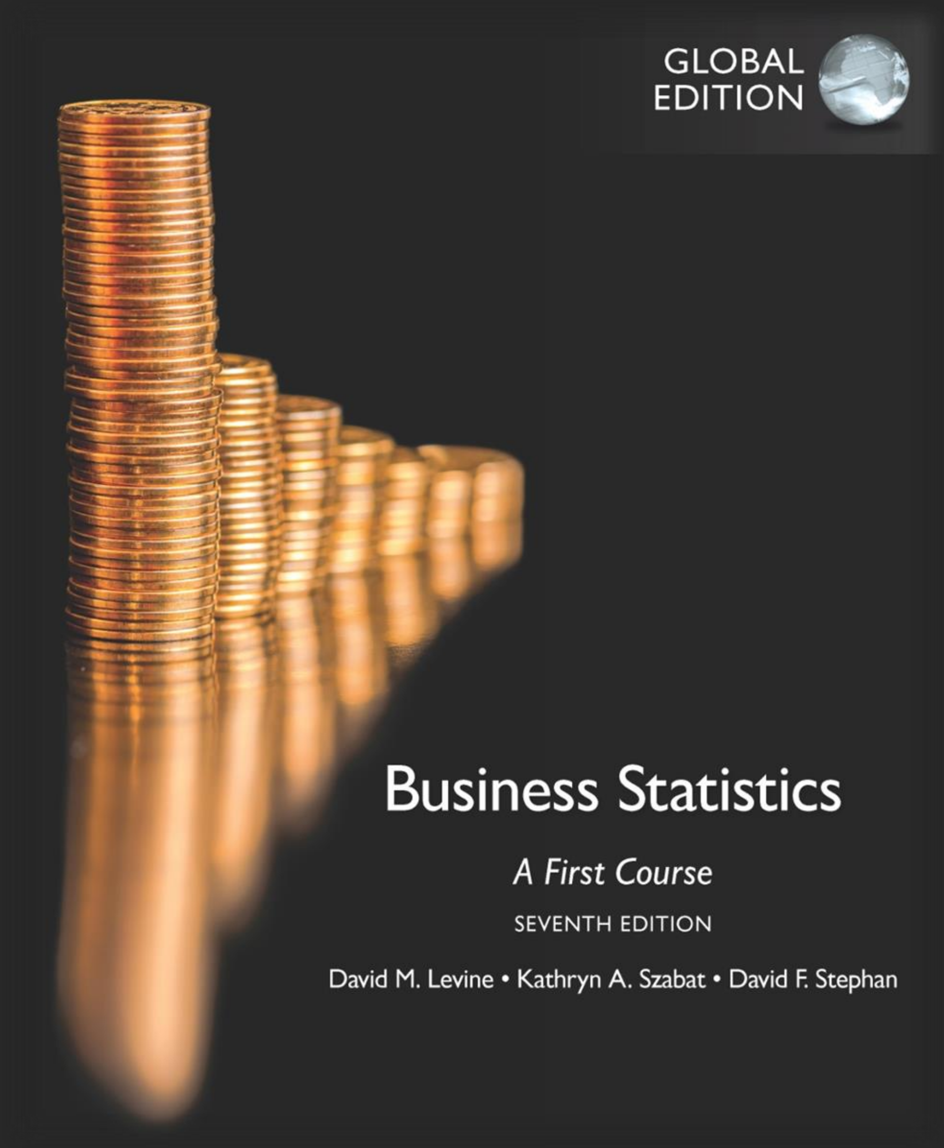


Confidence Interval Estimation

Objectives

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

Chapter Outline

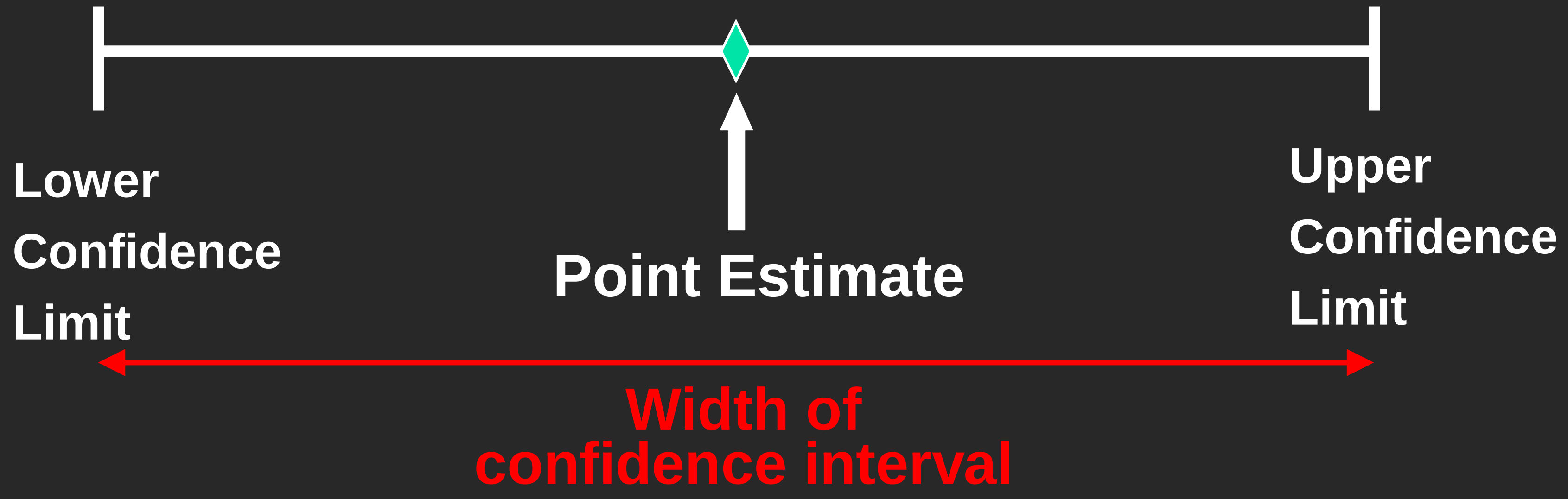
Content of this chapter

- 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
  - $368 \pm 1.96 * 15 / \sqrt{25} = (362.12, 373.88)$. 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 ≤ μ ≤ 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)*

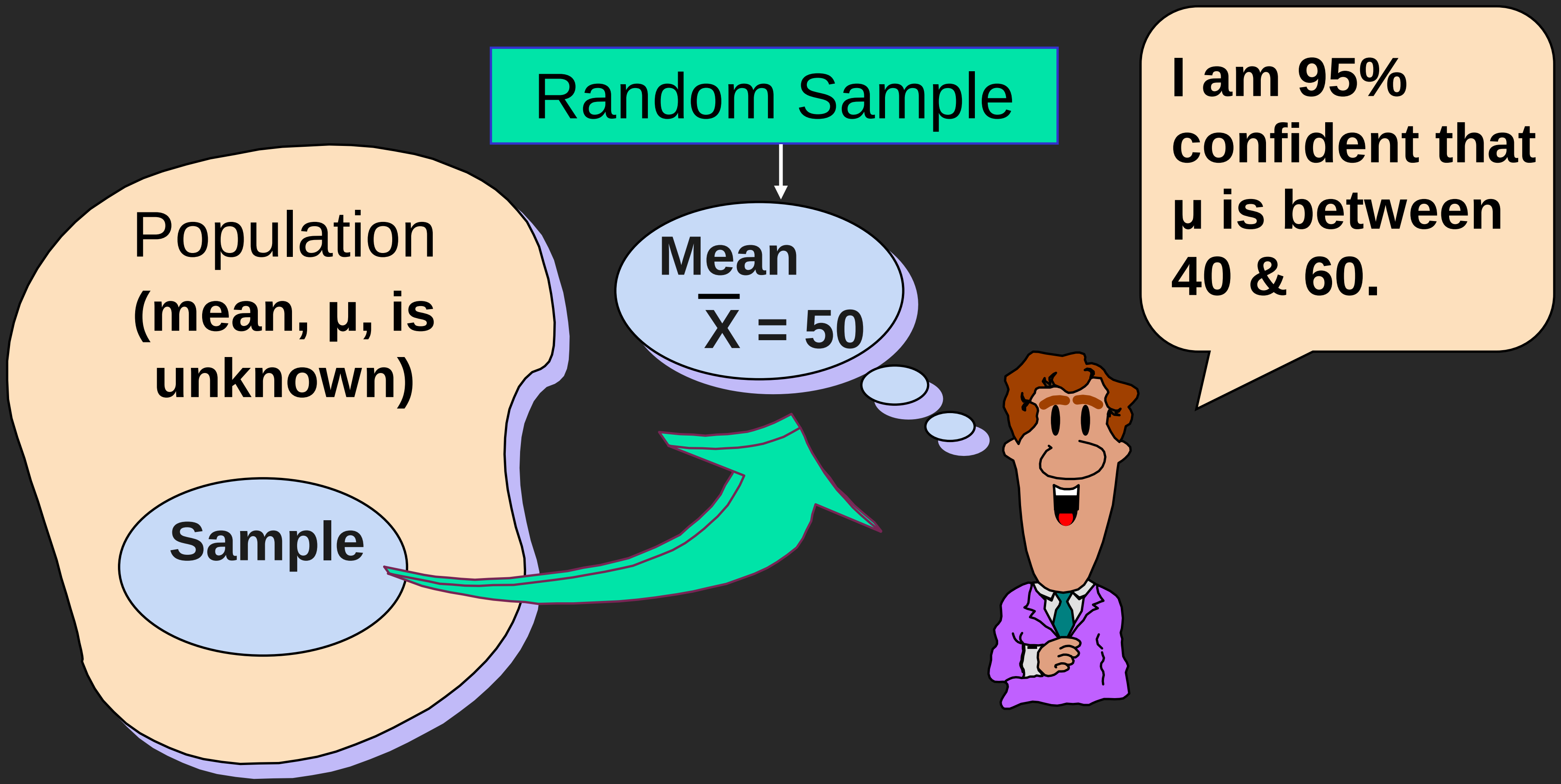
DCOVA

- 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



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%)

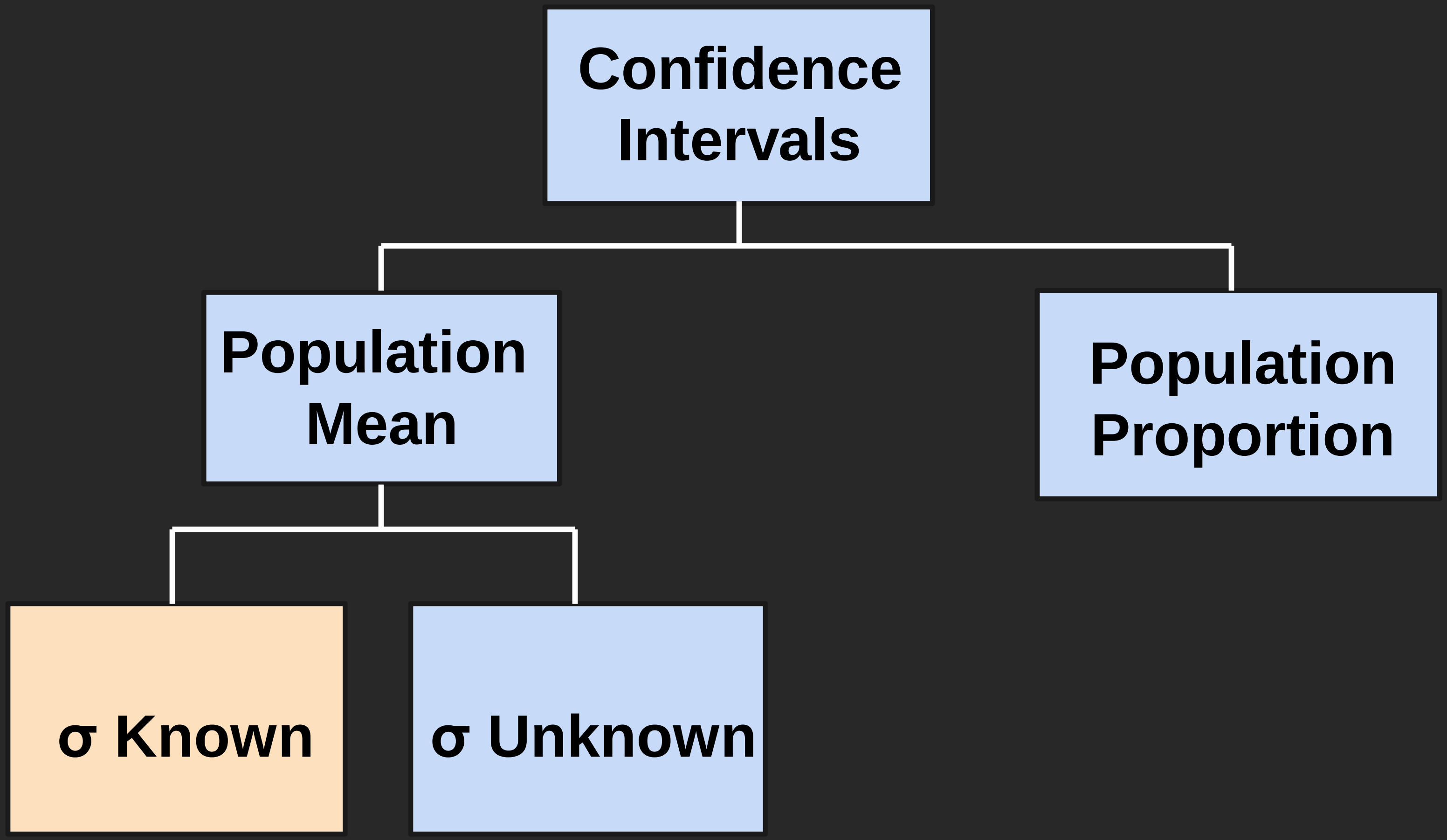
Confidence Level, $(1-\alpha)$ *(continued)*

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:

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

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

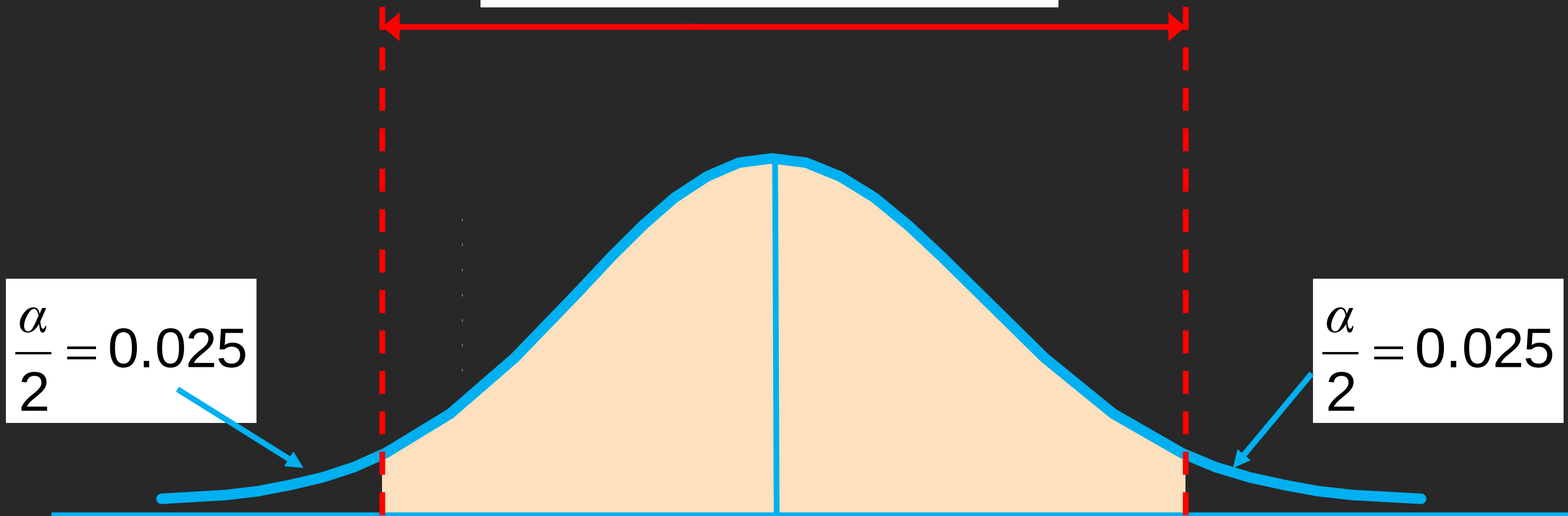
Finding the Critical Value, $Z_{\alpha/2}$

DCOVA

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

- Consider a 95% confidence interval:

$$1 - \alpha = 0.95 \text{ so } \alpha = 0.05$$



Z units:	$Z_{\alpha/2}$ **= -1.96**	**0**	$Z_{\alpha/2}$ **= 1.96**
X units:	**Lower Confidence Limit**	**Point Estimate**	**Upper Confidence Limit**

Common Levels of Confidence

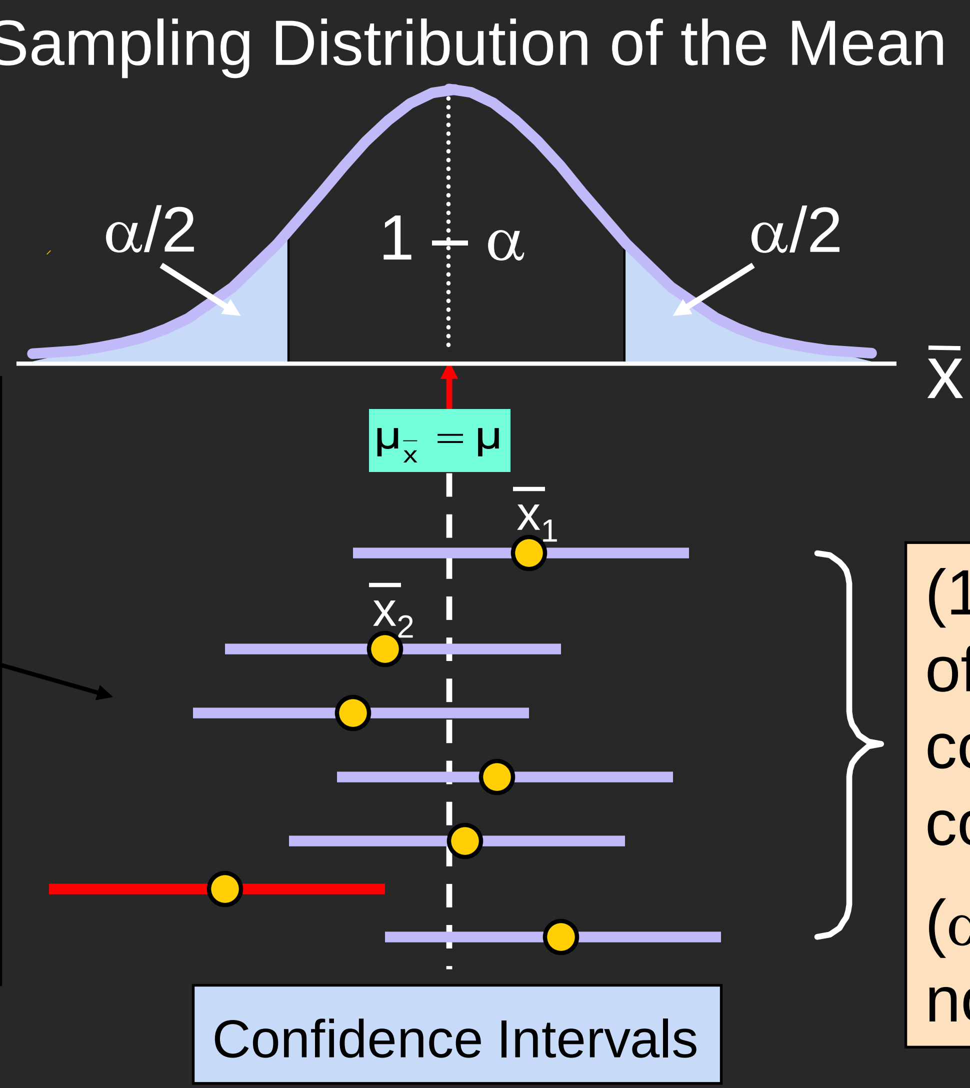
DCOVA

- 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;

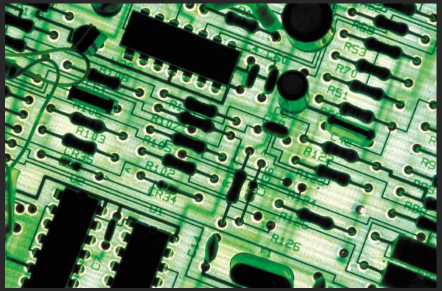
(α)100% do not.

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 *(continued)*

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.

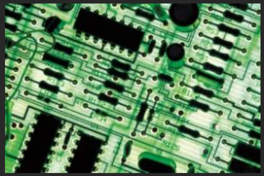
- Solution:

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

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

$$= 2.20 \pm 0.2068$$

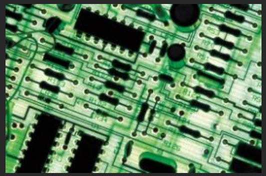
$$1.9932 \le \mu \le 2.4068$$



Interpretation

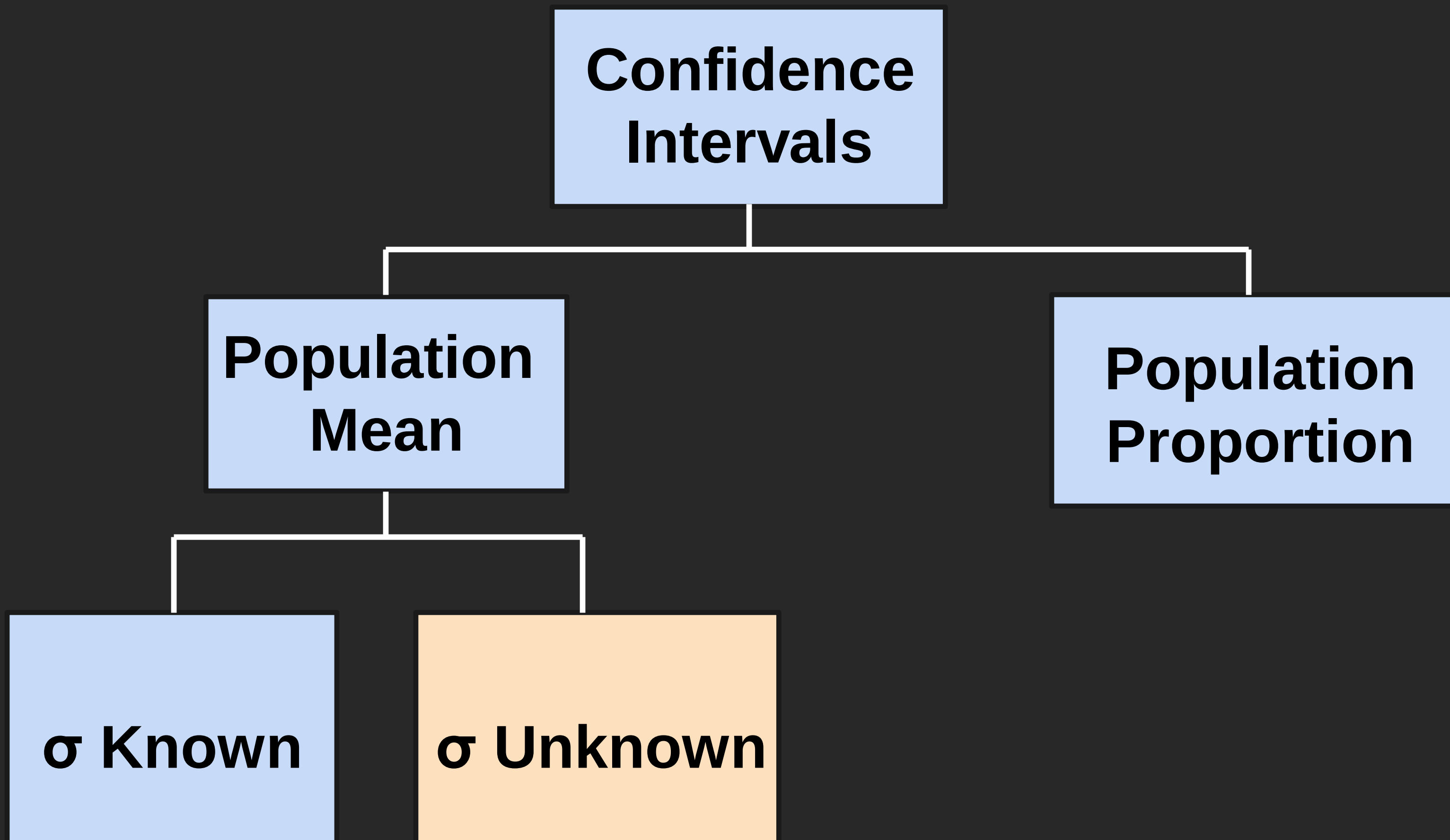
DCOVA

- 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
  - If population is not normal, use large sample (n > 30)
- 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

$$d.f. = n - 1$$

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)

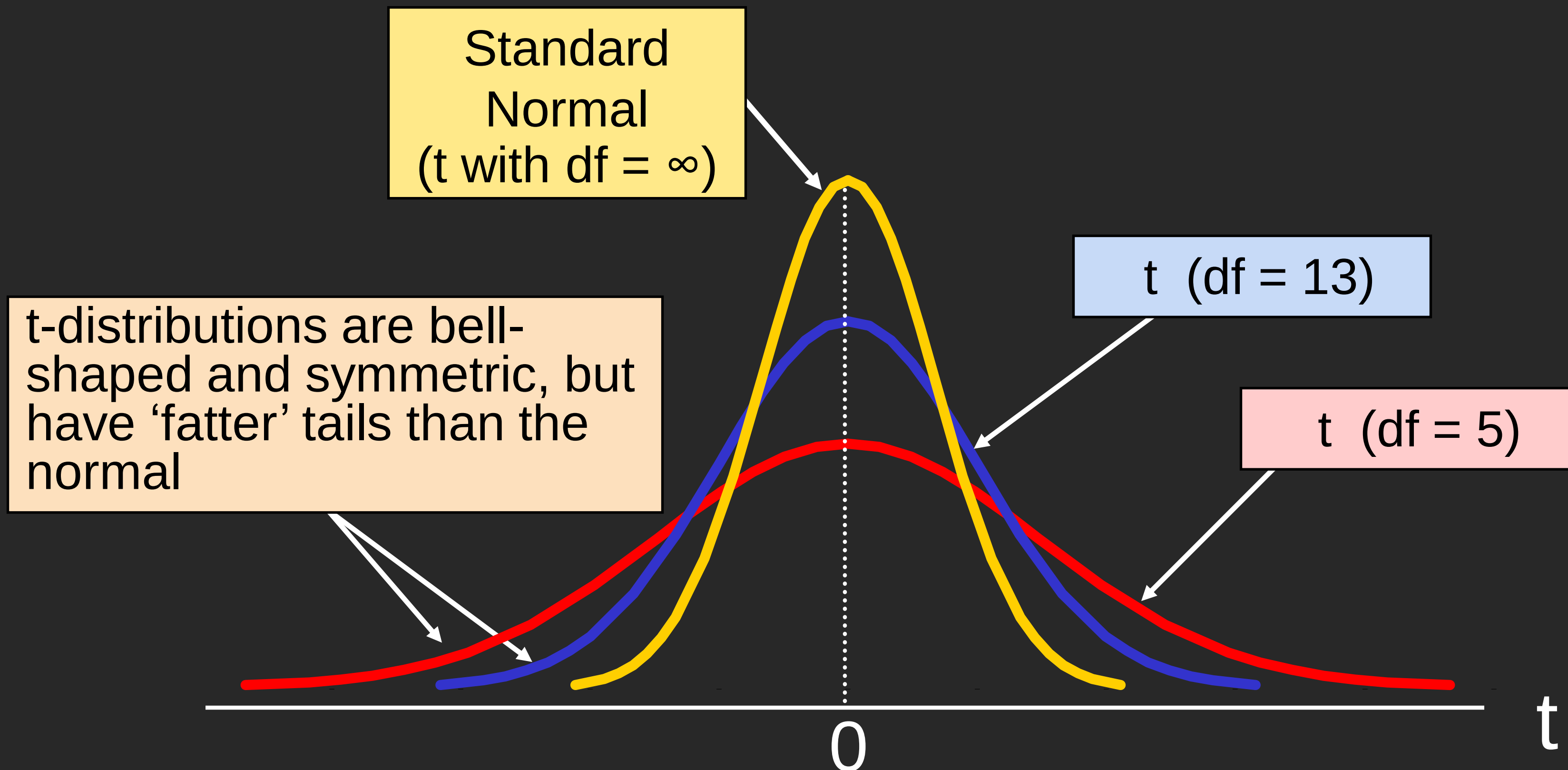
Here, n = 3, so degrees of freedom = n – 1 = 3 – 1 = 2

(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



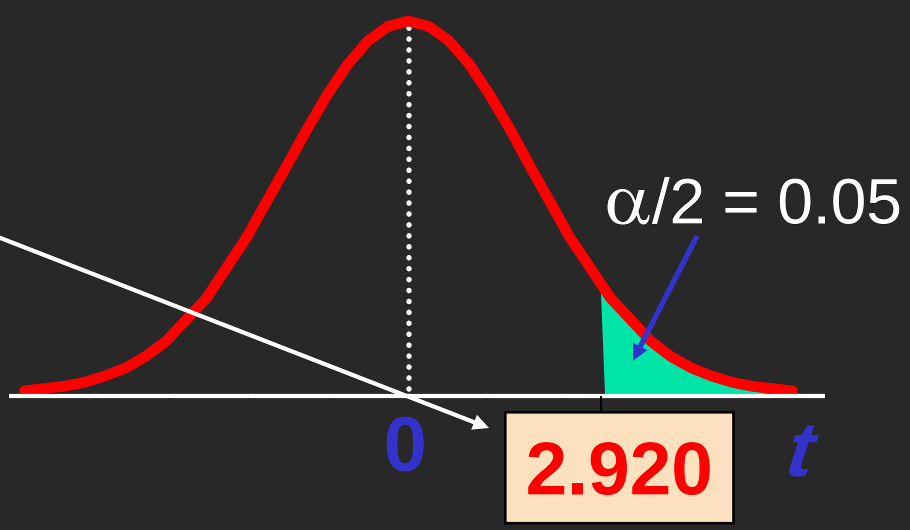
Student's t Table

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
$\alpha = 0.10$
$\alpha/2 = 0.05$

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 ⟶ Z as n increases

Example of t distribution confidence interval

DCOVA

A random sample of n = 25 has $\overline{X}$ = 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

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

$$46.698 \le \mu \le 53.302$$

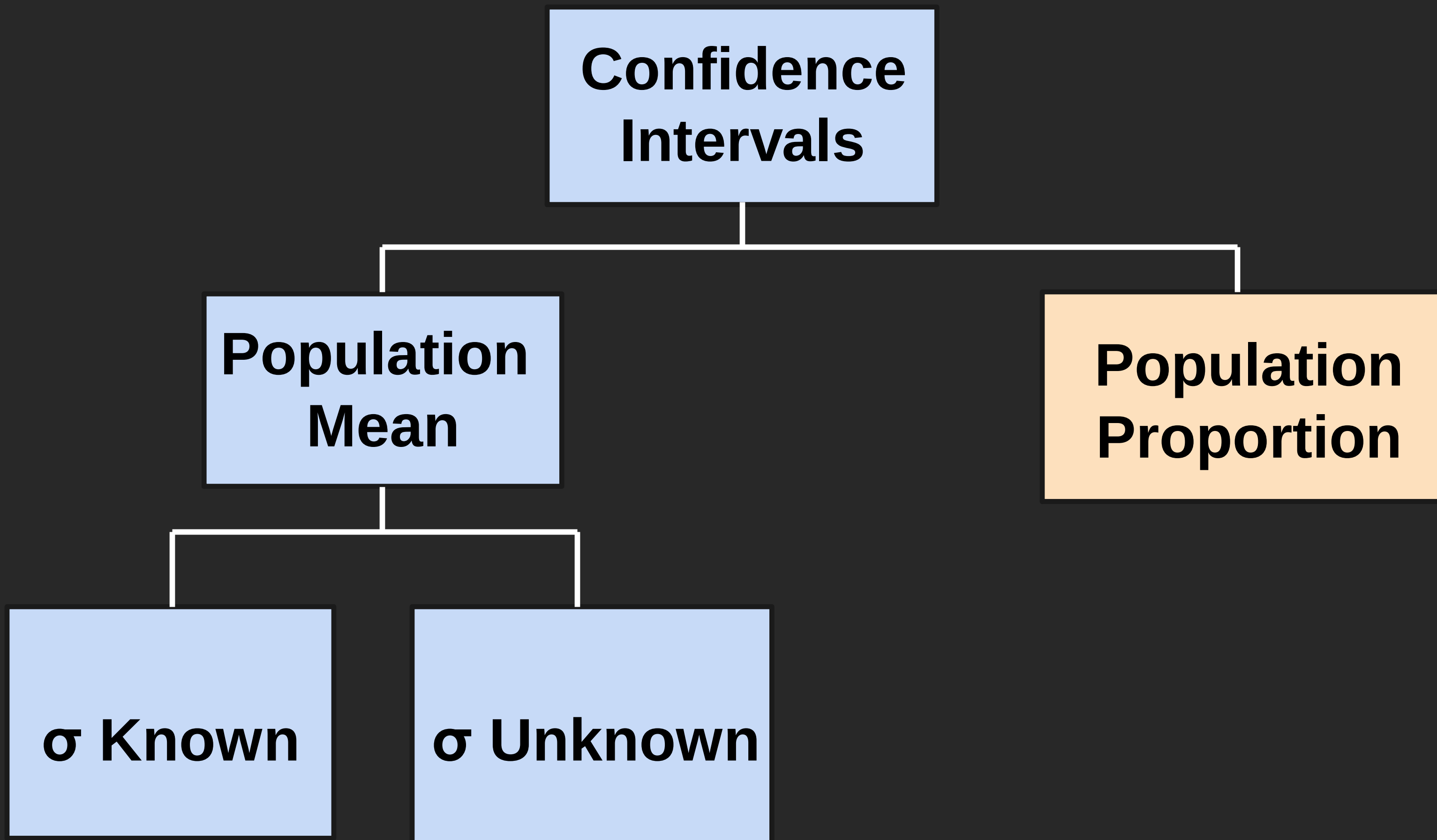
Example of t distribution confidence interval *(continued)*

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

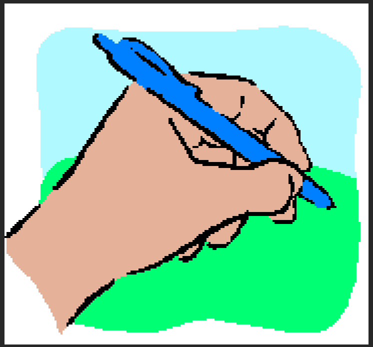
$$p \pm Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

- 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 true proportion of left-handers



Example *(continued)*

DCOVA

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

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

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

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

$$= 0.1651 \;\le\; \pi \le\; 0.3349$$

Interpretation

DCOVA

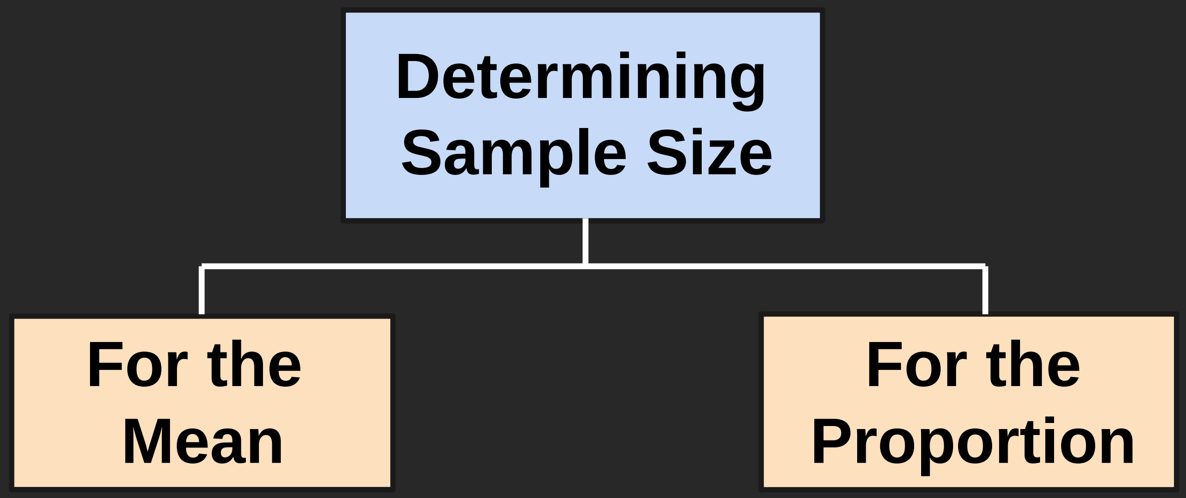
- We are 95% confident that the true percentage of left-handers in the population is between

  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

DCOVA



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

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

Determining Sample Size *(continued)*

DCOVA

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

Required Sample Size Example

DCOVA

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

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)*

DCOVA

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 defectives in a large population within ±3%, with 95% confidence?

(Assume a pilot sample yields p = 0.12)

Required Sample Size Example

(continued)

DCOVA

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

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 population mean and the population proportion
- The determination of the sample size necessary to develop a confidence interval for the population mean or population proportion

Bootstrapping Is A Method To Use When Population Is Not Normal

DCOVA

To estimate a population mean using bootstrapping, you would:

1. Select a random sample of size n without replacement from a population of size N.
2. Resample the initial sample by selecting n values with replacement from the initial sample.
3. Compute $\bar{X}$ from this resample.
4. Repeat steps 2 & 3 m different times.
5. Construct the resampling distribution of $\bar{X}$.
6. Construct an ordered array of the entire set of resampled $\bar{X}$'s.
7. In this ordered array find the value that cuts off the smallest $\alpha/2(100\%)$ and the value that cuts off the largest $\alpha/2(100\%)$. These values provide the lower and upper limits of the bootstrap confidence interval estimate of μ.

Bootstrapping Requires The Use of Software As Minitab or JMP

DCOVA

- Typically a very large number (thousands) of resamples are used.

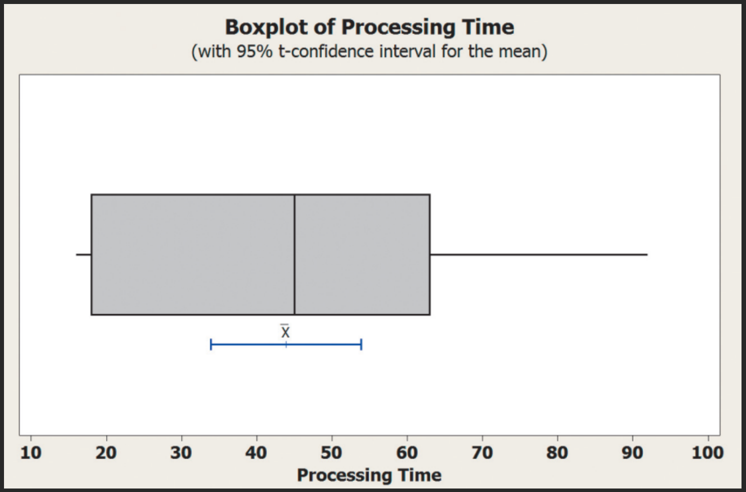
- Software is needed to:
  - Automate the resampling process
  - Calculate the appropriate sample statistic
  - Create the ordered array
  - Find the lower and upper confidence limits

Bootstrapping Example -- Processing Time of Life Insurance Applications

DCOVA

Sample of 27 times taken without replacement from population

73 19 16 64 28 28 31 90 60 56 31 56 22 18 45 48 17 17 17 91 92 63 50 51 69 16 17



From boxplot conclude population is not normal so t confidence interval is not appropriate.

Use bootstrapping to form a confidence interval for μ.

Comparing the original sample to the first resample with replacement

DCOVA

Sample of 27 times taken without replacement from population

73 19 16 64 28 28 31 90 60 56 31 56 22 18 45 48 17 17 17 91 92 63 50 51 69 16 17

The initial bootstrap resample omits some values (18, 45, 50, 63, and 91) that appear in the initial sample above. Note that the value of 73 appears twice even though it appears only once in the initial sample.

16 16 16 17 17 17 17 17 19 22 28 31 31 51 56 56 60 60 64 64 64 64 69 **73 73** 90 92

The Ordered Array of Sample Means for 100 Resamples

DCOVA

Fifth Smallest

31.5926	33.9259	35.4074	36.5185	**36.6296**	36.9630	37.0370	37.0741
37.1481	37.3704	37.9259	38.1111	38.1481	38.2222	38.2963	38.7407
38.8148	38.8519	38.8889	39.0000	39.1852	39.3333	39.3704	39.6667
40.1481	40.5185	40.6296	40.9259	40.9630	41.2593	41.2963	41.7037
41.8889	42.0741	42.1111	42.1852	42.8519	43.0741	43.1852	43.3704
43.4444	43.7037	43.8148	43.8519	43.8519	43.9259	43.9630	44.1481
44.4074	44.5556	44.7778	45.0000	45.4444	45.5185	45.5556	45.6667
45.7407	45.8519	45.9630	45.9630	46.0000	46.1111	46.2963	46.2963
46.3333	46.3333	46.4815	46.6667	46.7407	46.9630	47.0741	47.2222
47.2963	47.3704	47.4815	47.4815	47.5556	47.6667	47.8519	48.5185
48.8889	49.0000	49.2222	49.4444	49.4815	49.4815	49.6296	49.6296
49.7407	50.2963	50.4074	50.5926	50.9259	51.4074	51.4815	**51.5926**
51.9259	52.3704	53.4074	54.3333				

Fifth Largest

Finding a 90% Bootstrap CI for The Population Mean

DCOVA

- To find the 90% CI for 100 resamples we need to find the $0.05(100) = 5^{th}$ smallest and the 5^{th} largest values.

- From the table the 5^{th} smallest value is 36.6296 and the 5^{th} largest value is 51.5926.

- The 90% bootstrap CI is (36.6296, 51.5926)

Topic Summary

- In this topic we discussed
- The concept of bootstrapping.
- When bootstrapping might be appropriate to use.
- How to use bootstrapping to form a confidence interval for a population mean.