Section 9.1
The Logic in Constructing Confidence Intervals for a Population Mean When the Population Standard Deviation Is Known

Objectives
1. Compute a point estimate of the population mean
2. Construct and interpret a confidence interval for the population mean assuming that the population standard deviation is known
3. Understand the role of margin of error in constructing the confidence interval
4. Determine the sample size necessary for estimating the population mean within a specified margin of error

A point estimate is the value of a statistic that estimates the value of a parameter.

For example, the sample mean, $\bar{x}$, is a point estimate of the population mean $\mu$.

Parallel Example 1: Computing a Point Estimate

Pennies minted after 1982 are made from 97.5% zinc and 2.5% copper. The following data represent the weights (in grams) of 17 randomly selected pennies minted after 1982:

$2.46, 2.47, 2.49, 2.48, 2.50, 2.44, 2.46, 2.45, 2.49, 2.47, 2.45, 2.46, 2.45, 2.46, 2.47, 2.44, 2.45$

Treat the data as a simple random sample. Estimate the population mean weight of pennies minted after 1982.

Solution

The sample mean is

$$\bar{x} = \frac{2.46 + 2.47 + \cdots + 2.45}{17} \approx 2.464$$

The point estimate of $\mu$ is 2.464 grams.

A confidence interval for an unknown parameter consists of an interval of numbers.

The level of confidence represents the expected proportion of intervals that will contain the parameter if a large number of different samples is obtained. The level of confidence is denoted $(1-\alpha)$-100%.
For example, a 95% level of confidence (\(\alpha = 0.05\)) implies that if 100 different confidence intervals are constructed, each based on a different sample from the same population, we will expect 95 of the intervals to contain the parameter and 5 to not include the parameter.

- Confidence interval estimates for the population mean are of the form
  \[
  \text{Point estimate} \pm \text{margin of error}.
  \]

- The margin of error of a confidence interval estimate of a parameter is a measure of how accurate the point estimate is.

The margin of error depends on three factors:
1. **Level of confidence:** As the level of confidence increases, the margin of error also increases.
2. **Sample size:** As the size of the random sample increases, the margin of error decreases (consequence of the Law of Large Numbers).
3. **Standard deviation of the population:** The more spread there is in the population, the wider our interval will be for a given level of confidence.

**From chapter 8, we know that.....**

The shape of the distribution of all possible sample means will be normal, provided the population is normal or approximately normal (given the sample size is large (\(n \geq 30\)) enough), with

- mean \(\mu_x = \mu\)
- and standard deviation \(\sigma_x = \frac{\sigma}{\sqrt{n}}\)

Because \(\bar{x}\) is normally distributed, we know 95% of all sample means lie within 1.96 standard deviations of the population mean, \(\mu\), and 2.5% of the sample means lie in each tail.
95% of all sample means are in the interval

$$\mu - 1.96 \frac{\sigma}{\sqrt{n}} < \bar{x} < \mu + 1.96 \frac{\sigma}{\sqrt{n}}$$

With a little algebraic manipulation, we can rewrite this inequality and obtain:

$$\bar{x} - 1.96 \sigma_{\bar{x}} < \mu < \bar{x} + 1.96 \sigma_{\bar{x}}$$

It is common to write the 95% confidence interval as

$$\bar{x} \pm 1.96 \sigma_{\bar{x}}$$

so that it is of the form

Point estimate ± margin of error.

We will use Minitab to simulate obtaining 30 simple random samples of size $n=8$ from a population that is normally distributed with $\mu=50$ and $\sigma=10$. Construct a 95% confidence interval for each sample. How many of the samples result in intervals that contain $\mu=50$?

<table>
<thead>
<tr>
<th>Sample</th>
<th>Mean</th>
<th>95.0% CI</th>
<th>Does not capture mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>47.07</td>
<td>(40.14, 54.00)</td>
<td></td>
</tr>
<tr>
<td>C2</td>
<td>49.33</td>
<td>(42.40, 56.26)</td>
<td></td>
</tr>
<tr>
<td>C3</td>
<td>50.62</td>
<td>(43.69, 57.54)</td>
<td></td>
</tr>
<tr>
<td>C4</td>
<td>47.91</td>
<td>(40.98, 54.84)</td>
<td></td>
</tr>
<tr>
<td>C5</td>
<td>44.31</td>
<td>(37.38, 51.24)</td>
<td></td>
</tr>
<tr>
<td>C6</td>
<td>51.50</td>
<td>(44.57, 58.43)</td>
<td></td>
</tr>
<tr>
<td>C7</td>
<td>52.47</td>
<td>(45.54, 59.40)</td>
<td></td>
</tr>
<tr>
<td>C8</td>
<td>59.62</td>
<td>(52.69, 66.54)</td>
<td></td>
</tr>
<tr>
<td>C9</td>
<td>43.49</td>
<td>(36.56, 50.44)</td>
<td></td>
</tr>
<tr>
<td>C10</td>
<td>55.45</td>
<td>(48.52, 62.38)</td>
<td></td>
</tr>
<tr>
<td>C11</td>
<td>50.08</td>
<td>(43.15, 57.01)</td>
<td></td>
</tr>
<tr>
<td>C12</td>
<td>56.37</td>
<td>(49.44, 63.30)</td>
<td></td>
</tr>
<tr>
<td>C13</td>
<td>49.05</td>
<td>(42.12, 55.98)</td>
<td></td>
</tr>
<tr>
<td>C14</td>
<td>47.34</td>
<td>(40.41, 54.27)</td>
<td></td>
</tr>
<tr>
<td>C15</td>
<td>50.33</td>
<td>(43.40, 57.25)</td>
<td></td>
</tr>
</tbody>
</table>

Note that 28 out of 30, or 93%, of the confidence intervals contain the population mean $\mu=50$.

In general, for a 95% confidence interval, any sample mean that lies within 1.96 standard errors of the population mean will result in a confidence interval that contains $\mu$.

Whether a confidence interval contains $\mu$ depends solely on the sample mean, $\bar{x}$. 
Interpretation of a Confidence Interval

A \((1-\alpha)\cdot100\%\) confidence interval indicates that, if we obtained many simple random samples of size \(n\) from the population whose mean, \(\mu\), is unknown, then approximately \((1-\alpha)\cdot100\%\) of the intervals will contain \(\mu\).

For example, if we constructed a 99% confidence interval with a lower bound of 52 and an upper bound of 71, we would interpret the interval as follows: “We are 99% confident that the population mean, \(\mu\), is between 52 and 71.”

Constructing a \((1-\alpha)\cdot100\%\) Confidence Interval for \(\mu\), \(\sigma\) Known

Suppose that a simple random sample of size \(n\) is taken from a population with unknown mean, \(\mu\), and known standard deviation \(\sigma\). A \((1-\alpha)\cdot100\%\) confidence interval for \(\mu\) is given by

\[
\text{Lower Bound: } \bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}
\]
\[
\text{Upper Bound: } \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}
\]

where \(z_{\alpha/2}\) is the critical Z-value.

Note: The sample size must be large (\(n \geq 30\)) or the population must be normally distributed.

Parallel Example 3: Constructing a Confidence Interval

Construct a 99% confidence interval about the population mean weight (in grams) of pennies minted after 1982. Assume \(\sigma = 0.02\) grams.

\[
\begin{align*}
\text{Data} & : 2.46 & 2.47 & 2.49 & 2.48 & 2.50 & 2.44 & 2.46 & 2.45 & 2.49 \\
\text{Weight (in grams) of Pennies} & : 2.47 & 2.45 & 2.46 & 2.47 & 2.44 & 2.45 & 2.47 & 2.44 & 2.45 \\
\end{align*}
\]

• \(z_{0.005} = 2.575\)
• Lower bound:
\[
\bar{x} - z_{0.005} \cdot \frac{\sigma}{\sqrt{n}} = 2.464 - 2.575 \left( \frac{0.02}{\sqrt{17}} \right) = 2.464 - 0.012 = 2.452
\]
• Upper bound:
\[
\bar{x} + z_{0.005} \cdot \frac{\sigma}{\sqrt{n}} = 2.464 + 2.575 \left( \frac{0.02}{\sqrt{17}} \right) = 2.464 + 0.012 = 2.476
\]

We are 99% confident that the mean weight of pennies minted after 1982 is between 2.452 and 2.476 grams.
The margin of error, \( E \), in a \((1-\alpha)\cdot100\%\) confidence interval in which \( \sigma \) is known is given by

\[
E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}
\]

where \( n \) is the sample size.

Note: We require that the population from which the sample was drawn be normally distributed or the samples size \( n \) be greater than or equal to 30.

### Parallel Example 5: Role of the Level of Confidence in the Margin of Error

Construct a 90\% confidence interval for the mean weight of pennies minted after 1982. Comment on the effect that decreasing the level of confidence has on the margin of error.

- \( z_{0.05} = 1.645 \)
- Lower bound:
  \[
  \bar{x} - z_{0.05} \cdot \frac{\sigma}{\sqrt{n}} = 2.464 - 1.645 \left( \frac{0.02}{\sqrt{17}} \right)
  = 2.464 - 0.008 = 2.456
  
  \]
- Upper bound:
  \[
  \bar{x} + z_{0.05} \cdot \frac{\sigma}{\sqrt{n}} = 2.464 + 1.645 \left( \frac{0.02}{\sqrt{17}} \right)
  = 2.464 + 0.008 = 2.472
  
  We are 90\% confident that the mean weight of pennies minted after 1982 is between 2.456 and 2.472 grams.

Notice that the margin of error decreased from 0.012 to 0.008 when the level of confidence decreased from 99\% to 90\%. The interval is therefore wider for the higher level of confidence. (because of the increased error)

<table>
<thead>
<tr>
<th>Confidence Level</th>
<th>Margin of Error</th>
<th>Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>90%</td>
<td>0.008</td>
<td>(2.456, 2.472)</td>
</tr>
<tr>
<td>99%</td>
<td>0.012</td>
<td>(2.452, 2.476)</td>
</tr>
</tbody>
</table>

### Parallel Example 6: Role of Sample Size in the Margin of Error

Suppose that we obtained a simple random sample of pennies minted after 1982. Construct a 99\% confidence interval with \( n=35 \). Assume the larger sample size results in the same sample mean, 2.464. The standard deviation is still \( \sigma=0.02 \). Comment on the effect increasing sample size has on the width of the interval.

- \( z_{0.01} = 2.575 \)
- Lower bound:
  \[
  \bar{x} - z_{0.01} \cdot \frac{\sigma}{\sqrt{n}} = 2.464 - 2.575 \left( \frac{0.02}{\sqrt{35}} \right)
  = 2.464 - 0.009 = 2.455
  
  \]
- Upper bound:
  \[
  \bar{x} + z_{0.01} \cdot \frac{\sigma}{\sqrt{n}} = 2.464 + 2.575 \left( \frac{0.02}{\sqrt{35}} \right)
  = 2.464 + 0.009 = 2.473
  
  We are 99\% confident that the mean weight of pennies minted after 1982 is between 2.455 and 2.473 grams.
Notice that the margin of error decreased from 0.012 to 0.009 when the sample size increased from 17 to 35. The interval is therefore narrower for the larger sample size.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Margin of Error</th>
<th>Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>0.012</td>
<td>(2.452, 2.476)</td>
</tr>
<tr>
<td>35</td>
<td>0.009</td>
<td>(2.455, 2.473)</td>
</tr>
</tbody>
</table>

### Determining the Sample Size $n$

The sample size required to estimate the population mean, $\mu$, with a level of confidence $(1-\alpha)$-100% with a specified margin of error, $E$, is given by

$$n = \left( \frac{z_{\alpha/2} \cdot \sigma}{E} \right)^2$$

where $n$ is rounded up to the nearest whole number.

Back to the pennies. How large a sample would be required to estimate the mean weight of a penny manufactured after 1982 within 0.005 grams with 99% confidence? Assume $\sigma = 0.02$.

### Parallel Example 7: Determining the Sample Size

- $z_{0.005} = 2.575$
- $\sigma = 0.02$
- $E = 0.005$

$$n = \left( \frac{z_{0.005} \cdot \sigma}{E} \right)^2 = \left( \frac{2.575(0.02)}{0.005} \right)^2 = 106.09$$

Rounding up, we find $n = 107$. 

### Zintervals Using calculator

- If necessary, enter data into L1
- Press STAT, highlight TESTS, and select 7: Zinterval
- If data are raw, highlight DATA. Make sure List is set to L1, Freq set to 1.
- If summary stats are known, highlight STAT and enter in standard deviation $\sigma$, mean point estimate $\bar{x}$, and $n$.
- Enter confidence level after C-level
- Highlight Calculate, and press ENTER

### Section 9.2

Confidence Intervals about a Population Mean When the Population Standard Deviation is Unknown

**Objectives**

1. Know the properties of Student’s $t$-distribution
2. Determine $t$-values
3. Construct and interpret a confidence interval for a population mean
Suppose that a simple random sample of size \( n \) is taken from a population. If the population from which the sample is drawn follows a normal distribution, the distribution of

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

follows Student’s \( t \)-distribution with \( n-1 \) degrees of freedom where \( \bar{x} \) is the sample mean and \( s \) is the sample standard deviation.

**Parallel Example 1: Comparing the Standard Normal Distribution to the \( t \)-Distribution Using Simulation**

a) Obtain 1,000 simple random samples of size \( n=5 \) from a normal population with \( \mu=50 \) and \( \sigma=10 \).
b) Determine the sample mean and sample standard deviation for each of the samples.
c) Compute \( z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \) and \( t = \frac{\bar{x} - \mu}{s / \sqrt{n}} \) for each sample.
d) Draw a histogram for both \( z \) and \( t \).

**CONCLUSIONS:**

- The histogram for \( z \) is symmetric and bell-shaped with the center of the distribution at 0 and virtually all the rectangles between -3 and 3. In other words, \( z \) follows a standard normal distribution.
- The histogram for \( t \) is also symmetric and bell-shaped with the center of the distribution at 0, but the distribution of \( t \) has longer tails (i.e., \( t \) is more dispersed), so it is unlikely that \( t \) follows a standard normal distribution. The additional spread in the distribution of \( t \) can be attributed to the fact that we use \( s \) to find \( t \) instead of \( \sigma \). Because the sample standard deviation is itself a random variable (rather than a constant such as \( \sigma \)), we have more dispersion in the distribution of \( t \).

**Properties of the \( t \)-Distribution**

1. The \( t \)-distribution is different for different degrees of freedom.
2. The \( t \)-distribution is centered at 0 and is symmetric about 0.
3. The area under the curve is 1. The area under the curve to the right of 0 equals the area under the curve to the left of 0 equals 1/2.
4. As \( t \) increases without bound, the graph approaches, but never equals, zero. As \( t \) decreases without bound, the graph approaches, but never equals, zero.
Properties of the $t$-Distribution

5. The area in the tails of the $t$-distribution is a little greater than the area in the tails of the standard normal distribution, because we are using $s$ as an estimate of $\sigma$, thereby introducing further variability into the $t$-statistic.

6. As the sample size $n$ increases, the density curve of $t$ gets closer to the standard normal density curve. This result occurs because, as the sample size $n$ increases, the values of $s$ get closer to the values of $\sigma$, by the Law of Large Numbers.

Parallel Example 2: Finding $t$-values

Find the $t$-value such that the area under the $t$-distribution to the right of the $t$-value is 0.2 assuming 10 degrees of freedom. That is, find $t_{0.20}$ with 10 degrees of freedom.

Solution

The figure to the left shows the graph of the $t$-distribution with 10 degrees of freedom. The unknown value of $t$ is labeled, and the area under the curve to the right of $t$ is shaded. The value of $t_{0.20}$ with 10 degrees of freedom is 0.8791.

The TI84 has an invT feature which finds the value of $t$ given an area left of the unknown $t$-value and the degrees of freedom.

Constructing a $(1-\alpha)$-100% Confidence Interval for $\mu$, $\sigma$ Unknown

Suppose that a simple random sample of size $n$ is taken from a population with unknown mean $\mu$ and unknown standard deviation $\sigma$. A $(1-\alpha)$-100% confidence interval for $\mu$ is given by

$$
\bar{x} - t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \quad \text{Lower bound:} \quad \bar{x} + t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \quad \text{Upper bound:}
$$

Note: The interval is exact when the population is normally distributed. It is approximately correct for nonnormal populations, provided that $n$ is large enough.
Parallel Example 3: Constructing a Confidence Interval about a Population Mean

The pasteurization process reduces the amount of bacteria found in dairy products, such as milk. The following data represent the counts of bacteria in pasteurized milk (in CFU/mL) for a random sample of 12 pasteurized glasses of milk. Data courtesy of Dr. Michael Lee, Professor, Joliet Junior College.

Construct a 95% confidence interval for the bacteria count.

<table>
<thead>
<tr>
<th>Sample</th>
<th>CFU/mL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.99</td>
</tr>
<tr>
<td>2</td>
<td>1.95</td>
</tr>
<tr>
<td>3</td>
<td>9.22</td>
</tr>
<tr>
<td>4</td>
<td>9.41</td>
</tr>
<tr>
<td>5</td>
<td>4.14</td>
</tr>
<tr>
<td>6</td>
<td>1.79</td>
</tr>
<tr>
<td>7</td>
<td>3.63</td>
</tr>
<tr>
<td>8</td>
<td>3.49</td>
</tr>
<tr>
<td>9</td>
<td>6.95</td>
</tr>
<tr>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>12</td>
<td>0.38</td>
</tr>
</tbody>
</table>

NOTE: Each observation is in tens of thousand. So, 9.06 represents 9.06 x 10^4.

Solution:

Checking Normality and Existence of Outliers

In Tests, Select 8: Tinterval

\[ \bar{x} = 6.41 \quad \text{and} \quad s = 4.55 \]
\[ \alpha = 0.05, \quad n = 12, \quad \text{so} \quad t_{0.025} = 2.201 \]

Lower bound:
\[ 6.41 - 2.201 \times \frac{4.55}{\sqrt{12}} = 3.52 \]

Upper bound:
\[ 6.41 + 2.201 \times \frac{4.55}{\sqrt{12}} = 9.30 \]

The 95% confidence interval for the mean bacteria count in pasteurized milk is (3.52, 9.30).

Parallel Example 5: The Effect of Outliers

Suppose a student miscalculated the amount of bacteria and recorded a result of 2.3 x 10^5. We would include this value in the data set as 23.0.

What effect does this additional observation have on the 95% confidence interval?
Boxplot of CFU/mL

Solution: Checking Normality and Existence of Outliers

A point estimate is an unbiased estimator of the parameter. The point estimate for the population proportion is

$$\hat{p} = \frac{x}{n}$$

where $x$ is the number of individuals in the sample with the specified characteristic and $n$ is the sample size.

Solution

- $x = 7.69$ and $s = 6.34$
- $\alpha = 0.05$, $n = 13$, so $t_{0.05} = 2.179$

Lower bound:

$$7.69 - 2.179 \times \frac{6.34}{\sqrt{13}} = 3.86$$

Upper bound:

$$7.69 + 2.179 \times \frac{6.34}{\sqrt{13}} = 11.52$$

The 95% confidence interval for the mean bacteria count in pasteurized milk, including the outlier is $(3.86, 11.52)$.

CONCLUSIONS:

- With the outlier, the sample mean is larger because the sample mean is not resistant
- With the outlier, the sample standard deviation is larger because the sample standard deviation is not resistant
- Without the outlier, the width of the interval decreased from 7.66 to 5.78.

<table>
<thead>
<tr>
<th>$\bar{x}$</th>
<th>$s$</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without Outlier</td>
<td>6.41</td>
<td>4.55</td>
</tr>
<tr>
<td>With Outlier</td>
<td>7.69</td>
<td>6.34</td>
</tr>
</tbody>
</table>

Objectives

1. Obtain a point estimate for the population proportion
2. Construct and interpret a confidence interval for the population proportion
3. Determine the sample size necessary for estimating a population proportion within a specified margin of error

In July of 2008, a Quinnipiac University Poll asked 1783 registered voters nationwide whether they favored or opposed the death penalty for persons convicted of murder. 1123 were in favor.

Obtain a point estimate for the proportion of registered voters nationwide who are in favor of the death penalty for persons convicted of murder.
Obtain a point estimate for the proportion of registered voters nationwide who are in favor of the death penalty for persons convicted of murder.

\[ \hat{p} = \frac{1123}{1783} = 0.63 \]

### Solution

Obtain a point estimate for the proportion of registered voters nationwide who are in favor of the death penalty for persons convicted of murder.

\[ \hat{p} = \frac{1123}{1783} = 0.63 \]

For a simple random sample of size \( n \), the sampling distribution of \( \hat{p} \) is approximately normal with mean \( \mu_\hat{p} = p \) and standard deviation \( \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} \), provided that \( np(1-p) \geq 10 \).

### Sampling Distribution of \( \hat{p} \)

Note: We also require that each trial be independent when sampling from finite populations.

### Constructing a (1-\( \alpha \)) 100% Confidence Interval for a Population Proportion

Suppose that a simple random sample of size \( n \) is taken from a population. A (1-\( \alpha \)) 100\% confidence interval for \( p \) is given by the following quantities

- **Lower bound:** \( \hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \)
- **Upper bound:** \( \hat{p} + z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \)

### Parallel Example 2: Constructing a Confidence Interval for a Population Proportion

In July of 2008, a Quinnipiac University Poll asked 1783 registered voters nationwide whether they favored or opposed the death penalty for persons convicted of murder. 1123 were in favor.

Obtain a 90% confidence interval for the proportion of registered voters nationwide who are in favor of the death penalty for persons convicted of murder.

\[ \hat{p} = \frac{1123}{1783} = 0.63 \]

- **\( \hat{p} = 0.63 \)**
- **\( n\hat{p}(1-\hat{p}) = 1783(0.63)(1-0.63) = 415.6 \geq 10 \) and the sample size is definitely less than 5\% of the population size**
- **\( \alpha = 0.10 \) so \( z_{0.05} = 1.645 \)**
- **Lower bound:** \( 0.63 - 1.645 \cdot \sqrt{\frac{0.63(1-0.63)}{1783}} = 0.61 \)
- **Upper bound:** \( 0.63 + 1.645 \cdot \sqrt{\frac{0.63(1-0.63)}{1783}} = 0.65 \)

### Solution

We are 90\% confident that the proportion of registered voters who are in favor of the death penalty for those convicted of murder is between 0.61 and 0.65.
Sample size needed for a specified margin of error, $E$, and level of confidence $(1 - \alpha)$:

$$n = \hat{p}(1 - \hat{p})\left(\frac{z_{\alpha/2}}{E}\right)^2$$

**Problem:** The formula uses $\hat{p}$ which depends on $n$, the quantity we are trying to determine!

Two possible solutions:
1. Use an estimate of $p$ based on a pilot study or an earlier study.
2. Let $\hat{p} = 0.5$ which gives the largest possible value of $n$ for a given level of confidence and a given margin of error.

The sample size required to obtain a $(1 - \alpha)$-100% confidence interval for $p$ with a margin of error $E$ is given by

$$n = \hat{p}(1 - \hat{p})\left(\frac{z_{\alpha/2}}{E}\right)^2$$

(rounded up to the next integer), where $\hat{p}$ is a prior estimate of $p$. If a prior estimate of $p$ is unavailable, the sample size required is

$$n = 0.25\left(\frac{z_{\alpha/2}}{E}\right)^2$$

A sociologist wanted to determine the percentage of residents of America that only speak English at home. What size sample should be obtained if she wishes her estimate to be within 3 percentage points with 90% confidence assuming she uses the 2000 estimate obtained from the Census 2000 Supplementary Survey of 82.4%?

**Parallel Example 4: Determining Sample Size**

- $E = 0.03$
- $z_{0.05} = 1.645$
- $\hat{p} = 0.824$
- $n = 0.824(1 - 0.824)\left(1.645^2\right) = 436.04$

We round this value up to 437. The sociologist must survey 437 randomly selected American residents.