interval for the mean winter weight (including carry-on luggage) of Frontier Airlines passengers.

c. The new FAA recommendations are 190 pounds for summer and 195 pounds for winter. Comment on these recommendations in light of the confidence interval estimates from Parts (a) and (b).

9.48 Example 9.3 gave the following airborne times (in minutes) for 10 randomly selected flights from San Francisco to Washington Dulles airport:

270 256 267 285 274 275 266 258 271 281

a. Compute and interpret a 90% confidence interval for the mean airborne time for flights from San Francisco to Washington Dulles.

b. Give an interpretation of the 90% confidence level associated with the interval estimate in Part (a).

c. If a flight from San Francisco to Washington Dulles is scheduled to depart at 10 A.M., what would you recommend for the published arrival time? Explain.

9.49 Fat content (in grams) for seven randomly selected hot dogs that were rated as very good by Consumer Reports (www.consumerreports.org) is shown below. Is it reasonable to use this data and the $t$ confidence interval of this section to construct a confidence interval for the mean fat content of hot dogs rated as very good by Consumer Reports? Explain why or why not.

14 15 11 10 6 15 16

9.50 Five students visiting the student health center for a free dental examination during National Dental Hygiene Month were asked how many months had passed since their last visit to a dentist. Their responses were as follows:

6 17 11 22 29

Assuming that these five students can be considered a random sample of all students participating in the free checkup program, construct a 95% confidence interval for the mean number of months elapsed since the last visit to a dentist for the population of students participating in the program.

9.51 The Bureau of Alcohol, Tobacco, and Firearms (BATF) has been concerned about lead levels in California wines. In a previous testing of wine specimens, lead levels ranging from 50 to 700 parts per billion were recorded. How many wine specimens should be tested if the BATF wishes to estimate the true mean lead level for California wines to within 10 parts per billion with 95% confidence?

9.52 The formula described in this section for determining sample size corresponds to a confidence level of 95%. What would be the appropriate formula for determining sample size when the desired confidence level is 90%? 98%?
The purpose of most surveys and many research studies is to produce estimates of population characteristics. One way of providing such an estimate is to construct and report a confidence interval for the population characteristic of interest.

Communicating the Results of Statistical Analyses

When using sample data to estimate a population characteristic, a point estimate or a confidence interval estimate might be used. Confidence intervals are generally preferred because a point estimate by itself does not convey any information about the accuracy of the estimate. For this reason, whenever you report the value of a point estimate, it is a good idea to also include an estimate of the bound on the error of estimation.

Reporting and interpreting a confidence interval estimate requires a bit of care. First, always report both the confidence interval and the confidence level associated with the method used to produce the interval. Then, remember that both the confidence interval and the confidence level should be interpreted. A good strategy is to begin with an interpretation of the confidence interval in the context of the problem and then to follow that with an interpretation of the confidence level. For example, if a 90% confidence interval for \( p \), the proportion of students at a particular university who own a laptop computer, is (.56, .78), we might say

- **We can be 90% confident that between 56% and 78% of the students at this university own laptops.**
- **We have used a method to produce this estimate that is successful in capturing the actual population proportion 90% of the time.**

When providing an interpretation of a confidence interval, remember that the interval is an estimate of a population characteristic and be careful not to say that the interval applies to individual values in the population or to the values of sample statistics. For example, if a 99% confidence interval for \( \mu \), the mean amount of ketchup in bottles labeled as 12 ounces, is (11.94, 11.98), this does not tell us that 99% of 12-ounce ketchup bottles contain between 11.94 and 11.98 ounces of ketchup. Nor does it tell us that 99% of samples of the same size would have sample means in this particular range. The confidence interval is an estimate of the **mean** for all bottles in the population of interest.

Interpreting the Results of Statistical Analyses

Unfortunately, there is no customary way of reporting the estimates of population characteristics in published sources. Possibilities include

- confidence interval
- estimate ± bound on error
- estimate ± standard error
If the population characteristic being estimated is a population mean, then you may also see

\[ \text{sample mean } \pm \text{ sample standard deviation} \]

If the interval reported is described as a confidence interval, a confidence level should accompany it. These intervals can be interpreted just as we have interpreted the confidence intervals in this chapter, and the confidence level specifies the long-run error rate associated with the method used to construct the interval (for example, a 95% confidence level specifies a 5% long-run error rate).

A form particularly common in news articles is estimate ± bound on error, where the bound on error is also sometimes called the margin of error. The bound on error reported is usually two times the standard deviation of the estimate. This method of reporting is a little less formal than a confidence interval and, if the sample size is reasonably large, is roughly equivalent to reporting a 95% confidence interval. You can interpret these intervals as you would a confidence interval with approximate confidence level of 95%.

You must use care in interpreting intervals reported in the form of an estimate ± standard error. Recall from Section 9.2 that the general form of a confidence interval is

\[ \text{estimate } \pm (\text{critical value})(\text{standard deviation of the estimate}) \]

In journal articles, the estimated standard deviation of the estimate is usually referred to as the standard error. The critical value in the confidence interval formula was determined by the form of the sampling distribution of the estimate and by the confidence level. Note that the reported form, estimate ± standard error, is equivalent to a confidence interval with the critical value set equal to 1. For a statistic whose sampling distribution is (approximately) normal (such as the mean of a large sample or a large-sample proportion), a critical value of 1 corresponds to an approximate confidence level of about 68%. Because a confidence level of 68% is rather low, you may want to use the given information and the confidence interval formula to convert to an interval with a higher confidence level.

When researchers are trying to estimate a population mean, they sometimes report sample mean ± sample standard deviation. Be particularly careful here. To convert this information into a useful interval estimate of the population mean, you must first convert the sample standard deviation to the standard error of the sample mean (by dividing by \( \sqrt{n} \)) and then use the standard error and an appropriate critical value to construct a confidence interval.

For example, suppose that a random sample of size 100 is used to estimate the population mean. If the sample resulted in a sample mean of 500 and a sample standard deviation of 20, you might find the published results summarized in any of the following ways:

- 95% confidence interval for the population mean: (496.08, 503.92)
- mean ± bound on error: 500 ± 4
- mean ± standard error: 500 ± 2
- mean ± standard deviation: 500 ± 20

### What to Look For in Published Data

Here are some questions to ask when you encounter interval estimates in research reports:

- Is the reported interval a confidence interval, mean ± bound on error, mean ± standard error, or mean ± standard deviation? If the reported interval is not a
confidence interval, you may want to construct a confidence interval from the
given information.

• What confidence level is associated with the given interval? Is the choice of con-
fidence level reasonable? What does the confidence level say about the long-run
error rate of the method used to construct the interval?

• Is the reported interval relatively narrow or relatively wide? Has the population
characteristic been estimated precisely?

For example, the article “Use of a Cast Compared with a Functional Ankle
Brace After Operative Treatment of an Ankle Fracture” (Journal of Bone and Joint
Surgery [2003]: 205–211) compared two different methods of immobilizing an ankle
after surgery to repair damage from a fracture. The article includes the following
statement:

The mean duration (and standard deviation) between the operation and re-
turn to work was 63 ± 13 days (median, sixty-three days; range, thirty three
to ninety-eight days) for the cast group and 65 ± 19 days (median, sixty-two
days; range, eight to 131 days) for the brace group; the difference was not
significant.

This is an example of a case where we must be careful—the reported intervals are
of the form estimate ± standard deviation. We can use this information to construct
a confidence interval for the mean time between surgery and return to work for each
method of immobilization. One hundred patients participated in the study, with 50
wearing a cast after surgery and 50 wearing an ankle brace (random assignment was
used to assign patients to treatment groups). Because the sample sizes are both large,
we can use the $t$ confidence interval formula

$$\text{mean} \pm (t \text{ critical value})\left(\frac{s}{\sqrt{n}}\right)$$

Each sample has df = 50 − 1 = 49. The closest df value in Appendix Table 3 is
for df = 40, and the corresponding $t$ critical value for a 95% confidence level is 2.02.
The corresponding intervals are

Cast: $63 \pm 2.02\left(\frac{13}{\sqrt{50}}\right) = 63 \pm 3.71 = (59.29, 66.71)$

Brace: $65 \pm 2.02\left(\frac{19}{\sqrt{50}}\right) = 65 \pm 5.43 = (59.57, 70.43)$

The chosen confidence level of 95% implies that the method used to construct
each of the intervals has a 5% long-run error rate. Assuming that it is reasonable to
view these samples as representative of the patient population, we can interpret these
intervals as follows: We can be 95% confident that the mean return-to-work time for
those treated with a cast is between 59.29 and 66.71 days, and we can be 95% con-
fident that the mean return-to-work time for those treated with an ankle brace is
between 59.57 and 70.43 days. These intervals are relatively wide, indicating that the
values of the treatment means have not been estimated as precisely as we might like.
This is not surprising, given the sample sizes and the variability in each sample. Note
that the two intervals overlap. This supports the statement that the difference be-
tween the two immobilization methods was not significant. Formal methods for di-
rectly comparing two groups, covered in Chapter 11, could be used to further inves-
tigate this issue.
A Word to the Wise: Cautions and Limitations

When working with point and confidence interval estimates, here are a few things you need to keep in mind:

1. In order for an estimate to be useful, we must know something about accuracy. You should beware of point estimates that are not accompanied by a bound on error or some other measure of accuracy.

2. A confidence interval estimate that is wide indicates that we don’t have very precise information about the population characteristics being estimated. Don’t be fooled by a high confidence level if the resulting interval is wide. High confidence, while desirable, is not the same thing as saying we have precise information about the value of a population characteristic.

The width of a confidence interval is affected by the confidence level, the sample size, and the standard deviation of the statistic used (for example, \( \hat{p} \) or \( \bar{x} \)) as the basis for constructing the interval. The best strategy for decreasing the width of a confidence interval is to take a larger sample. It is far better to think about this before collecting data and to use the required sample size formulas to determine a sample size that will result in a confidence interval estimate that is narrow enough to provide useful information.

3. The accuracy of estimates depends on the sample size, not the population size. This may be counter to intuition, but as long as the sample size is small relative to the population size (\( n \) less than 10% of the population size), the bound on error for estimating a population proportion with 95% confidence is approximately \( 2 \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \) and for estimating a population mean with 95% confidence is approximately \( 2 \frac{s}{\sqrt{n}} \).

Note that each of these involves the sample size \( n \), and both bounds decrease as the sample size increases. Neither approximate bound on error depends on the population size.

The size of the population does need to be considered if sampling is without replacement and the sample size is more than 10% of the population size. In this case, a finite population correction factor \( \sqrt{\frac{N - n}{N - 1}} \) is used to adjust the bound on error (the given bound is multiplied by the correction factor). Since this correction factor is always less than 1, the adjusted bound on error is smaller.

4. Assumptions and “plausibility” conditions are important. The confidence interval procedures of this chapter require certain assumptions. If these assumptions are met, the confidence intervals provide us with a method for using sample data to estimate population characteristics with confidence. When the assumptions associated with a confidence interval procedure are in fact true, the confidence level specifies a correct success rate for the method. However, assumptions (such as the assumption of a normal population distribution) are rarely exactly met in practice. Fortunately, in most cases, as long as the assumptions are approximately met, the confidence interval procedures still work well.

In general, we can only determine if assumptions are “plausible” or approximately met, and that we are in the situation where we expect the inferential procedure to work reasonably well. This is usually confirmed by knowledge of the data collection process and by using the sample data to check certain “plausibility conditions.”
9.4 Interpreting and Communicating the Results of Statistical Analyses

The formal assumptions for the \( z \) confidence interval for a population proportion are

1. The sample is a random sample from the population of interest.
2. The sample size is large enough for the sampling distribution of \( \hat{p} \) to be approximately normal.
3. Sampling is without replacement.

Whether the random sample assumption is plausible will depend on how the sample was selected and the intended population. Plausibility conditions for the other two assumptions are the following:

\[
 n \hat{p} \geq 10 \quad \text{and} \quad n(1 - \hat{p}) \geq 10 \quad \text{(so the sampling distribution of \( \hat{p} \) is approximately normal), and}
\]

\( n \) is less than 10% of the population size (so that the formula for the standard deviation of \( \hat{p} \) provides a good approximation to the actual standard deviation).

The formal assumptions for the \( t \) confidence interval for a population mean are

1. The sample is a random sample from the population of interest.
2. The population distribution is normal, so that the distribution of

\[
 t = \frac{\bar{x} - \mu}{s/\sqrt{n}}
\]

has a \( t \) distribution.

The plausibility of the random sample assumption, as was the case for proportions, will depend on how the sample was selected and the population of interest. The plausibility conditions for the normal population distribution assumption are the following:

A normal probability plot of the data is reasonably straight (indicating that the population distribution is approximately normal), or

The data distribution is approximately symmetric and there are no outliers.

This may be confirmed by looking at a dotplot, boxplot, stem-and-leaf display, or histogram of the data.

Alternatively, if \( n \) is large (\( n \geq 30 \)), the sampling distribution of \( \bar{x} \) will be approximately normal even for nonnormal population distributions. This implies that use of the \( t \) interval is appropriate even if population normality is not plausible.

In the end, you must decide that the assumptions are met or that they are “plausible” and that the inferential method used will provide reasonable results. This is also true for the inferential methods introduced in the chapters that follow.

5. Watch out for the “\( \pm \)” when reading published reports. Don’t fall into the trap of thinking confidence interval every time you see a \( \pm \) in an expression. As was discussed earlier in this section, published reports are not consistent, and in addition to confidence intervals, it is common to see estimate \( \pm \) standard error and estimate \( \pm \) sample standard deviation reported.
9.53 The following quote is from the article “Credit Card Debt Rises Faster for Seniors” (USA Today, July 28, 2009):

The study, which will be released today by Demos, a liberal public policy group, shows that low- and middle-income consumers 65 and older carried $10,235 in average credit card debt last year.

What additional information would you want in order to evaluate the accuracy of this estimate? Explain.

9.54 Authors of the news release titled “Major Gaps Still Exist Between the Perception and the Reality of Americans’ Internet Security Protections, Study Finds” (The National Cyber Security Alliance) estimated the proportion of Americans who claim to have a firewall installed on their computer to protect them from computer hackers to be .80 based on a survey conducted by the Zogby market research firm. They also estimated the proportion of those who actually have a firewall installed to be .42, based on checkups performed by Norton’s PC Help software. The following quote is from the news release:

For the study, NCSA commissioned a Zogby survey of more than 3000 Americans and Symantec conducted checkups of 400 Americans’ personal computers performed by PC Help by Norton (www.norton.com/tuneup). The Zogby poll has a margin of error of ±1.6% and the checkup has a margin of error of ±5%.

Explain why the margins of error for the two estimated proportions are different.

9.55 The paper “The Curious Promiscuity of Queen Honey Bees (Apis mellifera): Evolutionary and Behavioral Mechanisms” (Annals of Zoology Fennici [2001]:255–265) describes a study of the mating behavior of queen honeybees. The following quote is from the paper:

Queens flew for an average of 24.2 ± 9.21 minutes on their mating flights, which is consistent with previous findings. On those flights, queens effectively mated with 4.6 ± 3.47 males (mean ± SD).

a. The intervals reported in the quote from the paper were based on data from the mating flights of $n = 30$ queen honeybees. One of the two intervals reported is stated to be a confidence interval for a population mean. Which interval is this? Justify your choice.

b. Use the given information to construct a 95% confidence interval for the mean number of partners on a mating flight for queen honeybees. For purposes of this exercise, assume that it is reasonable to consider these 30 queen honeybees as representative of the population of queen honeybees.

**ACTIVITY 9.1 Getting a Feel for Confidence Level**

**Technology Activity (Applet):** Open the applet (available at www.cengage.com/statistics/POD4e) called ConfidenceIntervals. You should see a screen like the one shown here.

**Getting Started:** If the “Method” box does not say “Means,” use the drop-down menu to select Means. In the box just below, select “t” from the drop-down menu. This applet will select a random sample from a specified normal population distribution and then use the sample to construct a confidence interval for the population mean. The interval is then plotted on the display, and you can see if the resulting interval contains the actual value of the population mean.
For purposes of this activity, we will sample from a normal population with mean 100 and standard deviation 5. We will begin with a sample size of \( n = 10 \). In the applet window, set \( \mu = 100 \), \( \sigma = 5 \), and \( n = 10 \). Leave the conf-level box set at 95%. Click the “Recalculate” button to rescale the picture on the right. Now click on the sample button. You should see a confidence interval appear on the display on the right-hand side. If the interval contains the actual mean of 100, the interval is drawn in green; if 100 is not in the confidence interval, the interval is shown in red. Your screen should look something like the following.

**Simulating Confidence Intervals**

Part 1: Click on the “Sample” button several more times, and notice how the confidence interval estimate changes from sample to sample. Also notice that at the bottom of the left-hand side of the display, the applet is keeping track of the proportion of all the intervals calculated so far that include the actual value of \( \mu \). If we were to construct a large number of intervals, this proportion should closely approximate the capture rate for the confidence interval method.

To look at more than one interval at a time, change the “Intervals” box from 1 to 100, and then click the sample button. You should see a screen similar to the one at the top right of this page, with 100 intervals in the display on the right-hand side. Again, intervals containing 100 (the value of \( \mu \) in this case) will be green and those that do not contain 100 will be red. Also note that the capture proportion on the left-hand side has also been updated to reflect what happened with the 100 newly generated intervals.

Continue generating intervals until you have seen at least 1000 intervals, and then answer the following question:

a. How does the proportion of intervals constructed that contain \( \mu = 100 \) compare to the stated confidence level of 95%? On how many intervals was your proportion based? (Note—if you followed the instructions, this should be at least 1000.)

Experiment with three other confidence levels of your choice, and then answer the following question:

b. In general, is the proportion of computed \( t \) confidence intervals that contain \( \mu = 100 \) close to the stated confidence level?

Part 2: When the population is normal but \( \sigma \) is unknown, we construct a confidence interval for a population mean using a \( t \) critical value rather than a \( z \) critical value. How important is this distinction?

Let’s investigate. Use the drop-down menu to change the box just below the method box that’s says “Means” from “\( t \)” to “\( z \)” with \( s \).” The applet will now construct intervals using the sample standard deviation, but will use a \( z \) critical value rather than the \( t \) critical value.

Use the applet to construct at least 1000 95% intervals, and then answer the following question:

c. Comment on how the proportion of the computed intervals that include the actual value of the population mean compares to the stated confidence level of 95%. Is this surprising? Explain why or why not.

Now experiment with some different samples sizes. What happens when \( n = 20 \), \( n = 50 \), \( n = 100 \)? Use what you have learned to write a paragraph explaining what these simulations tell you about the advisability of using a \( z \) critical value in the construction of a confidence interval for \( \mu \) when \( \sigma \) is unknown.
Technology Activity (Applet): This activity presumes that you have already worked through Activity 9.1.

Background: In Section 9.2, it was suggested that a confidence interval of the form

\[
\hat{p}_{\text{mod}} \pm (z \text{ critical value}) \sqrt{\frac{\hat{p}_{\text{mod}}(1 - \hat{p}_{\text{mod}})}{n}}
\]

where \( \hat{p}_{\text{mod}} = \frac{\text{successes} + 2}{n + 4} \) is an alternative to the usual large-sample \( z \) confidence interval. This alternative interval is preferred by many statisticians because, in repeated sampling, the proportion of intervals constructed that include the actual value of the population proportion, \( p \), tends to be closer to the stated confidence level. In this activity, we will explore how the “capture rates” for the two different interval estimation methods compare.

Open the applet (available at www.cengage.com/statistics/POD4e) called ConfidenceIntervals. You should see a screen like the one shown.

Simulating Confidence Intervals

1. How does the proportion of intervals constructed that include \( p = .3 \), the population proportion, compare to .95? Does this surprise you? Explain.

Now use the drop-down menu to change “Large Sample \( z \)” to “Modified.” Now the applet will construct the alternative confidence interval that is based on \( \hat{p}_{\text{mod}} \). Use the applet to construct a large number (at least 1000) of 95% confidence intervals.

2. How does the proportion of intervals constructed that include \( p = .3 \), the population proportion, compare to .95? Is this proportion closer to .95 than was the case for the large-sample \( z \) interval?

3. Experiment with different combinations of values of sample size and population proportion \( p \). Can you find a combination for which the large sample \( z \) interval has a capture rate that is close to 95%? Can you find a combination for which it has a capture rate that is even farther from 95% than it was for \( n = 40 \) and \( p = .3 \)? How does the modified interval perform in each of these cases?

ACTIVITY 9.3 Verifying Signatures on a Recall Petition

Background: In 2003, petitions were submitted to the California Secretary of State calling for the recall of Governor Gray Davis. Each of California’s 58 counties then had to report the number of valid signatures on the petitions from that county so that the State could determine whether there were enough valid signatures to certify the recall and set a date for the recall election. The following paragraph appeared in the San Luis Obispo Tribune (July 23, 2003):

In the campaign to recall Gov. Gray Davis, the secretary of state is reporting 16,000 verified signatures from San Luis Obispo County. In all, the County Clerk’s Office received 18,866 signatures on recall petitions and was instructed by the state to check a random sample of 567. Out of those, 84.48% were good. The verification process includes checking whether the signer is a registered voter and whether the address and signa-
2. How do you think that the reported figure of 16,000 verified signature for San Luis Obispo County was obtained?

3. Based on your confidence interval from Step 1, explain why you think that the reported figure of 16,000 verified signatures is or is not reasonable.

**ACTIVITY 9.4  A Meaningful Paragraph**

Write a meaningful paragraph that includes the following six terms: sample, population, confidence level, estimate, mean, margin of error.

A “meaningful paragraph” is a coherent piece writing in an appropriate context that uses all of the listed words. The paragraph should show that you understand the meanings of the terms and their relationship to one another. A sequence of sentences that just define the terms is not a meaningful paragraph. When choosing a context, think carefully about the terms you need to use. Choosing a good context will make writing a meaningful paragraph easier.

**Summary of Key Concepts and Formulas**

<table>
<thead>
<tr>
<th>TERM OR FORMULA</th>
<th>COMMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Point estimate</strong></td>
<td>A single number, based on sample data, that represents a plausible value of a population characteristic.</td>
</tr>
<tr>
<td><strong>Unbiased statistic</strong></td>
<td>A statistic that has a sampling distribution with a mean equal to the value of the population characteristic to be estimated.</td>
</tr>
<tr>
<td><strong>Confidence interval</strong></td>
<td>An interval that is computed from sample data and provides a range of plausible values for a population characteristic.</td>
</tr>
<tr>
<td><strong>Confidence level</strong></td>
<td>A number that provides information on how much “confidence” we can have in the method used to construct a confidence interval estimate. The confidence level specifies the percentage of all possible samples that will produce an interval containing the value of the population characteristic.</td>
</tr>
<tr>
<td>[ \hat{p} \pm (z \text{ critical value}) \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} ]</td>
<td>A formula used to construct a confidence interval for ( p ) when the sample size is large.</td>
</tr>
<tr>
<td>[ n = \frac{p(1 - p)(1.96)^2}{B^2} ]</td>
<td>A formula used to compute the sample size necessary for estimating ( p ) to within an amount ( B ) with 95% confidence. (For other confidence levels, replace 1.96 with an appropriate ( z ) critical value.)</td>
</tr>
<tr>
<td>[ \bar{x} \pm (z \text{ critical value}) \frac{\sigma}{\sqrt{n}} ]</td>
<td>A formula used to construct a confidence interval for ( \mu ) when ( \sigma ) is known and either the sample size is large or the population distribution is normal.</td>
</tr>
<tr>
<td>[ \bar{x} \pm (t \text{ critical value}) \frac{s}{\sqrt{n}} ]</td>
<td>A formula used to construct a confidence interval for ( \mu ) when ( \sigma ) is unknown and either the sample size is large or the population distribution is normal.</td>
</tr>
</tbody>
</table>
A formula used to compute the sample size necessary for estimating \( \mu \) to within an amount \( B \) with 95% confidence. (For other confidence levels, replace 1.96 with an appropriate \( z \) critical value.)

\[
 n = \left( \frac{1.96 \sigma}{B} \right)^2
\]

Chapter Review Exercises 9.56 - 9.73

9.56 According to an AP-Ipsos poll (June 15, 2005), 42% of 1001 randomly selected adult Americans made plans in May 2005 based on a weather report that turned out to be wrong.

a. Construct and interpret a 99% confidence interval for the proportion of Americans who made plans in May 2005 based on an incorrect weather report.

b. Do you think it is reasonable to generalize this estimate to other months of the year? Explain.

9.57  "Tongue Piercing May Speed Tooth Loss, Researchers Say" is the headline of an article that appeared in the San Luis Obispo Tribune (June 5, 2002). The article describes a study of 52 young adults with pierced tongues. The researchers found receding gums, which can lead to tooth loss, in 18 of the participants. Construct a 95% confidence interval for the proportion of young adults with pierced tongues who have receding gums. What assumptions must be made for use of the \( z \) confidence interval to be appropriate?

9.58 In a study of 1710 schoolchildren in Australia (Herald Sun, October 27, 1994), 1060 children indicated that they normally watch TV before school in the morning. (Interestingly, only 35% of the parents said their children watched TV before school!) Construct a 95% confidence interval for the true proportion of Australian children who say they watch TV before school. What assumption about the sample must be true for the method used to construct the interval to be valid?

9.59 The authors of the paper "Short-Term Health and Economic Benefits of Smoking Cessation: Low Birth Weight" (Pediatrics [1999]: 1312–1320) investigated the medical cost associated with babies born to mothers who smoke. The paper included estimates of mean medical cost for low-birth-weight babies for different ethnic groups. For a sample of 654 Hispanic low-birth-weight babies, the mean medical cost was $55,007 and the standard error (\( s/\sqrt{n} \)) was $3011. For a sample of 13 Native American low-birth-weight babies, the mean and standard error were $73,418 and $29,577, respectively. Explain why the two standard errors are so different.

9.60 The article "Consumers Show Increased Liking for Diesel Autos" (USA Today, January 29, 2003) reported that 27% of U.S. consumers would opt for a diesel car if it ran as cleanly and performed as well as a car with a gas engine. Suppose that you suspect that the proportion might be different in your area and that you want to conduct a survey to estimate this proportion for the adult residents of your city. What is the required sample size if you want to estimate this proportion to within .05 with 95% confidence? Compute the required sample size first using .27 as a preliminary estimate of \( p \) and then using the conservative value of .5. How do the two sample sizes compare? What sample size would you recommend for this study?

9.61 In the article "Fluoridation Brushed Off by Utah" (Associated Press, August 24, 1998), it was reported that a small but vocal minority in Utah has been successful in keeping fluoride out of Utah water supplies despite evidence that fluoridation reduces tooth decay and despite the fact that a clear majority of Utah residents favor fluoridation. To support this statement, the article included the result of a survey of Utah residents that found 65% to be in favor of fluoridation. Suppose that this result was based on a random sample of 150 Utah residents. Construct and interpret a 90% confidence interval for \( p \), the true proportion of Utah residents who favor fluoridation. Is this interval consistent with the statement that fluoridation is favored by a clear majority of residents?

9.62 Seventy-seven students at the University of Virginia were asked to keep a diary of conversations with their mothers, recording any lies they told during these
An Associated Press article on potential violent behavior reported the results of a survey of 750 workers who were employed full time (San Luis Obispo Tribune, September 7, 1999). Of those surveyed, 125 indicated that they were so angered by a coworker during the past year that they felt like hitting the coworker (but didn’t). Assuming that it is reasonable to regard this sample of 750 as a random sample from the population of full-time workers, use this information to construct and interpret a 90% confidence interval estimate of $p$, the true proportion of full-time workers so angered in the last year that they wanted to hit a colleague.

The 1991 publication of the book *Final Exit*, which includes chapters on doctor-assisted suicide, caused a great deal of controversy in the medical community. The Society for the Right to Die and the American Medical Association quoted very different figures regarding the proportion of primary-care physicians who have participated in some form of doctor-assisted suicide for terminally ill patients (USA Today, July 1991). Suppose that a survey of physicians is to be designed to estimate this proportion to within .05 with 95% confidence. How many primary-care physicians should be included in the random sample?

A manufacturer of small appliances purchases plastic handles for coffeepots from an outside vendor. If a handle is cracked, it is considered defective and must be discarded. A large shipment of plastic handles is received. The proportion of defective handles $p$ is of interest. How many handles from the shipment should be inspected to estimate $p$ to within 0.1 with 95% confidence?

An article in the *Chicago Tribune* (August 29, 1999) reported that in a poll of residents of the Chicago suburbs, 43% felt that their financial situation had improved during the past year. The following statement is from the article: “The findings of this Tribune poll are based on interviews with 930 randomly selected suburban residents. The sample included suburban Cook County plus DuPage, Kane, Lake, McHenry, and Will Counties. In a sample of this size, one can say with 95% certainty that results will differ by no more than 3% from results obtained if all residents had been included in the poll.”

Comment on this statement. Give a statistical argument to justify the claim that the estimate of 43% is within 3% of the true proportion of residents who feel that their financial situation has improved.

A manufacturer of college textbooks is interested in estimating the strength of the bindings produced by a particular binding machine. Strength can be measured by recording the force required to pull the pages from the binding. If this force is measured in pounds, how many books should be tested to estimate the mean force required to break the binding to within 0.1 pounds with 95% confidence? Assume that $\sigma$ is known to be 0.8 pound.

Recent high-profile legal cases have many people reevaluating the jury system. Many believe that juries in criminal trials should be able to convict on less than a unanimous vote. To assess support for this idea, investigators asked each individual in a random sample of Californians whether they favored allowing conviction by a 10–2 verdict in criminal cases not involving the death penalty. The Associated Press (San Luis Obispo Telegram-Tribune, September 13, 1995) reported that 71% supported the 10–2 verdict. Suppose that the sample size for this survey was $n = 900$. Compute and interpret a 99% confidence interval for the proportion of Californians who favor the 10–2 verdict.

The confidence intervals presented in this chapter give both lower and upper bounds on plausible values for the population characteristic being estimated. In some instances, only an upper bound or only a lower bound is appropriate. Using the same reasoning that gave the large sample interval in Section 9.3, we can say that when $n$ is large, 99% of all samples have

$$
\mu < \bar{x} + 2.33 \frac{s}{\sqrt{n}} 
$$

(because the area under the $z$ curve to the left of 2.33 is .99). Thus, $\bar{x} + 2.33 \frac{s}{\sqrt{n}}$ is a 99% upper confidence bound for $\mu$. Use the data of Example 9.9 to calculate the 99% upper confidence bound for the true mean wait time for bypass patients in Ontario.

**Bold exercises answered in back**

**Data set available online**

**Video Solution available**
9.70  The Associated Press (December 16, 1991) reported that in a random sample of 507 people, only 142 correctly described the Bill of Rights as the first 10 amendments to the U.S. Constitution. Calculate a 95% confidence interval for the proportion of the entire population that could give a correct description.

9.71  When \( n \) is large, the statistic \( s \) is approximately unbiased for estimating \( \sigma \) and has approximately a normal distribution. The standard deviation of this statistic when the population distribution is normal is \( \sigma_s = \frac{\sigma}{\sqrt{2n}} \), which can be estimated by \( \frac{s}{\sqrt{2n}} \). A large-sample confidence interval for the population standard deviation \( \sigma \) is then

\[
s \pm (z \text{ critical value}) \frac{s}{\sqrt{2n}}
\]

Use the data of Example 9.9 to obtain a 95% confidence interval for the true standard deviation of waiting time for angiography.

9.72  The interval from \(-2.33\) to \(1.75\) captures an area of .95 under the \( z \) curve. This implies that another large-sample 95% confidence interval for \( \mu \) has lower limit \( \bar{x} - 2.33 \frac{\sigma}{\sqrt{n}} \) and upper limit \( \bar{x} + 1.75 \frac{\sigma}{\sqrt{n}} \). Would you recommend using this 95% interval over the 95% interval \( \bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}} \) discussed in the text? Explain. (Hint: Look at the width of each interval.)

9.73  The eating habits of 12 bats were examined in the article “Foraging Behavior of the Indian False Vampire Bat” (Biotropica [1991]: 63–67). These bats consume insects and frogs. For these 12 bats, the mean time to consume a frog was \( \bar{x} = 21.9 \) minutes. Suppose that the standard deviation was \( s = 7.7 \) minutes. Construct and interpret a 90% confidence interval for the mean sup-pertime of a vampire bat whose meal consists of a frog. What assumptions must be reasonable for the one-sam-ple \( t \) interval to be appropriate?

**Graphing Calculator Explorations**

**EXPLORATION 9.1  The Confidence Interval for a Population Proportion**

Because confidence intervals are widely used, it will come as no surprise to you that your calculator may have a built-in capability to determine a confidence interval for a population proportion. Once again, you will need to navigate your calculator’s menu system, looking for key words such as INTR for “interval,” or possibly TESTS. (Confidence intervals are frequently associated with “hypothesis tests,” a topic to come later in Chapter 10.) Once you find the right menu, look for words like “1” and “prop” and “z”—together these key words should indicate a one-sample \( z \) confidence interval for a proportion.

Once you select the correct choice, you will be presented with a screen for providing the information needed to calculate a confidence interval: the number of successes, the sample size, and the confidence level. Two representative screens are given in Figure 9.10.

From Example 9.4, supply the following information: \( x = 567 \), \( n = 1031 \), and the C-Level = .95. Move your cursor down to Execute or Calculate, and press the Enter, Execute, or Calculate button, depending on your calculator. The confidence interval should appear immediately. Again, two representative screens are given in Figure 9.11.
Notice that the formatting in these screens is slightly different. Which of these is the “right” format? Probably neither one! Recall a previous graphing calculator exploration in Chapter 3, where we discussed the differences between a calculator’s presentation of information and the appropriate way to communicate this information. Check with your instructor to determine her or his preferences about what format information is required.

EXPLORATION 9.2 A Confidence Interval for a Population Mean

Finding the confidence interval for a single mean on your calculator will require you to navigate the menu system much as you did to find the confidence interval for a proportion. The confidence interval for the mean will have some added challenges, however:

1. You will need to decide whether to base the interval on the $z$ or $t$ distribution, and
2. You may use previous calculations of the sample mean and standard deviation, or the calculator will evaluate these statistics from data contained in a List.

We will use the data from Example 9.10 to construct a 99% confidence interval. We have entered the data in List1 and are ready to proceed. Since the population standard deviation is not known, we will use the $t$ confidence interval. For the $t$ confidence interval, the normality of the population becomes an issue. The original data are at hand, and we can assess the plausibility of a normal population via a normal probability plot. Figure 9.12 shows the normal probability plot.

After verifying the plausibility of the normality of the population, we can now construct the confidence interval. In this case, we have a choice of entering the sample calculations or letting the calculator evaluate the sample mean and standard deviation. Based on that choice, we see one of the screens shown in Figure 9.13.

Figure 9.14 shows one calculator’s version of the confidence interval.

![FIGURE 9.11](image)

*Calculator screens for confidence intervals.*

![FIGURE 9.12](image)

*(a) Edit window for normal probability plot; (b) normal probability plot.*

![FIGURE 9.13](image)

*(a) Confidence interval input screen for raw data; (b) confidence interval input screen for summary statistics.*

![FIGURE 9.14](image)

*Calculated confidence interval.*