Chapter 7

Sampling Distributions

Learning Objectives

In this chapter, you learn:

- The concept of the sampling distribution
- To compute probabilities related to the sample mean and the sample proportion
- The importance of the Central Limit Theorem

Sampling Distributions

DCOVA

- A sampling distribution is a distribution of all of the possible values of a sample statistic for a given size sample selected from a population.

- For example, suppose you sample 50 students from your college regarding their mean GPA. If you obtained many different samples of 50, you will compute a different mean for each sample. We are interested in the distribution of all potential mean GPAs we might calculate for any given sample of 50 students.

Developing a Sampling Distribution

DCOVA

- **Assume there is a population …**
- Population size N=4
- Random variable, X, is age of individuals
- Values of X: 18, 20, 22, 24 (years)

Developing a Sampling Distribution

(continued)

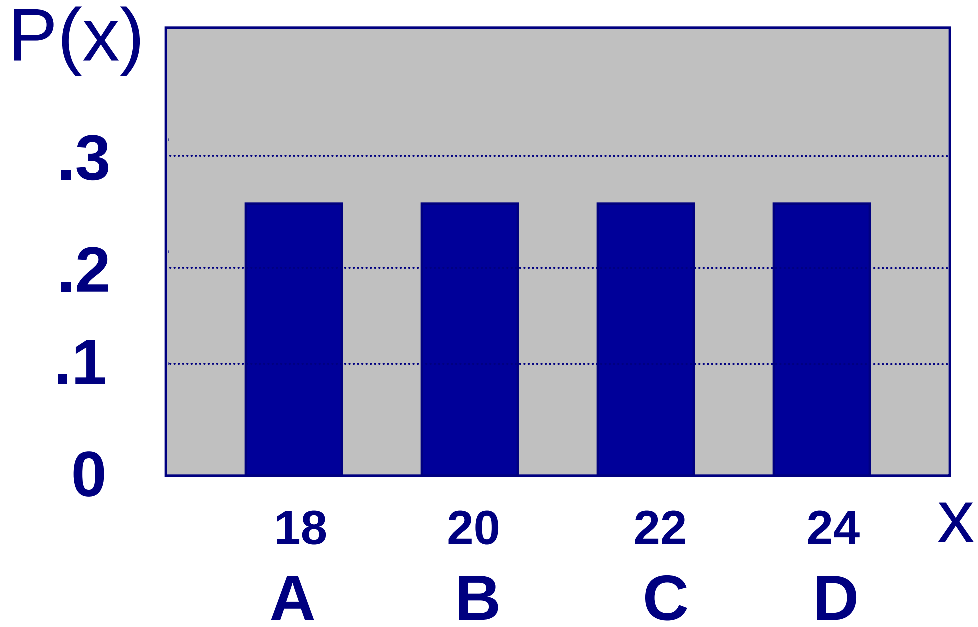
DCOVA

Summary Measures for the Population Distribution:

$$\mu = \frac{\sum X_i}{N}$$

$$= \frac{18 + 20 + 22 + 24}{4} = 21$$

$$\sigma = \sqrt{\frac{\sum (X_i - \mu)^2}{N}} = 2.236$$



Uniform Distribution

Developing a Sampling Distribution *(continued)*

DCOVA

Now consider all possible samples of size n=2

1st Obs	2nd Observation			
	18	20	22	24
18	18,18	18,20	18,22	18,24
20	20,18	20,20	20,22	20,24
22	22,18	22,20	22,22	22,24
24	24,18	24,20	24,22	24,24

16 possible samples (sampling with replacement)

16 Sample Means

1st Obs	2nd Observation			
	18	20	22	24
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

Developing a Sampling Distribution

DCOVA

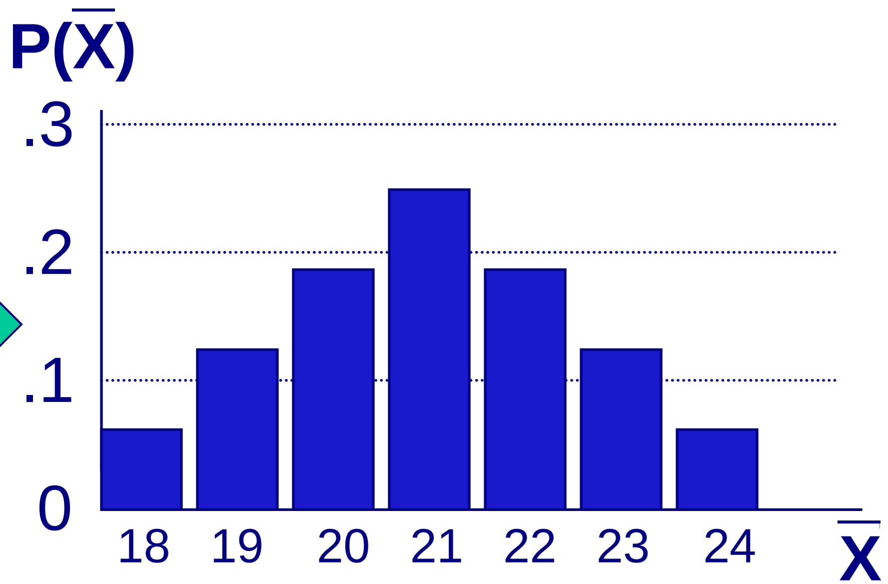
(continued)

Sampling Distribution of All Sample Means

16 Sample Means

1st Obs	2nd Observation			
	18	**20**	**22**	**24**
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

Sample Means Distribution



(no longer uniform)

Developing a Sampling Distribution

DCOVA

(continued)

Summary Measures of this Sampling Distribution:

$$\mu_{\bar{X}} = \frac{18 + 19 + 19 + \cdots + 24}{16} = 21$$

$$\sigma_{\bar{X}} = \sqrt{\frac{(18-21)^2 + (19-21)^2 + \cdots + (24-21)^2}{16}} = 1.58$$

Note: Here we divide by 16 because there are 16 different samples of size 2.

Comparing the Population Distribution to the Sample Means Distribution

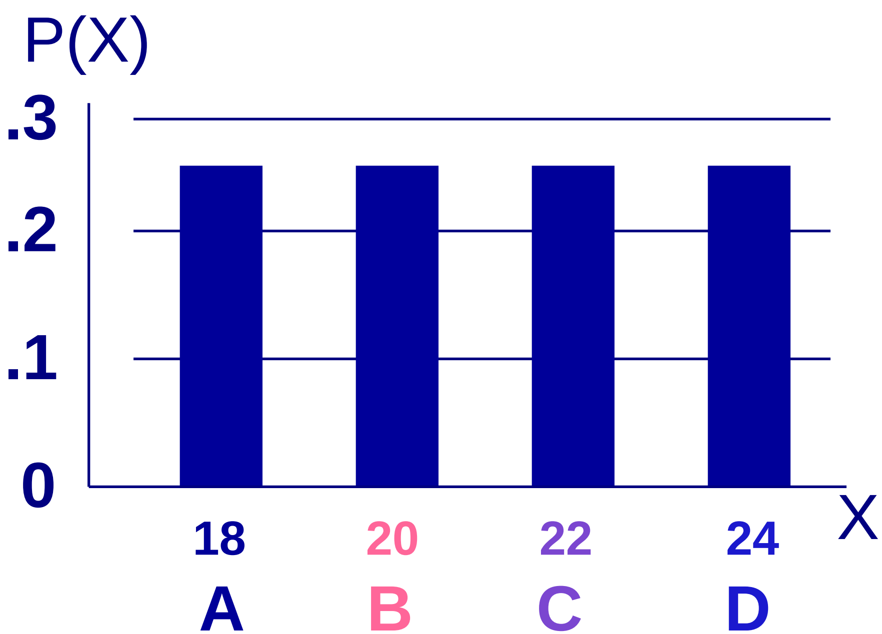
DCOVA

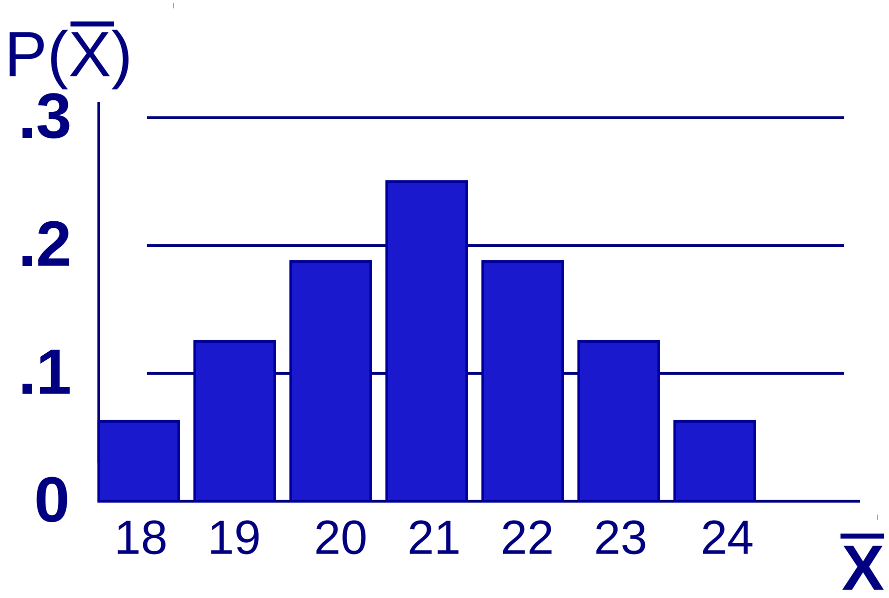
Population
N = 4

$\mu = 21 \qquad \sigma = 2.236$

Sample Means Distribution
n = 2

$\mu_{\bar{X}} = 21 \qquad \sigma_{\bar{X}} = 1.58$





Sample Mean Sampling Distribution: Standard Error of the Mean

DCOVA

- Different samples of the same size from the same population will yield different sample means
- A measure of the variability in the mean from sample to sample is given by the Standard Error of the Mean:

  (This assumes that sampling is with replacement or sampling is without replacement from an infinite population)

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

- Note that the standard error of the mean decreases as the sample size increases

Sample Mean Sampling Distribution: If the Population is Normal

DCOVA

- If a population is normal with mean μ and standard deviation σ, the sampling distribution of $\bar{X}$ is also normally distributed with

$$\boxed{\mu_{\bar{X}} = \mu} \quad \text{and} \quad \boxed{\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}}$$

Z-value for Sampling Distribution of the Mean

DCOVA

- Z-value for the sampling distribution of $\overline{X}$:

$$Z = \frac{(\overline{X} - \mu_{\overline{X}})}{\sigma_{\overline{X}}} = \frac{(\overline{X} - \mu)}{\frac{\sigma}{\sqrt{n}}}$$

where:

$\overline{X}$ = sample mean

μ = population mean

σ = population standard deviation

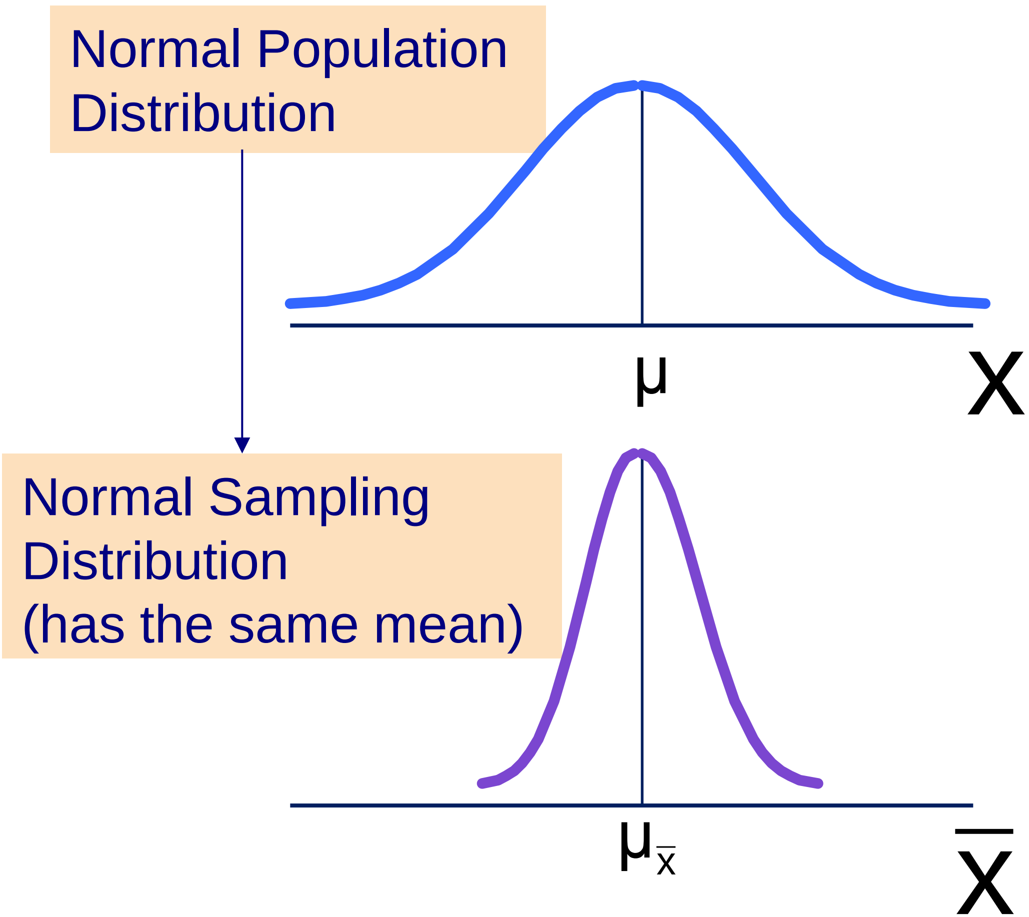
n = sample size

Sampling Distribution Properties

DCOVA

- $$\mu_{\bar{X}} = \mu$$

(i.e. $\bar{X}$ is unbiased)



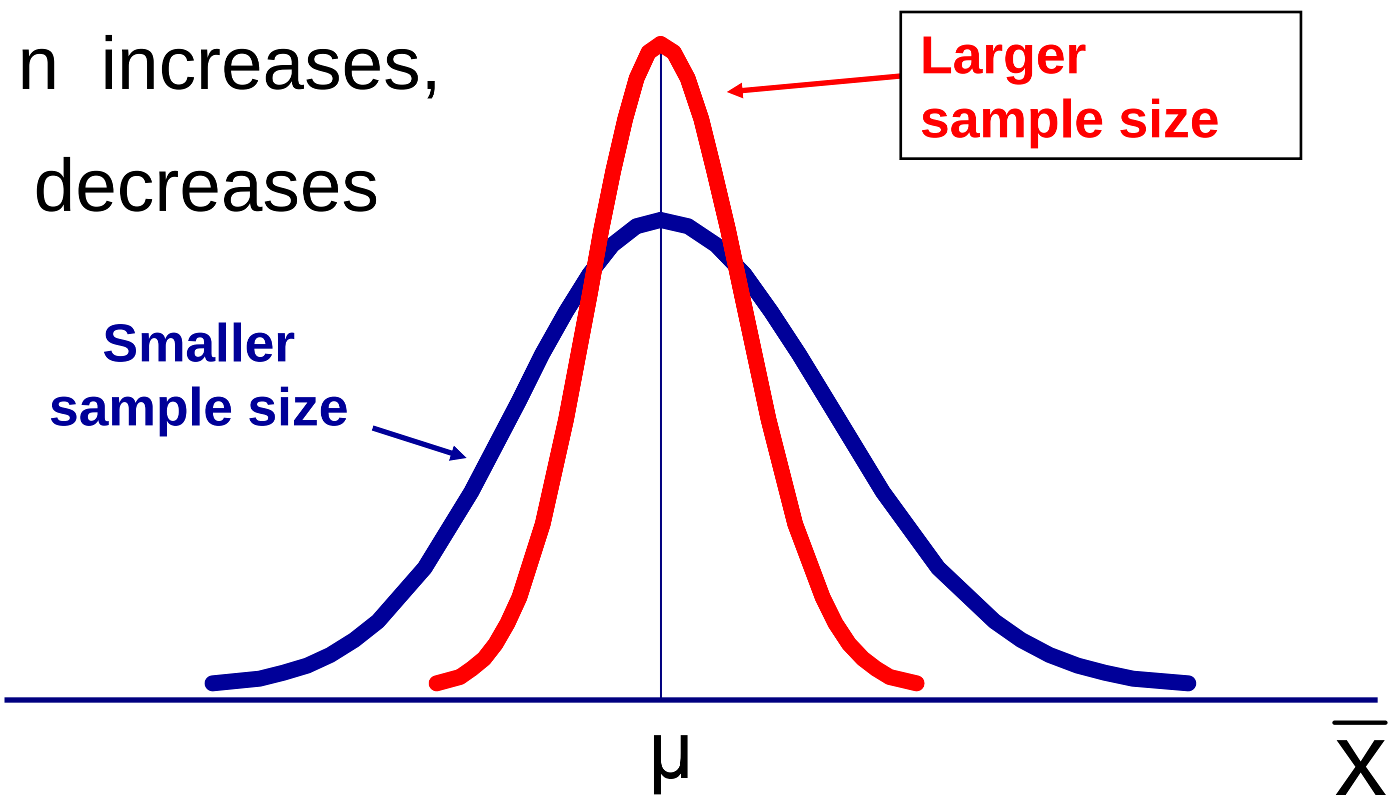
Sampling Distribution Properties

(continued)

DCOVA

As n increases,

$\sigma_{\bar{x}}$ decreases



Determining An Interval Including A Fixed Proportion of the Sample Means

DCOVA

Find a symmetrically distributed interval around μ that will include 95% of the sample means when μ = 368, σ = 15, and n = 25.

- Since the interval contains 95% of the sample means 5% of the sample means will be outside the interval
- Since the interval is symmetric 2.5% will be above the upper limit and 2.5% will be below the lower limit.
- From the standardized normal table, the Z score with 2.5% (0.0250) below it is -1.96 and the Z score with 2.5% (0.0250) above it is 1.96.

Determining An Interval Including A Fixed Proportion of the Sample Means

(continued)

DCOVA

- Calculating the lower limit of the interval

$$\bar{X}_L = \mu + Z\frac{\sigma}{\sqrt{n}} = 368 + (-1.96)\frac{15}{\sqrt{25}} = 362.12$$

- Calculating the upper limit of the interval

$$\bar{X}_U = \mu + Z\frac{\sigma}{\sqrt{n}} = 368 + (1.96)\frac{15}{\sqrt{25}} = 373.88$$

- 95% of all sample means of sample size 25 are between 362.12 and 373.88

Sample Mean Sampling Distribution: If the Population is not Normal DCOVA

- We can apply the Central Limit Theorem:
  - Even if the population is not normal,
  - …sample means from the population will be approximately normal as long as the sample size is large enough.

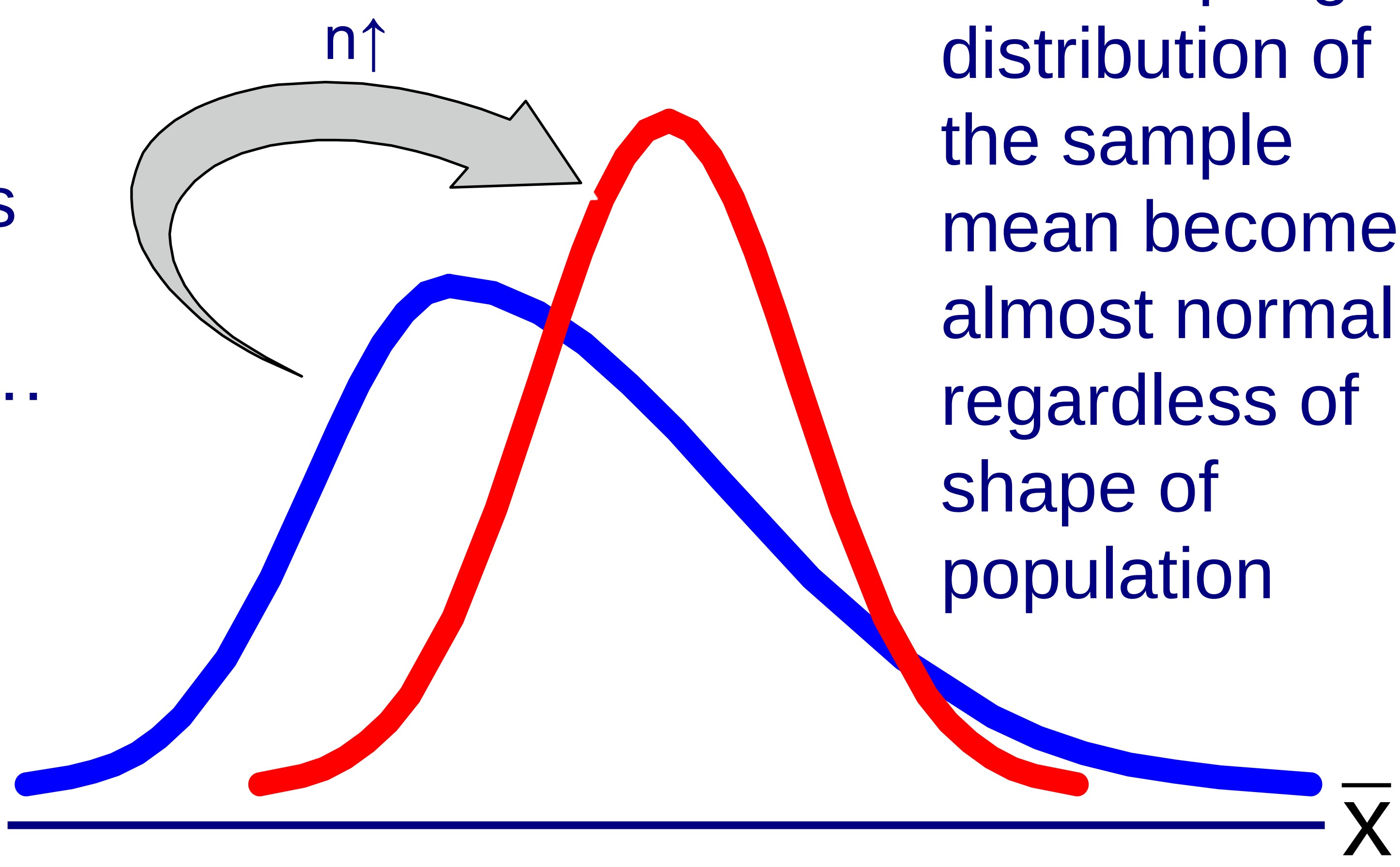
Properties of the sampling distribution:

$$\boxed{\mu_{\bar{x}} = \mu} \quad \text{and} \quad \boxed{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}}$$

Central Limit Theorem

DCOVA

As the sample size gets large enough…



the sampling distribution of the sample mean becomes almost normal regardless of shape of population

Sample Mean Sampling Distribution: If the Population is not Normal

(continued)

DCOVA

Sampling distribution properties:

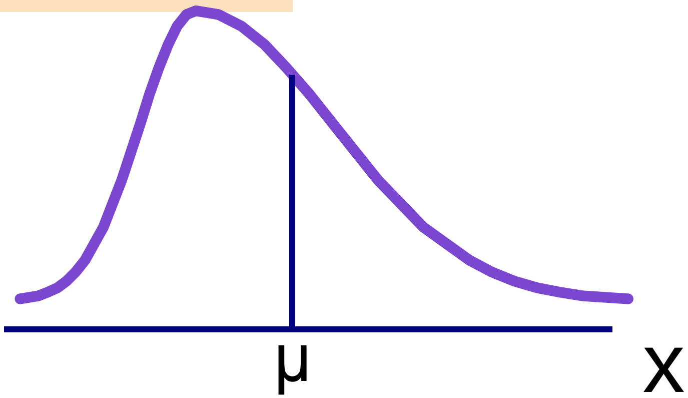
Central Tendency

$$\mu_{\bar{X}} = \mu$$

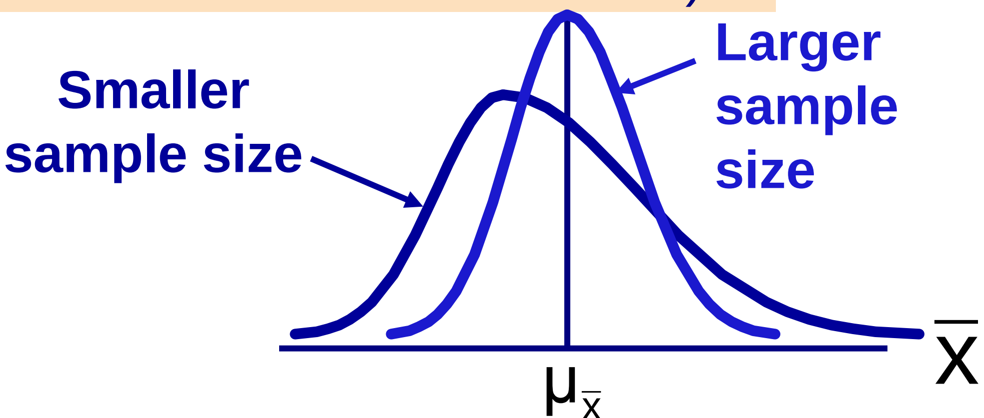
Variation

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Population Distribution



Sampling Distribution
(becomes normal as n increases)



How Large is Large Enough?

DCOVA

- For most distributions, $n > 30$ will give a sampling distribution that is nearly normal
- For fairly symmetric distributions, $n > 15$
- For normal population distributions, the sampling distribution of the mean is always normally distributed

Example

DCOVA

- Suppose a population has mean $\mu = 8$ and standard deviation $\sigma = 3$. Suppose a random sample of size $n = 36$ is selected.

- What is the probability that the sample mean is between 7.8 and 8.2?

Example

Solution:

(continued)

DCOVA

- Even if the population is not normally distributed, the central limit theorem can be used (n > 30)
- … so the sampling distribution of $\bar{X}$ is approximately normal
- … with mean $\mu_{\bar{X}} = 8$
- …and standard deviation $\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{3}{\sqrt{36}} = 0.5$

Example

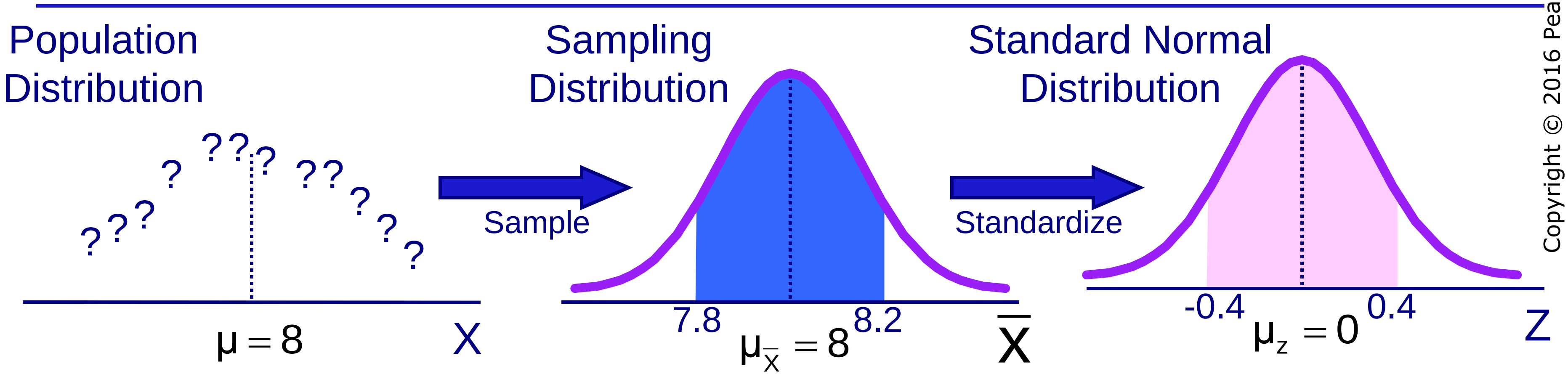
(continued)

DCOVA

Solution (continued):

$$P(7.8 < \overline{X} < 8.2) = P\left(\frac{7.8-8}{3/\sqrt{36}} < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < \frac{8.2-8}{3/\sqrt{36}} \right)$$

$$= P(-0.4 < Z < 0.4) = 0.6554 - 0.3446 = \boxed{0.3108}$$



Population Proportions

DCOVA

π = the proportion of the population having some characteristic

- Sample proportion (p) provides an estimate of π:

$$p = \frac{X}{n} = \frac{\text{number of items in the sample having the characteristic of interest}}{\text{sample size}}$$

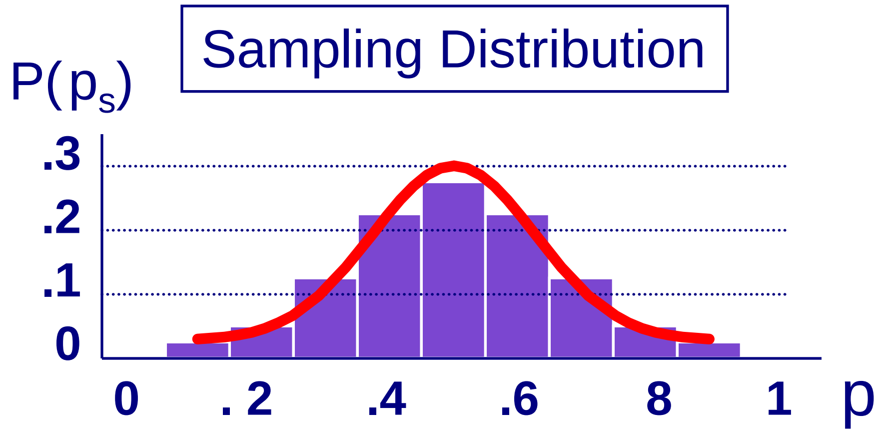
- $0 \leq p \leq 1$
- p is approximately distributed as a normal distribution when n is large

(assuming sampling with replacement from a finite population or without replacement from an infinite population)

Sampling Distribution of p

DCOVA

- Approximated by a normal distribution if:
  - $n\pi \geq 5$ and $n(1-\pi) \geq 5$



where

$$\mu_p = \pi$$

and

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

(where π = population proportion)

Z-Value for Proportions

DCOVA

Standardize p to a Z value with the formula:

$$Z = \frac{p - \pi}{\sigma_p} = \frac{p - \pi}{\sqrt{\dfrac{\pi(1-\pi)}{n}}}$$

Example

DCOVA

- If the true proportion of voters who support Proposition A is $\pi = 0.4$, what is the probability that a sample of size 200 yields a sample proportion between 0.40 and 0.45?

- i.e.: **if $\pi = 0.4$ and $n = 200$, what is**

  $P(0.40 \le p \le 0.45)$?

Example

(continued)

DCOVA

- **if $\pi = 0.4$ and $n = 200$, what is $P(0.40 \le p \le 0.45)$?**

Find σ_p :

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0.4(1-0.4)}{200}} = 0.03464$$

Convert to standardized normal:

$$P(0.40 \le p \le 0.45) = P\left(\frac{0.40-0.40}{0.03464} \le Z \le \frac{0.45-0.40}{0.03464}\right)$$

$$= P(0 \le Z \le 1.44)$$

Example

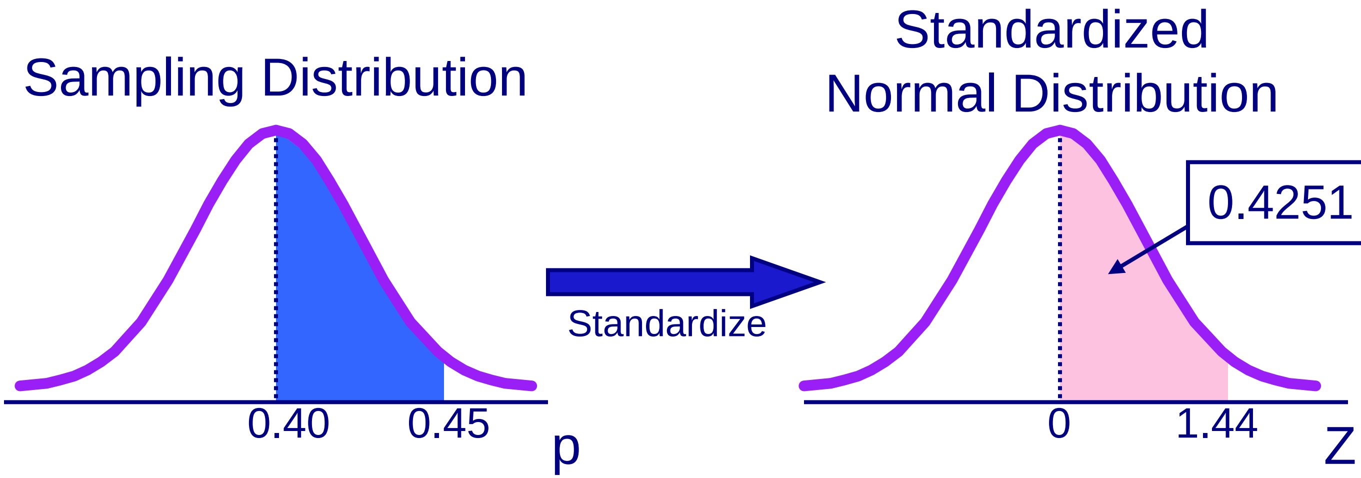
(continued)

DCOVA

- if $\pi = 0.4$ and $n = 200$, what is $P(0.40 \leq p \leq 0.45)$?

Utilize the cumulative normal table:

$P(0 \leq Z \leq 1.44) = 0.9251 - 0.5000 = \boxed{0.4251}$



Chapter Summary

In this chapter we discussed

- Sampling distributions
- The sampling distribution of the mean
  - For normal populations
  - Using the Central Limit Theorem
- The sampling distribution of a proportion
- Calculating probabilities using sampling distributions

Confidence Intervals

- Confidence Intervals for the Population Mean, μ
  - when Population Standard Deviation σ is Known
  - when Population Standard Deviation σ is Unknown

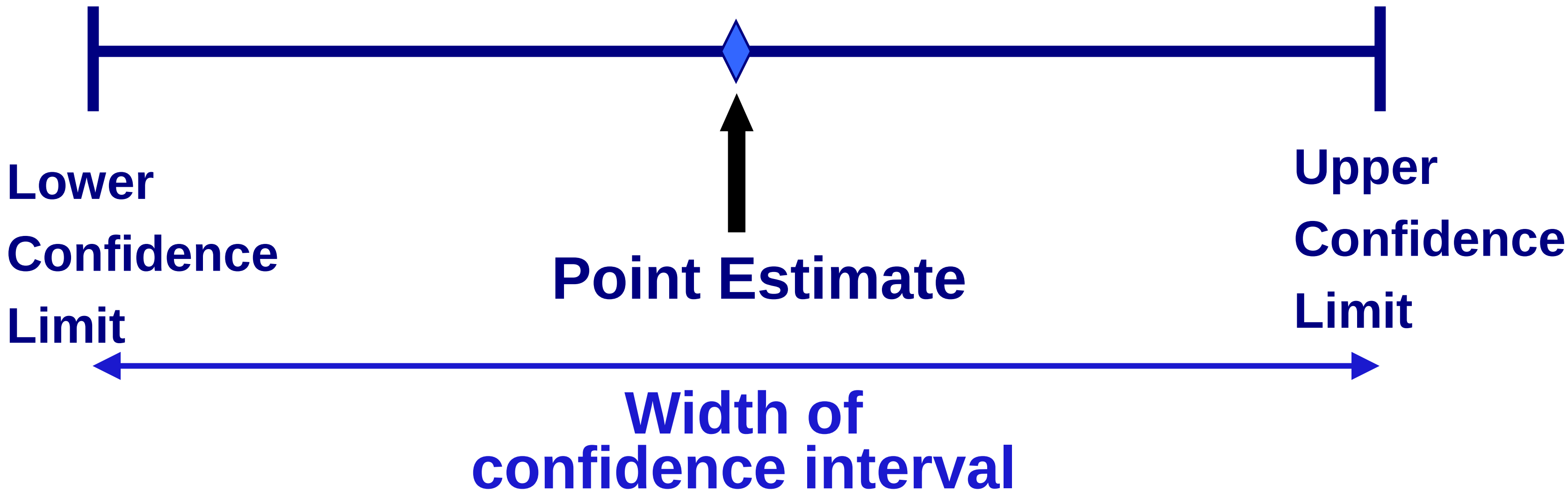
- Confidence Intervals for the Population Proportion, π

- 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

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 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
  - 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

DCOVA

(continued)

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

DCOVA

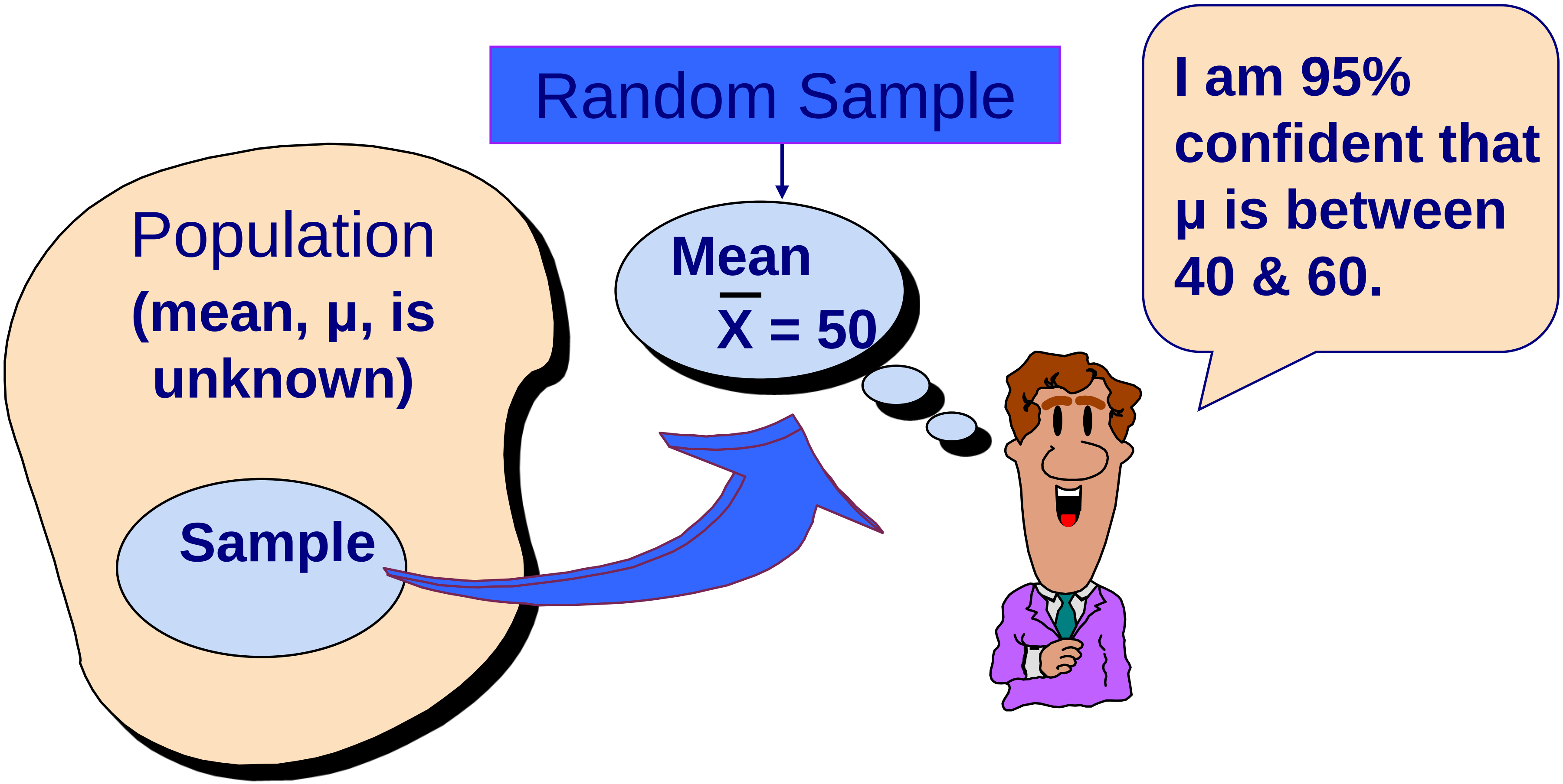
(continued)

- 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

DCOVA



Confidence Level

DCOVA

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

Confidence Level, $(1-\alpha)$

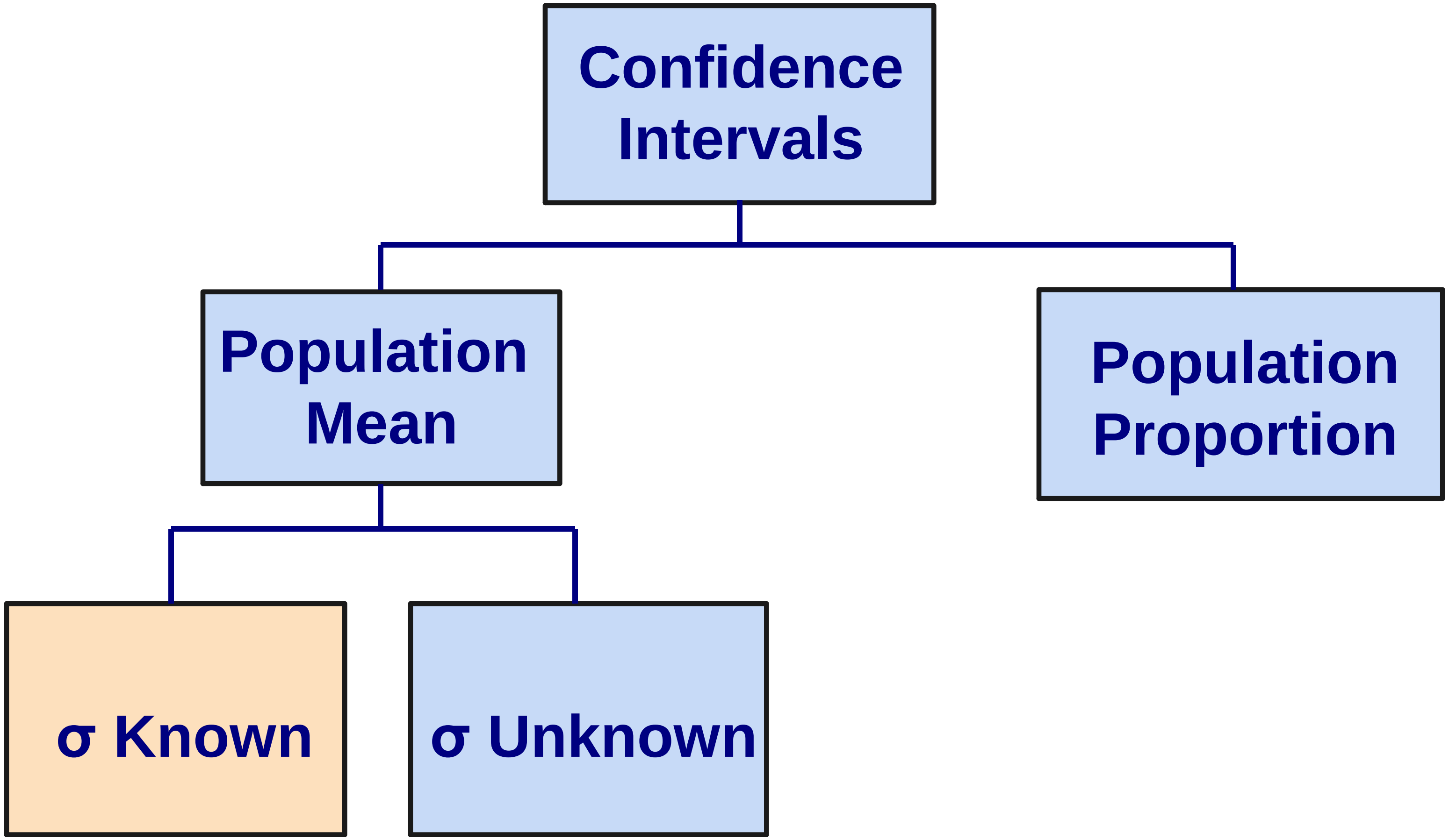
DCOVA

(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

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
- 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

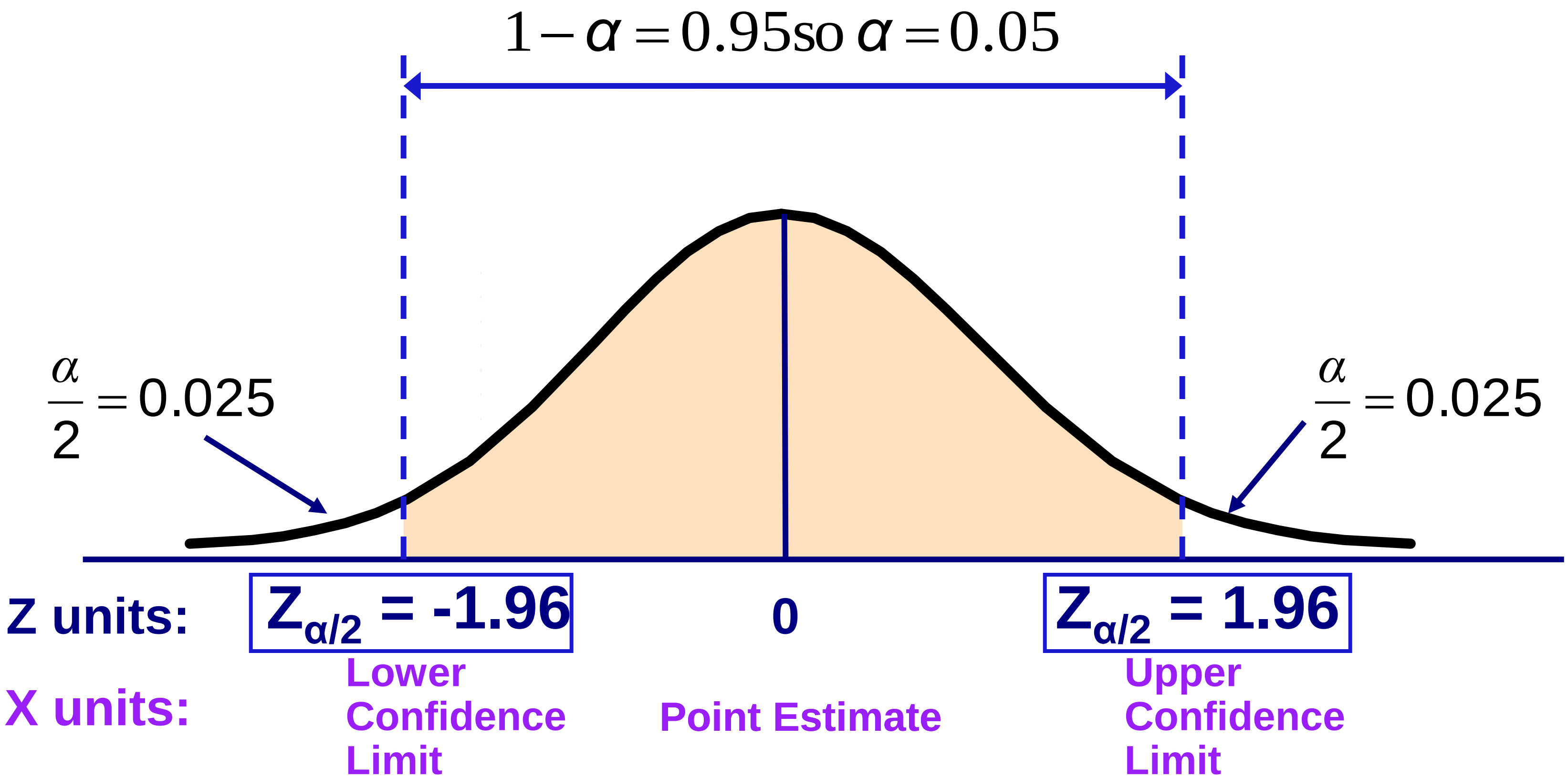
$\sigma/\sqrt{n}$ is the standard error

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

DCOVA

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

- Consider a 95% confidence interval:



Common Levels of Confidence

DCOVA

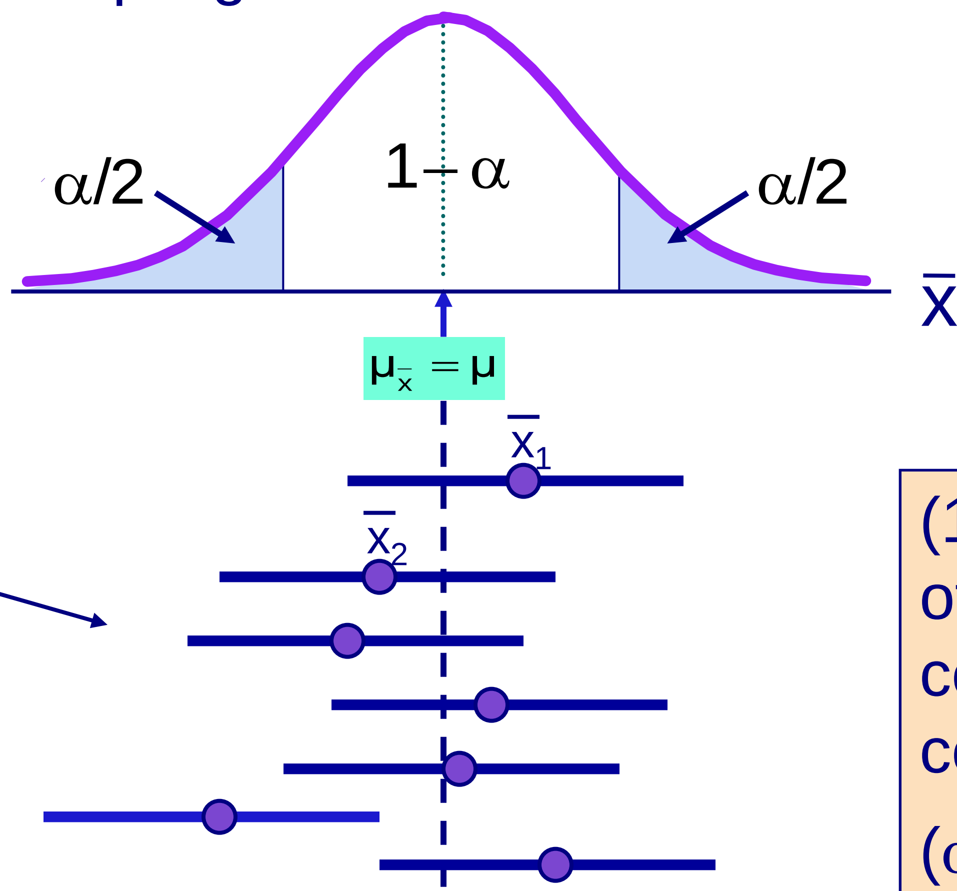
- Commonly used confidence levels are 90%, 95%, and 99%

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

Sampling Distribution of the Mean



Intervals extend from

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

to

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

Confidence Intervals

$(1-\alpha)$x100% of intervals constructed contain μ;

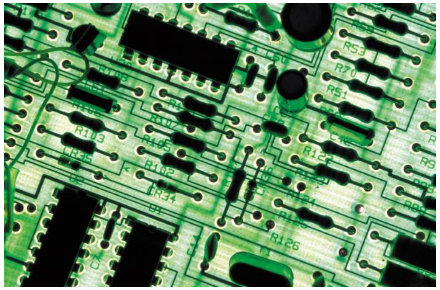
(α)x100% do not.

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)

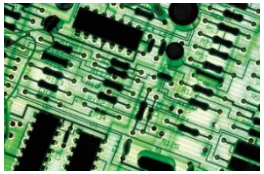
- A sample of 11 circuits from a large normal population has a mean resistance of 2.20 ohms. We know from past testing that the population standard deviation is 0.35 ohms.
- Solution:

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

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

$$= 2.20 \pm 0.2068$$

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



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
- 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)

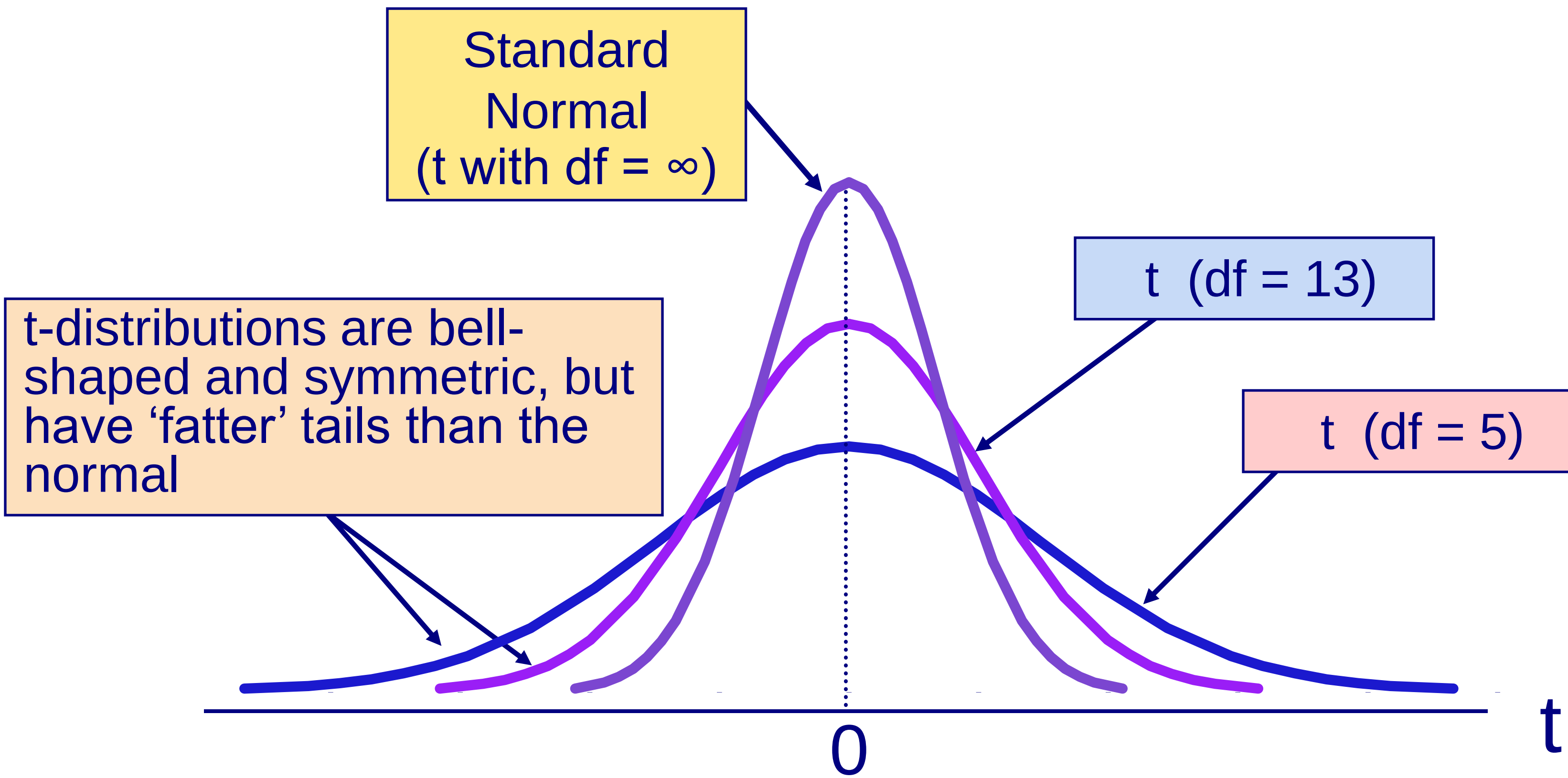
Here, n = 3, so degrees of freedom = n – 1 = 3 – 1 = 2

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

Student’s t Distribution

DCOVA

Note: $t \longrightarrow Z$ as n increases



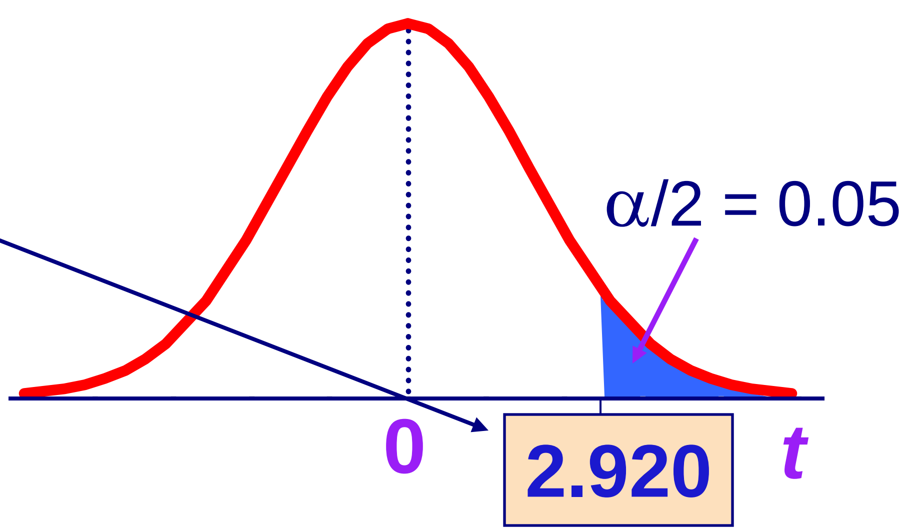
Student's t Table

DCOVA

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

The body of the table contains t values, not probabilities

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



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 $\longrightarrow$ Z as n increases

Example of t distribution confidence interval

DCOVA

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

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

$$46.698 \leq \mu \leq 53.302$$

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

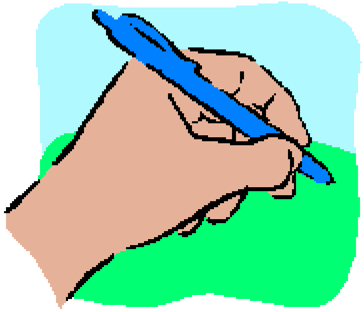
$$p \pm Z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}$$

- 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

DCOVA

(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.

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

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

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

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

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.

Online Topic

Sampling From Finite Populations

Learning Objectives

In this topic, you learn:

- To know when finite population corrections are needed
- To know how to utilize finite population correction factors in calculating standard errors

Finite Population Correction Factors

DCOVA

- Used to calculate the standard error of both the sample mean and the sample proportion

- Needed when the sample size, n, is more than 5% of the population size N (i.e. n / N > 0.05)

- The Finite Population Correction Factor Is:

$$\text{fpc} = \sqrt{\frac{N-n}{N-1}}$$

Using The fpc In Calculating Standard Errors

DCOVA

Standard Error of the Mean for Finite Populations

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

Standard Error of the Proportion for Finite Populations

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}\sqrt{\frac{N-n}{N-1}}$$

Using The fpc Reduces The Standard Error

DCOVA

- The fpc is always less than 1
- So when it is used it reduces the standard error
- Resulting in more precise estimates of population parameters

Using fpc With The Mean - Example

DCOVA

Suppose a random sample of size 100 is drawn from a population of size 1,000 with a standard deviation of 40.

Here n=100, N=1,000 and 100/1,000 = 0.10 > 0.05.

So using the fpc for the standard error of the mean we get:

$$\sigma_{\overline{X}} = \frac{40}{\sqrt{100}}\sqrt{\frac{1000-100}{1000-1}} = 3.8$$

Topic Summary

In this topic we discussed

- When a finite population correction should be used.
- How to utilize a finite population correction factor in calculating the standard error of both a sample mean and a sample proportion