Basic Business Statistics
11^{th} Edition

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

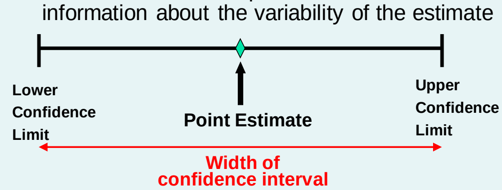
Chapter Outline

Content of this chapter

- Confidence Intervals for the Population Mean, μ
  - when Population Standard Deviation σ is Known
  - when Population Standard Deviation σ is Unknown
- Confidence Intervals for the Population Proportion, π
- 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)
Mean	μ	$\bar{X}$
Proportion	π	p

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
  - $368 \pm 1.96 * 15 / \sqrt{25} = (362.12, 373.88)$ contains 95% of the sample means
  - When you don't know μ, you use $\bar{X}$ to estimate μ
    - If $\bar{X} = 362.3$ the interval is $362.3 \pm 1.96 * 15 / \sqrt{25} = (356.42, 368.18)$
    - Since $356.42 \le \mu \le 368.18$ the interval based on this sample makes a correct statement about μ.

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

Confidence Interval Example

(continued)

Sample #	$\bar{X}$	Lower Limit	Upper Limit	Contain μ?
1	362.30	356.42	368.18	Yes
2	369.50	363.62	375.38	Yes
3	360.00	354.12	365.88	No
4	362.12	356.24	368.00	Yes
5	373.88	368.00	379.76	Yes

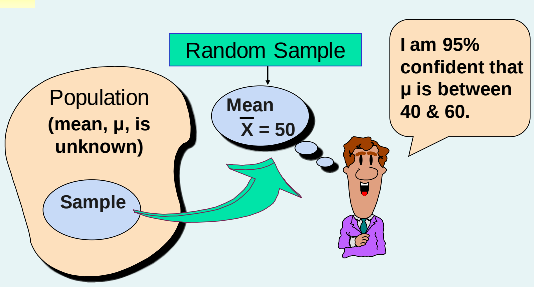
Confidence Interval Example

(continued)

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

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

Estimation Process



General Formula

- The general formula for all confidence intervals is:

Point Estimate ± (Critical Value)(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

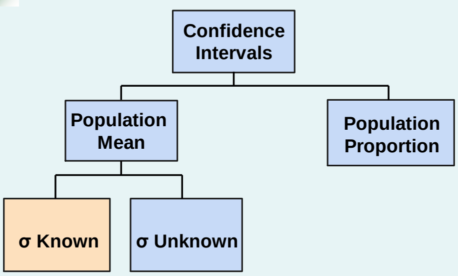
- Confidence Level
  - Confidence the interval will contain the unknown population parameter
  - A percentage (less than 100%)

Confidence Level, $(1-\alpha)$

(continued)

- 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



Confidence Interval for μ (σ Known)

- Assumptions
  - Population standard deviation σ is known
  - Population is normally distributed
  - If population is not normal, use large sample
- Confidence interval estimate:

$$\overline{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

where $\overline{X}$ is the point estimate

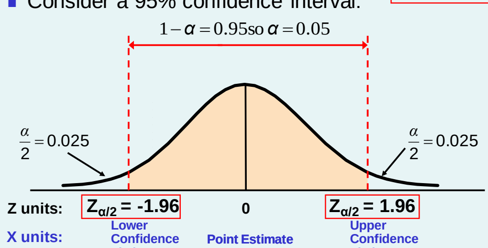
$Z_{\alpha/2}$ is the normal distribution critical value for a probability of $\alpha/2$ in each tail

$\sigma/\sqrt{n}$ is the standard error

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

- Consider a 95% confidence interval:

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



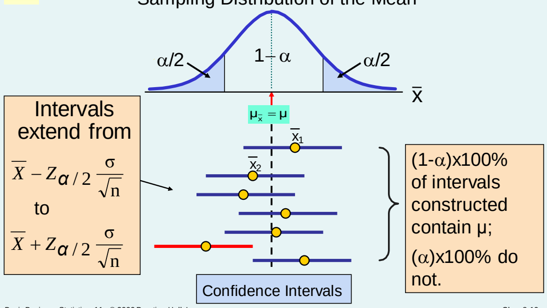
Common Levels of Confidence

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

Confidence Level	*Confidence Coefficient,* $1-\alpha$	$Z_{\alpha/2}$ *value*
80%	0.80	1.28
90%	0.90	1.645
95%	0.95	1.96
98%	0.98	2.33
99%	0.99	2.58
99.8%	0.998	3.08
99.9%	0.999	3.27

Intervals and Level of Confidence

Sampling Distribution of the Mean



Basic Business Statistics, 11e © 2009 Prentice-Hall, Inc.. Chap 8-19

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.

Basic Business Statistics, 11e © 2009 Prentice-Hall, Inc.. Chap 8-20

Example

(continued)

- 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:

$$\bar{X} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$
$$= 2.20 \pm 1.96(0.35/\sqrt{11})$$
$$= 2.20 \pm 0.2068$$
$$\boxed{1.9932 \le \mu \le 2.4068}$$

Basic Business Statistics, 11e © 2009 Prentice-Hall, Inc.. Chap 8-21

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

Basic Business Statistics, 11e © 2009 Prentice-Hall, Inc.. Chap 8-22

Confidence Intervals

- **Confidence Intervals**
  - **Population Mean**
    - **σ Known**
    - **σ Unknown**
  - **Population Proportion**

Basic Business Statistics, 11e © 2009 Prentice-Hall, Inc.. Chap 8-23

Do You Ever Truly Know σ?

- Probably not!
- In virtually all real world business situations, σ is not known.
- If there is a situation where σ is known then μ is also known (since to calculate σ you need to know μ.)
- If you truly know μ there would be no need to gather a sample to estimate it.

Basic Business Statistics, 11e © 2009 Prentice-Hall, Inc.. Chap 8-24

Confidence Interval for μ (σ Unknown)

- If the population standard deviation σ 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

Confidence Interval for μ (σ Unknown)

(continued)

- 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:

$$\overline{X} \pm t_{\alpha/2} \frac{S}{\sqrt{n}}$$

(where $t_{\alpha/2}$ is the critical value of the t distribution with n -1 degrees of freedom and an area of α/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?

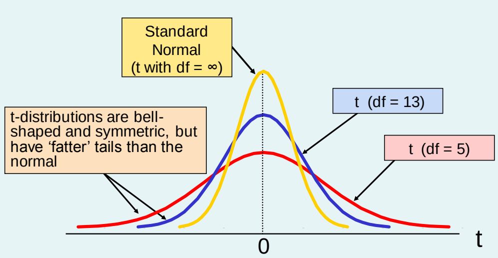
If the mean of these three values is 8.0, then X_3 must be 9 (i.e., X_3 is not free to vary)

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 ⟶ Z as n increases

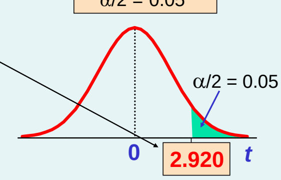


Student's t Table

	Upper Tail Area		
df	**.25**	**.10**	**.05**
1	1.000	3.078	6.314
2	0.817	1.886	2.920
3	0.765	1.638	2.353

The body of the table contains t values, not probabilities

Let: n = 3
df = *n* - 1 = 2
α = 0.10
α/2 = 0.05



Selected t distribution values

With comparison to the Z value

Confidence Level	t (10 d.f.)	t (20 d.f.)	t (30 d.f.)	Z (∞ d.f.)
0.80	1.372	1.325	1.310	1.28
0.90	1.812	1.725	1.697	1.645
0.95	2.228	2.086	2.042	1.96
0.99	3.169	2.845	2.750	2.58

Note: t ⟶ 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 μ

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

The confidence interval is

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

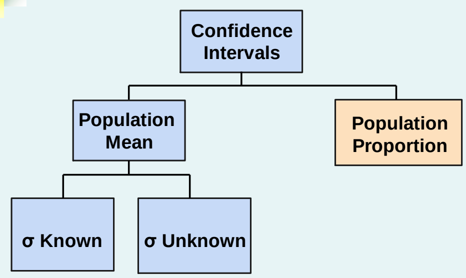
$46.698 \le \mu \le 53.302$

Example of t distribution confidence interval

(continued)

- 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



Confidence Intervals for the Population Proportion, π

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

Confidence Intervals for the Population Proportion, π

(continued)

- 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

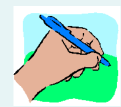
- Upper and lower confidence limits for the population proportion are calculated with the formula

$$\boxed{p \pm Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}}$$

- 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



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.

$$p \pm Z_{\alpha/2}\sqrt{p(1-p)/n}$$
$$= 25/100 \pm 1.96\sqrt{0.25(0.75)/100}$$
$$= 0.25 \pm 1.96(0.0433)$$
$$\boxed{0.1651 \le \pi \le 0.3349}$$

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

Determining Sample Size

- **For the Mean**
- **For the Proportion**

Sampling Error

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

Determining Sample Size

For the Mean

Sampling error (margin of error)

$$\bar{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \qquad e = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Determining Sample Size

(continued)

Determining Sample Size

For the Mean

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

Now solve for n to get

$$n = \frac{Z_{\alpha/2}^{2}\,\sigma^{2}}{e^{2}}$$

Determining Sample Size

(continued)

- To determine the required sample size for the mean, you must know:
  - The desired level of confidence (1 - α), which determines the critical value, $Z_{\alpha/2}$
  - The acceptable sampling error, e
  - The standard deviation, σ

Required Sample Size Example

If σ = 45, what sample size is needed to estimate the mean within ± 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 σ is unknown

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

Determining Sample Size

(continued)

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

(continued)

- To determine the required sample size for the proportion, you must know:
  - The desired level of confidence (1 - α), which determines the critical value, $Z_{\alpha/2}$
  - The acceptable sampling error, e
  - The true proportion of events of interest, π
    - π can be estimated with a pilot sample if necessary (or conservatively use 0.5 as an estimate of π)

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

(continued)

Solution:

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

e = 0.03

p = 0.12, so use this to estimate π

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

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

(continued)

- 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 isoften 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:

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

(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

$N = 1000, \quad n = 80, \quad \bar{X} = 87.6, \quad S = 22.3$

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

$$= (1000)(87.6) \pm (1000)(1.9905)\frac{22.3}{\sqrt{80}}\sqrt{\frac{1000-80}{1000-1}}$$

$$= 87{,}600 \pm 4{,}762.48$$

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 = audited value - original value

Confidence Interval for Total Difference

(continued)

- Confidence interval estimate:

$$N\bar{D} \pm N(t_{\alpha/2})\frac{S_D}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

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_α 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

Chapter 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

