Chapter 8

Chapter 8
Confidence Interval Estimation

Chapter Topics

• Estimation Process
• Point Estimates
• Interval Estimates
• Confidence Interval Estimation for the Mean (σ Known)
• Determining Sample Size
• Confidence Interval Estimation for the Mean (σ Unknown)

(continued)

• Confidence Interval Estimation and Sample Size Determination for the Proportion
• Confidence Interval Estimation for Population Total
• Confidence Interval Estimation for Total Difference in the Population
• Estimation and Sample Size Determination for Finite Population (CD-ROM Topic)
• Confidence Interval Estimation and Ethical Issues
Estimation Process

Population  Random Sample

Mean, \( \mu \), is unknown

Sample

\[ \bar{X} = 50 \]

I am 95% confident that \( \mu \) is between 40 & 60.

Point Estimates

<table>
<thead>
<tr>
<th>Estimate Population Parameters …</th>
<th>with Sample Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean ( \mu )</td>
<td>( \bar{X} )</td>
</tr>
<tr>
<td>Proportion ( p )</td>
<td>( P_s )</td>
</tr>
<tr>
<td>Variance ( \sigma^2 )</td>
<td>( S^2 )</td>
</tr>
<tr>
<td>Difference ( \mu_1 - \mu_2 )</td>
<td>( \bar{X}_1 - \bar{X}_2 )</td>
</tr>
</tbody>
</table>

Interval Estimates

- Provide Range of Values
- Take into consideration variation in sample statistics from sample to sample
- Based on observation from 1 sample
- Give information about closeness to unknown population parameters
- Stated in terms of level of confidence
- Never 100% sure
Confidence Interval Estimates

Confidence Intervals

- Mean
- Proportion

- \( \sigma \) Known
- \( \sigma \) Unknown

Confidence Interval for \( \mu \) (\( \sigma \) Known)

Assumptions
- Population standard deviation is known
- Population is normally distributed
- If population is not normal, use large sample

Confidence Interval Estimate

\[
X - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq X + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}
\]

- \( e = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \) is called the sampling error or margin of error

Elements of Confidence Interval Estimation

- Level of Confidence
  - Confidence that the interval will contain the unknown population parameter
- Precision (Range)
  - Closeness to the unknown parameter
- Cost
  - Cost required to obtain a sample of size \( n \)
Level of Confidence

- Denoted by $100(1-\alpha)\%$
- A Relative Frequency Interpretation
  - In the long run, $100(1-\alpha)\%$ of all the confidence intervals that can be constructed will contain (bracket) the unknown parameter
- A Specific Interval Will Either Contain or Not Contain the Parameter

Interval and Level of Confidence

Confidence Intervals

$\mu - Z_{\alpha/2} \sigma_{\bar{x}}$ to $\mu + Z_{\alpha/2} \sigma_{\bar{x}}$

$\bar{x}$ is the sample mean

Interval and Level of Confidence

Example

Confidence Interval Estimate for the Mean

<table>
<thead>
<tr>
<th>Population Standard Deviation</th>
<th>Sample Mean</th>
<th>Sample Size</th>
<th>Confidence Level</th>
<th>Standard Error of the Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>195000</td>
<td>215000</td>
<td>15</td>
<td>99%</td>
<td>50348.7835</td>
</tr>
</tbody>
</table>

Z Value $= -2.57583451$

Interval Half Width $= 129690.1343$

Interval Lower Limit $= 85309.86569$

Interval Upper Limit $= 344690.1343$

PHStat output

85309.9 344690.1

$\mu$ <<

The 99% CI for the population mean:

85309.9 < $\mu$ < 344690.1
Example: Interpretation

If all possible samples of size 15 are taken and the corresponding 99% confidence intervals are constructed, 99% of the confidence intervals that are constructed will contain the true unknown population mean.

We are 99% confident that the population average number of shares traded on the NASDAQ is between 85309.9 and 344690.1.

Using the confidence interval method on repeated sampling, the probability that we will have constructed a confidence interval that will contain the unknown population mean is 99%.

Obtaining Confidence Interval in PHStat

- PHStat | Confidence Interval | Estimates for the Mean, Sigma Known

Factors Affecting Interval Width (Precision)

- Data Variation
  - Measured by $\sigma$
- Sample Size
  - $\sigma_{X} = \frac{\sigma}{\sqrt{n}}$
- Level of Confidence
  - $100(1 - \alpha)\%$

Intervals Extend from $\bar{X} - Z\sigma_{X}$ to $\bar{X} + Z\sigma_{X}$
Determining Sample Size (Cost)

**Too Big:**
- Requires more resources

**Too small:**
- Won’t do the job

---

**Determining Sample Size for Mean**

What sample size is needed to be 90% confident of being correct within ± 5? A pilot study suggested that the standard deviation is 45.

\[
 n = \frac{Z^2 \sigma^2}{\text{Error}^2} = \frac{1.645^2 (45^2)}{5^2} = 219.2 \approx 220
\]

Round Up

---

**Determining Sample Size for Mean in PHStat**

- PHStat | Sample Size | Determination for the Mean ...
- Example in Excel Spreadsheet

<table>
<thead>
<tr>
<th>Sample Size Determination</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data</strong></td>
</tr>
<tr>
<td>Population Standard Deviation</td>
</tr>
<tr>
<td>Sampling Error</td>
</tr>
<tr>
<td>Confidence Level</td>
</tr>
<tr>
<td><strong>Intermediate Calculations</strong></td>
</tr>
<tr>
<td>Z Value</td>
</tr>
<tr>
<td>Calculated Sample Size</td>
</tr>
<tr>
<td><strong>Result</strong></td>
</tr>
<tr>
<td>Sample Size Needed</td>
</tr>
</tbody>
</table>
Assumptions
- Population standard deviation is unknown
- Population is normally distributed
- If population is not normal, use large sample
- Use Student’s t Distribution

Confidence Interval Estimate
\[
\bar{X} - t_{\alpha/2, n-1} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\alpha/2, n-1} \frac{S}{\sqrt{n}}
\]

Student’s t Distribution
- Standardized Normal
- Bell-Shaped Symmetric
- ‘Fatter’ Tails

Student’s t Table

<table>
<thead>
<tr>
<th>df</th>
<th>.25</th>
<th>.10</th>
<th>.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.000</td>
<td>3.078</td>
<td>6.314</td>
</tr>
<tr>
<td>2</td>
<td>2.776</td>
<td>2.353</td>
<td>2.920</td>
</tr>
<tr>
<td>3</td>
<td>2.998</td>
<td>2.306</td>
<td>2.353</td>
</tr>
</tbody>
</table>

Let: 
- \( n = 3 \)
- \( df = n - 1 = 2 \)
- \( \alpha = .10 \)
- \( \alpha/2 = .05 \)
Example
A random sample of $n = 25$ has $\bar{X} = 50$ and $S = 8$.
Set up a 95% confidence interval estimate for $\mu$.

$$\bar{X} - t_{\alpha/2, n-1} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\alpha/2, n-1} \frac{S}{\sqrt{n}}$$

$$50 - 2.0639 \frac{8}{\sqrt{25}} \leq \mu \leq 50 + 2.0639 \frac{8}{\sqrt{25}}$$

$$46.69 \leq \mu \leq 53.30$$

We are 95% confident that the unknown true population mean is somewhere between 46.69 and 53.30.

Confidence Interval for $\mu$ (unknown) in PHStat

- PHStat | Confidence Interval | Estimate for the Mean, Sigma Unknown
- Example in Excel Spreadsheet

Microsoft Excel Worksheet

<table>
<thead>
<tr>
<th>Data</th>
<th>Sample Standard Deviation</th>
<th>Sample Mean</th>
<th>Sample Size</th>
<th>Confidence Level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>25</td>
<td>95%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Degrees of Freedom</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T Value</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.063898137</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Interval Half Width</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.302237019</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Interval Lower Limit</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>46.70</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Interval Upper Limit</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>53.30</td>
</tr>
</tbody>
</table>

Intermediate Calculations

Confidence Interval Estimate for Proportion

- Assumptions
  - Two categorical outcomes
  - Population follows binomial distribution
  - Normal approximation can be used if $np \geq 5$ and $n(1-p) \geq 5$
- Confidence Interval Estimate
  - $p_\hat{} \pm Z_\alpha : \sqrt{\frac{p_\hat{}(1-p_\hat{})}{n}} \leq p \leq p_\hat{} + Z_\alpha : \sqrt{\frac{p_\hat{}(1-p_\hat{})}{n}}$
Example

A random sample of 400 voters showed that 32 preferred Candidate A. Set up a 95% confidence interval estimate for \( p \).

\[
\hat{p} - Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}
\]

\[
0.08 - 1.96 \sqrt{\frac{0.08(1-0.08)}{400}} \leq p \leq 0.08 + 1.96 \sqrt{\frac{0.08(1-0.08)}{400}}
\]

\[
0.053 \leq p \leq 0.107
\]

We are 95% confident that the proportion of voters who prefer Candidate A is somewhere between 0.053 and 0.107.

Confidence Interval Estimate for Proportion in PHStat

- PHStat | Confidence Interval | Estimate for the Proportion ...
- Example in Excel Spreadsheet

Microsoft Excel

Worksheet

Confidence Interval Estimate for the Mean

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Successes</td>
<td>32</td>
</tr>
<tr>
<td>Confidence Level</td>
<td>95%</td>
</tr>
<tr>
<td>Hypothesis Calculation</td>
<td></td>
</tr>
<tr>
<td>Sample Proportion</td>
<td>0.08</td>
</tr>
<tr>
<td>Error</td>
<td>0.05</td>
</tr>
<tr>
<td>Standard Error of the Proportion</td>
<td>0.01356466</td>
</tr>
<tr>
<td>Interval Lower Limit</td>
<td>0.053413794</td>
</tr>
<tr>
<td>Interval Upper Limit</td>
<td>0.106586206</td>
</tr>
<tr>
<td>Confidence Interval</td>
<td>95%</td>
</tr>
</tbody>
</table>

Determining Sample Size for Proportion

What sample size is needed to be within ±5% with 90% confidence if past studies show about 30% are defective?

\[
n = \frac{Z^2 \hat{p}(1-\hat{p})}{\text{Error}^2} = \frac{1.645^2(0.3)(0.7)}{0.05^2} = 227.3 \approx 228
\]

Round Up
Determining Sample Size for Proportion in PHStat

- PHStat | Sample Size | Determination for the Proportion …
- Example in Excel Spreadsheet

Microsoft Excel Worksheet

<table>
<thead>
<tr>
<th>Sample Size Determination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Intermediate Calculations

- Z Value: 1.644853
- Calculated Sample Size: 227.265477
- Result: Sample Size Needed: 228

Confidence Interval for Population Total Amount

- Point Estimate
  - $N \bar{X}$
- Confidence Interval Estimate
  
  $N \bar{X} \pm N \left( t_{\alpha/2, n-1} \right) \frac{S}{\sqrt{n}} \sqrt{\frac{(N-n)}{(N-1)}}$

Confidence Interval for Population Total: Example

An auditor is faced with a population of 1000 vouchers and wishes to estimate the total value of the population of vouchers. A sample of 50 vouchers is selected with the average voucher amount of $1076.39, standard deviation of $273.62. Set up the 95% confidence interval estimate of the total amount for the population of vouchers.
Example Solution

\[ N = 1000 \quad n = 50 \quad \bar{X} = $1076.39 \quad S = $273.62 \]

\[ N\bar{X} \pm N \left( t_{\alpha/2, n-1} \right) \frac{S}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \]

\[ = (1000)(1076.39) \pm (1000)(2.0096) \frac{273.62}{\sqrt{50}} \sqrt{\frac{1000-50}{1000-1}} \]

\[ = 1,076,390 \pm 75,830.85 \]

The 95% confidence interval for the population total amount of the vouchers is between $1,000,559.15 and $1,152,220.85.

Example Solution in PHStat

- PHStat | Confidence Intervals | Estimate for the Population Total
- Excel Spreadsheet for the Voucher Example

Confidence Interval for Total Difference in the Population

- Point Estimate
  \[ \bar{D} = \frac{\sum D_i}{n} \]

- Confidence Interval Estimate
  \[ N\bar{D} \pm N \left( t_{\alpha/2, n-1} \right) \frac{S_D}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \]

- where
  \[ S_D = \sqrt{\frac{\sum (D_i - \bar{D})^2}{n-1}} \]
Estimation for Finite Population (CD-ROM Topic)

- Samples are Selected Without Replacement
  - Confidence interval for the mean ($\sigma$ unknown)
    \[ \bar{x} \pm t_{\alpha/2,n-1} \frac{s}{\sqrt{n}} \frac{(N-n)}{(N-1)} \]
  - Confidence interval for proportion
    \[ p_\hat{x} \pm Z_{\alpha/2} \sqrt{\frac{p_\hat{x}(1-p_\hat{x})}{n}} \frac{(N-n)}{(N-1)} \]

Sample Size ($n$) Determination for Finite Population (CD-ROM Topic)

- Samples are Selected Without Replacement
  - When estimating the mean
    \[ n_0 = \frac{Z^2 \sigma^2}{e^2} \]
  - When estimating the proportion
    \[ n_0 = \frac{Z^2 p(1-p)}{e^2} \]

Ethical Issues

- Confidence Interval (Reflects Sampling Error) Should Always Be Reported Along with the Point Estimate
- The Level of Confidence Should Always Be Reported
- The Sample Size Should Be Reported
- An Interpretation of the Confidence Interval Estimate Should Also Be Provided
Chapter Summary

- Illustrated Estimation Process
- Discussed Point Estimates
- Addressed Interval Estimates
- Discussed Confidence Interval Estimation for the Mean (σ known)
- Addressed Determining Sample Size
- Discussed Confidence Interval Estimation for the Mean (σ unknown)

Chapter Summary (continued)

- Discussed Confidence Interval Estimation for the Proportion
- Addressed Confidence Interval Estimation for Population Total
- Discussed Confidence Interval Estimation for Total Difference in the Population
- Addressed Estimation and Sample Size Determination for Finite Population (CD-ROM Topic)
- Addressed Confidence Interval Estimation and Ethical Issues