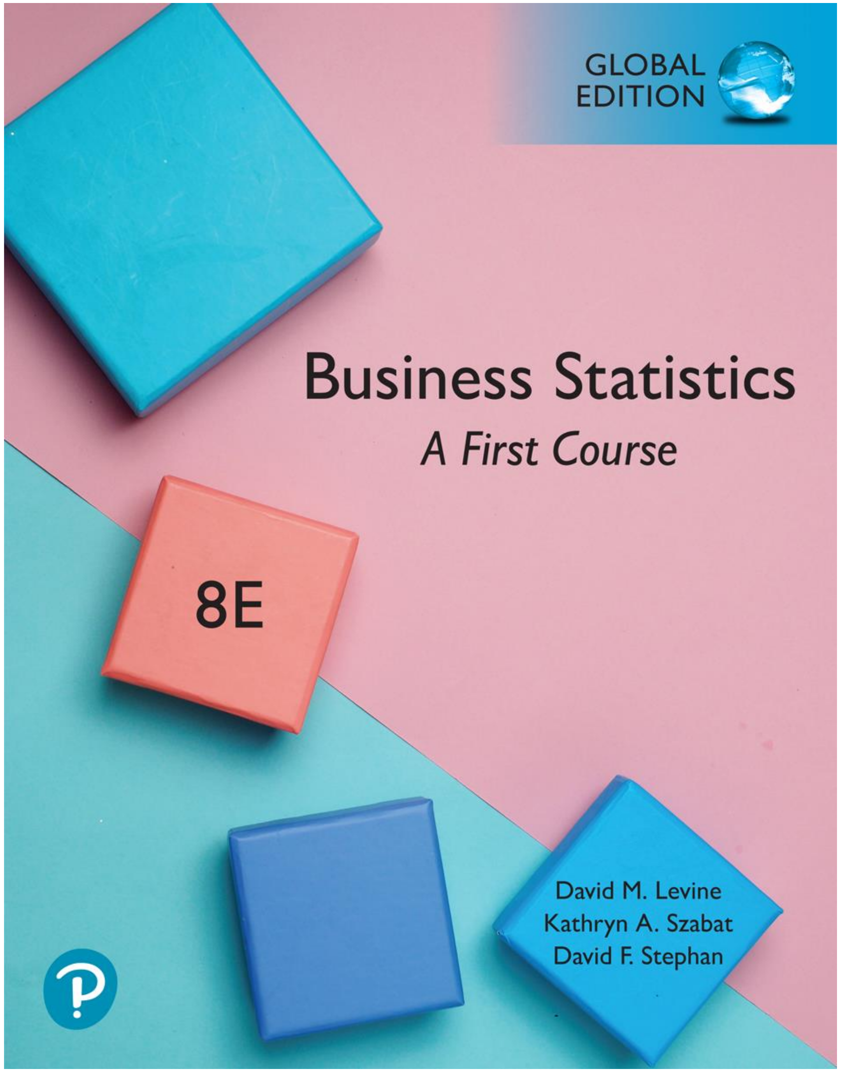


Chapter 7

Sampling Distributions

Objectives

In this chapter, you learn:

- The concept of the sampling distribution.
- To calculate 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 sample size 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 size 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 samples of 50 students.

Developing a Sampling Distribution

DCOVA

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

Developing a Sampling Distribution

(continued)

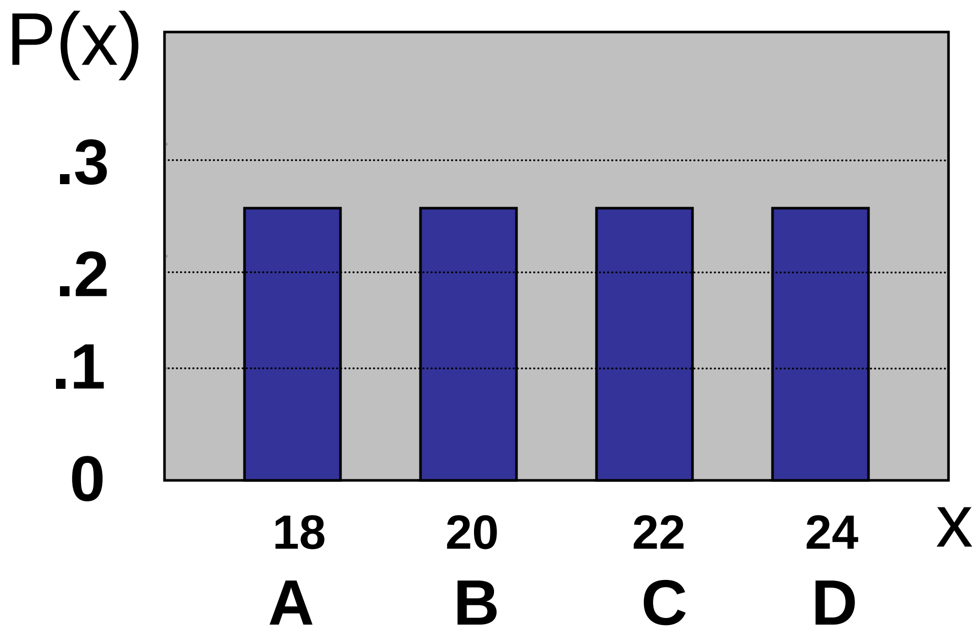
DCOVA

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)

Now consider all possible samples of size n=2.

DCOVA

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

(continued)

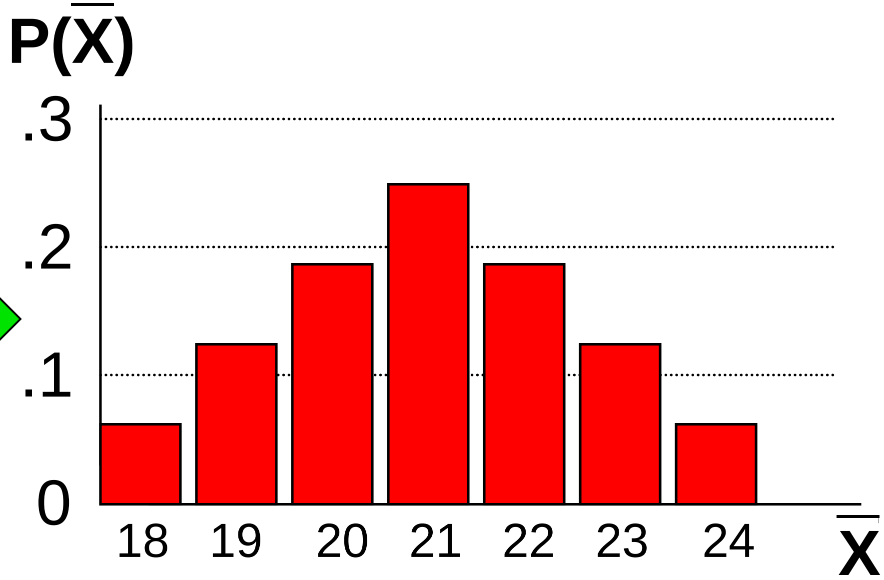
DCOVA

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

(continued)

DCOVA

Summary Measures of this Sampling Distribution:

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

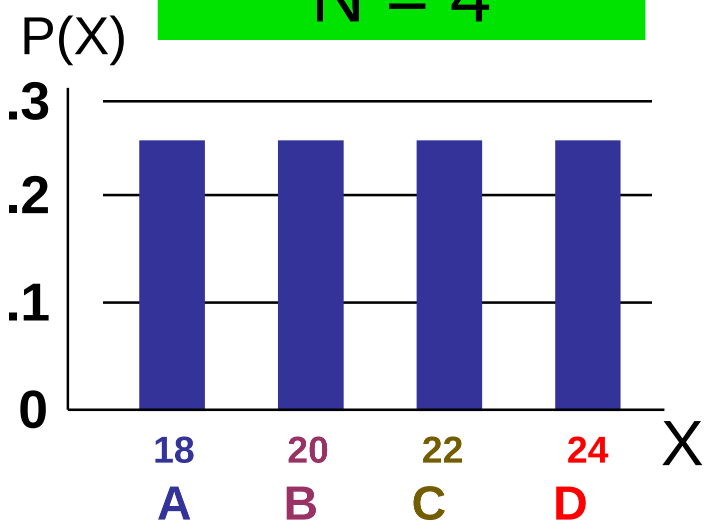
$$\sigma_{\bar{X}} = \sqrt{\frac{(18-21)^2+(19-21)^2+\cdots+(24-21)^2}{16}} = 1.58$$

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

Comparing the Population Distribution to the Sample Means Distribution

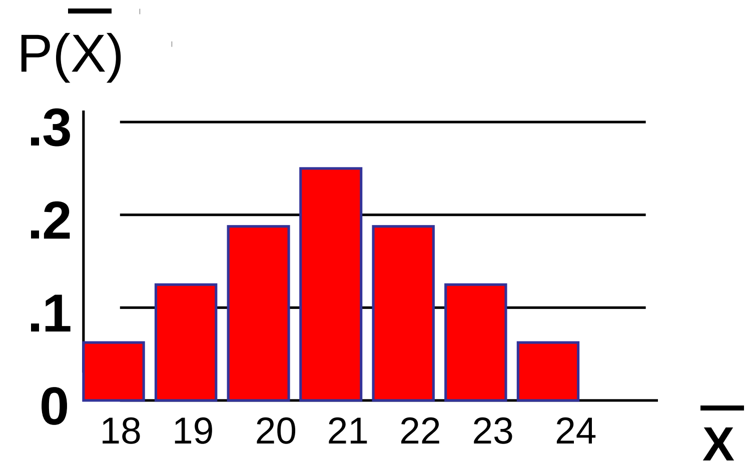
DCOVA

Population
N = 4



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

Sample Means Distribution
n = 2



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

Since $\mu_{\bar{X}} = \mu$, $\bar{X}$ is an unbiased estimator of μ.

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.)

$$\boxed{\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}}$$

- 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:

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

and

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

Z-value for Sampling Distribution of the Mean

DCOVA

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

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

where: $\bar{X}$ = sample mean

μ = population mean

σ = population standard deviation

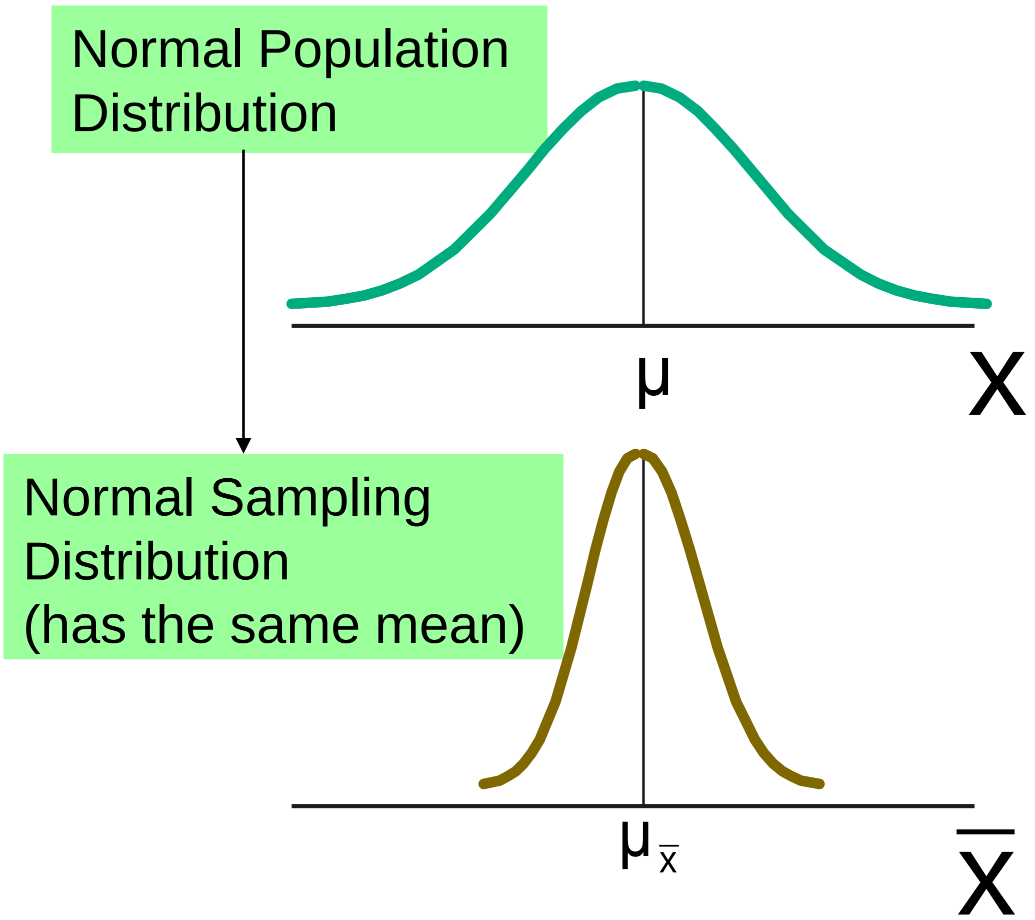
n = sample size

Sampling Distribution Properties

DCOVA

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

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



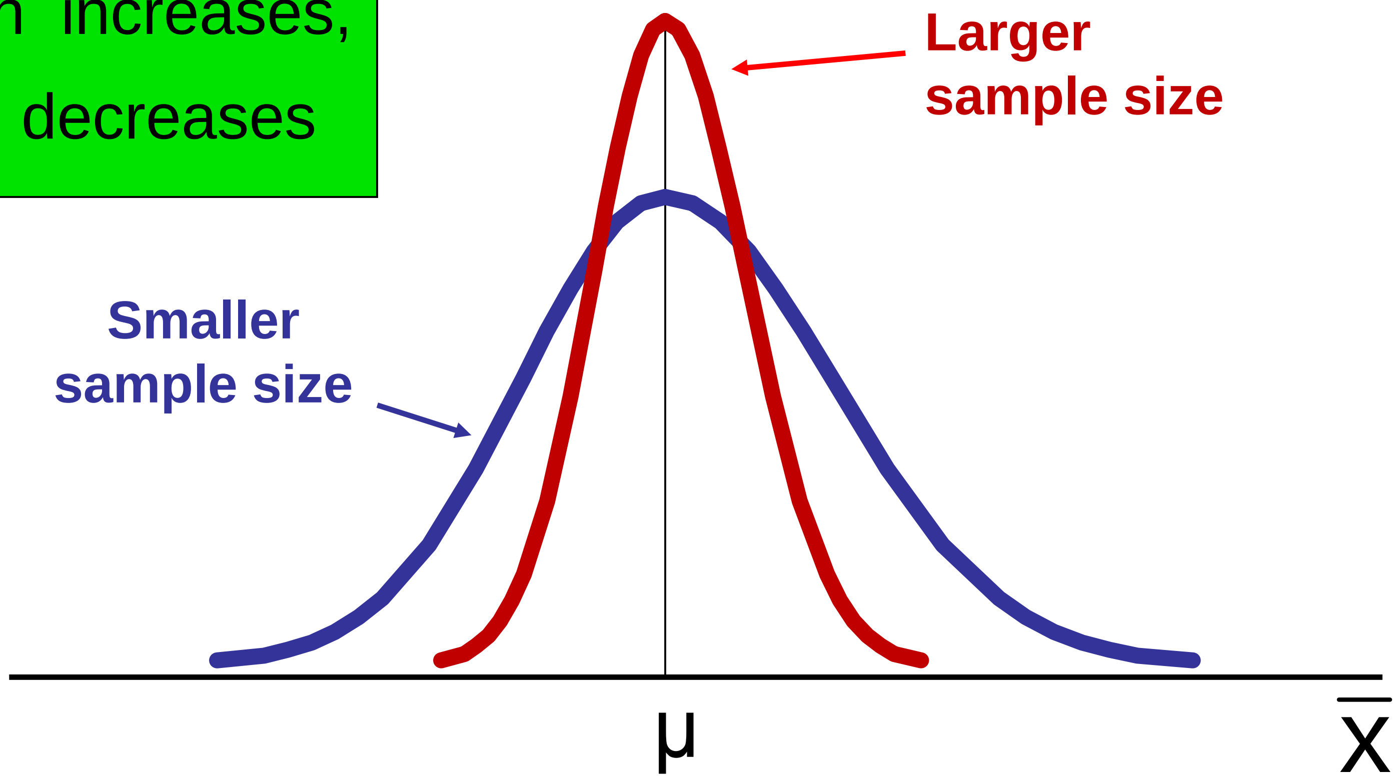
Sampling Distribution Properties

(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:

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

- Based on samples of size 25, the sample means in 95% of all samples 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:

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

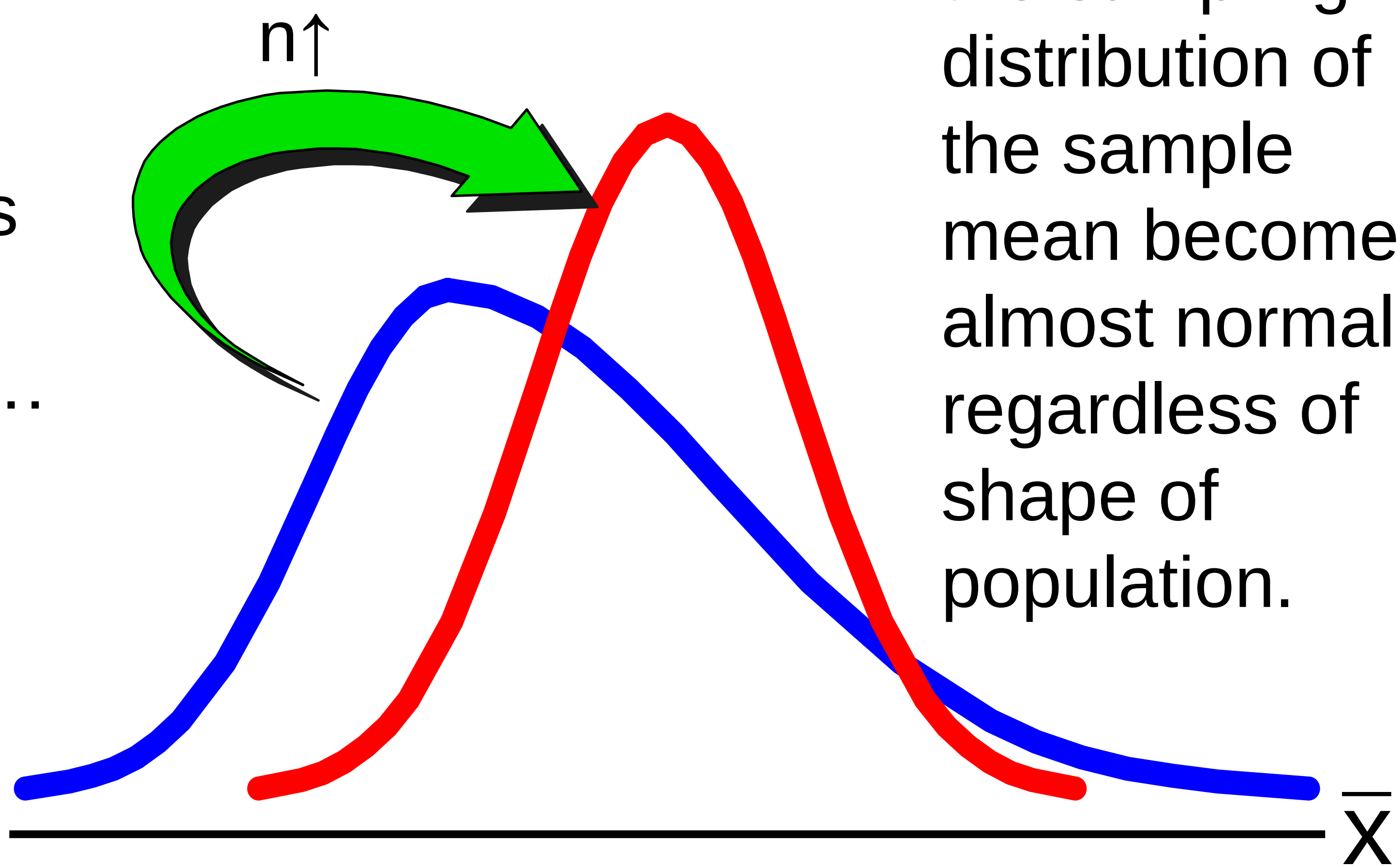
and

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

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

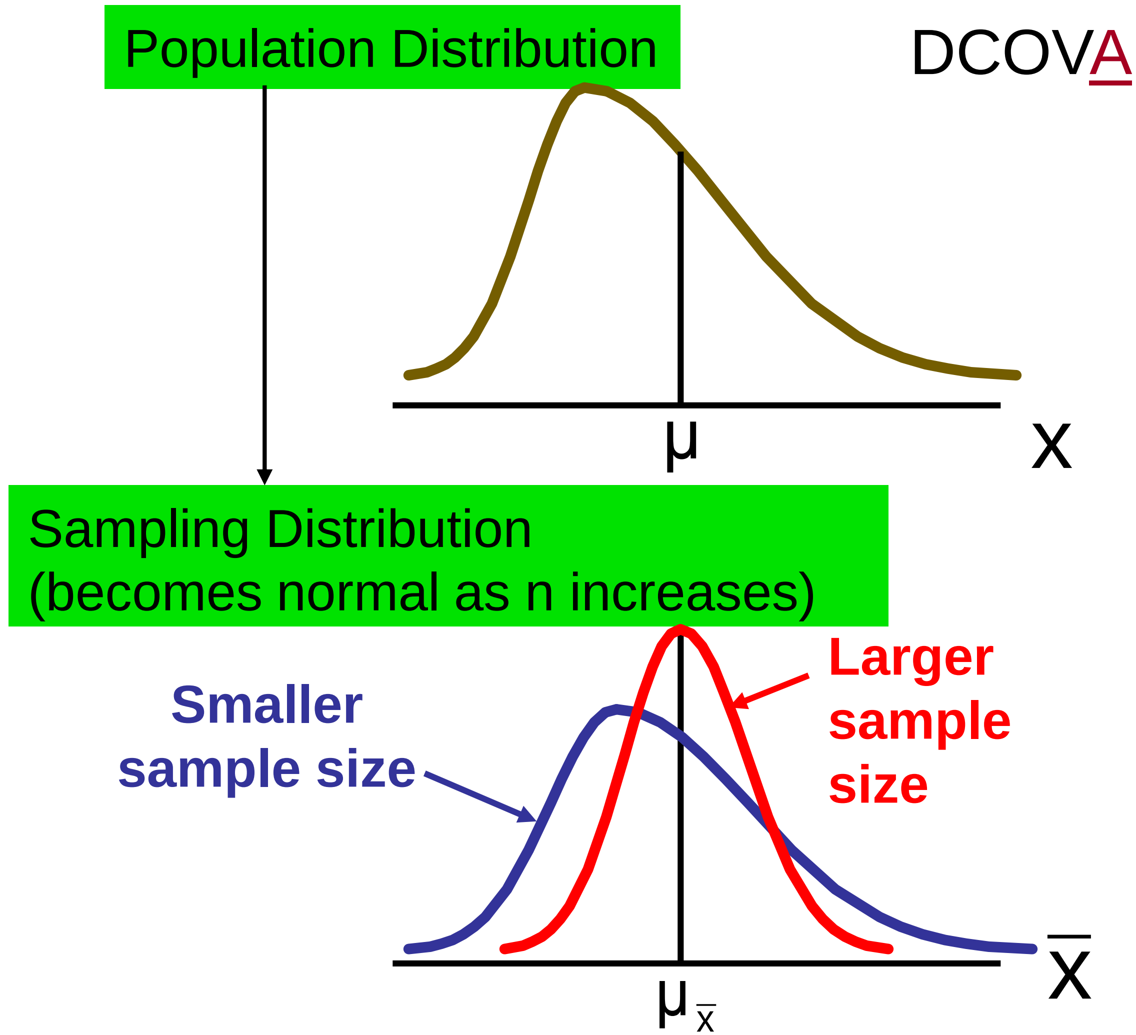
Sampling distribution properties:

Central Tendency

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

Variation

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



How Large is Large Enough?

DCOVA

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

Example

DCOVA

- Suppose the wait time to be seated at a restaurant has a mean of $\mu = 8$ minutes and a standard deviation of $\sigma = 3$ minutes. And random sample of size.

- 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

(continued)

DCOVA

Solution:

- 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 $\mu_{\bar{X}} = 8$ minutes.
- …and standard deviation $\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{3}{\sqrt{36}} = 0.5$ minutes.

Example

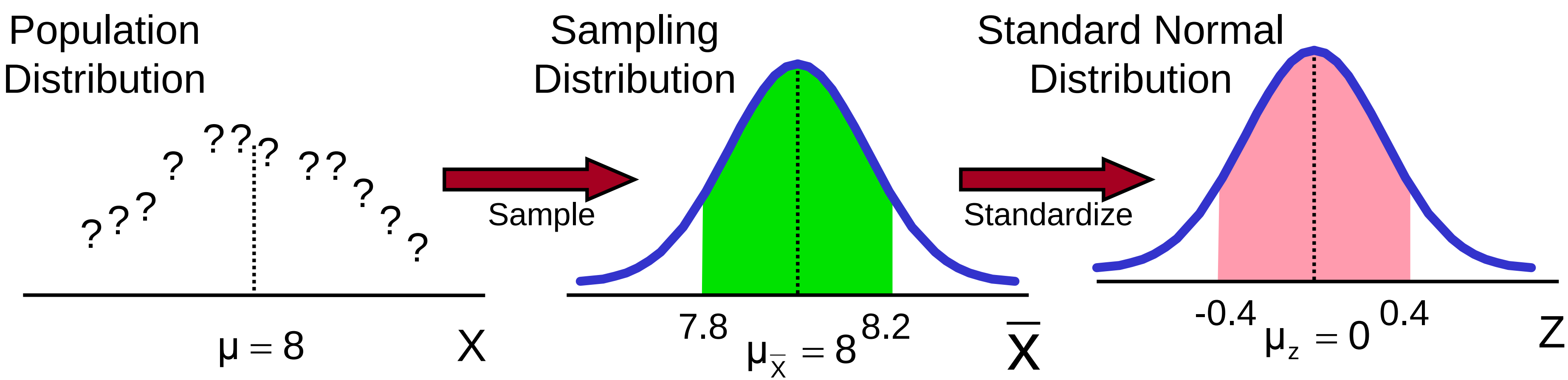
(continued)

DCOVA

Solution (continued):

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

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



Population Proportions

DCOVA

π = the proportion of the population having some characteristic.

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

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

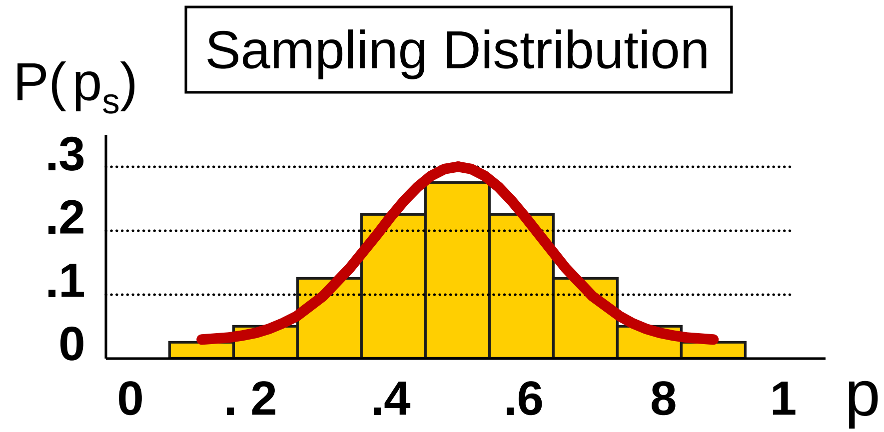
- $0 \leq p \leq 1$.
- p is approximately distributed as a normal distribution when n is large.

(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**

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

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

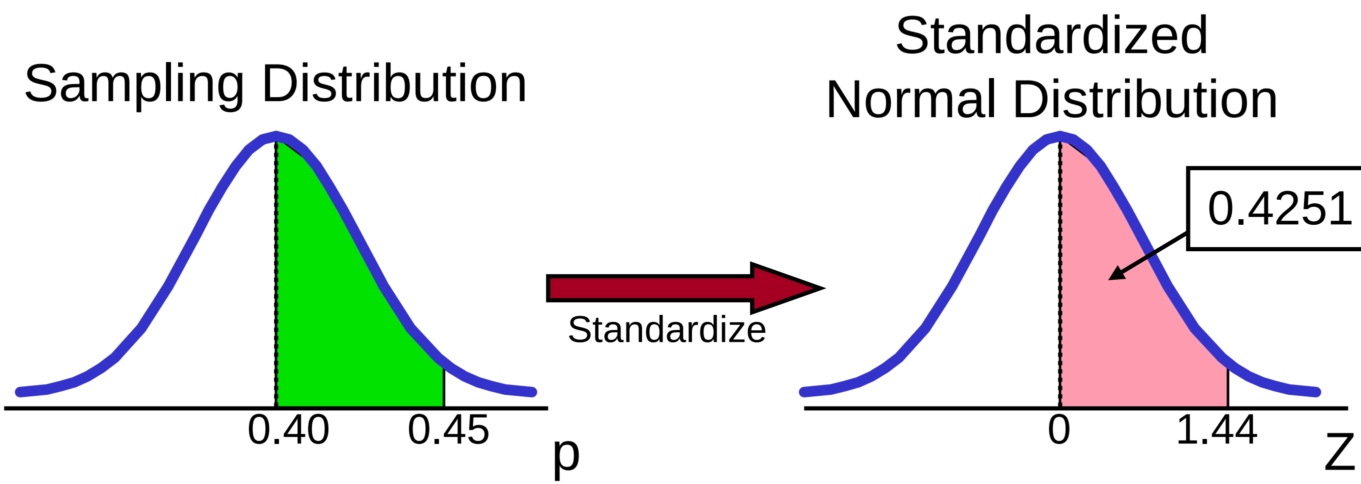
(continued)

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:

- The concept of a sampling distribution.
- Calculating probabilities related to the sample mean and the sample proportion.
- The importance of the Central Limit Theorem.