Statistics
for
Economics & Business

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 estimate for the mean or proportion
- How to use confidence interval estimates in auditing
Chapter Outline

- 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

```
<table>
<thead>
<tr>
<th>Lower Confidence Limit</th>
<th>Point Estimate</th>
<th>Upper Confidence Limit</th>
</tr>
</thead>
</table>
```

Width of confidence interval
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>

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?
Confidence Interval Example

<table>
<thead>
<tr>
<th>Sample #</th>
<th>X</th>
<th>Lower Limit</th>
<th>Upper Limit</th>
<th>Contain µ?</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

Population (mean, \( \mu \), is unknown)

Random Sample

Mean \( \bar{X} = 50 \)

Sample

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

General Formula

- The general formula for all confidence intervals is:

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

Where:

- **Point Estimate** is the sample statistic estimating the population parameter of interest
- **Critical Value** is a table value based on the sampling distribution of the point estimate and the desired confidence level
- **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%)

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

(continued)
Confidence Intervals

Confidence Intervals

Population Mean

σ Known

σ Unknown

Population Proportion

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

Assumptions
- Population standard deviation σ 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:

$Z_{\alpha/2} = \pm 1.96$

1 - $\alpha = 0.95$ so $\alpha = 0.05$

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>99.8%</td>
<td>0.998</td>
<td>3.08</td>
</tr>
<tr>
<td>99.9%</td>
<td>0.999</td>
<td>3.27</td>
</tr>
</tbody>
</table>
Intervals and Level of Confidence

Confidence Intervals

Intervals extend from
\[ \bar{X} - Z \frac{\alpha}{2} \frac{\sigma}{\sqrt{n}} \]

to
\[ \bar{X} + Z \frac{\alpha}{2} \frac{\sigma}{\sqrt{n}} \]

(1-\(\alpha\))x100% of intervals constructed contain \(\mu\); 
(\(\alpha\))x100% 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.
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.

Solution:

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

\[ = 2.20 \pm 1.96 \left( \frac{0.35}{\sqrt{11}} \right) \]

\[ = 2.20 \pm 0.2068 \]

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

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

- Population Mean
- Population Proportion
  - $\sigma$ 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.
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:
$$\bar{X} \pm t_{\alpha/2} \frac{S}{\sqrt{n}}$$

(where $t_{\alpha/2}$ 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 Distribution

Note: $t \rightarrow Z$ as $n$ increases

$t$-distributions are bell-shaped and symmetric, but have ‘fatter’ tails than the normal

Standard Normal (t with $df = \infty$)

$t$ $(df = 5)$

$t$ $(df = 13)$

Student’s t Table

<table>
<thead>
<tr>
<th>df</th>
<th>.10</th>
<th>.05</th>
<th>.025</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.078</td>
<td>6.314</td>
<td>12.706</td>
</tr>
<tr>
<td>2</td>
<td>1.886</td>
<td><strong>2.920</strong></td>
<td>4.303</td>
</tr>
<tr>
<td>3</td>
<td>1.638</td>
<td>2.353</td>
<td>3.182</td>
</tr>
</tbody>
</table>

Let: $n = 3$
$df = n - 1 = 2$
$\alpha = 0.10$
$\alpha/2 = 0.05$

The body of the table contains $t$ values, not probabilities
Selected t distribution values

With comparison to the Z value

<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>Z (∞ 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: $t \rightarrow Z$ as $n$ increases

Example of t distribution confidence interval

A random sample of $n = 25$ has $\bar{X} = 50$ and $S = 8$. Form a 95% confidence interval for $\mu$

- d.f. = $n - 1 = 24$, so $t_{\alpha/2} = t_{0.025} = 2.0639$

The confidence interval is

$$\bar{X} \pm t_{\alpha/2} \frac{S}{\sqrt{n}} = 50 \pm (2.0639) \frac{8}{\sqrt{25}}$$

$$46.698 \leq \mu \leq 53.302$$
Example of t distribution confidence interval

Interpreting this interval requires the assumption that the population you are sampling from is approximately a normal distribution (especially since n is only 25).

This condition can be checked by creating a:
- Normal probability plot or
- Boxplot

Confidence Intervals

- Population Mean
  - \( \sigma \text{ known} \)
  - \( \sigma \text{ unknown} \)
- Population Proportion
An interval estimate for the population proportion (\( \pi \)) can be calculated by adding an allowance for uncertainty to the sample proportion (\( p \)).

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

\[
\sqrt{\frac{p(1-p)}{n}}
\]
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 \( X > 5 \) and \( n - X > 5 \)

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
Example (continued)

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

\[
p = \frac{25}{100} = 0.25 (0.75)\]

\[
p \pm Z_{\alpha/2} \sqrt{p(1-p)/n}
= 0.25 \pm 1.96 \sqrt{0.25(0.75)/100}
= 0.25 \pm 1.96 (0.0433)
= 0.25 \pm 0.09
\]

\[
0.1651 \leq \pi \leq 0.3349
\]

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

For the Mean

For the Proportion

Sampling Error

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

- The margin of error is also called \textit{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

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

Sampling error (margin of error)

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

(continued)

Now solve for \( n \) to get

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

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, \( \sigma \)

Required Sample Size Example

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

\[
\begin{align*}
    n &= \frac{Z^2 \sigma^2}{e^2} = \frac{(1.645)^2(45)^2}{5^2} = 219.19 \\
    \text{So the required sample size is } n &= 220
\end{align*}
\]

(Always round up)
If $\sigma$ is unknown

- If unknown, $\sigma$ can be estimated when determining the required sample size
  - 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

For the Proportion

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

Now solve for $n$ to get

\[ n = \frac{Z^2 \pi (1-\pi)}{e^2} \]
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 \( p = 0.12 \))
Required Sample Size Example

Solution:

For 95% confidence, use $Z_{\alpha/2} = 1.96$

e = 0.03

p = 0.12, so use this to estimate $\pi$

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

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
Applications in Auditing

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

Sampling is often more accurate for drawing conclusions about large populations
- Examining every item in a large population is subject to significant non-sampling error

Sampling allows auditors to combine, and then evaluate collectively, samples collected by different individuals

Confidence Interval for Population Total Amount

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

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

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

A firm has a population of 1,000 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.

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

Example Solution

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

\[ = 1,000(87.6) \pm 1,000(1.9905) \frac{22.3}{\sqrt{80}} \sqrt{\frac{1,000 - 80}{1,000 - 1}} \]

\[ = 87,600 \pm 4,762.48 \]

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 $N$:
  \[
  \text{Total Difference} = N \bar{D}
  \]

- Where the average difference, $\bar{D}$, is:
  \[
  \bar{D} = \frac{\sum_{i=1}^{n} D_i}{n}
  \]
  where $D_i = \text{audited value} - \text{original value}$

Confidence Interval for Total Difference (continued)

- Confidence interval estimate:
  \[
  N \bar{D} \pm N \left( t_{\alpha/2} \right) \frac{S_D}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}
  \]
  where
  \[
  S_D = \sqrt{\frac{\sum_{i=1}^{n} (D_i - \bar{D})^2}{n-1}}
  \]
One-Sided Confidence Intervals

- Application: find the upper bound for the proportion of items that do not conform with internal controls

\[ \text{Upper bound} = p + Z_{\alpha} \sqrt{\frac{p(1-p)}{n}} \sqrt{\frac{N-n}{N-1}} \]

- 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
Summary

- Introduced the concept of confidence intervals
- Discussed point estimates
- Developed confidence interval estimates
- Created confidence interval estimates for the mean ($\sigma$ known)
- Determined confidence interval estimates for the mean ($\sigma$ unknown)
- Created confidence interval estimates for the proportion
- Determined required sample size for mean and proportion confidence interval estimates with a desired margin of error

Summary (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
In this topic, you learn:

- When to use a finite population correction factor in calculating a confidence interval for either \( \mu \) or \( \pi \)
- How to use a finite population correction factor in calculating a confidence interval for either \( \mu \) or \( \pi \)
- How to use a finite population correction factor in calculating a sample size for a confidence interval for either \( \mu \) or \( \pi \)
Use a fpc when sampling more than 5% of the population \((n/N > 0.05)\)

Confidence Interval for \(\mu\) with a fpc

\[ \bar{X} \pm t_{\alpha/2} \frac{S}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \]

Confidence Interval for \(\pi\) with a fpc

\[ p \pm z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}} \sqrt{\frac{N-n}{N-1}} \]

A fpc simply reduces the standard error of either the sample mean or the sample proportion

Confidence Interval for \(\mu\) with a fpc

Suppose \(N = 1000, n = 100, \bar{X} = 50, s = 10\)

\[
95\% \text{ CI for } \mu: \bar{X} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} = 50 \pm 1.984 \frac{10}{\sqrt{100}} \sqrt{\frac{1000-100}{1000-1}} = 50 \pm 1.88 = (48.12, 51.88)
\]
Determining Sample Size with a fpc

- Calculate the sample size \( (n_0) \) without a fpc
  - For \( \mu \): \( n_0 = \frac{Z_{\alpha/2}^2 \sigma^2}{e^2} \)
  - For \( \pi \): \( n_0 = \frac{Z_{\alpha/2}^2 \pi (1-\pi)}{e^2} \)
- Apply the fpc utilizing the following formula to arrive at the final sample size \( (n) \).
  - \( n = \frac{n_0 N}{n_0 + (N-1)} \)

Topic Summary

- Described when to use a finite population correction in calculating a confidence interval for either \( \mu \) or \( \pi \)
- Examined the formulas for calculating a confidence interval for either \( \mu \) or \( \pi \) utilizing a finite population correction
- Examined the formulas for calculating a sample size in a confidence interval for either \( \mu \) or \( \pi \)