Chapter 8: Statistical Inference: Confidence Intervals

Section 8.1
What are Point and Interval Estimates of Population Parameters?
Learning Objectives

1. Point Estimate and Interval Estimate
2. Properties of Point Estimators
3. Confidence Intervals
4. Logic of Confidence Intervals
5. Margin of Error
6. Example
A **point estimate** is a **single number** that is our “best guess” for the parameter.

An **interval estimate** is an **interval of numbers** within which the parameter value is believed to fall.
Learning Objective 1: Point Estimate vs. Interval Estimate

- A *point estimate* doesn’t tell us how close the estimate is likely to be to the parameter.
- An *interval estimate* is more useful:
  - It incorporates a margin of error which helps us to gauge the accuracy of the point estimate.
Learning Objective 2: Properties of Point Estimators

Property 1: A good estimator has a sampling distribution that is centered at the parameter.

An estimator with this property is *unbiased*

- The sample mean is an unbiased estimator of the population mean.
- The sample proportion is an unbiased estimator of the population proportion.
Learning Objective 2: Properties of Point Estimators

Property 2: A good estimator has a *small standard error* compared to other estimators.

This means it tends to fall closer than other estimates to the parameter.

The sample mean has a smaller standard error than the sample median when estimating the population mean of a normal distribution.
Learning Objective 3: Confidence Interval

- A confidence interval is an interval containing the most believable values for a parameter.
- The probability that this method produces an interval that contains the parameter is called the confidence level.
  - This is a number chosen to be close to 1, most commonly 0.95.
Learning Objective 4: Logic of Confidence Intervals

- To construct a confidence interval for a population proportion, start with the **sampling distribution of a sample proportion**
  - Gives the possible values for the sample proportion and their probabilities
  - Is approximately a normal distribution for large random samples by the CLT
  - Has mean equal to the population proportion
  - Has standard deviation called the standard error
Fact: Approximately 95% of a normal distribution falls within 1.96 standard deviations of the mean.

- With probability 0.95, the sample proportion falls within about 1.96 standard errors of the population proportion.
- The distance of 1.96 standard errors is the margin of error in calculating a 95% confidence interval for the population proportion.
Learning Objective 5: Margin of Error

- The *margin of error* measures how accurate the point estimate is likely to be in estimating a parameter.
- It is a multiple of the standard error of the sampling distribution of the estimate when the sampling distribution is a normal distribution.
- The distance of 1.96 standard errors in the margin of error for a 95% confidence interval for a parameter from a normal distribution.
Learning Objective 6:
Example: CI for a Proportion

Example: The GSS asked 1823 respondents whether they agreed with the statement “It is more important for a wife to help her husband’s career than to have one herself”. 19% agreed. Assuming the standard error is 0.01, calculate a 95% confidence interval for the population proportion who agreed with the statement.

- Margin of error = $1.96 \times \text{se} = 1.96 \times 0.01 = 0.02$
- 95% CI = $0.19 \pm 0.02$ or (0.17 to 0.21)

We predict that the population proportion who agreed is somewhere between 0.17 and 0.21.
Chapter 8: Statistical Inference: Confidence Intervals

Section 8.2
How Can We Construct a Confidence Interval to Estimate a Population Proportion?
Learning Objectives

1. Finding the 95% Confidence Interval for a Population Proportion
2. Sample Size Needed for Large-Sample Confidence Interval for a Proportion
3. How Can We Use Confidence Levels Other than 95%?
4. What is the Error Probability for the Confidence Interval Method?
5. Summary
6. Effect of the Sample Size
7. Interpretation of the Confidence Level
Learning Objective 1: Finding the 95% Confidence Interval for a Population Proportion

- We symbolize a population proportion by $p$
- The point estimate of the population proportion is the *sample proportion*
- We symbolize the sample proportion by $\hat{p}$
Learning Objective 1: Finding the 95% Confidence Interval for a Population Proportion

- A 95% confidence interval uses a margin of error = 1.96(standard errors)
- CI = [point estimate ± margin of error] =

\[ \hat{p} \pm 1.96(\text{standard errors}) \]

for a 95% confidence interval
Learning Objective 1: Finding the 95% Confidence Interval for a Population Proportion

- The exact standard error of a sample proportion equals:

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

- This formula depends on the unknown population proportion, \( p \).
- In practice, we don’t know \( p \), and we need to estimate the standard error as

\[ se = \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \]
Learning Objective 1: Finding the 95% Confidence Interval for a Population Proportion

A 95% confidence interval for a population proportion $p$ is:

$$
\hat{p} \pm 1.96(\text{se}), \text{ with } \text{se} = \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}
$$
Learning Objective 1: Example 1

- In 2000, the GSS asked: “Are you willing to pay much higher prices in order to protect the environment?”
  - Of $n = 1154$ respondents, 518 were willing to do so
- Find and interpret a 95% confidence interval for the population proportion of adult Americans willing to do so at the time of the survey
Learning Objective 1: Example 1

\[ \hat{p} = \frac{518}{1154} = 0.45 \]

\[ se = \sqrt{\frac{(0.45)(0.55)}{1154}} = 0.015 \]

\[ \hat{p} \pm 1.96(se) = 1.96(0.015) \]

\[ = 0.45 \pm 0.03 = (0.42, 0.48) \]

TI Calculator

Press Stats,

EDIT
CALC
SHOW
6→2-PropZTest...
7:ZInterval...
8:TInterval...
9:2-SampZInt...
0:2-SampTInt...
1-PropZInt...
B+2-PropZInt...

1-PropZInt
x: 518
n: 1154
C-Level: .95
Calculate

1-PropZInt
(0.42018, 0.47757)
\hat{p} = .4488734835
n = 1154
Learning Objective 2:
Sample Size Needed for Large-Sample Confidence Interval for a Proportion

For the 95% confidence interval for a proportion $p$ to be valid, you should have at least 15 successes and 15 failures:

$$n\hat{p} \geq 15 \text{ and } n(1 - \hat{p}) \geq 15$$
Learning Objective 3: How Can We Use Confidence Levels Other than 95%?

- "95% confidence" means that there's a 95% chance that a sample proportion value occurs such that the confidence interval $\hat{p} \pm 1.96(se)$ contains the unknown value of the population proportion, $p$.

- With probability 0.05, the method produces a confidence interval that misses $p$. 
Learning Objective 3: How Can We Use Confidence Levels Other than 95%?

- In practice, the confidence level 0.95 is the most common choice.
- But, some applications require greater (or less) confidence.
- To increase the chance of a correct inference, we use a larger confidence level, such as 0.99.
Learning Objective 3: How Can We Use Confidence Levels Other than 95%?

- In using confidence intervals, we must compromise between the desired margin of error and the desired confidence of a correct inference.
- As the desired confidence level increases, the margin of error gets larger.
Learning Objective 3: Example 2

- A recent GSS asked “If the wife in a family wants children, but the husband decides that he does not want any children, is it all right for the husband to refuse to have children?
- Of 598 respondents, 366 said yes
- Calculate the 99% confidence interval

[Confidence interval calculation results: \( \hat{p} = 0.61 \), \( n = 598 \), \( 1 - \text{PropZInt} \) yields \( (0.56, 0.66) \)]
Learning Objective 3: Example 3

- Exit poll: Out of 1400 voters, 660 voted for the Democratic candidate.
- Calculate a 95% and a 99% Confidence Interval
Learning Objective 4:
What is the Error Probability for the Confidence Interval Method?

- The general formula for the confidence interval for a population proportion is:

  Sample proportion ± (z-score)(std. error)

  which in symbols is
Learning Objective 5: Summary: Confidence Interval for a Population Proportion, $p$

- A confidence interval for a population proportion $p$ is:

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

- Assumptions
  - Data obtained by randomization
  - A large enough sample size $n$ so that the number of success, $n\hat{p}$, and the number of failures, $n(1-\hat{p})$, are both at least 15
Learning Objective 6: Effects of Confidence Level and Sample Size on Margin of Error

- The *margin of error* for a confidence interval:
  - Increases as the confidence level increases
  - Decreases as the sample size increases
Learning Objective 7: Interpretation of the Confidence Level

- If we used the 95% confidence interval method to estimate many population proportions, then *in the long run about 95% of those intervals would give correct results, containing the population proportion*
Chapter 8: Statistical Inference: Confidence Intervals

Section 8.3
How Can We Construct a Confidence Interval to Estimate a Population Mean?
Learning Objectives

1. How to Construct a Confidence Interval for a Population Mean
2. Properties of the t Distribution
3. Formula for 95% Confidence Interval for a Population Mean
4. How Do We Find a t Confidence Interval for Other Confidence Levels?
5. If the Population is Not Normal, is the Method “Robust”? 
6. The Standard Normal Distribution is the \( t \) Distribution with \( df = \infty \)
Learning Objective 1:
How to Construct a Confidence Interval for a Population Mean

- Point estimate ± margin of error
- The sample mean is the point estimate of the population mean
- The exact standard error of the sample mean is $\sigma/\sqrt{n}$
- In practice, we estimate $\sigma$ by the sample standard deviation, $s$
Learning Objective 1:
How to Construct a Confidence Interval for a Population Mean

- For large n… from any population and also
- For small n from an underlying population that is normal…
- The confidence interval for the population mean is:

\[ \bar{x} \pm z \left( \frac{\sigma}{\sqrt{n}} \right) \]
Learning Objective 1: How to Construct a Confidence Interval for a Population Mean

- In practice, we don’t know the population standard deviation $\sigma$.
- Substituting the sample standard deviation $s$ for $\sigma$ to get $se = s/\sqrt{n}$ introduces extra error.
- To account for this increased error, we replace the z-score by a slightly larger score, the t-score.
Learning Objective 2: Properties of the t Distribution

- The t-distribution is bell shaped and symmetric about 0
- The probabilities depend on the degrees of freedom, \( df = n - 1 \)
- The t-distribution has thicker tails than the standard normal distribution, i.e., it is more spread out
Learning Objective 2: t Distribution

The t-distribution has thicker tails and is more spread out than the standard normal distribution.
Learning Objective 2:

**t Distribution**

<table>
<thead>
<tr>
<th>df</th>
<th>80%</th>
<th>90%</th>
<th>95%</th>
<th>98%</th>
<th>99%</th>
<th>99.8%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.078</td>
<td>6.314</td>
<td>12.706</td>
<td>31.821</td>
<td>63.657</td>
<td>318.3</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1.440</td>
<td>1.943</td>
<td>2.447</td>
<td>3.143</td>
<td>3.707</td>
<td>5.208</td>
</tr>
<tr>
<td>7</td>
<td>1.415</td>
<td>1.895</td>
<td>2.365</td>
<td>2.998</td>
<td>3.499</td>
<td>4.785</td>
</tr>
</tbody>
</table>

In the t table, row for df = 6
Learning Objective 3:
Formula for 95% Confidence Interval for a Population Mean

- When the standard deviation of the population is unknown, a 95% confidence interval for the population mean \( \mu \) is:

\[
\bar{x} \pm t_{0.025} \left( \frac{s}{\sqrt{n}} \right); \quad df = n - 1
\]

- To use this method, you need:
  - Data obtained by randomization
  - An approximately normal population distribution
Learning Objective 3:  
Example: eBay Auctions of Palm Handheld Computers

Do you tend to get a higher, or a lower, price if you give bidders the “buy-it-now” option?
  
- Consider some data from sales of the Palm M515 PDA (personal digital assistant)
  
- During the first week of May 2003, 25 of these handheld computers were auctioned off, 7 of which had the “buy-it-now” option
Learning Objective 3:  
Example: eBay Auctions of Palm Handheld Computers

**Summary of selling prices for the two types of auctions:**

<table>
<thead>
<tr>
<th>buy_now</th>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
<th>Minimum</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>no</td>
<td>18</td>
<td>231.61</td>
<td>21.94</td>
<td>178.0</td>
<td>221.25</td>
<td>240.0</td>
<td>246.75</td>
<td>255.0</td>
</tr>
<tr>
<td>yes</td>
<td>7</td>
<td>233.57</td>
<td>14.64</td>
<td>210.0</td>
<td>225.00</td>
<td>235.0</td>
<td>250.00</td>
<td>250.0</td>
</tr>
</tbody>
</table>

![Dotplot of selling_price vs buy_now](image)
Learning Objective 3: 
Example: eBay Auctions of Palm Handheld Computers

- Let μ denote the population mean for the “buy-it-now” option.
- The estimate of μ is the sample mean: \( \bar{x} = $233.57 \)
- The sample standard deviation: \( s = $14.64 \)
- Table B df=6, with 95% Confidence: \( t = 2.447 \)

\[
\bar{x} \pm t_{0.025} \left( \frac{s}{\sqrt{n}} \right) = 233.57 \pm 2.447 \left( \frac{14.64}{\sqrt{7}} \right)
\]

\[
233.57 \pm 13.54 \text{ or } (220.03, 247.11)
\]
Learning Objective 3: Example: eBay Auctions of Palm Handheld Computers

- The 95% confidence interval for the mean sales price for the bidding only option is:
  \[(220.70, 242.52)\]

- Notice that the two intervals overlap a great deal:
  - “Buy-it-now”: \[(220.03, 247.11)\]
  - Bidding only: \[(220.70, 242.52)\]

- There is not enough information for us to conclude that one probability distribution clearly has a higher mean than the other
Learning Objective 3:  
Example: Small Sample $t$ Confidence Interval

A study of 7 American adults from an SRS yields an average height of 67.2 inches and a standard deviation of 3.9 inches. Assuming the heights are normally distributed, a 95% confidence interval for the average height of all American adults ($\mu$) is:

$$
\bar{x} \pm t^* \frac{s}{\sqrt{n}} = 67.2 \pm 2.447 \frac{3.9}{\sqrt{7}} = 67.2 \pm 3.607
$$

$$
= 63.593 \text{ to } 70.807
$$

“We are 95% confident that the average height of all American adults is between 63.6 and 70.8 inches.”
In a time use study, 20 randomly selected managers spend a mean of 2.4 hours each day on paperwork. The standard deviation of the 20 times is 1.3 hours. Construct the 95% confidence interval for the mean paperwork time of all managers.

95% CI = (1.79 < µ < 3.01)

Note that our calculation assumes that the distribution of times is normally distributed.
Learning Objective 4:
How Do We Find a $t$- Confidence Interval for Other Confidence Levels?

- The 95% confidence interval uses $t_{0.025}$ since 95% of the probability falls between $-t_{0.025}$ and $t_{0.025}$

- For 99% confidence, the error probability is 0.01 with 0.005 in each tail and the appropriate t-score is $t_{0.005}$

- To get other confidence intervals use the appropriate t-value from Table B
Learning Objective 4: How Do We Find a $t$-Confidence Interval for Other Confidence Levels?

**TABLE 7.5: Part of Table B Displaying t-Scores for Large df Values**

The $z$-score of 1.96 is the $t$-score $t_{0.025}$ with right-tail probability of 0.025 and $df = \infty$.

<table>
<thead>
<tr>
<th>Confidence Level</th>
<th>80%</th>
<th>90%</th>
<th>95%</th>
<th>98%</th>
<th>99%</th>
<th>99.8%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right-Tail Probability</td>
<td>t_{0.10}</td>
<td>t_{0.05}</td>
<td>t_{0.025}</td>
<td>t_{0.010}</td>
<td>t_{0.005}</td>
<td>t_{0.001}</td>
</tr>
<tr>
<td>df</td>
<td>1</td>
<td>3.078</td>
<td>6.314</td>
<td>12.706</td>
<td>31.821</td>
<td>63.657</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>1.310</td>
<td>1.697</td>
<td>2.042</td>
<td>2.457</td>
<td>2.750</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>1.303</td>
<td>1.684</td>
<td>2.021</td>
<td>2.423</td>
<td>2.704</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>1.299</td>
<td>1.676</td>
<td>2.009</td>
<td>2.403</td>
<td>2.678</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>1.296</td>
<td>1.671</td>
<td>2.000</td>
<td>2.390</td>
<td>2.660</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>1.292</td>
<td>1.664</td>
<td>1.990</td>
<td>2.374</td>
<td>2.639</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>1.290</td>
<td>1.660</td>
<td>1.984</td>
<td>2.364</td>
<td>2.626</td>
</tr>
<tr>
<td></td>
<td>\infty</td>
<td>1.282</td>
<td>1.645</td>
<td>1.960</td>
<td>2.326</td>
<td>2.576</td>
</tr>
</tbody>
</table>
Learning Objective 5:
If the Population is Not Normal, is the Method “Robust”?

- A basic assumption of the confidence interval using the $t$-distribution is that the population distribution is normal.
- Many variables have distributions that are far from normal.
- We say the $t$-distribution is a *robust method* in terms of the normality assumption.
Learning Objective 5: If the Population is Not Normal, is the Method “Robust”?

- How problematic is it if we use the \( t \)-confidence interval even if the population distribution is not normal?
  - For large random samples, it’s not problematic because of the Central Limit Theorem
- What if \( n \) is small?
  - Confidence intervals using \( t \)-scores usually work quite well except for when extreme outliers are present. The method is robust
Learning Objective 6:
The Standard Normal Distribution is the $t$-Distribution with $df = \infty$

<table>
<thead>
<tr>
<th>Right-Tail Probability</th>
<th>80%</th>
<th>90%</th>
<th>95%</th>
<th>98%</th>
<th>99%</th>
<th>99.8%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$df$</td>
<td>$t_{.100}$</td>
<td>$t_{.050}$</td>
<td>$t_{.025}$</td>
<td>$t_{.010}$</td>
<td>$t_{.005}$</td>
<td>$t_{.001}$</td>
</tr>
<tr>
<td>1</td>
<td>3.078</td>
<td>6.314</td>
<td>12.706</td>
<td>31.821</td>
<td>63.657</td>
<td>318.3</td>
</tr>
<tr>
<td>30</td>
<td>1.310</td>
<td>1.697</td>
<td>2.042</td>
<td>2.457</td>
<td>2.750</td>
<td>3.385</td>
</tr>
<tr>
<td>40</td>
<td>1.303</td>
<td>1.684</td>
<td>2.021</td>
<td>2.423</td>
<td>2.704</td>
<td>3.307</td>
</tr>
<tr>
<td>50</td>
<td>1.299</td>
<td>1.676</td>
<td>2.009</td>
<td>2.403</td>
<td>2.678</td>
<td>3.261</td>
</tr>
<tr>
<td>60</td>
<td>1.296</td>
<td>1.671</td>
<td>2.000</td>
<td>2.390</td>
<td>2.660</td>
<td>3.232</td>
</tr>
<tr>
<td>80</td>
<td>1.292</td>
<td>1.664</td>
<td>1.990</td>
<td>2.374</td>
<td>2.639</td>
<td>3.195</td>
</tr>
<tr>
<td>100</td>
<td>1.290</td>
<td>1.660</td>
<td>1.984</td>
<td>2.364</td>
<td>2.626</td>
<td>3.174</td>
</tr>
<tr>
<td>$\infty$</td>
<td>1.282</td>
<td>1.645</td>
<td>1.960</td>
<td>2.326</td>
<td>2.576</td>
<td>3.090</td>
</tr>
</tbody>
</table>

The $z$-score of 1.96 is the $t$-score $t_{.025}$ with right-tail probability of 0.025 and $df = \infty$. 
Chapter 8: Statistical Inference: Confidence Intervals

Section 8.4
How Do We Choose the Sample Size for a Study?
Learning Objectives

1. Sample Size for Estimating a Population Proportion
2. Sample Size for Estimating a Population Mean
3. What Factors Affect the Choice of the Sample Size?
4. What if You Have to Use a Small n?
5. Confidence Interval for a Proportion with Small Samples
Learning Objective 1: Sample Size for Estimating a Population Proportion

To determine the sample size,

- First, we must decide on the desired margin of error
- Second, we must choose the confidence level for achieving that margin of error
- In practice, 95% confidence intervals are most common
Learning Objective 1: Sample Size for Estimating a Population Proportion

- The random sample size $n$ for which a confidence interval for a population proportion $\hat{p}$ has margin of error $m$ (such as $m = 0.04$) is

$$n = \frac{\hat{p}(1 - \hat{p})z^2}{m^2}$$

- In the formula for determining $n$, setting $\hat{p} = 0.50$ gives the largest value for $n$ out of all the possible values of $\hat{p}$.
Learning Objective 1:  
Example 1: Sample Size For Exit Poll

- A television network plans to predict the outcome of an election between two candidates – Levin and Sanchez
  - A poll one week before the election estimates 58% prefer Levin
- What is the sample size for which a 95% confidence interval for the population proportion has margin of error equal to 0.04?
Learning Objective 1:
Example 1: Sample Size For Exit Poll

- The z-score is based on the confidence level, such as \( z = 1.96 \) for 95% confidence.
- The 95% confidence interval for a population proportion \( p \) is:

\[
\hat{p} \pm 1.96 (se)
\]

- If the sample size is such that \( 1.96 (se) = 0.04 \), then the margin of error will be 0.04

\[
0.04 = 1.96 \sqrt{\hat{p}(1-\hat{p})/n}
\]

solve for \( n \):

\[
n = (1.96)^2 \frac{\hat{p}(1-\hat{p})}{(0.04)^2}
\]
Learning Objective 1:
Example 1: Sample Size For Exit Poll

- Using 0.58 as an estimate for \( p \)
  \[ n = (1.96)^2 (0.58)(0.42)/(0.04)^2 = 584.9 \]
  or \( n = 585 \)

- Without guessing,
  \[ n = (1.96)^2 (0.5)(0.5)/(0.04)^2 = 600.25 \]
  \( n=601 \) gives us a more conservative estimate (always round up)
Learning Objective 1: Example 2

- Suppose a soft drink bottler wants to estimate the proportion of its customers that drink another brand of soft drink on a regular basis.

- What sample size will be required to enable us to have a 99% confidence interval with a margin of error of 1%?

\[
 n = \frac{(0.5)(0.5)(2.58)^2}{(0.01)^2} = 16,641
\]

- Thus, we will need to sample at least 16,641 of the soft drink bottler’s customers.
Learning Objective 1: Example 3

- You want to estimate the proportion of home accident deaths that are caused by falls. How many home accident deaths must you survey in order to be 95% confident that your sample proportion is within 4% of the true population proportion?

- Answer: 601

\[ n = \left( \frac{z^*}{m} \right)^2 p^* (1 - p^*) = \left( \frac{1.96}{0.04} \right)^2 (0.5)(1 - 0.5) = 601 \]
Learning Objective 2: Sample Size for Estimating a Population Mean

The random sample size $n$ for which a confidence interval for a population mean has margin of error approximately equal to $m$ is

$$n = \frac{\sigma^2 z^2}{m^2}$$

where the $z$-score is based on the confidence level, such as $z=1.96$ for 95% confidence.
Learning Objective 2: Sample Size for Estimating a Population Mean

- In practice, you don’t know the value of the standard deviation, $\sigma$.
- You must substitute an educated guess for $\sigma$.
  - Sometimes you can use the sample standard deviation from a similar study.
  - When no prior information is known, a crude estimate that can be used is to divide the estimated range of the data by 6 since for a bell-shaped distribution we expect almost all of the data to fall within 3 standard deviations of the mean.
Learning Objective 2: Example 1

A social scientist plans a study of adult South Africans to investigate educational attainment in the black community.

How large a sample size is needed so that a 95% confidence interval for the mean number of years of education has margin of error equal to 1 year? Assume that the education values will fall within a range of 0 to 18 years.

- Crude estimate of $\sigma = \text{range}/6 = 18/6 = 3$

$$n = \frac{z^2 s^2}{m^2} = \frac{1.96^2 (3)^2}{1^2} = 35$$
Learning Objective 2:
Example 2

Find the sample size necessary to estimate the mean height of all adult males to within .5 in. if we want 99% confidence in our results. From previous studies we estimate $\sigma=2.8$.

$$n = \frac{(2.8)^2 (2.58)^2}{(.5)^2} = 208.7$$

Answer: 209 (always round up)
Learning Objective 3: What Factors Affect the Choice of the Sample Size?

- The first is the desired precision, as measured by the margin of error, $m$.
- The second is the confidence level.
- A third factor is the variability in the data.
  - If subjects have little variation (that is, $\sigma$ is small), we need fewer data than if they have substantial variation.
- A fourth factor is financial.
Chapter 8: Statistical Inference: Confidence Intervals

Section 8.5
How Do Computers Make New Estimation Methods Possible?
Learning Objectives

- The Bootstrap
Learning Objective 1: The Bootstrap: Using Simulation to Construct a Confidence Interval

- When it is difficult to derive a standard error or a confidence interval formula that works well you can use simulation.

- The bootstrap is a simulation method that resamples from the observed data. It treats the data distribution as if it were the population distribution.
Learning Objective 1: The Bootstrap: Using Simulation to Construct a Confidence Interval

- To use the bootstrap method
  - Resample, with replacement, $n$ observations from the data distribution
  - For the new sample of size $n$, construct the point estimate of the parameter of interest
  - Repeat process a very large number of times (e.g., selecting 10,000 separate samples of size $n$ and calculating the 10,000 corresponding parameter estimates)
Learning Objective 1: The Bootstrap: Using Simulation to Construct a Confidence Interval

Example:

- Suppose your data set includes the following:
  
  \[
  160.2, 160.8, 161.4, 162.0, 160.8, 162.0, 162.0, 161.8, 161.6, 161.8. \]

  This data has a mean of 161.44 and standard deviation of 0.63.

- Use the bootstrap method to find a 95% confidence interval for the population standard deviation
Learning Objective 1:  
The Bootstrap: Using Simulation to Construct a Confidence Interval

- Re-sample with replacement from this sample of size 10 and compute the standard deviation of the new sample
- Repeat this process 100,000 times. A histogram showing the distribution of 100,000 samples drawn from this sample is
Learning Objective 1: The Bootstrap Using Simulation to Construct a Confidence Interval

- Now, identify the middle 95% of these 100,000 sample standard deviations (take the 2.5th and 97.5th percentiles).

- For this example, these percentiles are 0.26 and 0.80.

- The 95% bootstrap confidence interval for $\sigma$ is (0.26, 0.80)