Dr. Saeed A. Dobbah Alghamdi

Department of Statistics

Faculty of Sciences

Building 90, Office 26F41

King Abdulaziz University

http://saalghamdy.kau.edu.sa

Dr. Saeed's Website

Statistics
for
Business & Economics

David R. Anderson
Dennis J. Sweeney
Thomas A. Williams

Interval Estimation

Chapter 8

Learning Objectives

LO1 Define confidence interval estimates.

LO2 Define level of confidence.

LO3 Compute a confidence interval for a population mean when the population standard deviation is known.

LO4 Compute a confidence interval for a population mean when the population standard deviation is unknown.

LO5 Compute a confidence interval for a population proportion.

LO6 Adjust a confidence interval for finite populations

LO7 Calculate the required sample size to estimate a population proportion or population mean.

Introduction

- The purpose of an interval estimate is to provide information about how close the point estimate might be to the value of the population parameter. In relatively simple cases, the general form of an interval estimate is

$$\text{Point estimate} \pm \text{Margin of error}$$

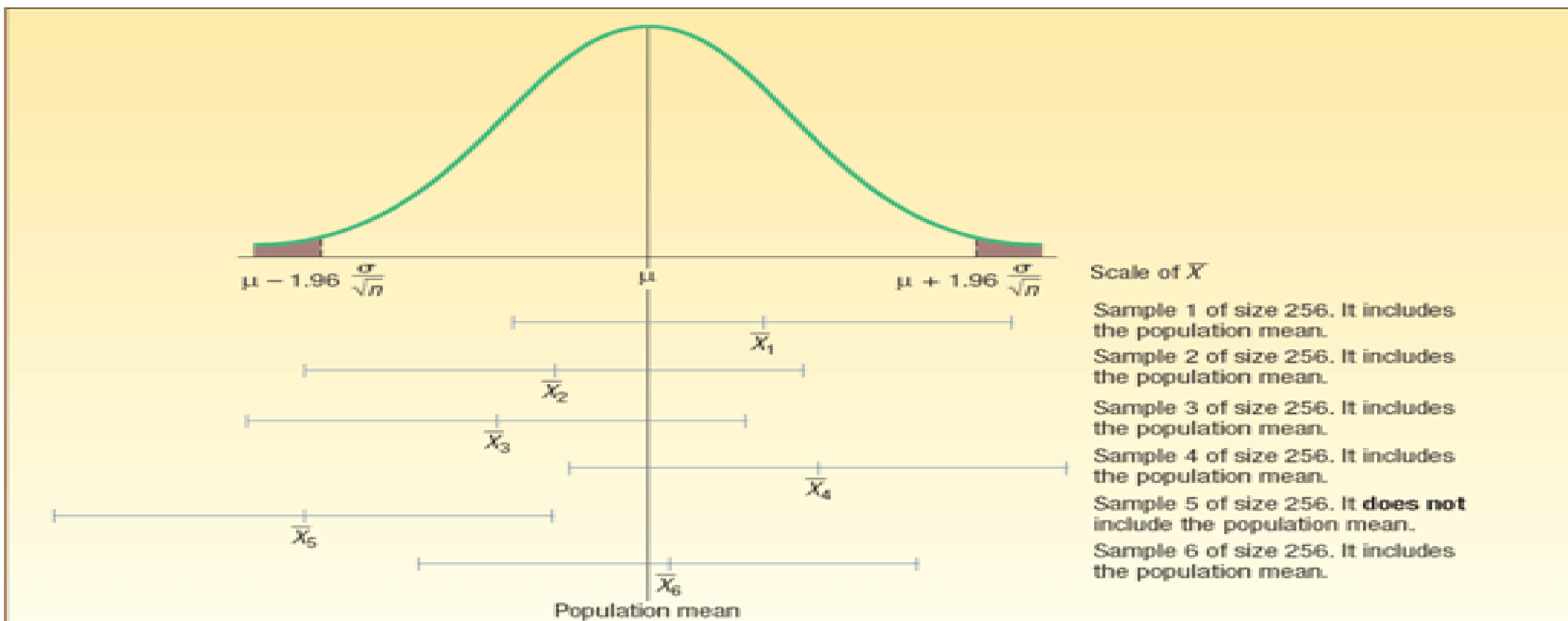
- A **confidence interval estimate** is a range of values constructed from sample data so that the population parameter is likely to occur within that range at a specified probability. The specified probability is called the level of confidence.
- The sampling distribution of the point estimates play roles in computing these interval estimates

- **The factors that determine the width of a confidence interval**
  1. The **sample size**, *n*.
  2. The **variability in the population**, usually σ estimated by *s*.
  3. The desired **level of confidence**.

Interval Estimates - Interpretation

For a 95% confidence interval, about 95% of the similarly constructed intervals will contain the parameter being estimated. Also 95% of the sample means for a specified sample size will lie within 1.96 standard deviations of the hypothesized population

The *95 percent confidence* refers to the middle 95 percent of the observations. Therefore, the remaining 5 percent are equally divided between the two tails.



Confidence Intervals for a Mean – σ Known

- The confidence interval for the population mean when σ

$$\bar{x} \pm z \frac{\sigma}{\sqrt{n}}$$

$\bar{x}$ – sample mean

z – z - value for a particular confidence level

σ – the population standard deviation

n – the number of observations in the sample

1. The width of the interval is determined by the level of confidence and the size of the standard error of the mean.
2. The standard error is affected by two values:
   - Standard deviation
   - Number of observations in the sample

Confidence Intervals for a Mean – σ Known

EXAMPLE

The American Management Association wishes to have information on the mean income of middle managers in the retail industry. A random sample of 256 managers reveals a sample mean of $45,420. The standard deviation of this population is $2,050. The association would like answers to the following questions:

1. What is the population mean?

   The sample mean of $45,420 is a *point estimate* of the unknown population mean.

2. What is a reasonable range of values for the population mean? (Use 95% confidence level)

$$\overline{X} \pm z\frac{\sigma}{\sqrt{n}} = \$45{,}420 \pm 1.96\frac{\$2{,}050}{\sqrt{256}} = \$45{,}420 \pm \$251$$

   The confidence limits are $45,169 and $45,671

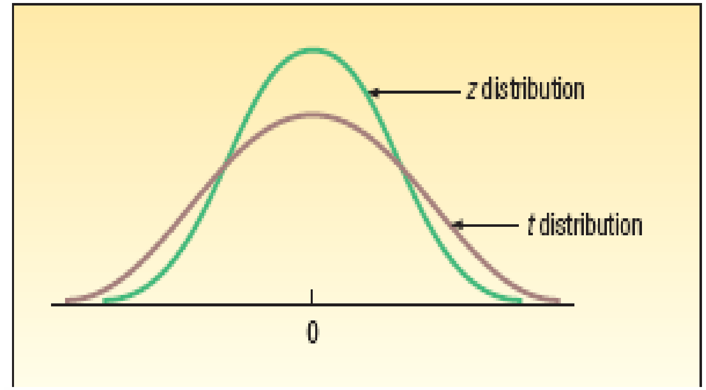
   The ±$251 is referred to as the margin of error

3. What do these results mean?

   If we select many samples of 256 managers, and for each sample we compute the mean and then construct a 95 percent confidence interval, we could expect about 95 percent of these confidence intervals to contain the *population* mean.

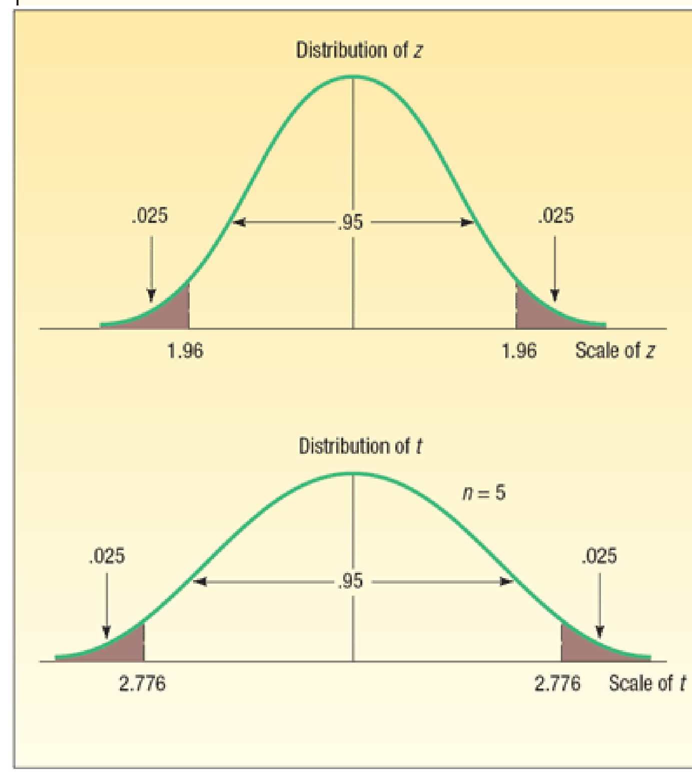
Confidence Intervals for a Mean – σ Unknown

In most sampling situations the population standard deviation (σ) is not known. In this case if the sample size is less than 30, we use **t-distribution** instead of **z-distribution**.



CHARACTERISTICS OF THE *t*-Distribution

1. It is, like the *z* distribution, a **continuous distribution**.
2. It is, like the *z* distribution, **bell-shaped** and **symmetrical**.
3. There is **not one *t* distribution**, but rather **a family of t distributions**. All *t* distributions have a mean of 0, but their standard deviations differ according to the sample size, *n*.
4. The ***t* distribution is more spread out** and ***flatter at the center than the standard normal distribution*** As the sample size increases, however, the *t* distribution approaches the standard normal distribution



Confidence Intervals for a Mean – σ Unknown

EXAMPLE

A tire manufacturer wishes to investigate the tread life of its tires. A sample of 10 tires driven 50,000 miles revealed a sample mean of 0.32 inch of tread remaining with a standard deviation of 0.09 inch.

- Construct a 95 percent confidence interval for the population mean.
- Would it be reasonable for the manufacturer to conclude that after 50,000 miles the population mean amount of tread remaining is 0.30 inches?

Compute the C.I.

using the t - dist. (since σ is unknown)

$$\bar{X} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$$

$$= \bar{X} \pm t_{.05/2,10-1}\frac{s}{\sqrt{n}}$$

$$= 0.32 \pm t_{.025,9}\frac{0.09}{\sqrt{10}}$$

$$= 0.32 \pm 2.262\frac{0.09}{\sqrt{10}}$$

$$= 0.32 \pm 0.064$$

$$= (0.256, 0.384)$$

Conclude : the manufacturer can be reasonably sure (95% confident) that the mean remaining tread depth is between 0.256 and 0.384 inches.

Confidence Intervals for a Mean – σ Unknown

EXAMPLE

The manager of the Inlet Square Mall, near Ft. Myers, Florida, wants to estimate the mean amount spent per shopping visit by customers. A sample of 20 customers reveals the following amounts spent.

$48.16	$42.22	$46.82	$51.45	$23.78	$41.86	$54.86
37.92	52.64	48.59	50.82	46.94	61.83	61.69
49.17	61.46	51.35	52.68	58.84	43.88	

Compute the C.I.

using the t - dist. (since σ is unknown)

$$\bar{X} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$$

$$= \bar{X} \pm t_{.05/2,20-1}\frac{s}{\sqrt{n}}$$

$$= 49.35 \pm t_{.025,19}\frac{9.01}{\sqrt{20}}$$

$$= 49.35 \pm 2.093\frac{9.01}{\sqrt{20}}$$

$$= 49.35 \pm 4.22$$

The endpoints of the confidence interval are $45.13 and $53.57

Conclude: It is reasonable that the population mean could be $50.

The value of $60 is not in the confidence interval. Hence, we conclude that the population mean is unlikely to be $60.

Confidence Interval Estimates for the Mean

Use *Z*-distribution

If the population standard deviation is known or the sample is greater than 30.

$$\bar{X} \pm z \frac{\sigma}{\sqrt{n}}$$

Use *t*-distribution

If the population standard deviation is unknown and the sample is less than 30.

$$\bar{X} \pm t \frac{s}{\sqrt{n}}$$

Confidence Interval for a Proportion (π)

The examples below illustrate the nominal scale of measurement:

1. The career services director at Southern Technical Institute reports that 80 percent of its graduates enter the job market in a position related to their field of study.

2. A company representative claims that 45 percent of Burger King sales are made at the drive-through window.

3. A survey of homes in the Chicago area indicated that 85 percent of the new construction had central air conditioning.

4. A recent survey of married men between the ages of 35 and 50 found that 63 percent felt that both partners should earn a living.

Confidence Interval for a Proportion (π)

Using the Normal Distribution to Approximate the Binomial Distribution

To develop a confidence interval for a proportion, we need to meet the following assumptions:

1. The binomial conditions have been met. Briefly, these conditions are:
   a. The sample data is the result of counts.
   b. There are only two possible outcomes.
   c. The probability of a success remains the same from trial to another.
   d. The trials are independent.
2. The values $n\,\pi$ and $n(1-\pi)$ should both be greater than or equal to 5. This condition allows us to invoke the central limit theorem and employ the standard normal distribution, that is, z, to complete a confidence interval.

Sample Proportion $\quad p = \frac{X}{n}$

Confidence Interval for a Population Proportion $\quad p \pm z\sqrt{\frac{p(1-p)}{n}}$

Confidence Interval for a Proportion (π)

EXAMPLE

The union representing the Bottle Blowers of America (BBA) is considering a proposal to merge with the Teamsters Union. According to BBA union bylaws, at least three-fourths of the union membership must approve any merger. A random sample of 2,000 current BBA members reveals 1,600 plan to vote for the merger proposal.

- What is the estimate of the population proportion?
- Develop a 95 percent confidence interval for the population proportion.
- Basing your decision on this sample information, can you conclude that the necessary proportion of BBA members favor the merger? Why?

First, compute the sample proportion:

$$p = \frac{x}{n} = \frac{1,600}{2000} = 0.80$$

Compute the 95% C.I.

$$\text{C.I.} = p \pm z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

$$= 0.80 \pm 1.96\sqrt{\frac{.80(1-.80)}{2,000}} = .80 \pm .018$$

$$= (0.782,\ 0.818)$$

Conclude: The merger proposal will likely pass because the interval estimate includes values greater than 75 percent of the union membership.

Finite-Population Correction Factor

- A population that has a fixed upper bound is said to be finite.
- For a finite population, where the total number of objects is *N* and the size of the sample is *n*, the following adjustment is made to the standard errors of the sample means and the proportion:

Standard Error of the Mean

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

Standard Error of the Proportion

$$\sigma_{p} = \sqrt{\frac{p(1-p)}{n}}\sqrt{\frac{N-n}{N-1}}$$

- However, if $n/N < .05$, the finite-population correction factor may be ignored. Why? See what happens to the value of the correction factor in the table below when the fraction n/N becomes smaller

Finite-Population Correction Factor for Selected Samples When the Population Is 1,000

Sample Size	Fraction of Population	Correction Factor
10	.010	.9955
25	.025	.9879
50	.050	.9752
100	.100	.9492
200	.200	.8949
500	.500	.7075

- The FPC approaches 1 when *n/N*

CI for Mean with FPC - Example

There are 250 families in Scandia, Pennsylvania. A random sample of 40 of these families revealed the mean annual church contribution was \$450 and the standard deviation of this was \$75.

Could the population mean be \$445 or \$425?

What is the population mean?

What is the best estimate of the population mean?

<u>Given in Problem:</u>

$N = 250$, $n = 40$, $s = \$75$

Since $n/N = 40/250 = 0.16$, the finite population correction factor must be used.

The population standard deviation is not known therefore use the t-distribution (may use the z-dist since n>30)

$$\bar{X} \pm t \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

$$= \$450 \pm t_{.10/2,40-1} \frac{\$75}{\sqrt{40}} \sqrt{\frac{250-40}{250-1}}$$

$$= \$450 \pm 1.685 \frac{\$75}{\sqrt{40}} \sqrt{\frac{250-40}{250-1}}$$

$$= \$450 \pm \$19.98\sqrt{.8434}$$

$$= \$450 \pm \$18.35$$

$$= (\$431.65,\ \ \$468.35)$$

It is likely that the population mean is more than \$431.65 but less than \$468.35. To put it another way, could the population mean be \$445? Yes, but it is not likely to be \$425 because the value \$445 is within the confidence interval while \$425 is not within the confidence interval.

Sample Size for Estimating a Population Mean

There are 3 factors that determine the size of a sample, none of which has any direct relationship to the size of the population.

- **The level of confidence desired.**
- **The margin of error the researcher will tolerate.**
- **The variation in the population being Studied.**

$$n = \left(\frac{z \cdot \sigma}{E} \right)^2$$

where:

n is the size of the sample.
z is the standard normal value corresponding to the desired level of confidence.
σ is the population standard deviation.
E is the maximum allowable error.

Sample Size for Estimating a Population Mean

EXAMPLE

A student in public administration wants to determine the mean amount members of city councils in large cities earn per month as remuneration for being a council member. The error in estimating the mean is to be less than $100 with a 95 percent level of confidence. The student found a report by the Department of Labor that estimated the standard deviation to be $1,000. What is the required sample size?

Given in the problem:

- *E*, the maximum allowable error, is $100
- The value of *z* for a 95 percent level of confidence is 1.96,
- The estimate of the standard deviation is $1,000.

$$
\begin{aligned}
n &= \left(\frac{z \cdot \sigma}{E}\right)^2 \\
&= \left(\frac{(1.96)(\$1{,}000)}{\$100}\right)^2 \\
&= (19.6)^2 \\
&= 384.16 \\
&= 385
\end{aligned}
$$

Sample Size for Estimating a Population Proportion

There are 3 factors that determine the size of a sample, none of which has any direct relationship to the size of the population.

- **The level of confidence desired.**
- **The margin of error the researcher will tolerate.**
- **The variation in the population being Studied.**

$$n = p(1-p)\left(\frac{Z}{E}\right)^2$$

where:

n is the size of the sample

z is the standard normal value corresponding to the desired level of confidence

E is the maximum allowable error

NOTE:

use p = 0.5 if no initial information on the probability of success is available

Sample Size for Estimating a Population Proportion

EXAMPLE 1

The American Kennel Club wanted to estimate the proportion of children that have a dog as a pet. If the club wanted the estimate to be within 3% of the population proportion, how many children would they need to contact? Assume a 95% level of confidence and that the club estimated that 30% of the children have a dog as a pet.

$$n = (.30)(.70)\left(\frac{1.96}{.03}\right)^2 = 897$$

EXAMPLE 2

A study needs to estimate the proportion of cities that have private refuse collectors. The investigator wants the margin of error to be within .10 of the population proportion, the desired level of confidence is 90 percent, and no estimate is available for the population proportion. What is the required sample size?

$$n = (.5)(1-.5)\left(\frac{1.65}{.10}\right)^2 = 68.0625$$

$$n = 69 \text{ cities}$$

Summary

Formulae

	Confidence Interval σ known	Confidence Interval σ unknown	Sample size
Population mean	$\bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$	$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$	$n = \frac{(z_{\alpha/2})^2 * \sigma^2}{E^2}$
Population Proportion	$\bar{p} \pm Z_{\alpha/2}\sqrt{\frac{\bar{p}\,(1-\bar{p})}{n}}$		$n = \frac{(z_{\alpha/2})^2 * \bar{p}\,(1-\bar{p})}{E^2}$

Summary

Interval estimate An estimate of a population parameter that provides an interval believed to contain the value of the parameter.

Margin of error The value added to and subtracted from a point estimate in order to develop an interval estimate of a population parameter.

σ known The case when historical data or other information provides a good value for the population standard deviation prior to taking a sample. The interval estimation procedure uses this known value of σ in computing the margin of error.

Confidence level The confidence associated with an interval estimate. For example, if an interval estimation procedure provides intervals such that 95% of the intervals formed using the procedure will include the population parameter, the interval estimate is said to be constructed at the 95% confidence level.

Confidence coefficient The confidence level expressed as a decimal value. For example, 0.95 is the confidence coefficient for a 95% confidence level.

Confidence interval Another name for an interval estimate.

σ unknown The more common case when no good basis exists for estimating the population standard deviation prior to taking the sample. The interval estimation procedure uses the sample standard deviation s in computing the margin of error.

Summary

***t* distribution** A family of probability distributions that can be used to develop an interval estimate of a population mean whenever the population standard deviation σ is unknown and is estimated by the sample standard deviation s.

Degrees of freedom A parameter of the t distribution. When the t distribution is used in the computation of an interval estimate of a population mean, the appropriate t distribution has n - 1 degrees of freedom, where n is the size of the simple random sample.

