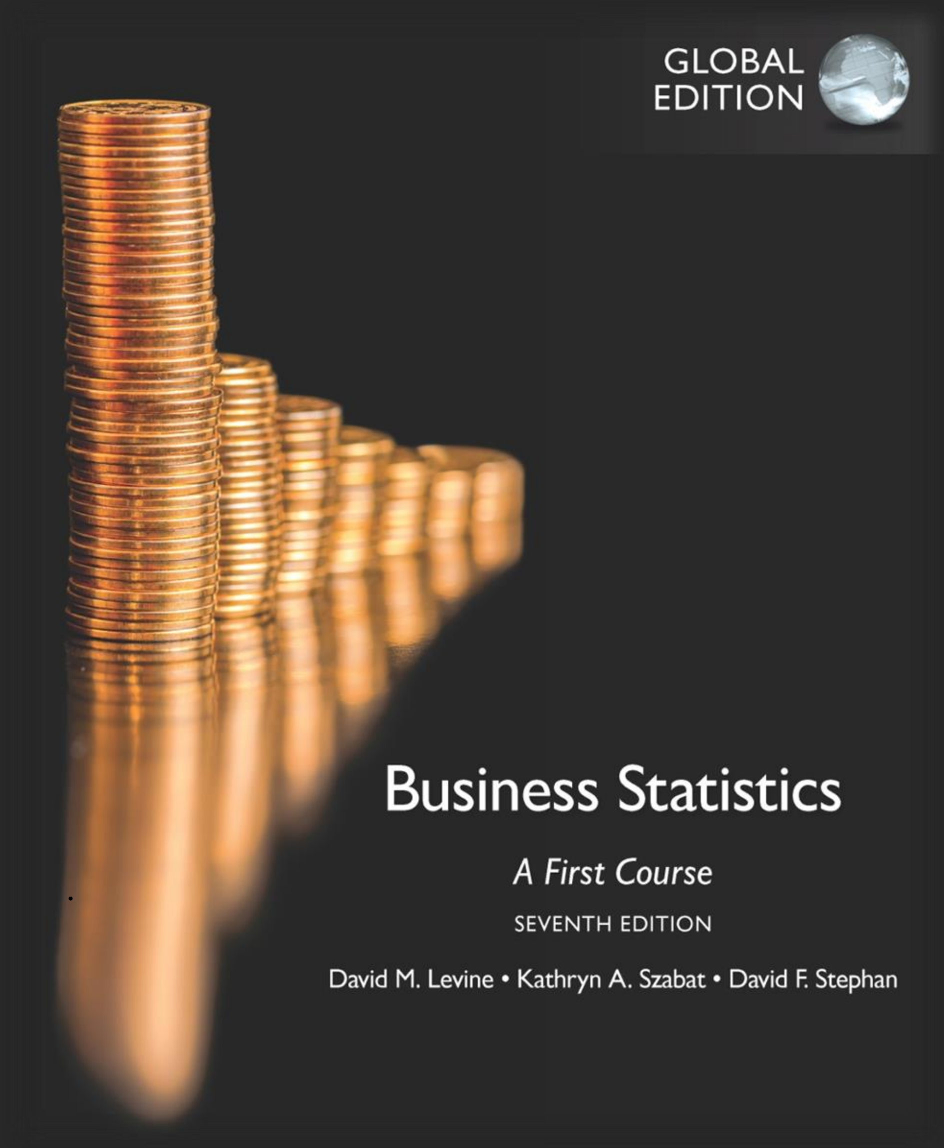


Chapter 7

Sampling Distributions

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

- 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 any sample of 50 students.

Developing a Sampling Distribution

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

Individual	A	B	C	D
Age	18	20	22	24

Developing a Sampling Distribution

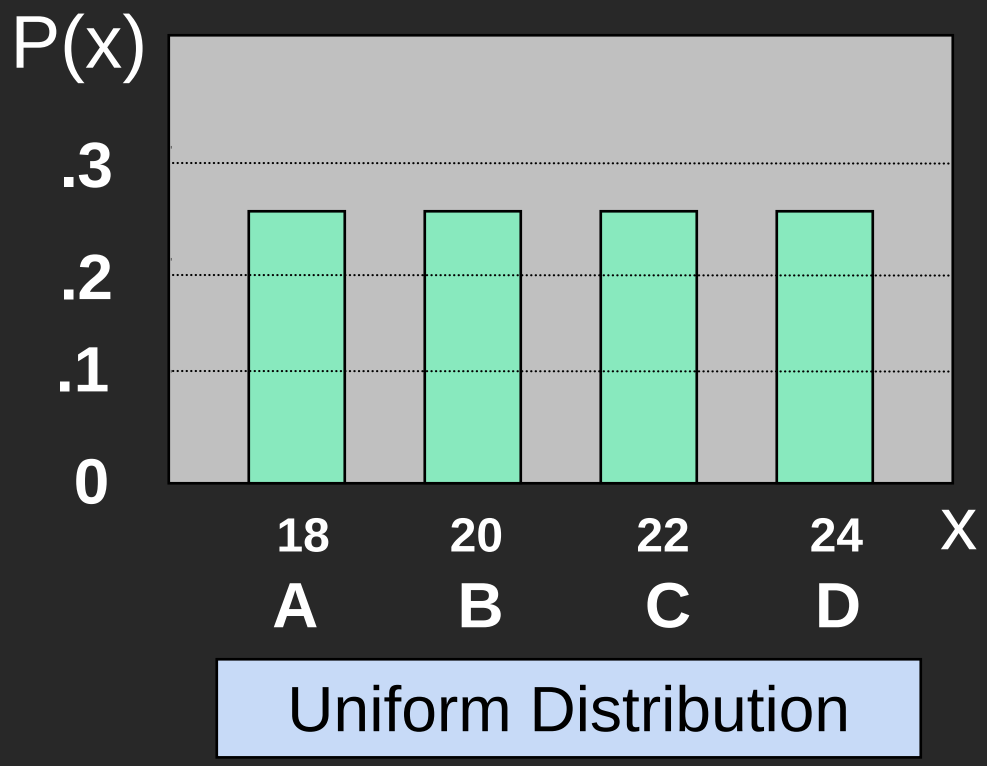
(continued)

Summary Measures for the Population Distribution:

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

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

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



Developing a Sampling Distribution

(continued)

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

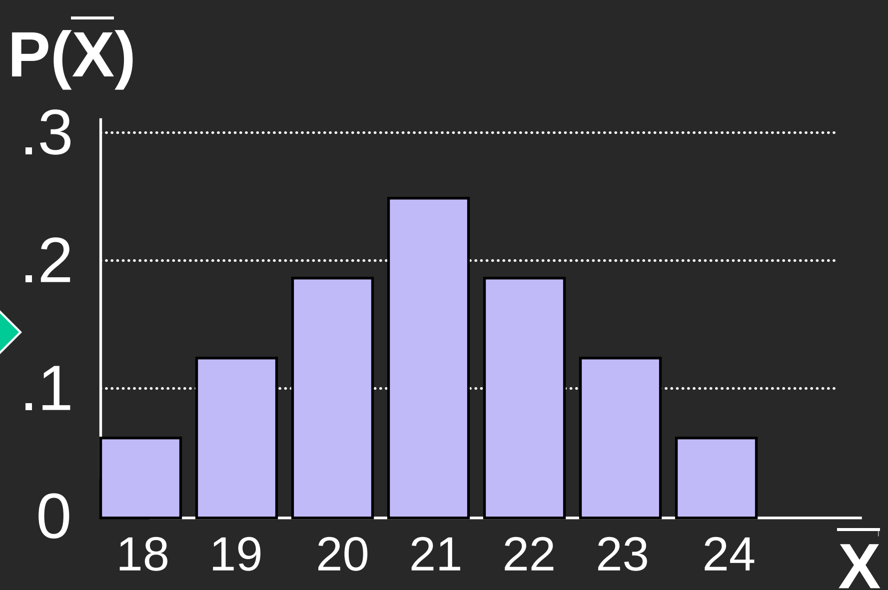
(continued)

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

(continued)

Summary Measures of this Sampling Distribution:

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

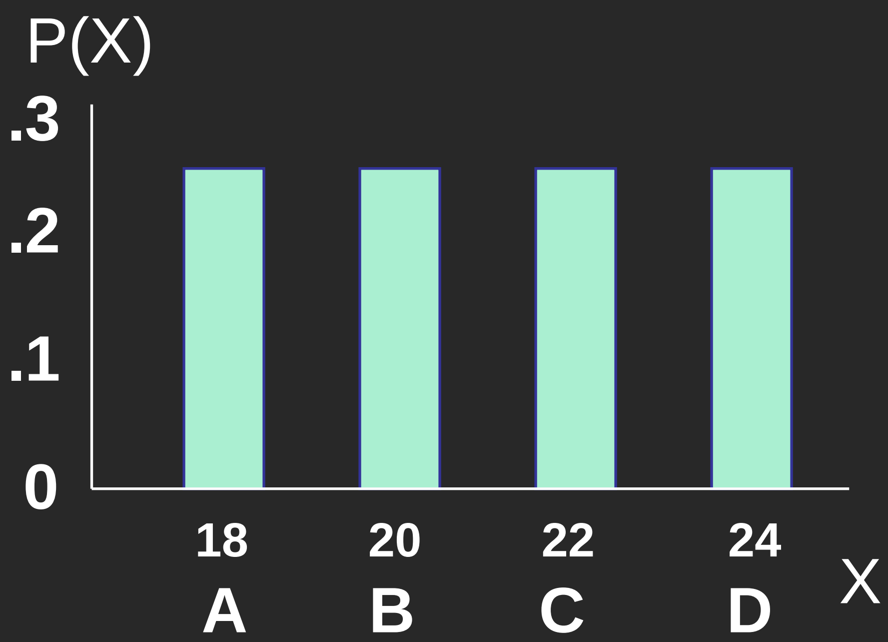
$$\sigma_{\bar{X}} = \sqrt{\frac{(18-21)^2 + (19-21)^2 + \cdots + (24-21)^2}{16}} = 1.58$$

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

Comparing the Population Distribution to the Sample Means Distribution

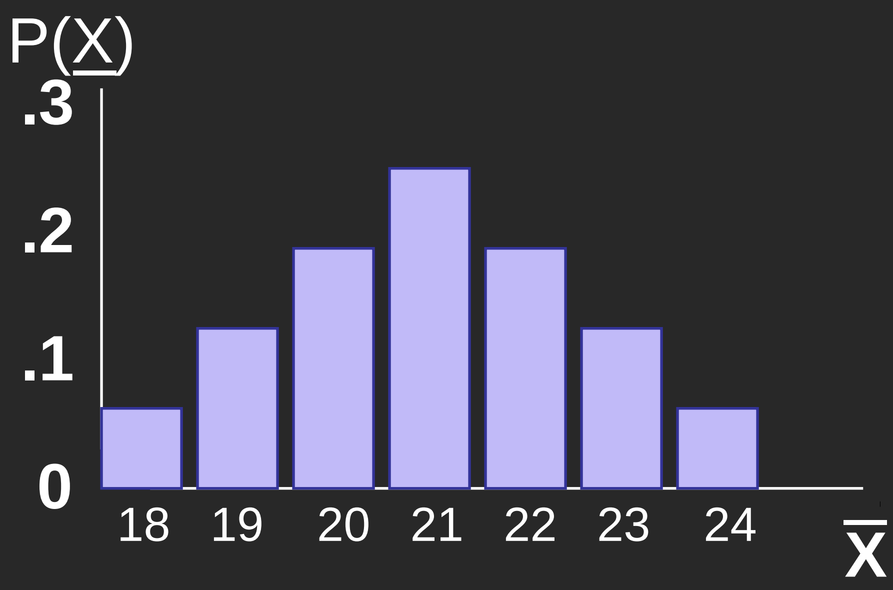
Population N = 4

$\mu = 21 \quad \sigma = 2.236$



Sample Means Distribution n = 2

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



Sample Mean Sampling Distribution: Standard Error of the Mean

- 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

- 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

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

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

where:

$\bar{X}$ = sample mean

μ = population mean

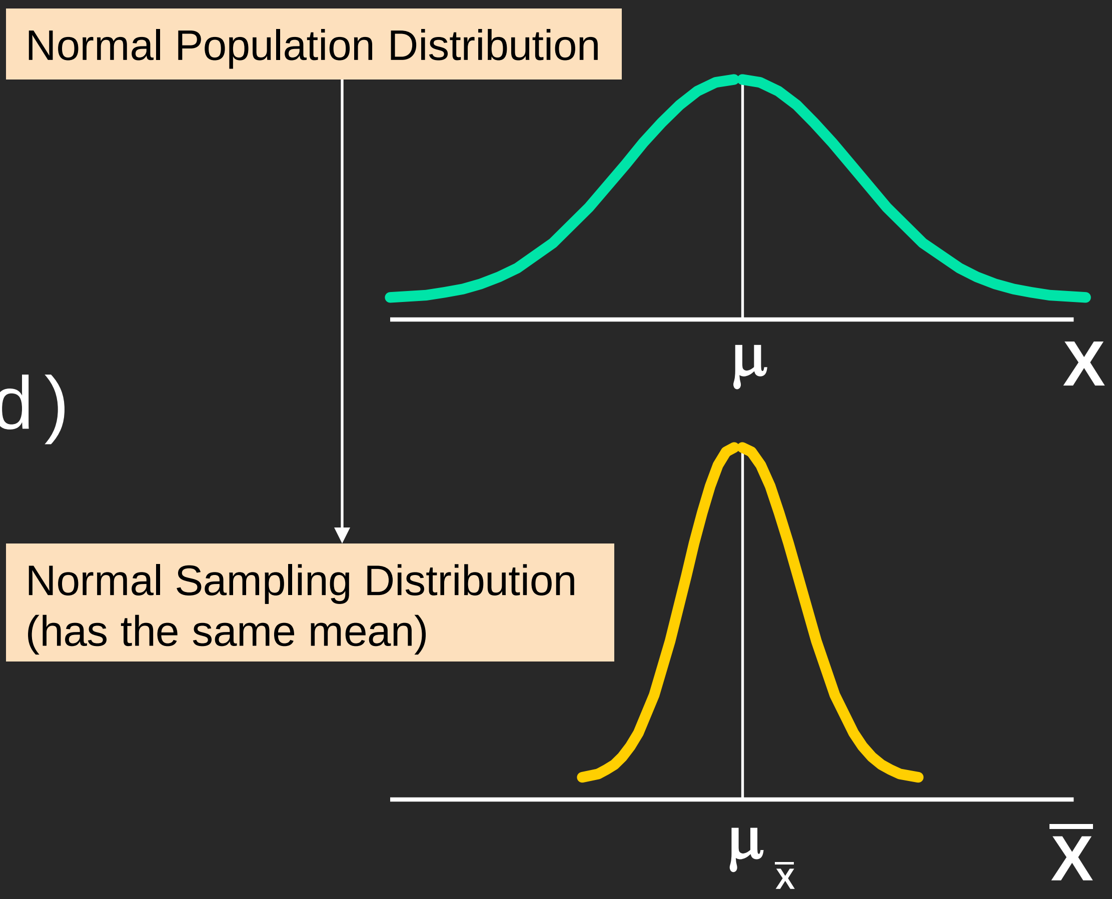
σ = population standard deviation

n = sample size

Sampling Distribution Properties

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

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

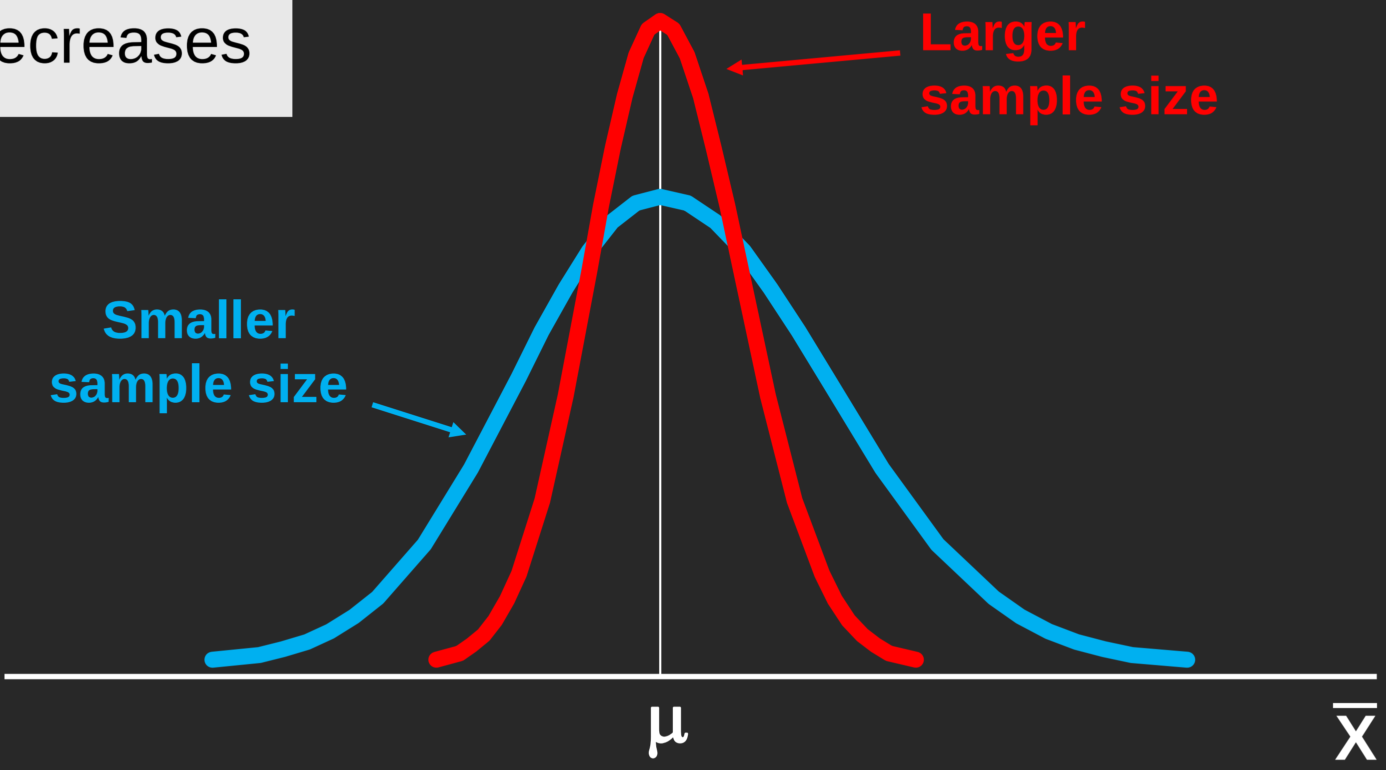


Sampling Distribution Properties

(continued)

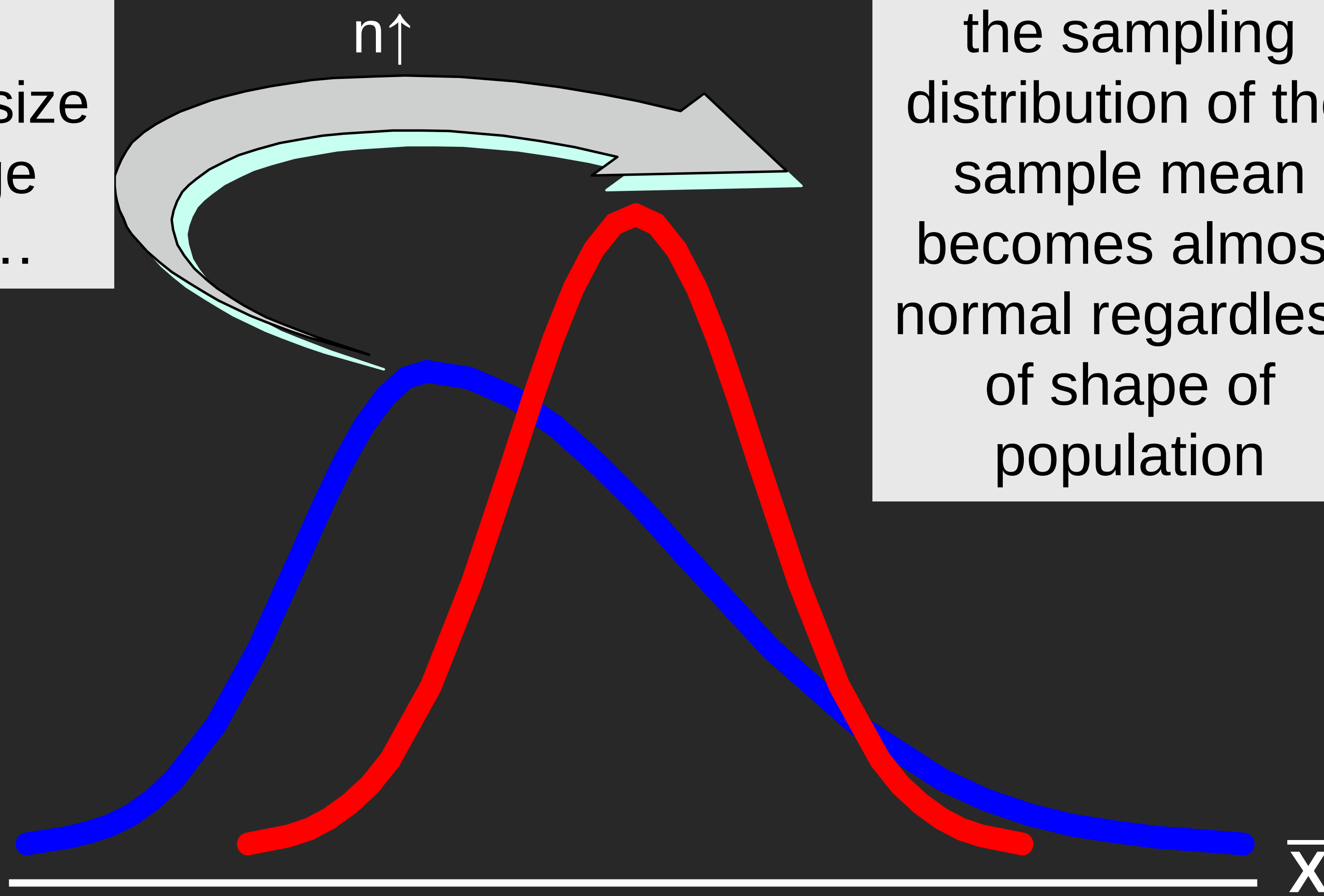
As n increases,

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



Central Limit Theorem

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

- 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:

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

Sample Mean Sampling Distribution: If the Population is NOT Normal

(continued)

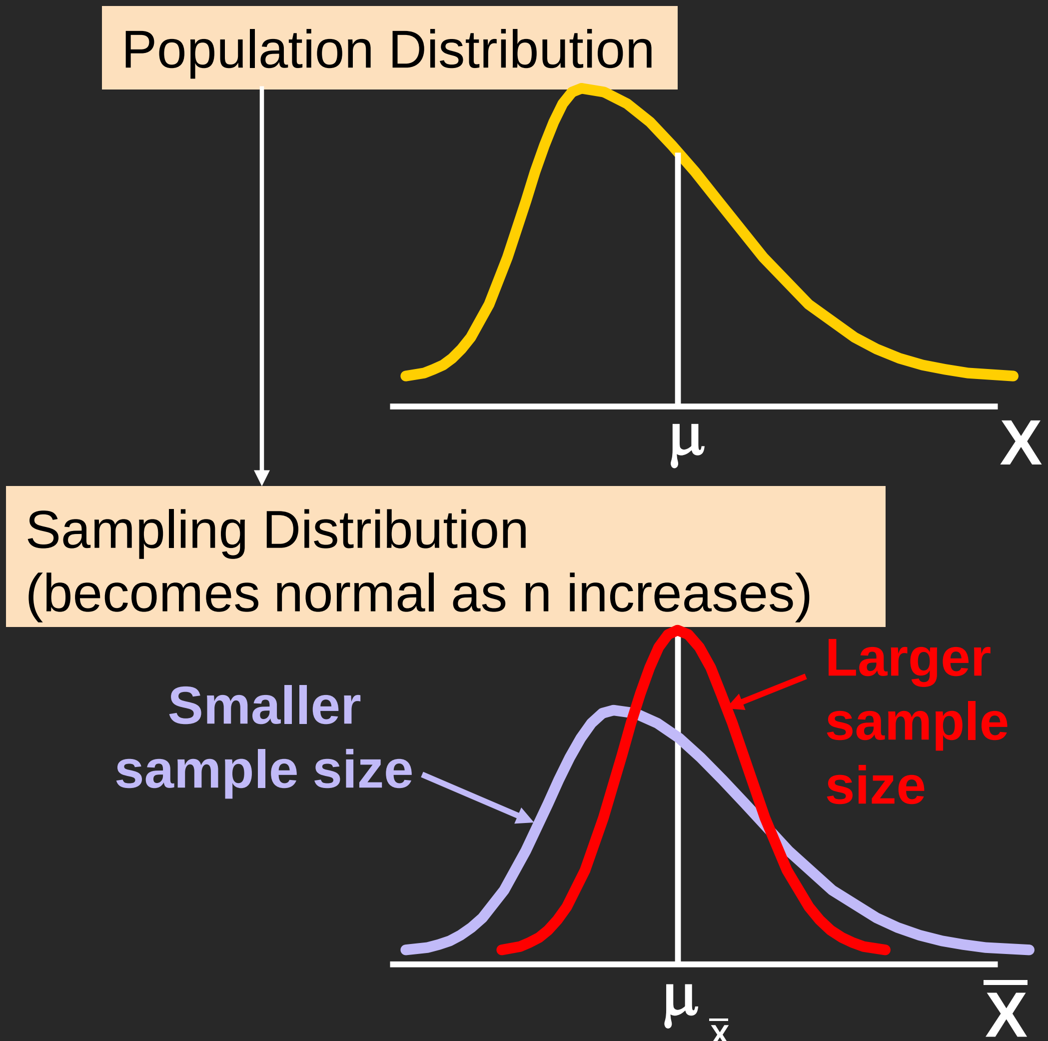
Sampling distribution properties:

Central Tendency

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

Variation

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



How Large is Large Enough?

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

Example

- 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

(continued)

Solution:

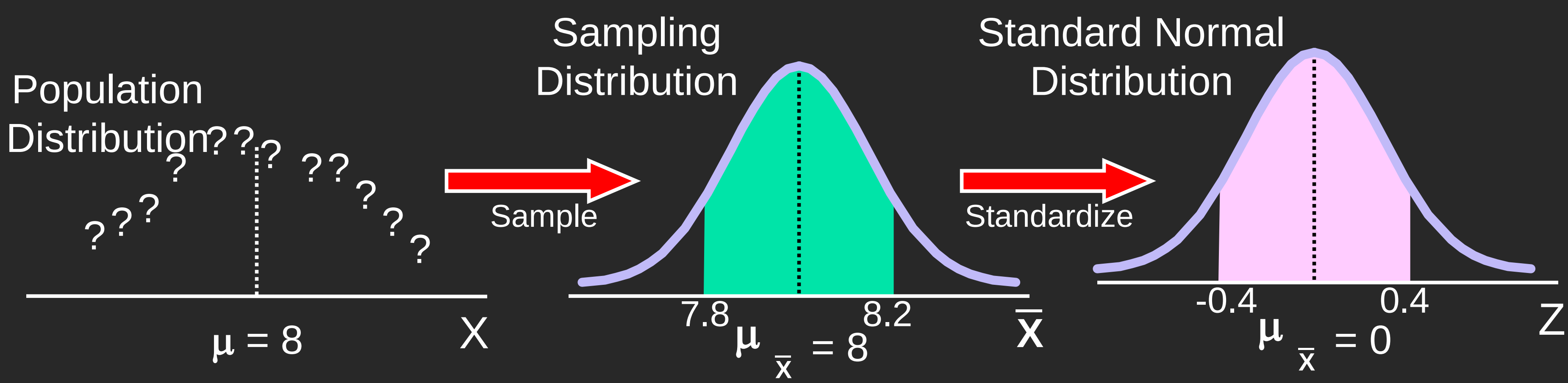
- Even if the population is not normally distributed, the central limit theorem can be used (n > 30)
- … so the sampling distribution of $\bar{x}$ is approximately normal
- … with mean $\mu_{\bar{x}} = 8$
- …and standard deviation $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{36}} = 0.5$

Example

(continued)

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

$$= P(-0.4 < Z < 0.4) = 0.6554 - 0.3446 = 0.3108$$



Population Proportions

π = the proportion of the population having some characteristic

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

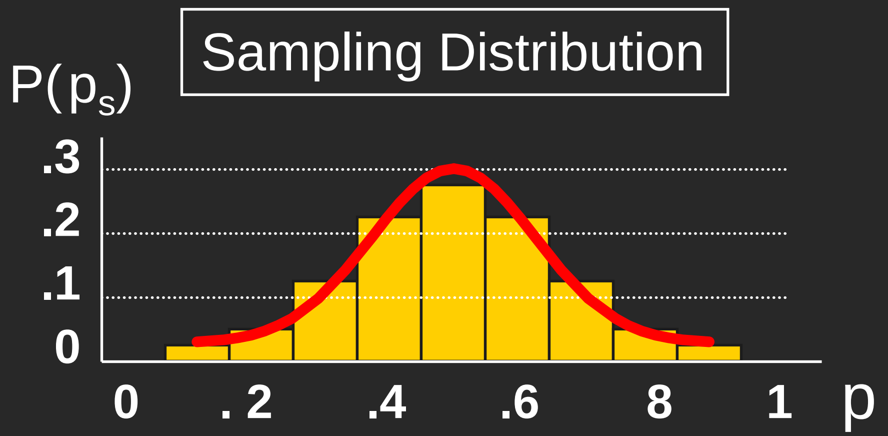
$$p = \frac{X}{n} = \frac{\text{number of items in the sample having the characteristic of interest}}{\text{sample size}}$$

- $0 \leq p \leq 1$
- 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

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



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:

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

Example

- 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)

- **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:

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

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

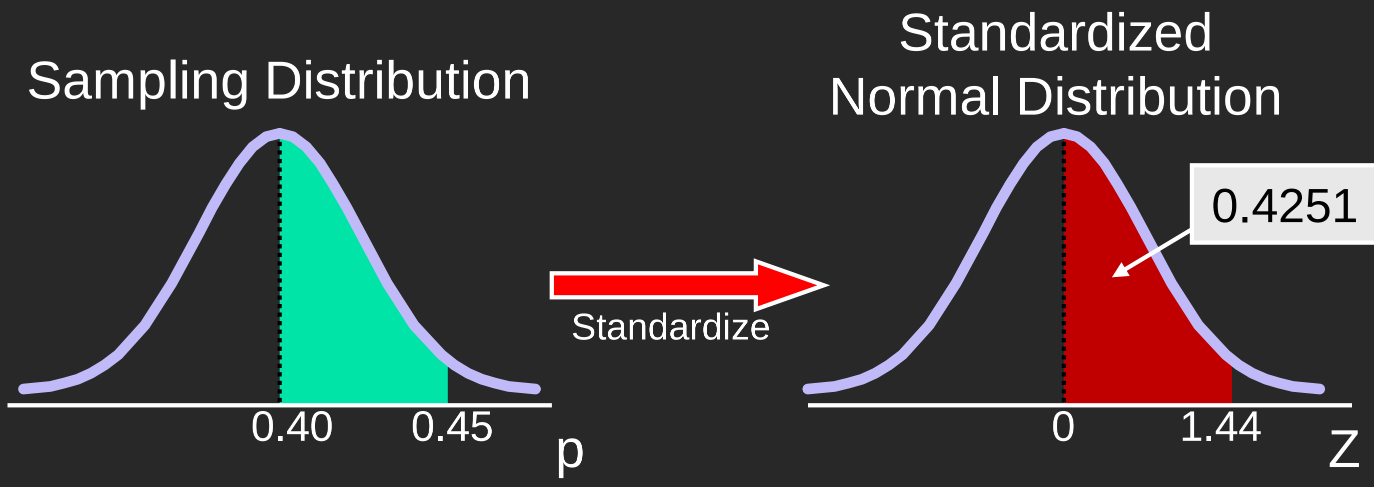
Example

(continued)

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

Utilize the cumulative normal table:

$P(0 \le Z \le 1.44) = 0.9251 - 0.5000 = 0.4251$



Chapter Summary

In this chapter we discussed:

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