

Sampling Distributions

Chapter Goals

After completing this chapter, you should be able to:

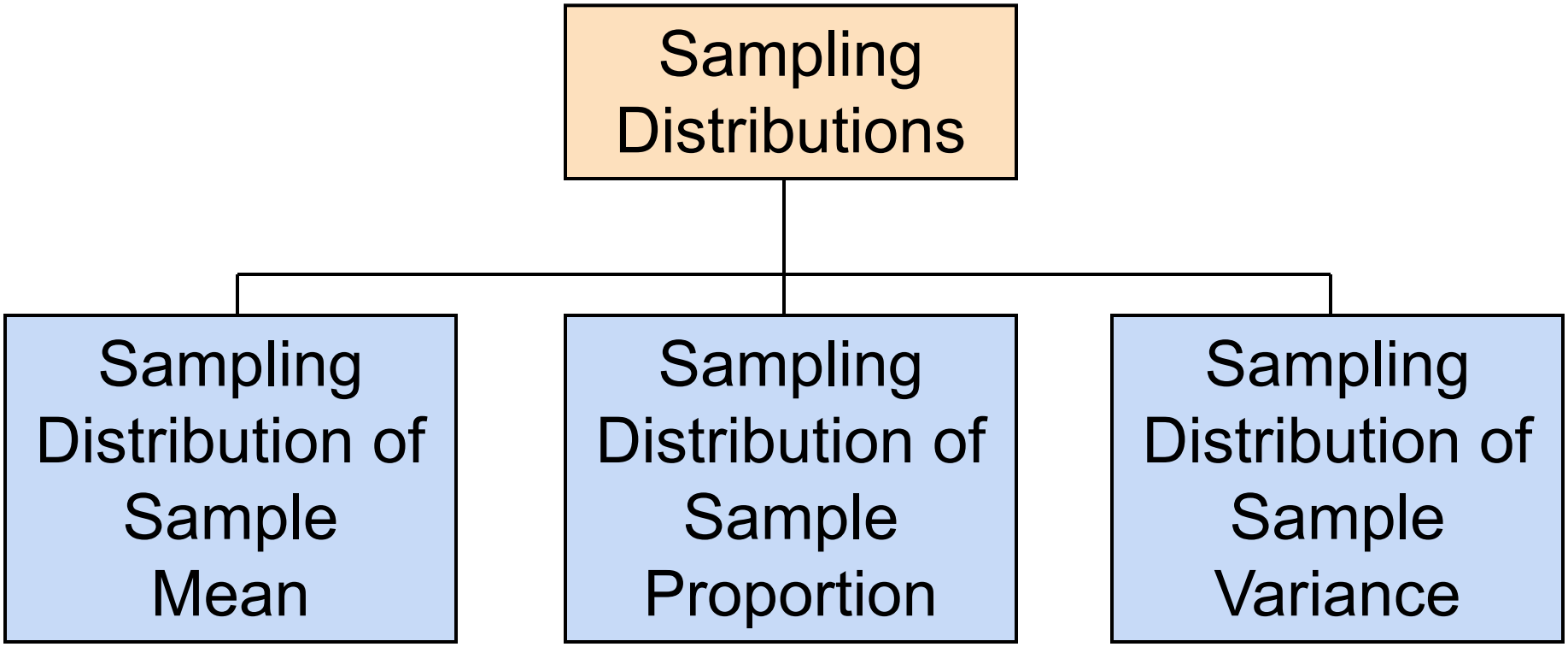
- Define the concept of a sampling distribution
- Determine the mean and standard deviation for the sampling distribution of the sample mean, $\bar{X}$
- Describe the Central Limit Theorem and its importance
- Determine the mean and standard deviation for the sampling distribution of the sample proportion $\hat{p}$



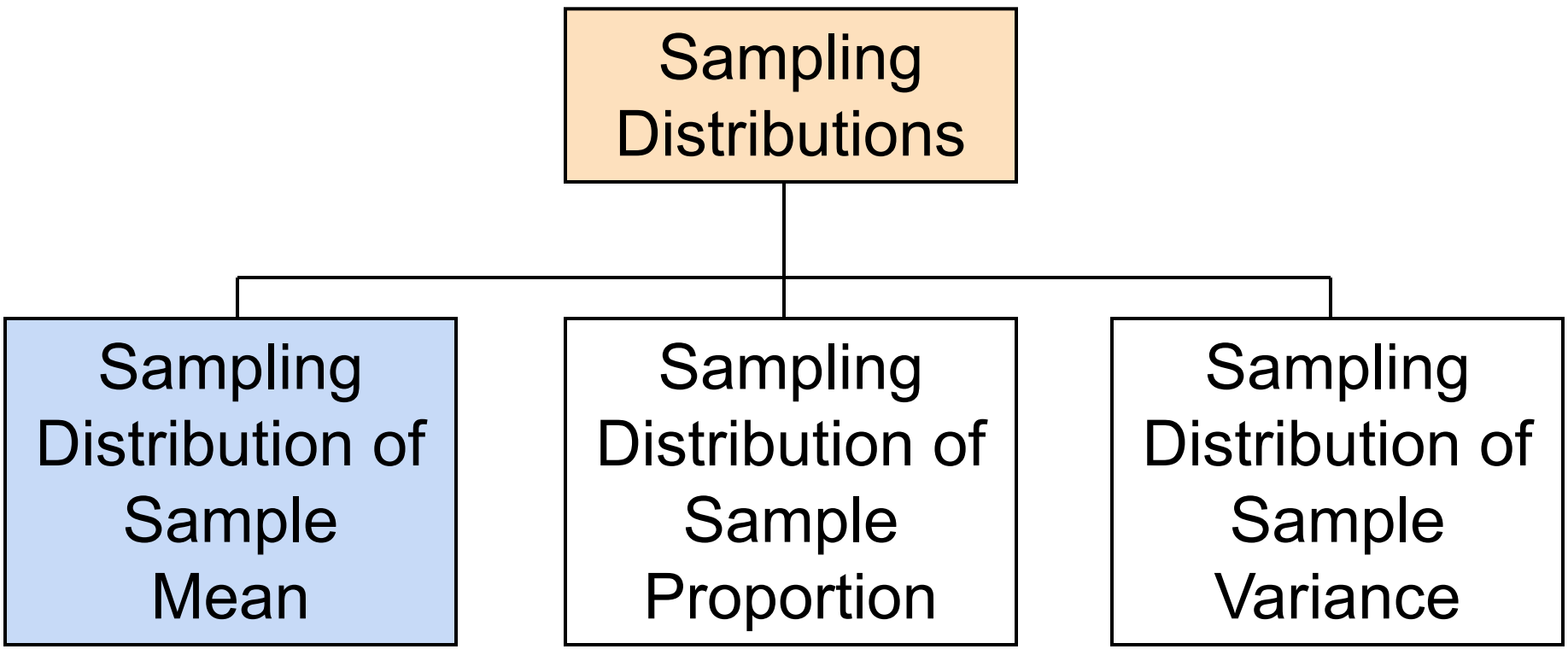
Sampling Distributions

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

Chapter Outline



Sampling Distributions of Sample Means



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)



Developing a Sampling Distribution

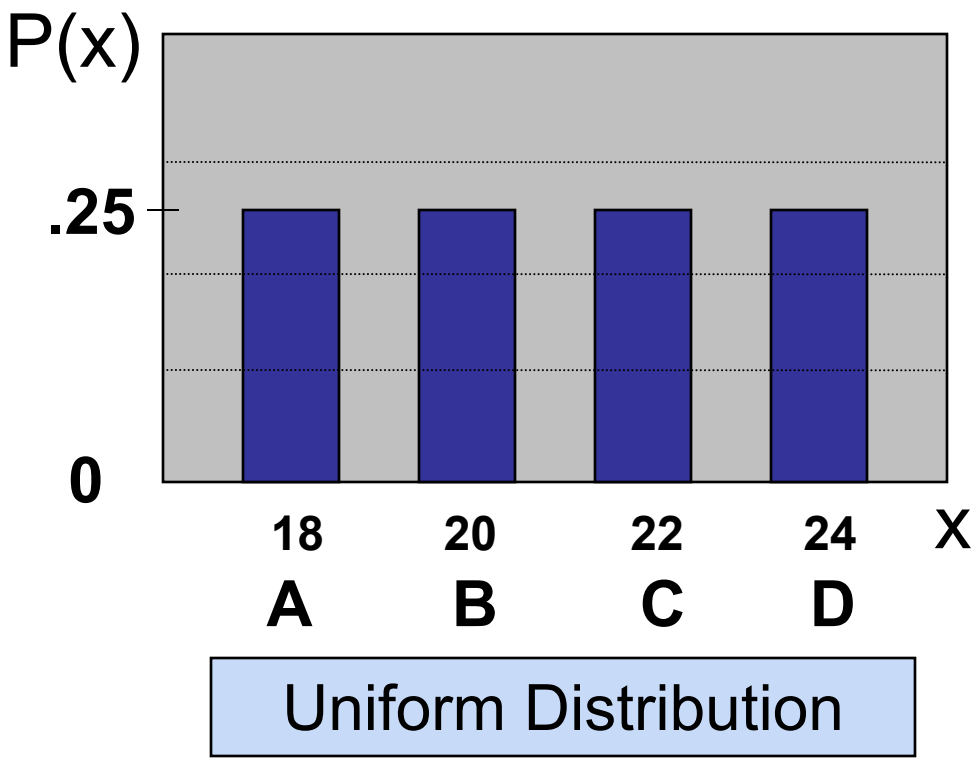
(continued)

Summary Measures for the Population Distribution:

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

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

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



Uniform Distribution

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

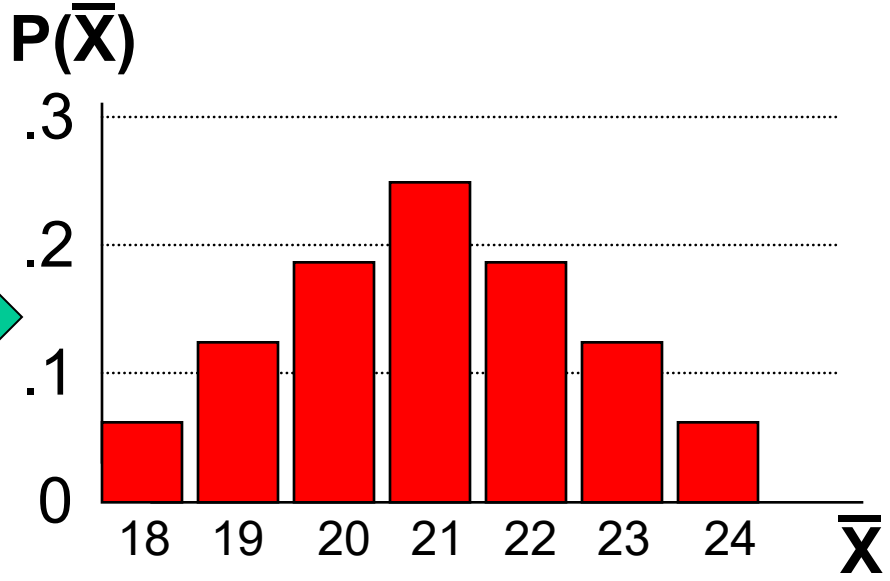
(continued)

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:

$$E(\bar{X}) = \frac{\sum \bar{X}_i}{N} = \frac{18 + 19 + 21 + \cdots + 24}{16} = 21 = \mu$$

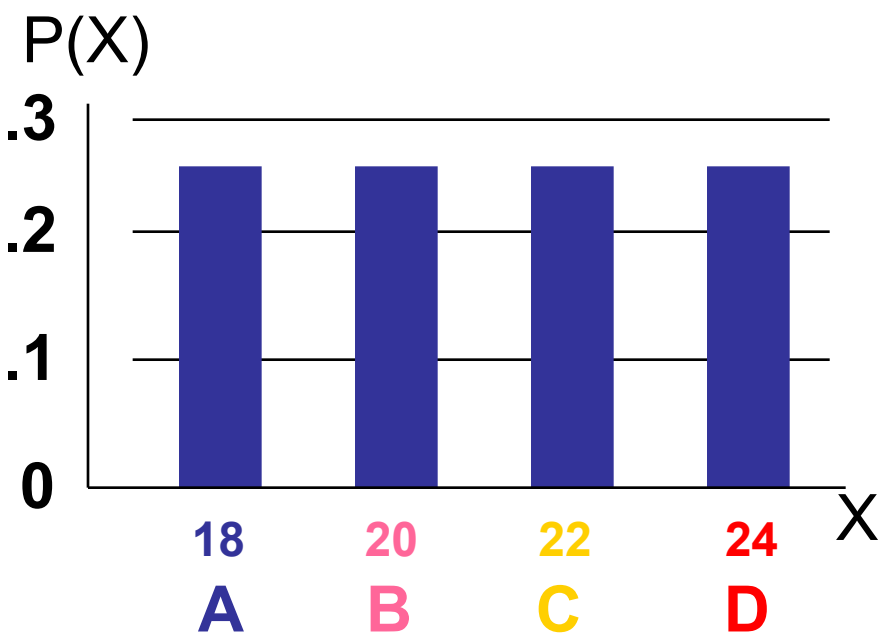
$$\sigma_{\bar{X}} = \sqrt{\frac{\sum (\bar{X}_i - \mu)^2}{N}}$$

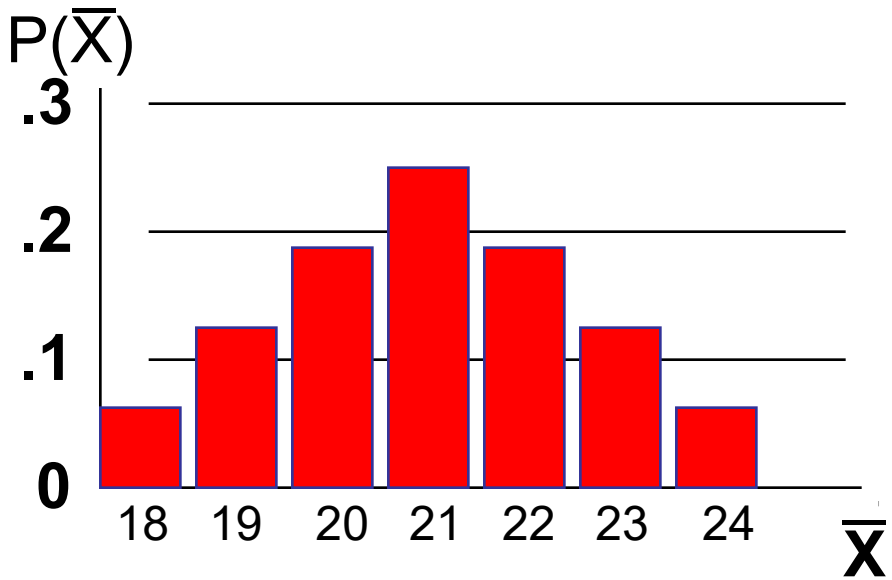
$$= \sqrt{\frac{(18-21)^2 + (19-21)^2 + \cdots + (24-21)^2}{16}} = 1.58$$

Comparing the Population with its Sampling Distribution

Population
N = 4
$\mu = 21 \quad \sigma = 2.236$

Sample Means Distribution
n = 2
$\mu_{\bar{X}} = 21 \quad \sigma_{\bar{X}} = 1.58$





Troy University Resources

Expected Value of Sample Mean

- Let $X_1, X_2, \ldots X_n$ represent a random sample from a population

- The sample mean value of these observations is defined as

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

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:

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

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

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 $\overline{X}$:

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

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

μ = population mean

σ = population standard deviation

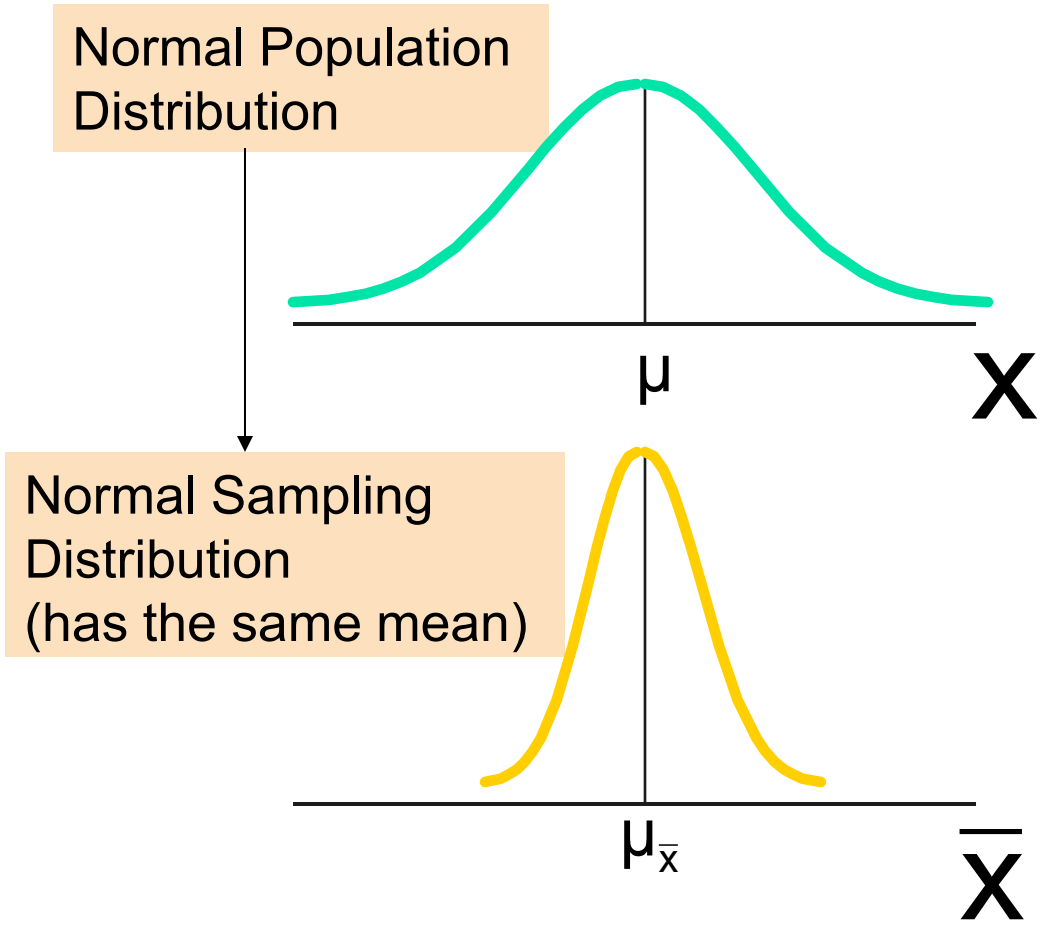
n = sample size

Troy University Resources

Sampling Distribution Properties

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

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



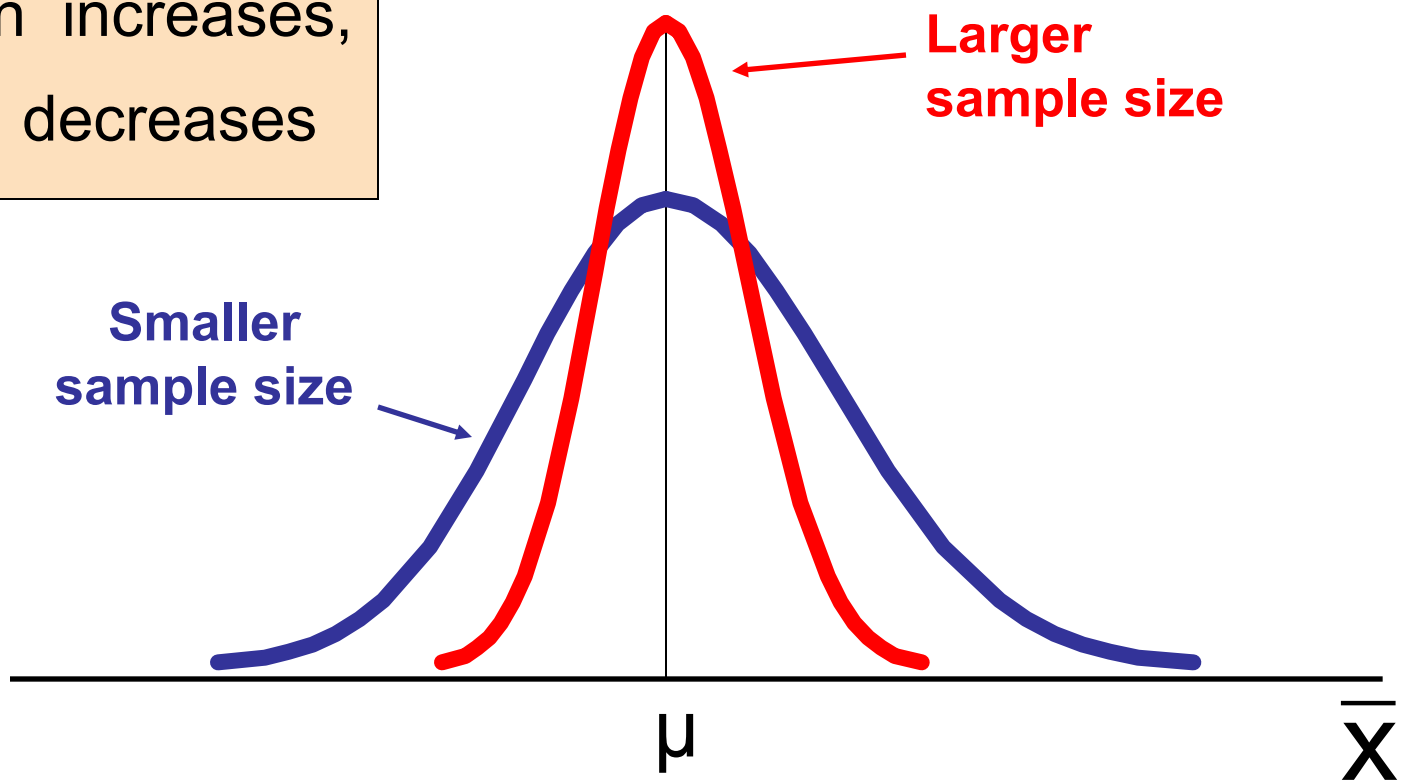
Sampling Distribution Properties

(continued)

- For sampling with replacement:

As n increases,

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



If the Population is **not** Normal

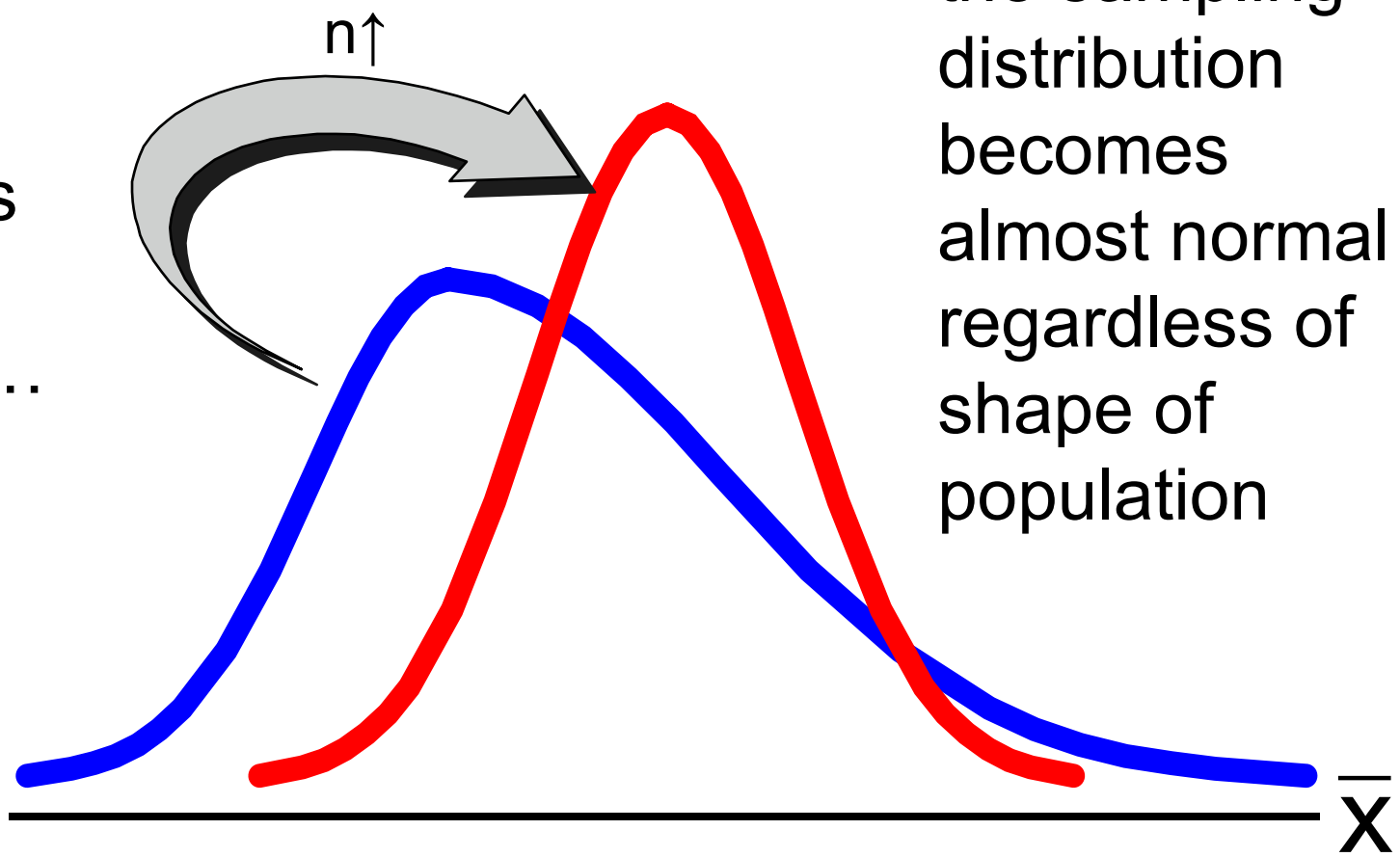
- We can apply the Central Limit Theorem:
  - Even if the population is not normal,
  - …sample means from the population will be approximately normal as long as the sample size is large enough.

Properties of the sampling distribution:

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

Central Limit Theorem

As the sample size gets large enough…



the sampling distribution becomes almost normal regardless of shape of population

If the Population is **not** Normal

(continued)

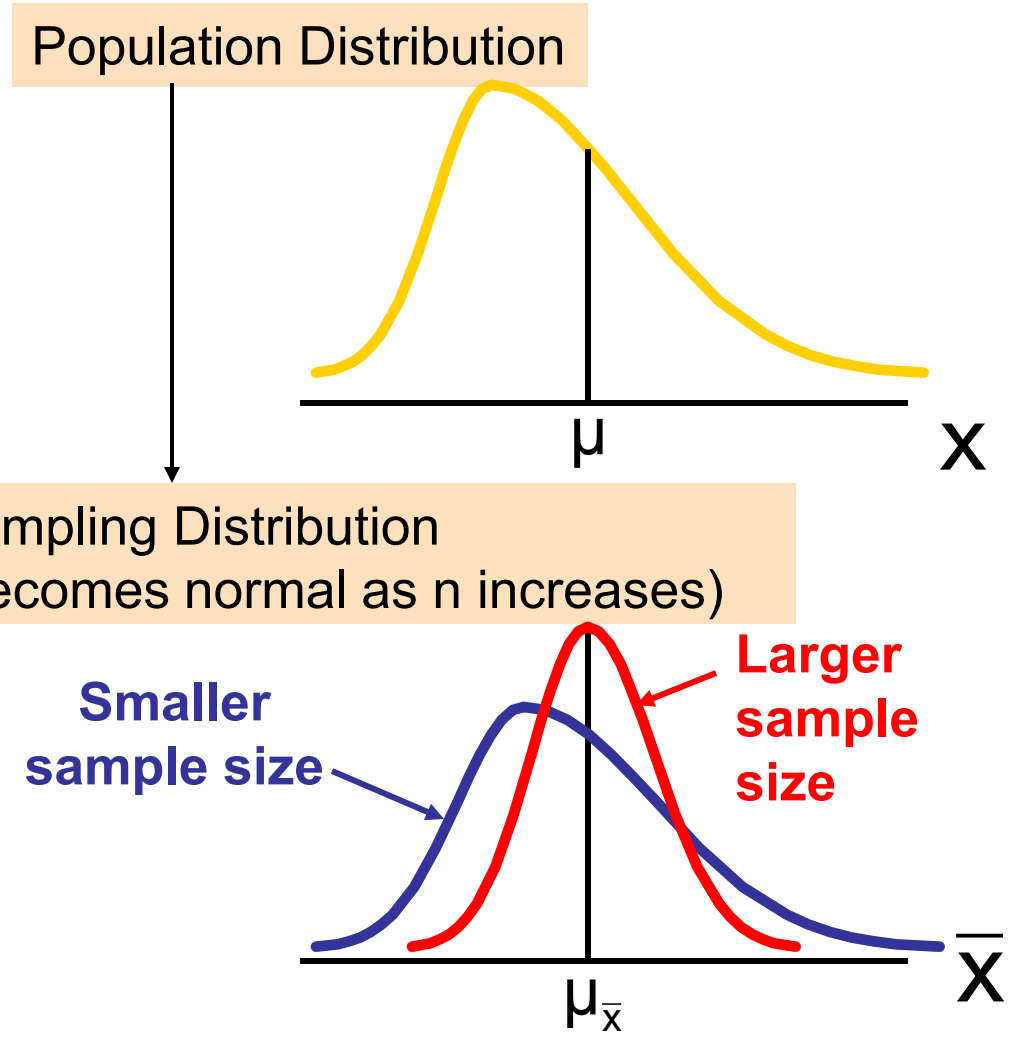
Sampling distribution properties:

Central Tendency

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

Variation

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



How Large is Large Enough?

- For most distributions, $n \geq 30$ will give a sampling distribution that is nearly normal
- For normal population distributions, 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 > 25)
- … so the sampling distribution of $\bar{X}$ is approximately normal
- … with mean $\mu_{\bar{x}} = 8$
- …and standard deviation $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{36}} = 0.5$

Troy University Resources

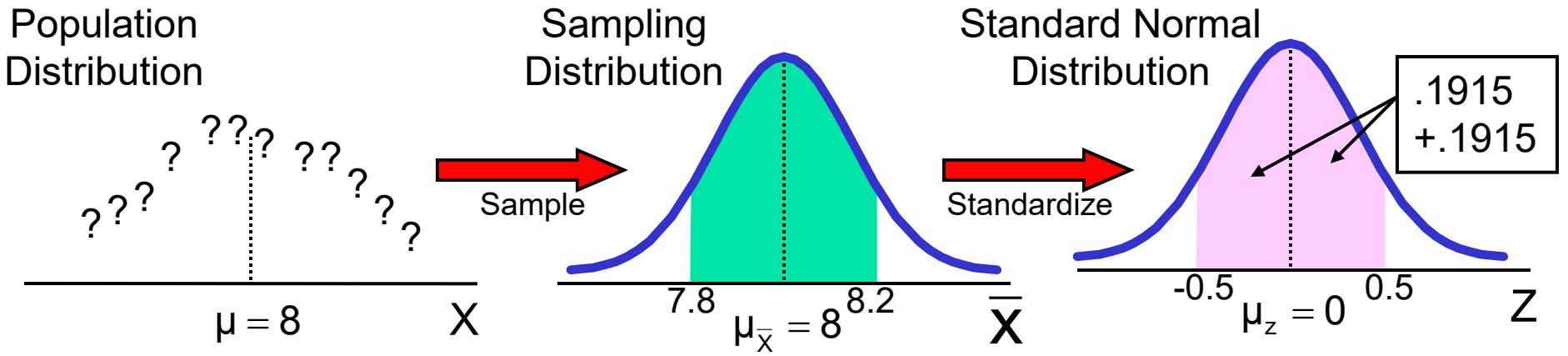
Example

(continued)

Solution (continued):

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

$$= P(-0.5 < Z < 0.5) = \boxed{0.3830}$$



Acceptance Intervals

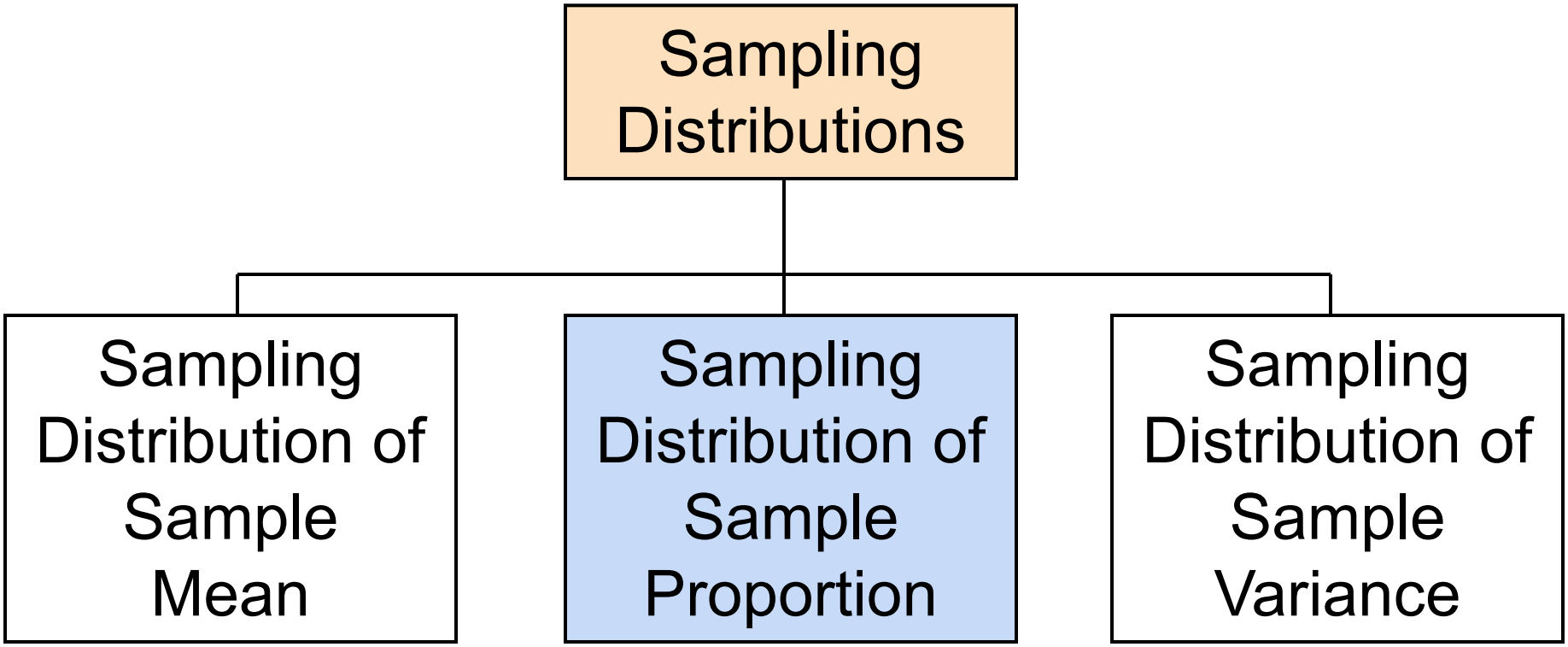
- Goal: determine a range within which sample means are likely to occur, given a population mean and variance
  - By the Central Limit Theorem, we know that the distribution of $\bar{X}$ is approximately normal if n is large enough, with mean μ and standard deviation $\sigma_{\bar{X}}$
  - Let $z_{\alpha/2}$ be the z-value that leaves area α/2 in the upper tail of the normal distribution (i.e., the interval - $z_{\alpha/2}$ to $z_{\alpha/2}$ encloses probability 1 – α)
  - Then

$$\mu \pm z_{\alpha/2}\sigma_{\bar{X}}$$

is the interval that includes $\bar{X}$ with probability 1 – α

Troy University Resources

Sampling Distributions of Sample Proportions



Population Proportions, P

P = the proportion of the population having some characteristic

- Sample proportion ($\hat{P}$) provides an estimate of P:

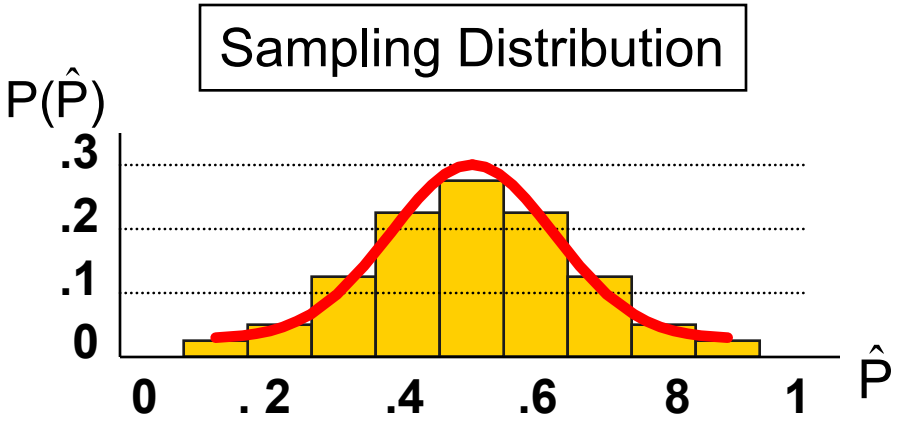
$$\hat{P} = \frac{X}{n} = \frac{\text{number of items in the sample having the characteristic of interest}}{\text{sample size}}$$

- $0 \le \hat{P} \le 1$
- $\hat{P}$ has a binomial distribution, but can be approximated by a normal distribution when $nP(1 - P) \ge 10$

Troy University Resources

Sampling Distribution of $\hat{P}$

- Normal approximation:



Properties:

$$E(\hat{P}) = p$$

and

$$\sigma^2_{\hat{P}} = Var\left(\frac{X}{n}\right) = \frac{P(1-P)}{n}$$

(where P = population proportion)

Z-Value for Proportions

Standardize $\hat{P}$ to a Z value with the formula:

$$Z = \frac{\hat{P} - P}{\sigma_{\hat{P}}} = \frac{\hat{P} - P}{\sqrt{\frac{P(1-P)}{n}}}$$

Example

- If the true proportion of voters who support Proposition A is P = .4, what is the probability that a sample of size 200 yields a sample proportion between .40 and .45?

- i.e.: **if P = .4 and n = 200, what is $P(.40 \leq \hat{P} \leq .45)$?**

Example

(continued)

- **if P = .4 and n = 200, what is $P(.40 \le \hat{P} \le .45)$?**

Find $\sigma_{\hat{P}}$:

$$\sigma_{\hat{P}} = \sqrt{\frac{P(1-P)}{n}} = \sqrt{\frac{.4(1-.4)}{200}} = .03464$$

Convert to standard normal:

$$P(.40 \le \hat{P} \le .45) = P\left(\frac{.40-.40}{.03464} \le Z \le \frac{.45-.40}{.03464}\right)$$
$$= P(0 \le Z \le 1.44)$$

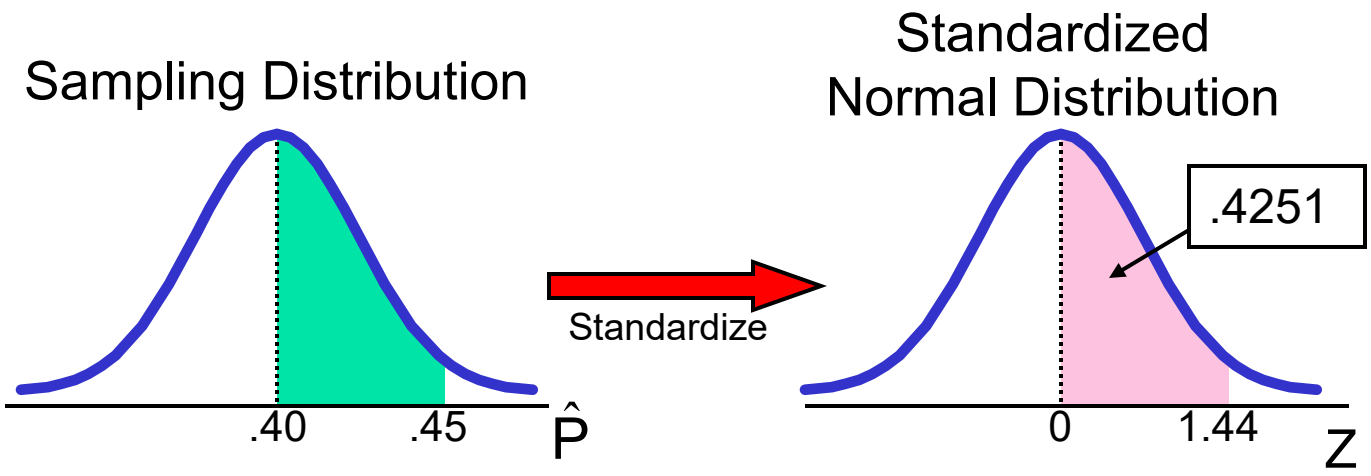
Troy University Resources

Example

(continued)

- **if p = .4 and n = 200, what is $P(.40 \le \hat{P} \le .45)$?**

Use standard normal table: $P(0 \le Z \le 1.44) =$ .4251





Chapter Summary

- Introduced sampling distributions
- Described the sampling distribution of sample means
  - For normal populations
  - Using the Central Limit Theorem
- Described the sampling distribution of sample proportions
- Calculated probabilities using sampling distributions