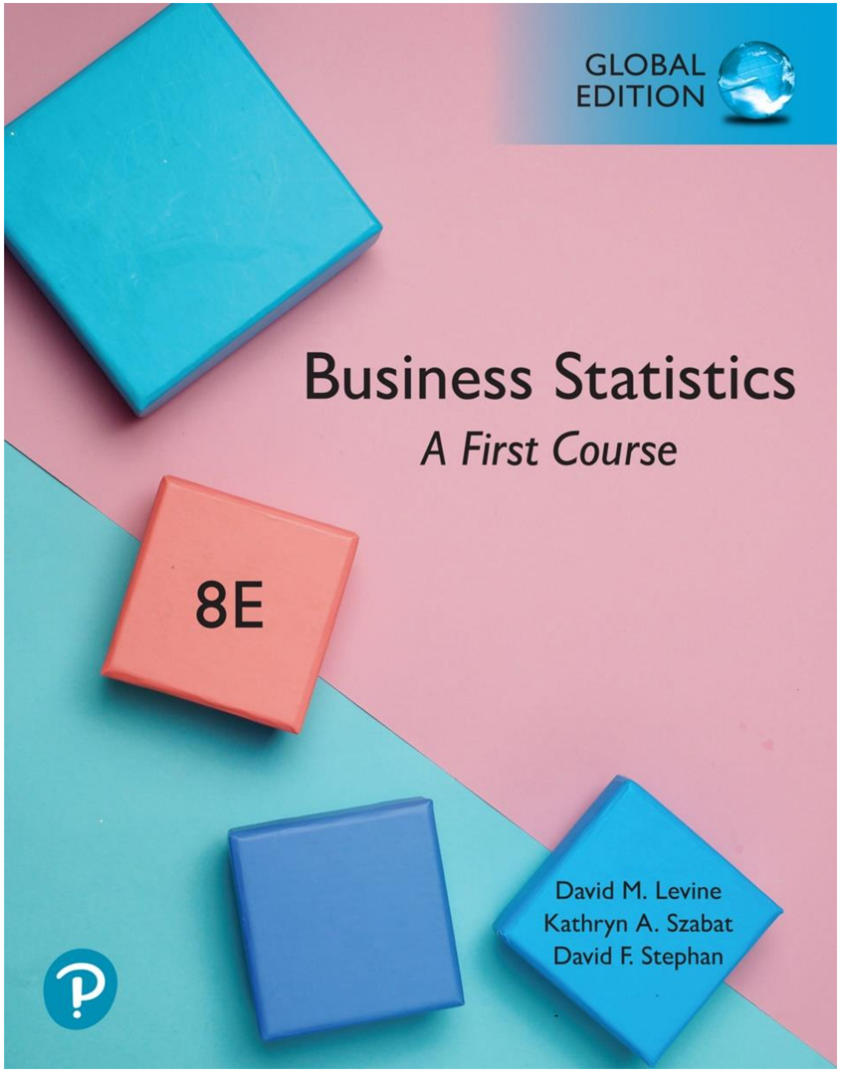


Chapter 8

Confidence Interval Estimation

Objectives

In this chapter, you learn:

- To construct and interpret confidence interval estimates for the mean and the proportion.
- To determine the sample size necessary to develop a confidence interval for the mean or proportion.

Chapter Outline

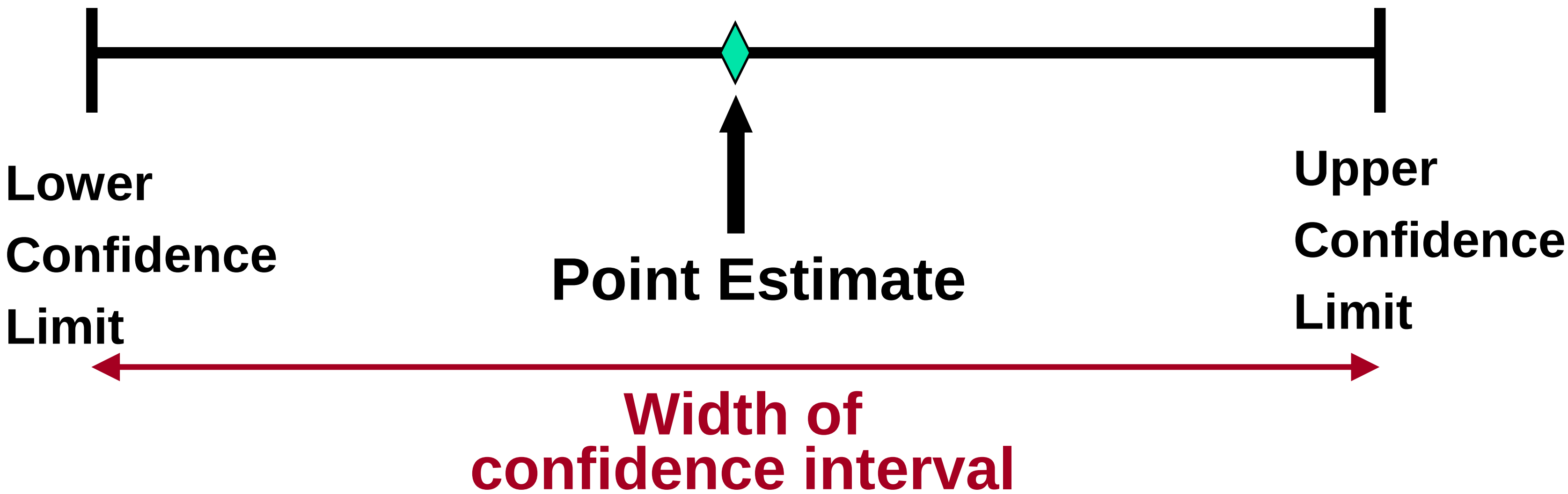
Content of this chapter

- 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

DCOVA

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



Point Estimates

DCOVA

We can estimate a Population Parameter …		with a Sample Statistic (a Point Estimate)
Mean	μ	$\bar{X}$
Proportion	π	p

Confidence Intervals

DCOVA

- 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

DCOVA

- 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

DCOVA

Cereal fill example

- Population has μ = 368 and σ = 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 of sample size 25.
  - 95% of the intervals formed in this manner will contain μ.
  - 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 $356.42 \leq 368 \leq 368.18$ 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)

DCOVA

Sample #	$\overline{X}$	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)

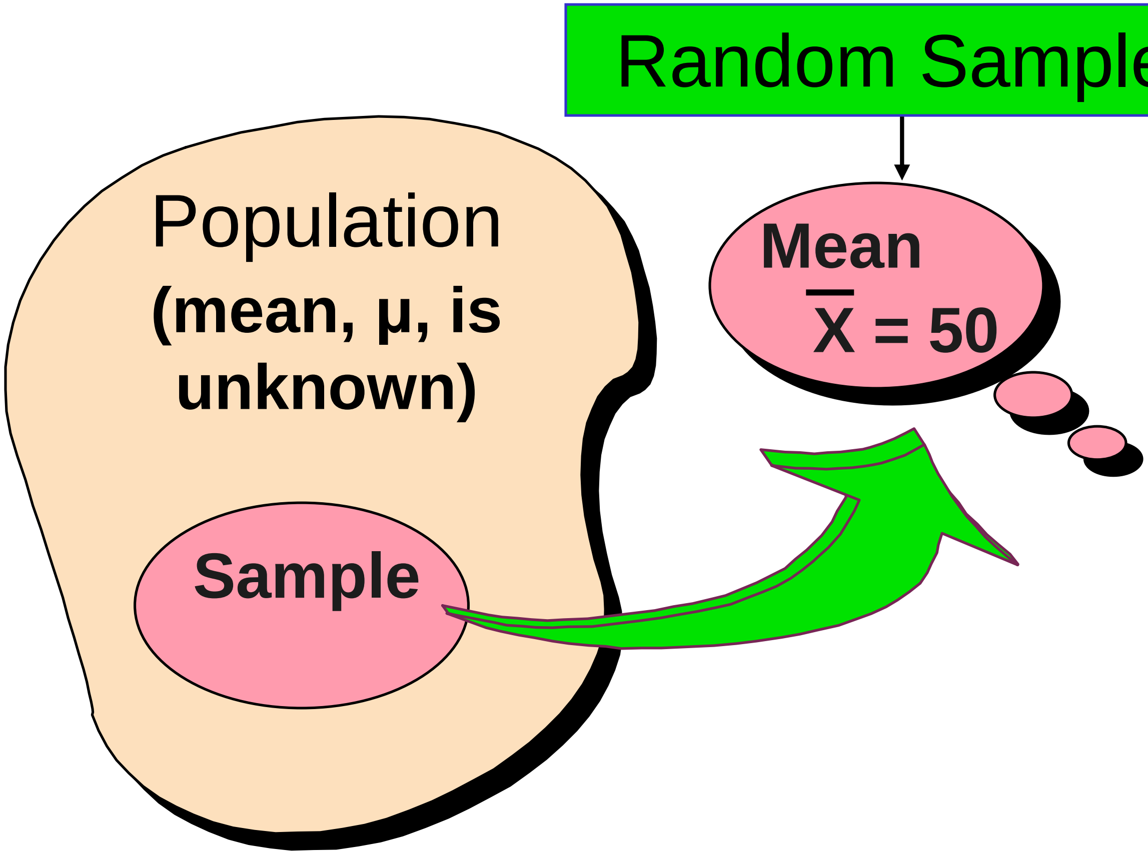
DCOVA

- 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

DCOVA



We can be 95% confident that μ is between 40 & 60.

General Formula

DCOVA

- 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

DCOVA

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

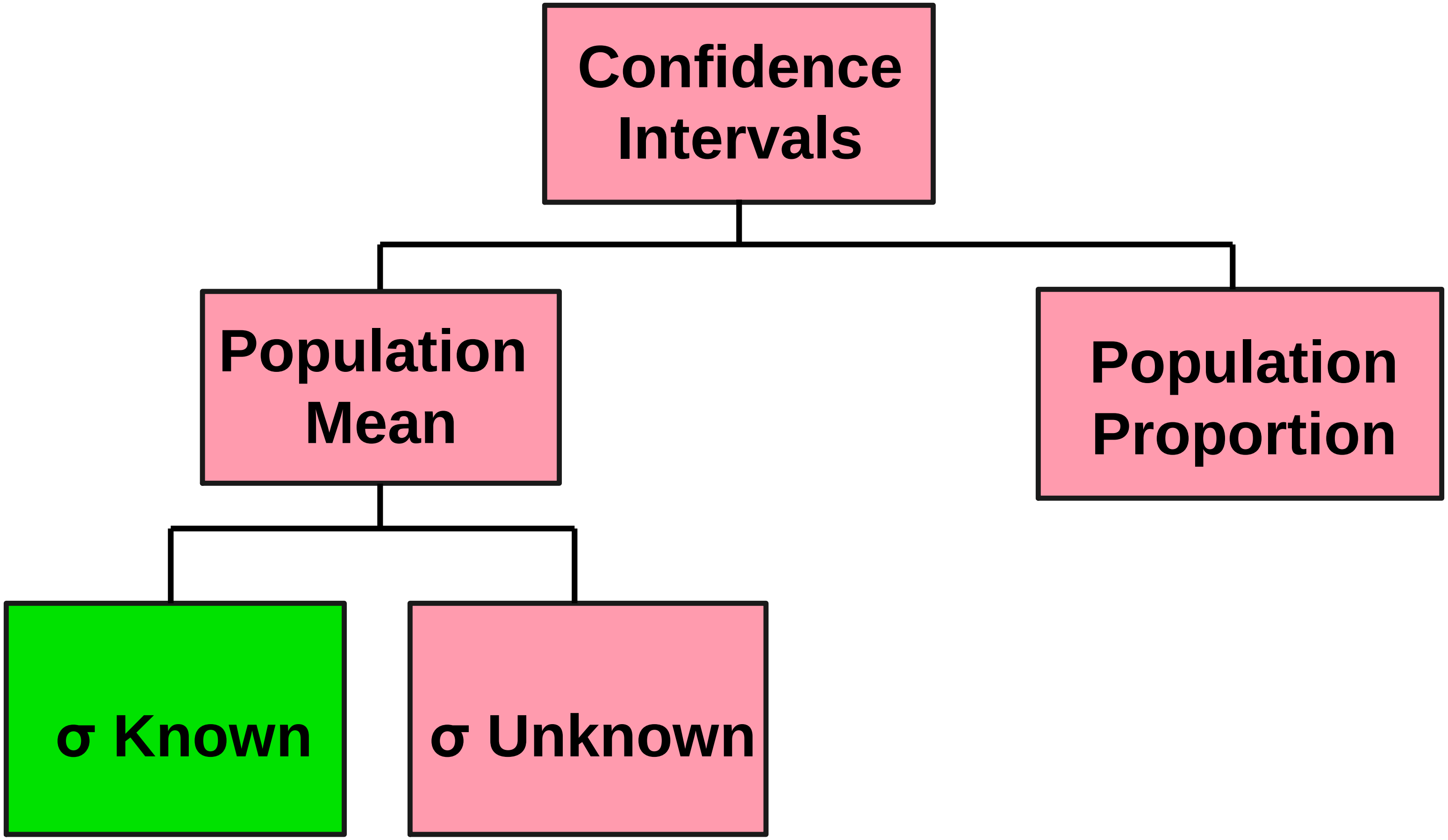
Confidence Level, $(1-\alpha)$ (continued)

DCOVA

- 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

DCOVA



Confidence Interval for μ (σ Known)

DCOVA

- Assumptions:
  - Population standard deviation σ is known.
  - Population is normally distributed.
  - If population is not normal, use large sample (n > 30).
- Confidence interval estimate:

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

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

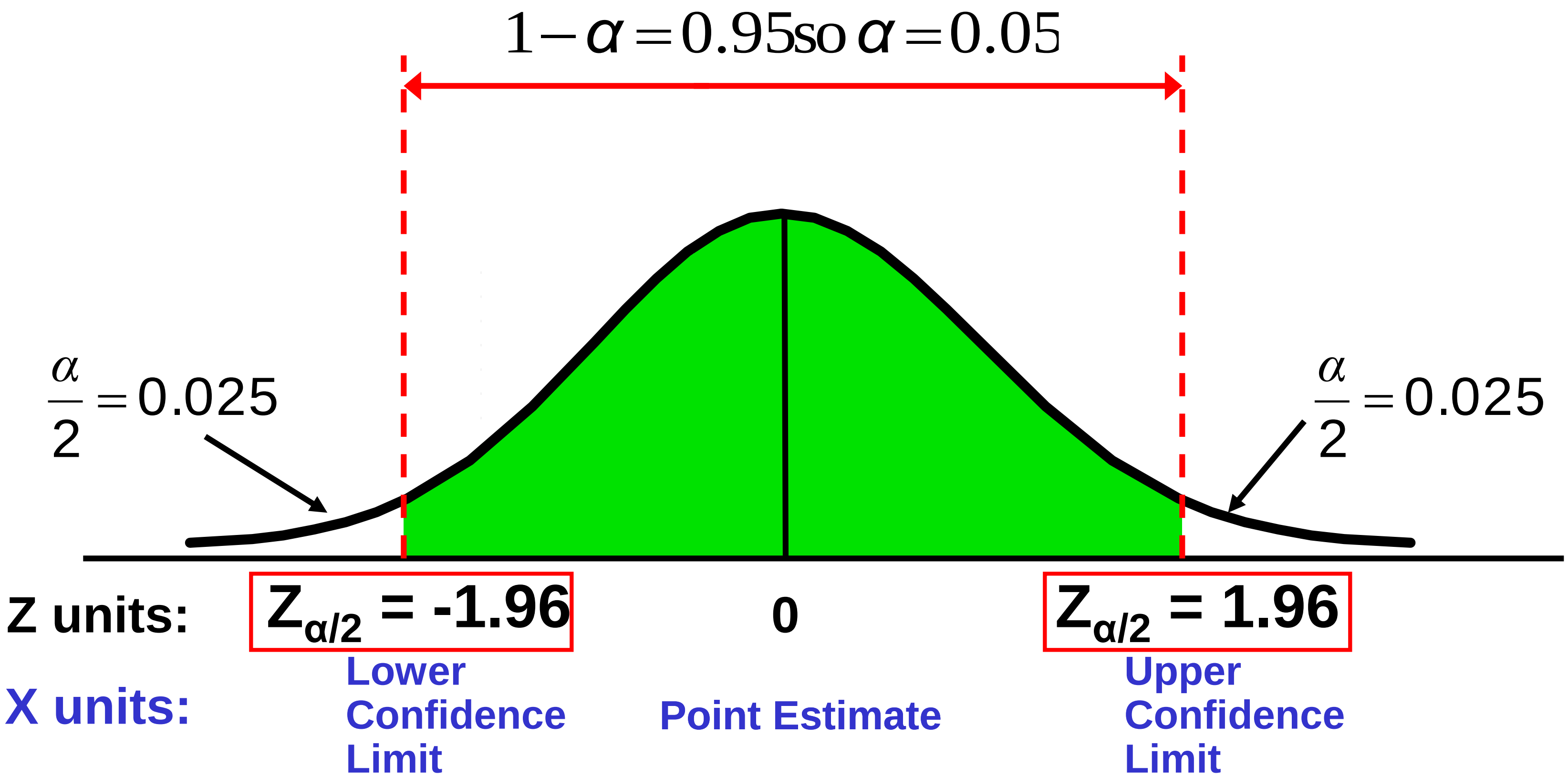
$\sigma/\sqrt{n}$ is the standard error

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

DCOVA

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

- Consider a 95% confidence interval:



Common Levels of Confidence

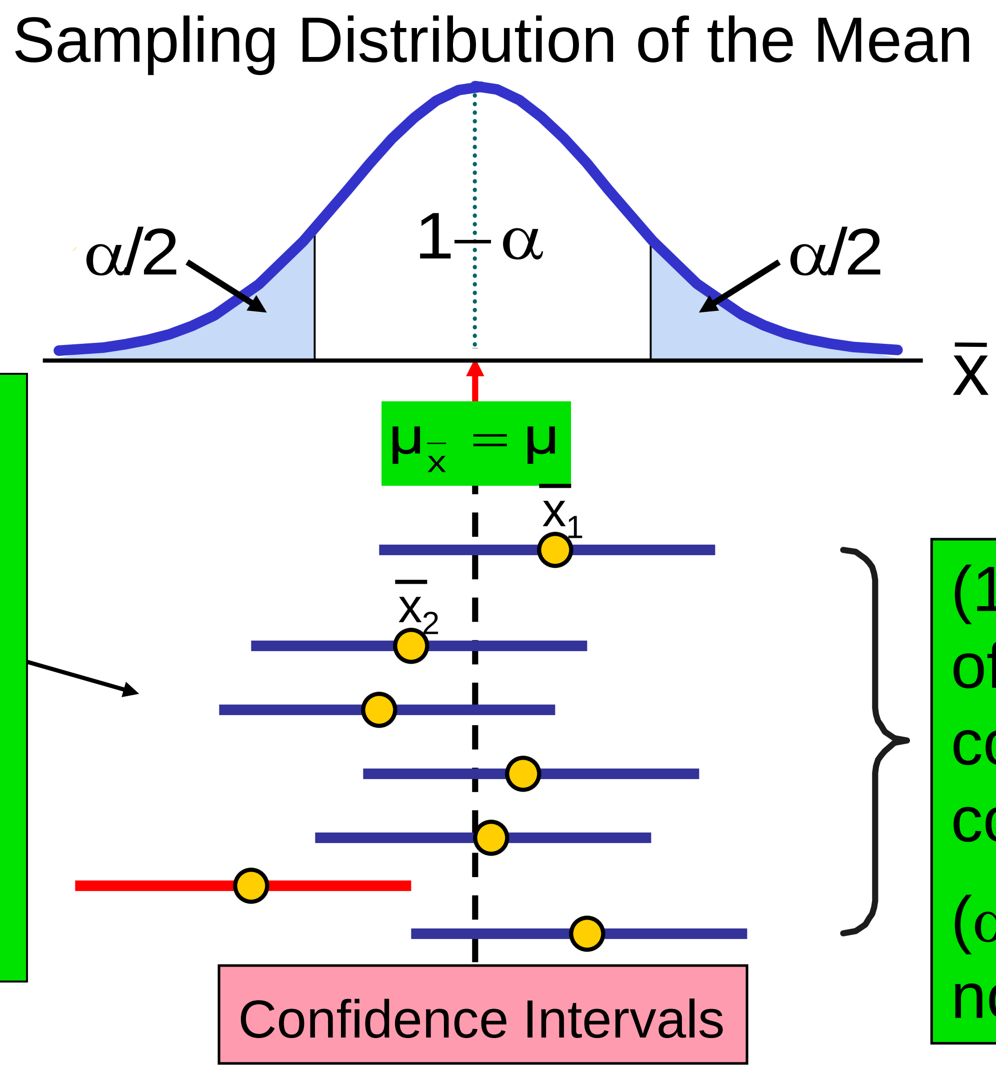
DCOVA

- 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

DCOVA



Intervals extend from

$$\bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

to

$$\bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

(1-α)100% of intervals constructed contain μ;

(α)100% do not.

Example

DCOVA

- 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

DCOVA

(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:

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

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

$$= 2.20 \pm 0.2068$$

$$\boxed{1.9932 \le \mu \le 2.4068}$$

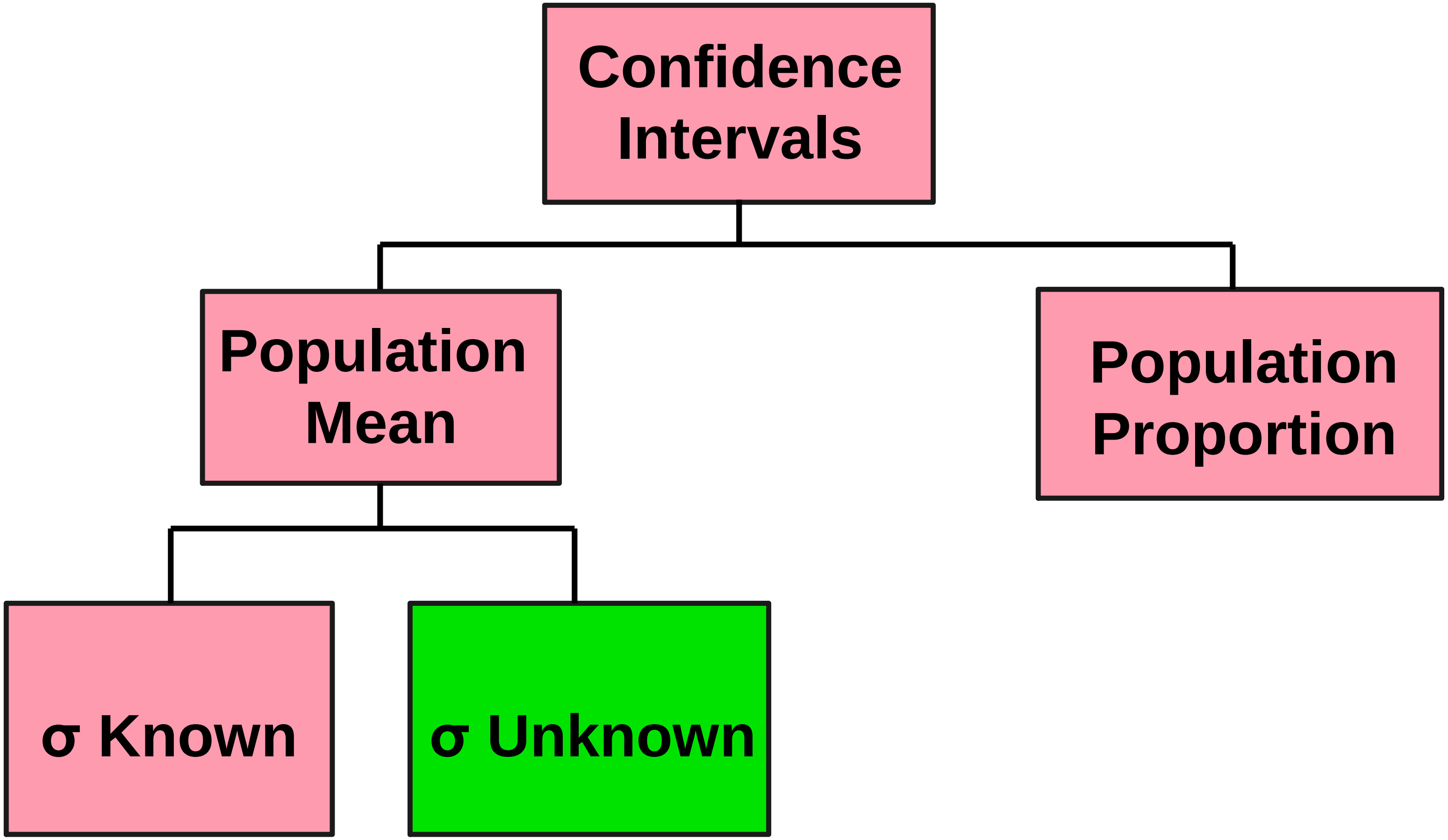
Interpretation

DCOVA

- 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

DCOVA



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.

Confidence Interval for μ (σ Unknown)

DCOVA

- 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)

DCOVA

- Assumptions:
  - Population standard deviation is unknown.
  - Population is normally distributed.
- 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 α/2 in each tail.)

Student's t Distribution

DCOVA

- 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.

$$\boxed{d.f. = n - 1}$$

Degrees of Freedom (df)

DCOVA

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)

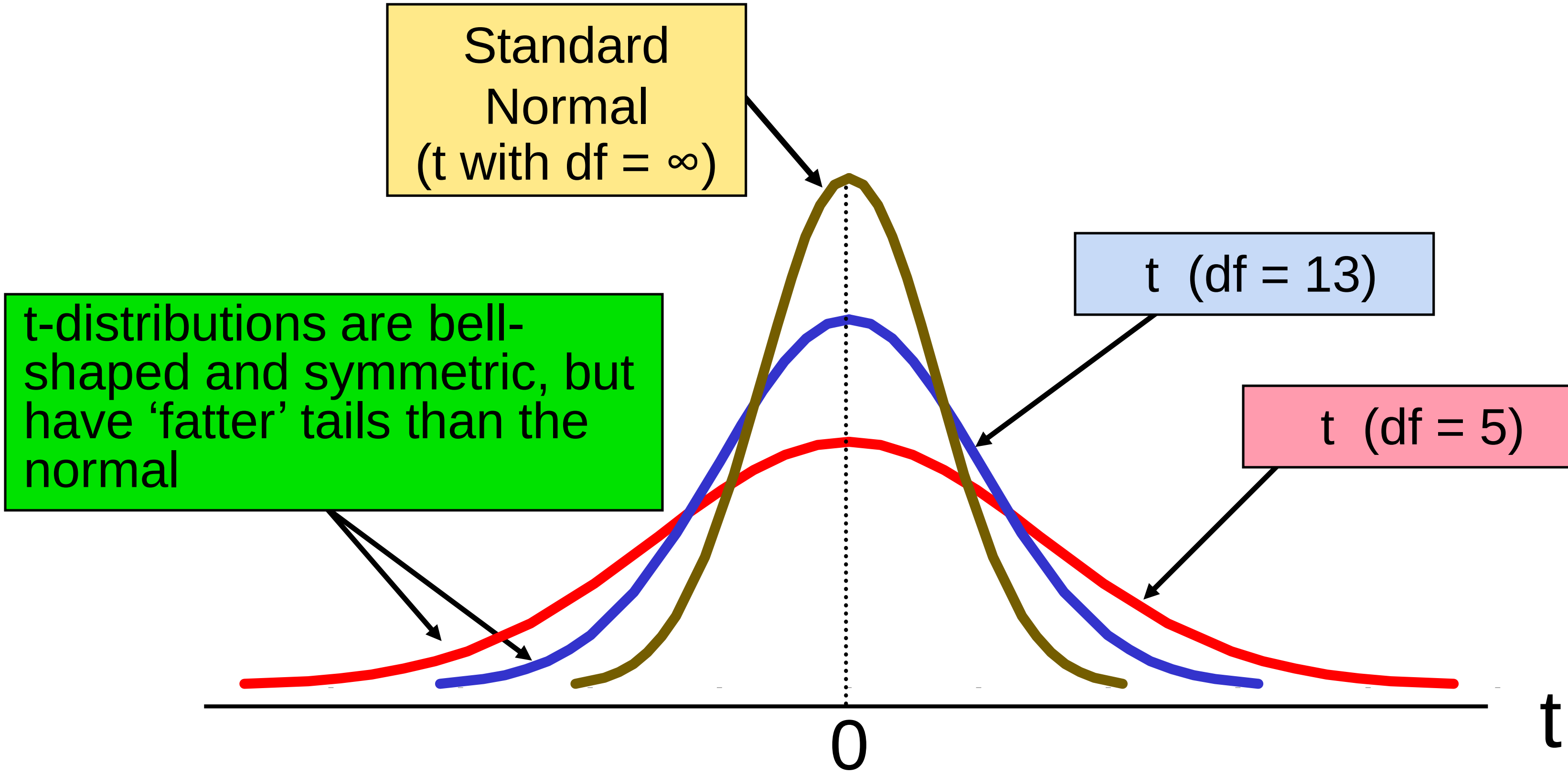
Here, n = 3, so degrees of freedom $= n - 1 = 3 - 1 = 2$.

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

Student’s t Distribution

DCOVA

Note: t ⟶ Z as n increases



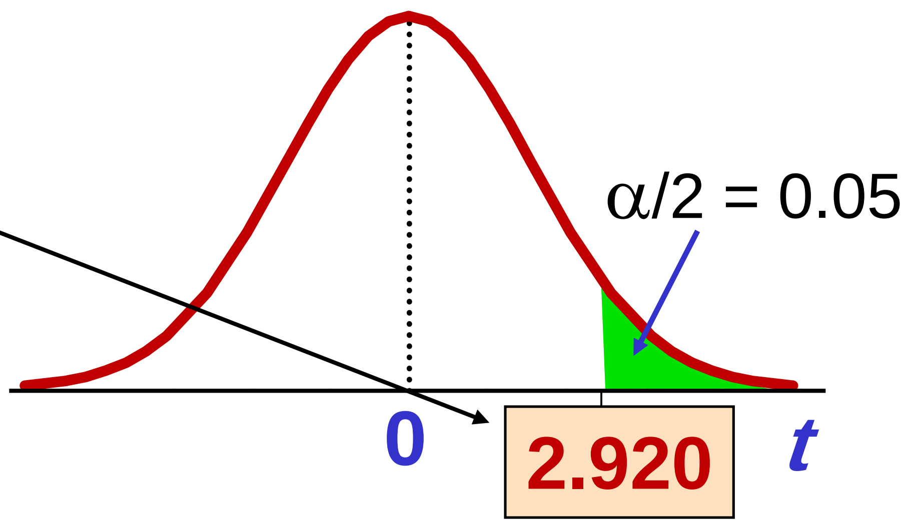
Student’s t Table

DCOVA

	Upper Tail Area		
df	.10	**.05**	.025
1	3.078	6.314	12.706
2	1.886	**2.920**	4.303
3	1.638	2.353	3.182

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

The body of the table contains t values, not probabilities



Selected t distribution values

DCOVA

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: $t \rightarrow Z$ as n increases

Example of t distribution confidence interval

DCOVA

A random sample of size n = 100 of travel times has a mean and standard deviation of $\bar{X}$ = 110.27 and S = 28.95. Form a 95% confidence interval for μ.

- d.f. = n – 1 = 99, so $t_{\alpha/2} = t_{0.025} = 1.9842$

The confidence interval is

$$\bar{X} \pm t_{\alpha/2}\frac{S}{\sqrt{n}} = 110.27 \pm (1.9842)\frac{28.95}{\sqrt{100}}$$

104.53 ≤ μ ≤ 116.01

Example of Excel, Minitab, & JMP Confidence Interval Output For Travel Time Sample

DCOVA

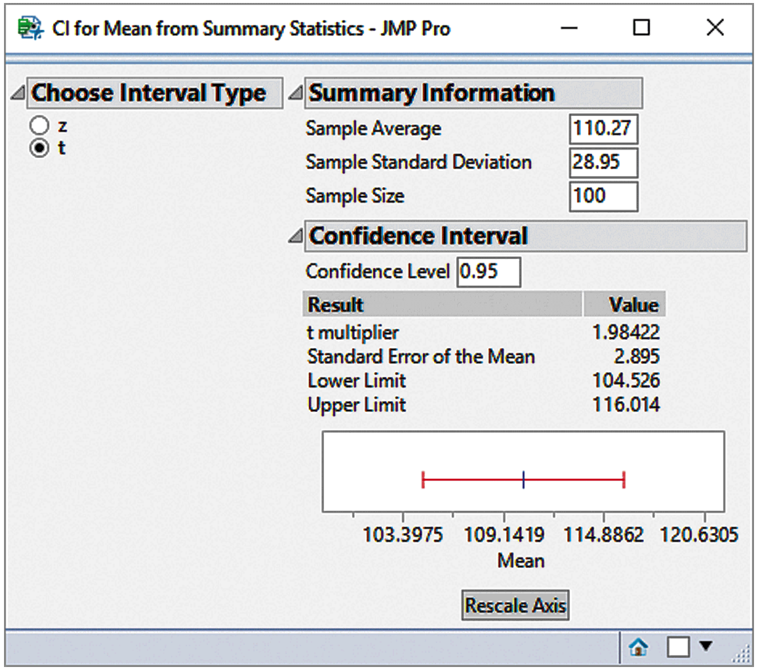
	A	B
1	**Confidence Interval Estimate for the Mean**	
2		
3	**Data**	
4	**Sample Standard Deviation**	**28.95**
5	**Sample Mean**	**110.27**
6	**Sample Size**	**100**
7	**Confidence Level**	**95%**
8		
9	Intermediate Calculations	
10	Standard Error of the Mean	2.895
11	Degrees of Freedom	99
12	*t* Value	1.9842
13	Interval Half Width	5.7443
14		
15	**Confidence Interval**	
16	**Interval Lower Limit**	**104.53**
17	**Interval Upper Limit**	**116.01**

One-Sample T

Descriptive Statistics

N	Mean	StDev	SE Mean	95% CI for μ
100	110.27	28.95	2.90	(104.53, 116.01)

μ: mean of Sample



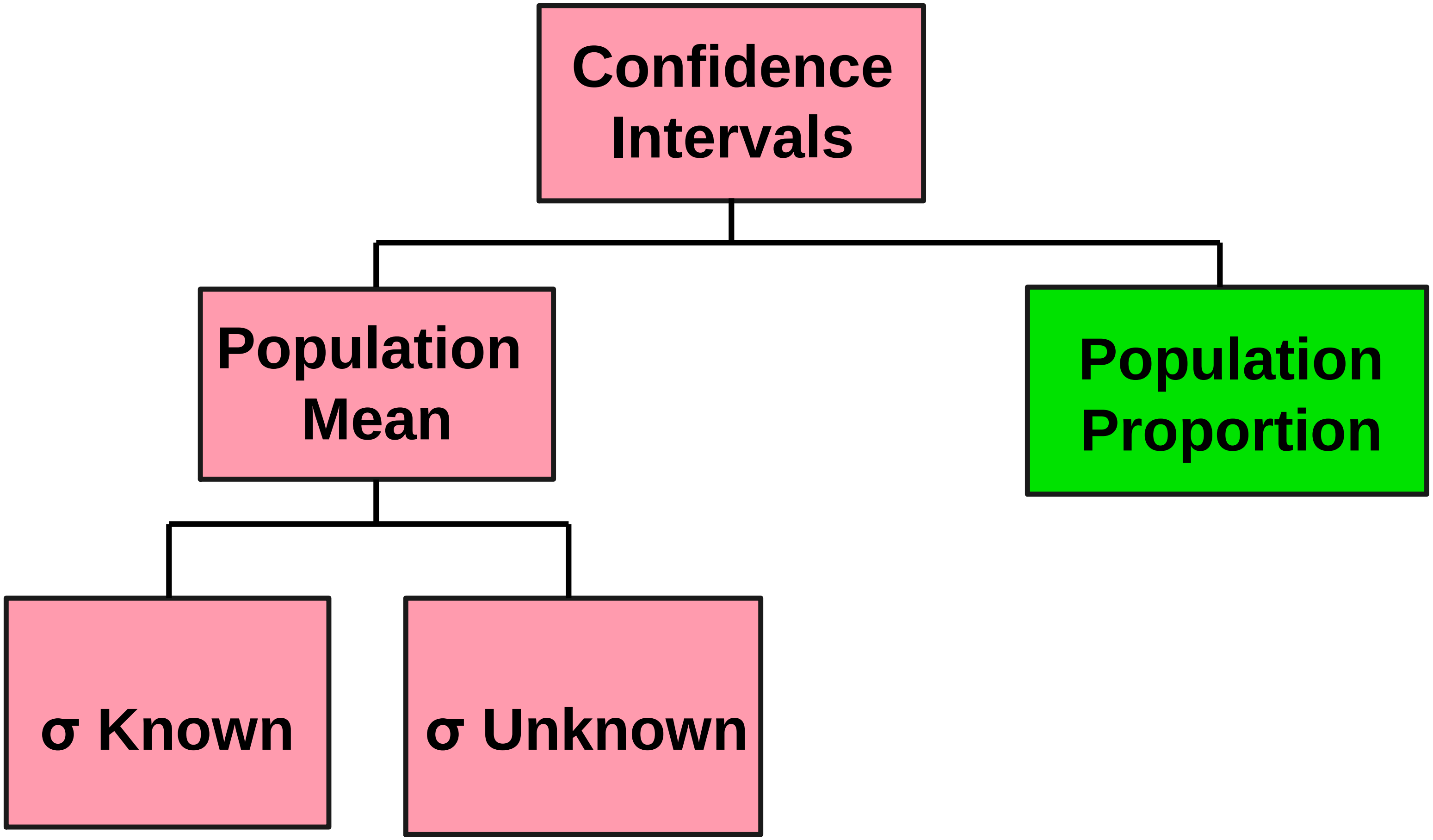
Example of t distribution confidence interval (continued)

DCOVA

- 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

DCOVA



Confidence Intervals for the Population Proportion, π

DCOVA

- 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)

DCOVA

- 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

DCOVA

- 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

DCOVA

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

Example

DCOVA

(continued)

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

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

$$= 25/100 \pm 1.96\sqrt{0.25(0.75)/100}$$

$$= 0.25 \pm 1.96(0.0433)$$

$$\boxed{= 0.1651 \leq \pi \leq 0.3349}$$

Interpretation

DCOVA

- We are 95% confident that the population percentage of left-handers is between

  16.51% and 33.49%.

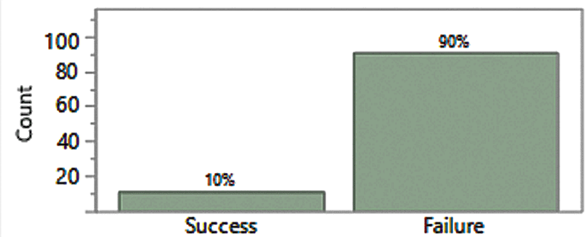
- Although the interval from 0.1651 to 0.3349 may or may not contain the population proportion, 95% of intervals formed from samples of size 100 in this manner will contain the population proportion.

Example of Excel, JMP, & Minitab Confidence Interval for π

DCOVA

	A	B
1	**Proportion of In-Error Sales Invoices**	
2		
3	**Data**	
4	**Sample Size**	**100**
5	**Number of Successes**	**10**
6	**Confidence Level**	**95%**
7		
8	Intermediate Calculations	
9	Sample Proportion	0.1
10	Z Value	-1.9600
11	Standard Error of the Proportion	0.03
12	Interval Half Width	0.0588
13		
14	**Confidence Interval**	
15	**Interval Lower Limit**	**0.0412**
16	**Interval Upper Limit**	**0.1588**

Summary Information



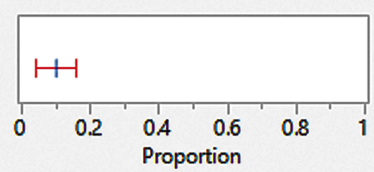
Number of Successes 10

Sample Size 100

Confidence Interval

Confidence Level 0.95

Result	Value
Estimated Proportion	0.1
Standard Error of Proportion	0.0588
Lower Limit	0.0412
Upper Limit	0.1588



Test and CI for One Proportion

Method

p: event proportion

Normal approximation method is used for this analysis.

Descriptive Statistics

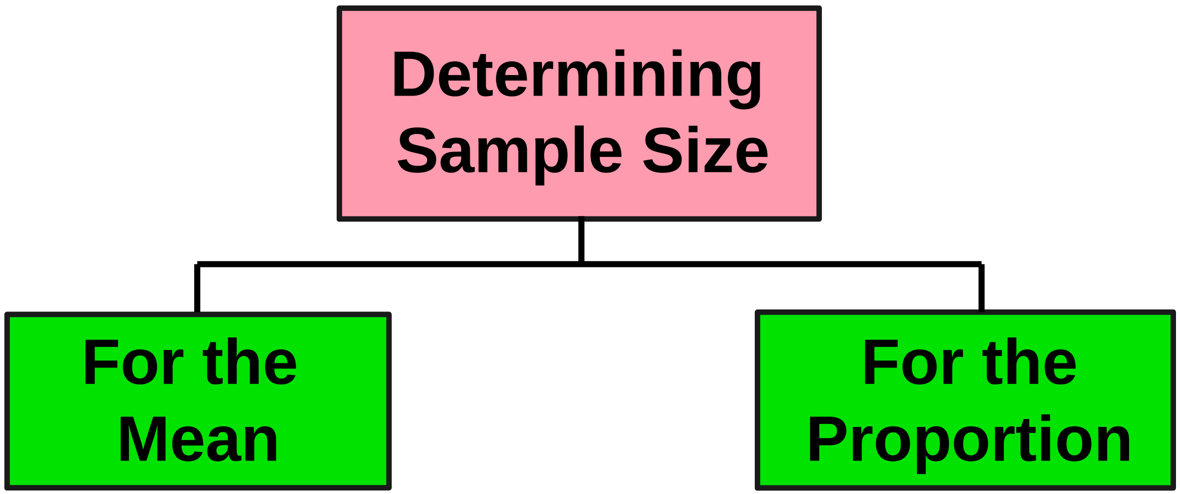
N	Event	Sample p	95% CI for p
100	10	0.100000	(0.041201, 0.158799)

Test

Null hypothesis	H_0: p = 0.5
Alternative hypothesis	H_1: p ≠ 0.5

Determining Sample Size

D<u>C</u>OVA



Sampling Error

DCOVA

- 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

DCOVA

Determining Sample Size

For the Mean

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

Sampling error (margin of error)

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

Determining Sample Size (continued)

DCOVA

Determining Sample Size

For the Mean

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

Now solve for n to get

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

Determining Sample Size (continued)

DCOVA

- 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

DCOVA

The population of download times for a particular video file has a standard deviation of σ = 25 seconds. If a random sample is taken, what sample size is needed to estimate μ within ± 5 seconds with 95% confidence?

$$n = \frac{Z^2 \sigma^2}{e^2} = \frac{(1.96)^2 (25)^2}{5^2} = 96.04$$

So the required sample size is **n = 97**

(Always round up)

Example Using Excel & Minitab For Calculating Sample Size For The Mean Video Download Time

DCOVA

	A	B
1	For the Mean Sales Invoice Amount	
2		
3	Data	
4	Population Standard Deviation	25
5	Sampling Error	5
6	Confidence Level	95%
7		
8	Intemediate Calculations	
9	Z Value	-1.9600
10	Calculated Sample Size	96.0365
11		
12	Result	
13	Sample Size Needed	97

Sample Size for Estimation

Method

Parameter	Mean
Distribution	Normal
Standard deviation	25 (population value)
Confidence level	95%
Confidence interval	Two-sided

Results

Margin of Error	Sample Size
5	97

If σ is unknown

DCOVA

- 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)

D<u>C</u>OVA

Determining Sample Size

For the Proportion

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

Now solve for n to get

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

Determining Sample Size

(continued)

DCOVA

- 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

DCOVA

How large a sample would be necessary to estimate the true proportion of sales invoices containing errors in a large population within ±7%, with 95% confidence?

(Assume a pilot sample yields $p = 0.15$.)

Required Sample Size Example

(continued)

DCOVA

Solution:

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

$e = 0.07$

$p = 0.15$, so use this to estimate π.

$$n = \frac{Z_{\alpha/2}^2 \pi(1-\pi)}{e^2} = \frac{(1.96)^2(0.15)(1-0.15)}{(0.03)^2} = 99.96$$

So use $n = 100$

Example Excel & Minitab Output For Calculating Sample Size For A Proportion

DCOVA

Excel Uses A Normal Approximation To Find n

	A	B
1	For the Proportion of In-Error Sales Invoices	
2		
3	Data	
4	Estimate of True Proportion	0.15
5	Sampling Error	0.07
6	Confidence Level	95%
7		
8	Intermediate Calculations	
9	Z Value	-1.9600
10	Calculated Sample Size	99.9563
11		
12	Result	
13	Sample Size Needed	100

Minitab Uses The Binomial Distribution To Find n

Sample Size for Estimation

Method

Parameter	Proportion
Distribution	Binomial
Proportion	0.15
Confidence level	95%
Confidence interval	Two-sided

Results

Margin of Error	Sample Size
0.07	141

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

In this chapter we discussed:

- The construction and interpretation of confidence interval estimates for the mean and the proportion.
- The determination of the sample size necessary to develop a confidence interval for the mean and the proportion.