Chapter 8
Confidence Interval Estimation

Learning Objectives

In this chapter, you learn:
- To construct and interpret confidence interval estimates for the mean and the proportion
- How to determine the sample size necessary to develop a confidence interval for the mean or proportion
- How to use confidence interval estimates in auditing

Chapter Outline

Content of this chapter
- Confidence Intervals for the Population Mean, \( \mu \)
  - when Population Standard Deviation \( \sigma \) is Known
  - when Population Standard Deviation \( \sigma \) is Unknown
- Confidence Intervals for the Population Proportion, \( \pi \)
- Determining the Required Sample Size

Point and Interval Estimates

- A point estimate is a single number,
- a confidence interval provides additional information about the variability of the estimate

Point Estimates

We can estimate a Population Parameter … with a Sample Statistic (a Point Estimate)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
<th>Sample Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>( \mu )</td>
<td>( \bar{X} )</td>
</tr>
<tr>
<td>Proportion</td>
<td>( \pi )</td>
<td>( p )</td>
</tr>
</tbody>
</table>

Confidence Intervals

- How much uncertainty is associated with a point estimate of a population parameter?
- An interval estimate provides more information about a population characteristic than does a point estimate
- Such interval estimates are called confidence intervals
Confidence Interval Estimate

- An interval gives a range of values:
  - Takes into consideration variation in sample statistics from sample to sample
  - Based on observations from 1 sample
  - Gives information about closeness to unknown population parameters
  - Stated in terms of level of confidence
  - e.g. 95% confident, 99% confident
  - Can never be 100% confident

Confidence Interval Example

Cereal fill example
- Population has \( \mu = 368 \) and \( \sigma = 15 \).
- If you take a sample of size \( n = 25 \) you know
  - \( 368 \pm 1.96 \times \frac{15}{\sqrt{25}} = (362.12, 373.88) \) contains 95% of the sample means
- When you don’t know \( \mu \), you use \( \bar{X} \) to estimate \( \mu \)
  - If \( \bar{X} = 362.3 \) the interval is \( 362.3 \pm 1.96 \times \frac{15}{\sqrt{25}} = (356.42, 368.18) \)
  - Since 356.42 \( \leq \mu \leq 368.18 \) the interval based on this sample makes a correct statement about \( \mu \).

- But what about the intervals from other possible samples of size 25?

<table>
<thead>
<tr>
<th>Sample #</th>
<th>( \bar{X} )</th>
<th>Lower Limit</th>
<th>Upper Limit</th>
<th>Contain ( \mu )?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>362.30</td>
<td>356.42</td>
<td>368.18</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>369.50</td>
<td>363.62</td>
<td>375.38</td>
<td>Yes</td>
</tr>
<tr>
<td>3</td>
<td>360.00</td>
<td>354.12</td>
<td>365.88</td>
<td>No</td>
</tr>
<tr>
<td>4</td>
<td>362.12</td>
<td>356.24</td>
<td>368.00</td>
<td>Yes</td>
</tr>
<tr>
<td>5</td>
<td>373.88</td>
<td>368.00</td>
<td>379.76</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Confidence Interval Example (continued)

- In practice you only take one sample of size \( n \)
- In practice you do not know \( \mu \) so you do not know if the interval actually contains \( \mu \)
- However you do know that 95% of the intervals formed in this manner will contain \( \mu \)
- Thus, based on the one sample, you actually selected you can be 95% confident your interval will contain \( \mu \) (this is a 95% confidence interval)

Note: 95% confidence is based on the fact that we used \( Z = 1.96 \).

Estimation Process

- The general formula for all confidence intervals is:

\[
\text{Point Estimate} \pm (\text{Critical Value})(\text{Standard Error})
\]

Where:
- \( \text{Point Estimate} \) is the sample statistic estimating the population parameter of interest
- \( \text{Critical Value} \) is a table value based on the sampling distribution of the point estimate and the desired confidence level
- \( \text{Standard Error} \) is the standard deviation of the point estimate
Confidence Level

- Confidence Level
- Confidence the interval will contain the unknown population parameter
- A percentage (less than 100%)

Confidence Level, (1-\(\alpha\))

- Suppose confidence level = 95%
- Also written \((1 - \alpha) = 0.95\), (so \(\alpha = 0.05\))
- A relative frequency interpretation:
  - 95% of all the confidence intervals that can be constructed will contain the unknown true parameter
  - A specific interval either will contain or will not contain the true parameter
  - No probability involved in a specific interval

Confidence Intervals

- Assumptions
  - Population standard deviation \(\sigma\) is known
  - Population is normally distributed
  - If population is not normal, use large sample
- Confidence interval estimate:
  \[
  \bar{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}
  \]
  where \(\bar{X}\) is the point estimate
  \(Z_{\alpha/2}\) is the normal distribution critical value for a probability of \(\alpha/2\) in each tail
  \(\sigma/\sqrt{n}\) is the standard error

Finding the Critical Value, \(Z_{\alpha/2}\)

- Consider a 95% confidence interval:
  \[
  Z_{\alpha/2} = \pm 1.96
  \]

Common Levels of Confidence

- Commonly used confidence levels are 90%, 95%, and 99%

<table>
<thead>
<tr>
<th>Confidence Level</th>
<th>Confidence Coefficient, (1 - \alpha)</th>
<th>(Z_{\alpha/2}) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>80%</td>
<td>0.80</td>
<td>1.28</td>
</tr>
<tr>
<td>90%</td>
<td>0.90</td>
<td>1.645</td>
</tr>
<tr>
<td>95%</td>
<td>0.95</td>
<td>1.96</td>
</tr>
<tr>
<td>98%</td>
<td>0.98</td>
<td>2.33</td>
</tr>
<tr>
<td>99%</td>
<td>0.99</td>
<td>2.58</td>
</tr>
<tr>
<td>99.9%</td>
<td>0.998</td>
<td>3.08</td>
</tr>
<tr>
<td>99.99%</td>
<td>0.999</td>
<td>3.27</td>
</tr>
</tbody>
</table>
Intervals and Level of Confidence

Sampling Distribution of the Mean

Intervals extend from

\[ \bar{x} - \frac{Z_{\alpha/2} \sigma}{\sqrt{n}} \] to \[ \bar{x} + \frac{Z_{\alpha/2} \sigma}{\sqrt{n}} \]

Confidence Intervals

\((1-\alpha)\times100\% \) of intervals constructed contain \( \mu \); \((\alpha)\times100\% \) do not.

Example

- A sample of 11 circuits from a large normal population has a mean resistance of 2.20 ohms. We know from past testing that the population standard deviation is 0.35 ohms.
- Determine a 95% confidence interval for the true mean resistance of the population.

Solution:

\[ \bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \]

\[ = 2.20 \pm 1.96(0.35/\sqrt{11}) \]

\[ = 2.20 \pm 0.2068 \]

\[ 1.9932 \leq \mu \leq 2.4068 \]

Interpretation

- We are 95% confident that the true mean resistance is between 1.9932 and 2.4068 ohms.
- Although the true mean may or may not be in this interval, 95% of intervals formed in this manner will contain the true mean.

Confidence Intervals

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

Do You Ever Truly Know \( \sigma \)?

- Probably not!
- In virtually all real world business situations, \( \sigma \) is not known.
- If there is a situation where \( \sigma \) is known then \( \mu \) is also known (since to calculate \( \sigma \) you need to know \( \mu \).)
- If you truly know \( \mu \) there would be no need to gather a sample to estimate it.
Confidence Interval for $\mu$ (\(\sigma\) Unknown)

- If the population standard deviation \(\sigma\) is unknown, we can substitute the sample standard deviation, \(S\).
- This introduces extra uncertainty, since \(S\) is variable from sample to sample.
- So we use the t distribution instead of the normal distribution.

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:

\[
\frac{\bar{X} \pm \frac{t_{\alpha/2}}{\sqrt{n}} S}{\sqrt{n}}
\]

(\(t_{\alpha/2}\) is the critical value of the t distribution with \(n - 1\) degrees of freedom and an area of \(\alpha/2\) in each tail)

Student's t Distribution

- The t is a family of distributions.
- The \(t_{\alpha/2}\) value depends on degrees of freedom (d.f.).
  - Number of observations that are free to vary after sample mean has been calculated.

\[
d.f. = n - 1
\]

Degrees of Freedom (df)

- Idea: Number of observations that are free to vary after sample mean has been calculated.

Example: Suppose the mean of 3 numbers is 8.0

Let \(X_1 = 7\)
Let \(X_2 = 8\)
What is \(X_3\)?

If the mean of these three values is 8.0, then \(X_3\) must be 9 (i.e., \(X_3\) is not free to vary)

Here, \(n = 3\), so degrees of freedom = \(n - 1 = 3 - 1 = 2\)

(2 values can be any numbers, but the third is not free to vary for a given mean)

Student's t Table

<table>
<thead>
<tr>
<th>df</th>
<th>Upper Tail Area</th>
<th>df</th>
<th>Upper Tail Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>.25</td>
<td>.10</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.000</td>
<td>3.078</td>
<td>6.314</td>
</tr>
<tr>
<td>2</td>
<td>.817</td>
<td>1.886</td>
<td>2.920</td>
</tr>
<tr>
<td>3</td>
<td>.765</td>
<td>1.638</td>
<td>2.353</td>
</tr>
<tr>
<td>Let: (n = 3) (\alpha = 0.10) (\alpha/2 = 0.05)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Student's t Table

- \(t\) distributions are bell-shaped and symmetric, but have 'fatter' tails than the normal.
- As \(n\) increases, \(t\) distribution approaches normal.

The body of the table contains \(t\) values, not probabilities.

\(\alpha/2 = 0.05\)

\(\alpha = 0.10\)

\(t = 2.920\)

\(n = 3\)
### Basic Business Statistics, 10/e
© 2006 Prentice Hall, Inc.

**Selected t distribution values**

<table>
<thead>
<tr>
<th>Confidence Level</th>
<th>t (10 d.f.)</th>
<th>t (20 d.f.)</th>
<th>t (30 d.f.)</th>
<th>t (∞ d.f.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.80</td>
<td>1.372</td>
<td>1.325</td>
<td>1.310</td>
<td>1.28</td>
</tr>
<tr>
<td>0.90</td>
<td>1.812</td>
<td>1.725</td>
<td>1.697</td>
<td>1.645</td>
</tr>
<tr>
<td>0.95</td>
<td>2.228</td>
<td>2.086</td>
<td>2.042</td>
<td>1.96</td>
</tr>
<tr>
<td>0.99</td>
<td>3.169</td>
<td>2.845</td>
<td>2.750</td>
<td>2.58</td>
</tr>
</tbody>
</table>

Note: t → Z as n increases

### Example of t distribution confidence interval

A random sample of n = 25 has X = 50 and S = 8. Form a 95% confidence interval for μ.

- d.f. = n – 1 = 24, so \( t_{0.025} = 2.0639 \)

The confidence interval is:

\[
\bar{X} \pm t_{a/2} \frac{S}{\sqrt{n}} = 50 \pm \frac{2.0639 \times 8}{\sqrt{25}}
\]

**46.698 \leq \mu \leq 53.302**

### Confidence Intervals for the Population Proportion, \( \pi \)

- An interval estimate for the population proportion ( \( \pi \) ) can be calculated by adding an allowance for uncertainty to the sample proportion ( \( p \) )

### Confidence Intervals for the Population Proportion, \( \pi \)

- Recall that the distribution of the sample proportion is approximately normal if the sample size is large, with standard deviation

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

- We will estimate this with sample data:
Confidence Interval Endpoints

- Upper and lower confidence limits for the population proportion are calculated with the formula:
  \[ p \pm Z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}} \]

  where:
  - \( Z_{\alpha/2} \) is the standard normal value for the level of confidence desired
  - \( p \) is the sample proportion
  - \( n \) is the sample size
  - Note: must have np > 5 and n(1-p) > 5

Example

- A random sample of 100 people shows that 25 are left-handed. Form a 95% confidence interval for the true proportion of left-handers.

  \[ p \pm Z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}} \]
  \[ = 25/100 \pm 1.96 \sqrt{0.25(0.75)/100} \]
  \[ = 0.25 \pm 1.96(0.0433) \]
  \[ = 0.1651 \leq \pi \leq 0.3349 \]

  Interpretation:
  - We are 95% confident that the true percentage of left-handers in the population is between 16.51% and 33.49%.
  - Although the interval from 0.1651 to 0.3349 may or may not contain the true proportion, 95% of intervals formed from samples of size 100 in this manner will contain the true proportion.

Determining Sample Size

- The required sample size can be found to reach a desired margin of error \( e \) with a specified level of confidence \( (1 - \alpha) \).

  - The margin of error is also called sampling error
    - the amount of imprecision in the estimate of the population parameter
    - the amount added and subtracted to the point estimate to form the confidence interval.
Determining Sample Size

For the Mean

\[
X \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = e = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}
\]

Sampling error (margin of error)

To determine the required sample size for the mean, you must know:

- The desired level of confidence \((1 - \alpha)\), which determines the critical value, \(Z_{\alpha/2}\)
- The acceptable sampling error, \(e\)
- The standard deviation, \(\sigma\)

Now solve for \(n\) to get

\[
n = \frac{Z_{\alpha/2}^2 \sigma^2}{e^2}
\]

If \(\sigma\) is unknown

If unknown, \(\sigma\) can be estimated when using the required sample size formula

- Use a value for \(\sigma\) that is expected to be at least as large as the true \(\sigma\)
- Select a pilot sample and estimate \(\sigma\) with the sample standard deviation, \(S\)

Required Sample Size Example

If \(\sigma = 45\), what sample size is needed to estimate the mean within \(\pm 5\) with 90% confidence?

\[
n = \frac{Z^2 \sigma^2}{e^2} = \frac{(1.645)^2 (45)^2}{5^2} = 219.19
\]

So the required sample size is \(n = 220\) (Always round up)

If \(\sigma\) is unknown

- If unknown, \(\sigma\) can be estimated when using the required sample size formula
  - Use a value for \(\sigma\) that is expected to be at least as large as the true \(\sigma\)
  - Select a pilot sample and estimate \(\sigma\) with the sample standard deviation, \(S\)
Determining Sample Size

To determine the required sample size for the proportion, you must know:

- The desired level of confidence (1 - \( \alpha \)), which determines the critical value, \( Z_{\alpha/2} \)
- The acceptable sampling error, \( e \)
- The true proportion of events of interest, \( \pi \)
  - \( \pi \) can be estimated with a pilot sample if necessary (or conservatively use 0.5 as an estimate of \( \pi \))

Required Sample Size Example

How large a sample would be necessary to estimate the true proportion defective in a large population within ±3% with 95% confidence?

(Assume a pilot sample yields \( p = 0.12 \))

Solution:

For 95% confidence, use \( Z_{\alpha/2} = 1.96 \)
- \( e = 0.03 \)
- \( p = 0.12 \), so use this to estimate \( \pi \)

\[
\begin{align*}
  n &= \frac{Z_{\alpha/2}^2 \pi(1 - \pi)}{e^2} \\
  &= \frac{(1.96)^2 (0.12)(1-0.12)}{(0.03)^2} \\
  &= 450.74
\end{align*}
\]

So use \( n = 451 \)

Applications in Auditing

- Six advantages of statistical sampling in auditing
  - Sampling is less time consuming and less costly
  - Sampling provides an objective way to calculate the sample size in advance
  - Sampling provides results that are objective and defensible.
    - Because the sample size is based on demonstrable statistical principles, the audit is defensible before one’s superiors and in a court of law.

- Sampling provides an estimate of the sampling error
  - Allows auditors to generalize their findings to the population with a known sampling error.
  - Can provide more accurate conclusions about the population

Confidence Interval for Population Total Amount

- Point estimate for a population of size \( N \):
  \[
  \text{Population total} = \hat{N} \overline{X}
  \]

- Confidence interval estimate:
  \[
  \overline{X} \pm N \left( t_{\alpha/2} / \sqrt{n} \right) \frac{S}{N-n} \sqrt{\frac{N-n}{N-1}}
  \]

(This is sampling without replacement, so use the finite population correction in the confidence interval formula)
**Confidence Interval for Population Total: Example**

A firm has a population of 1000 accounts and wishes to estimate the total population value. A sample of 80 accounts is selected with average balance of $87.6 and standard deviation of $22.3.

Find the 95% confidence interval estimate of the total balance.

**Example Solution**

The 95% confidence interval for the population total balance is $82,837.52 to $92,362.48.

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

$$= (1000)(87.6) \pm (1000)(1.990)(\frac{22.3}{\sqrt{80}}) \sqrt{\frac{1000-80}{1000-1}}$$

$$= 87600 \pm 4762.48$$

The 95% confidence interval for the population total balance is $82,837.52 to $92,362.48.

**Confidence Interval for Total Difference**

- **Point estimate for a population of size N:**
  $$\text{Total Difference} = \overline{D}$$
- **Where the average difference, \( D \), is:**
  $$\overline{D} = \frac{\sum D_i}{n}$$
  where \( D_i \) = audited value - original value

**Confidence Interval for Total Difference (continued)**

- **Confidence interval estimate:**
  $$\overline{N\overline{D}} \pm N\left(t_{\alpha/2}\right) \frac{SD}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$
  where
  $$SD = \sqrt{\frac{\sum (D_i - \overline{D})^2}{n-1}}$$

**One-Sided Confidence Intervals**

- **Application:** find the upper bound for the proportion of items that do not conform with internal controls
  $$\text{Upperbound} = p + Z_{\alpha} \sqrt{\frac{p(1-p)}{n}} \sqrt{\frac{N-n}{N-1}}$$
  where
  - \( Z_{\alpha} \) is the standard normal value for the level of confidence desired
  - \( p \) is the sample proportion of items that do not conform
  - \( n \) is the sample size
  - \( N \) is the population size

**Ethical Issues**

- A confidence interval estimate (reflecting sampling error) should always be included when reporting a 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

- Introduced the concept of confidence intervals
- Discussed point estimates
- Developed confidence interval estimates
- Created confidence interval estimates for the mean (σ known)
- Determined confidence interval estimates for the mean (σ unknown)
- Created confidence interval estimates for the proportion
- Determined required sample size for mean and proportion settings

(continued)

- Developed applications of confidence interval estimation in auditing
- Confidence interval estimation for population total
- Confidence interval estimation for total difference in the population
- One-sided confidence intervals for the proportion nonconforming
- Addressed confidence interval estimation and ethical issues