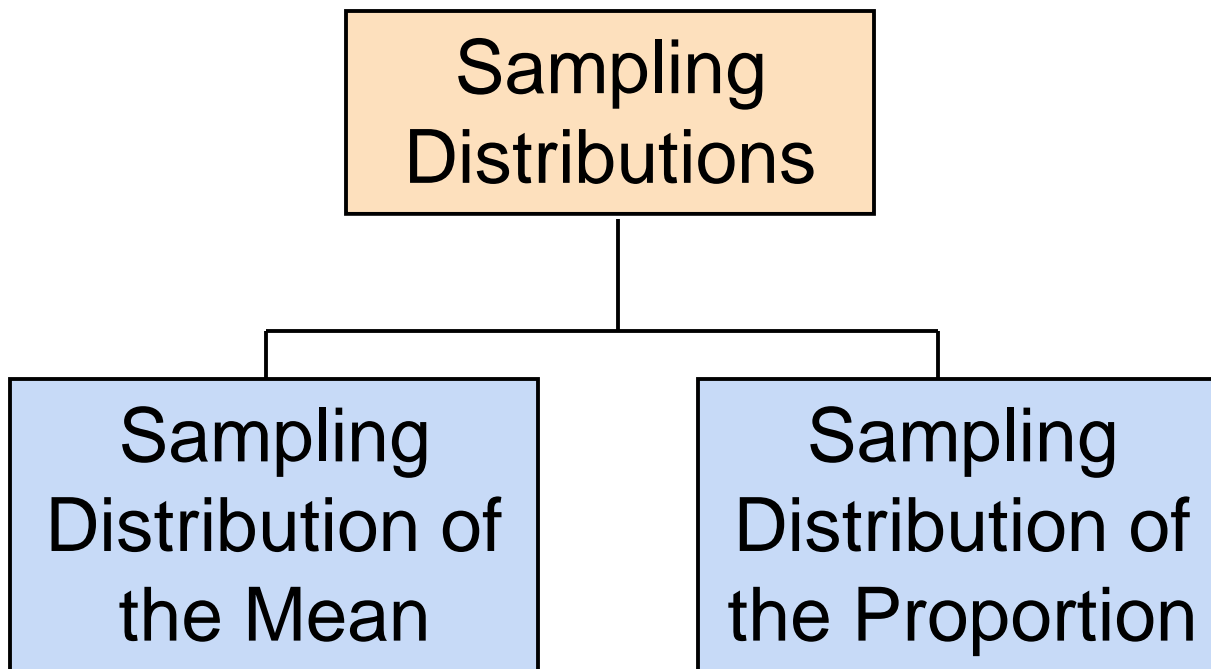




# Sampling Distributions



# 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

# 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)



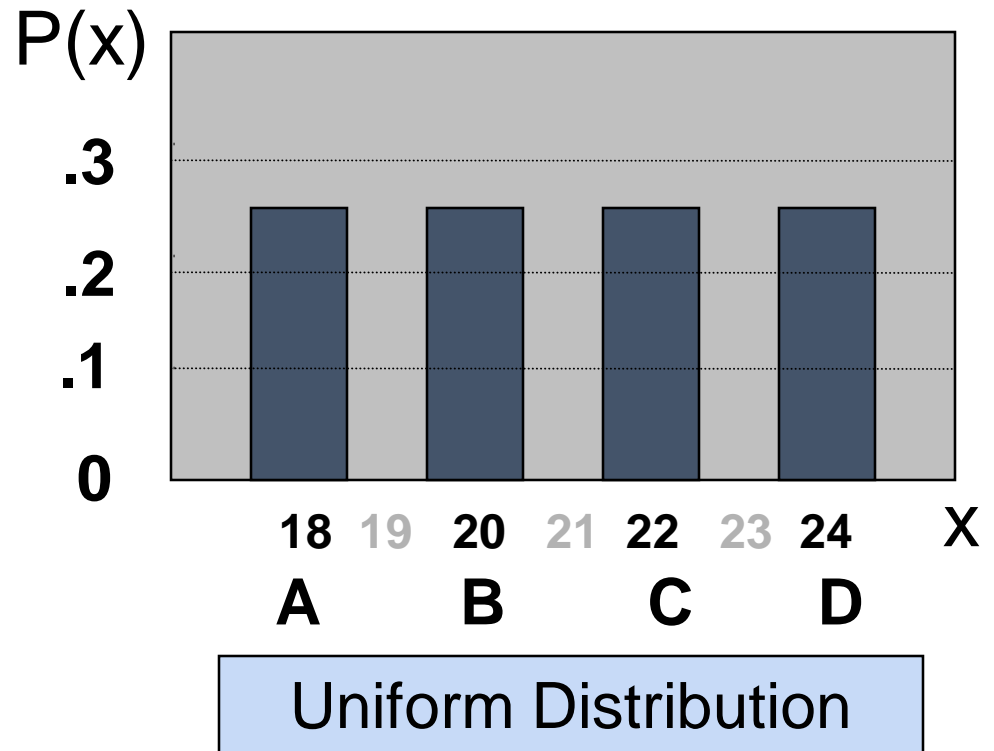
# Developing a Sampling Distribution

*(continued)*

Summary Measures for the Population Distribution:

$$\begin{aligned}\mu &= \frac{\sum X_i}{N} \\ &= \frac{18 + 20 + 22 + 24}{4} = 21\end{aligned}$$

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





# Developing a Sampling Distribution

Now consider all possible samples of size  $n=2$  (continued)

$n=2$

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

Sampling Distribution of All Sample

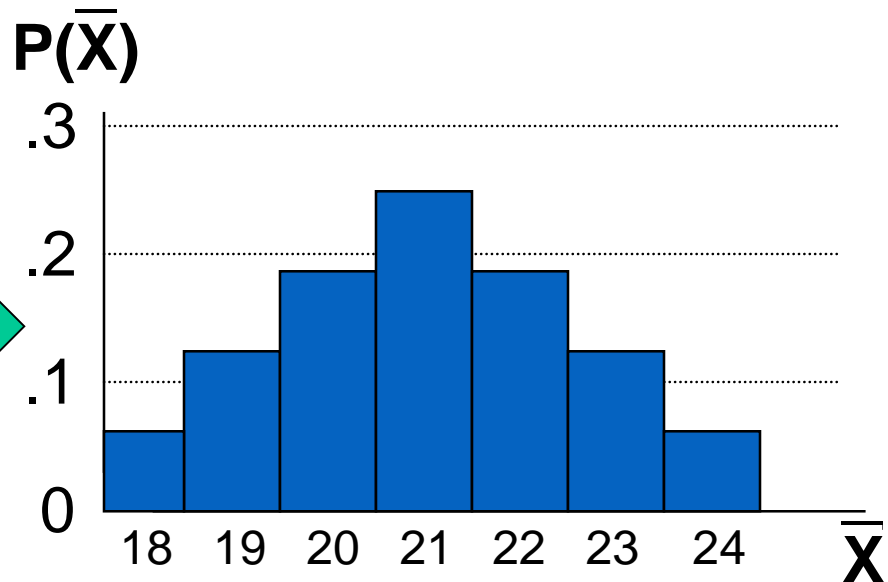
*(continued)*

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:

$$\mu_{\bar{X}} = \frac{\sum \bar{X}_i}{N} = \frac{18 + 19 + 21 + \dots + 24}{16} = 21$$

$$\begin{aligned} \sigma_{\bar{X}} &= \sqrt{\frac{\sum (\bar{X}_i - \mu_{\bar{X}})^2}{N}} \\ &= \sqrt{\frac{(18 - 21)^2 + (19 - 21)^2 + \dots + (24 - 21)^2}{16}} = 1.58 \end{aligned}$$

# Comparing the Population with its Sampling Distribution

Population

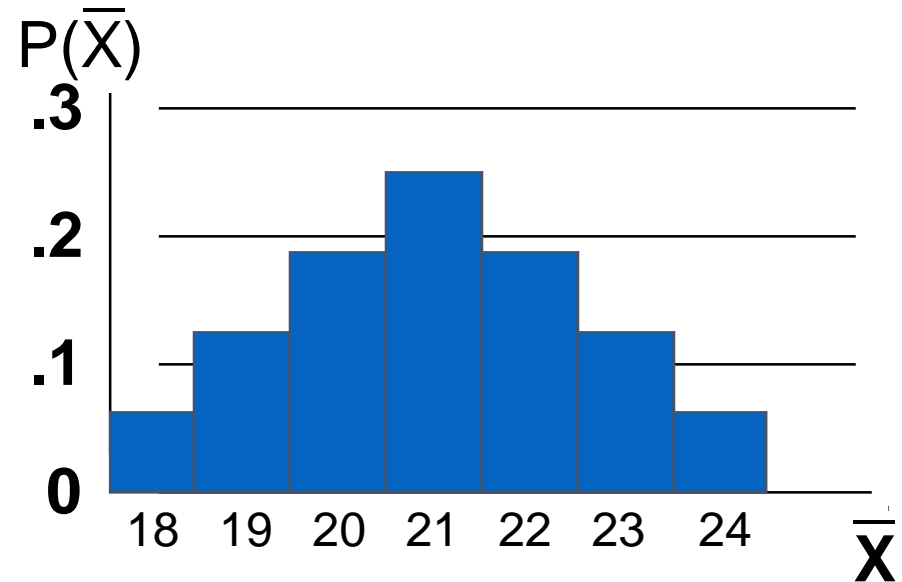
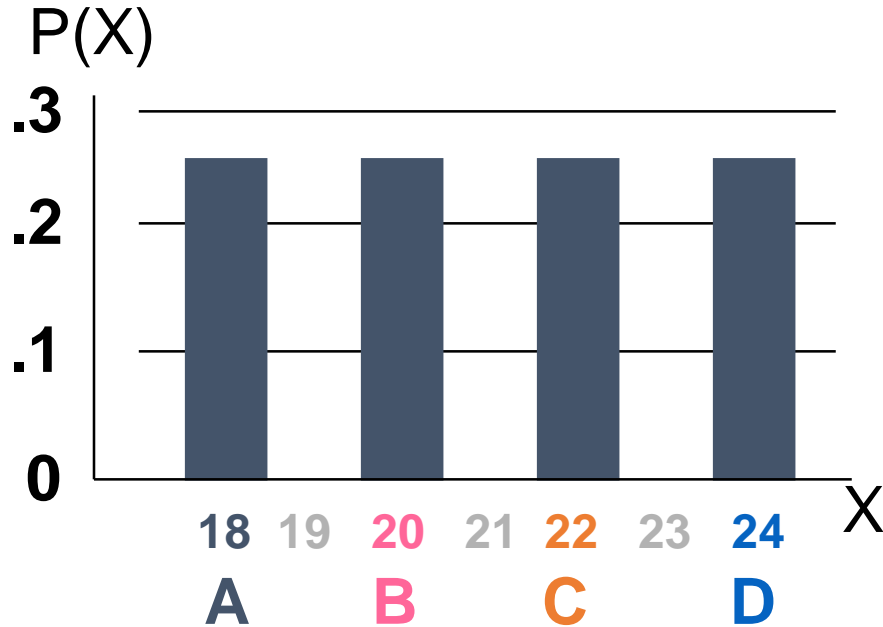
$$N = 4$$

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

Sample Means Distribution

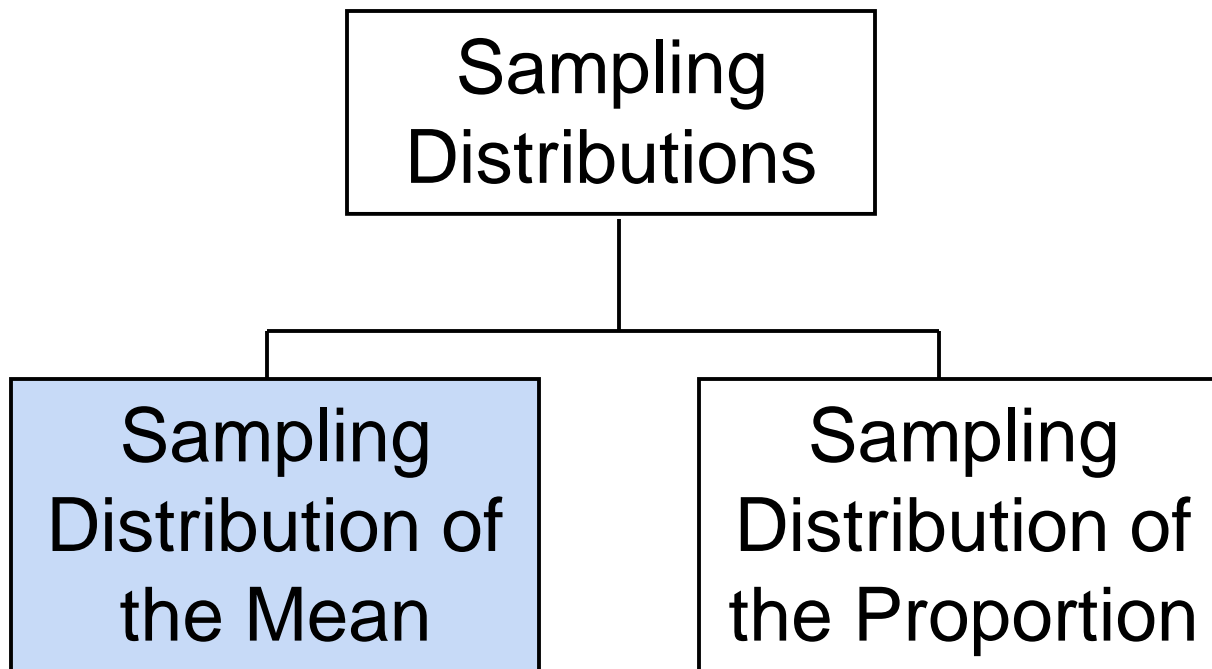
$$n = 2$$

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





# Sampling Distribution of the Mean





# 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



# If the Population is Normal

- If a population is **normal** with mean  $\mu$  and standard deviation  $\sigma$ , 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  $\bar{X}$  :

$\bar{X}$

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

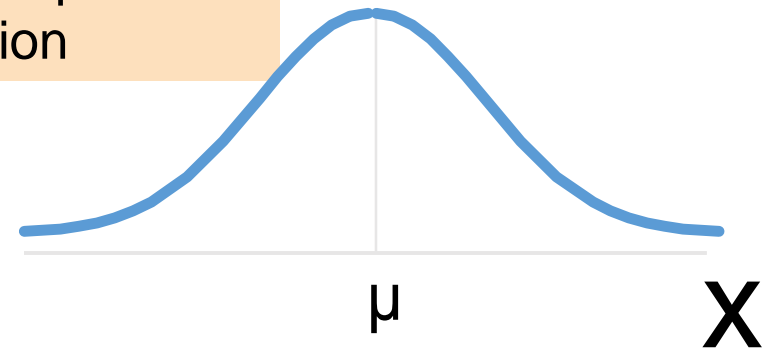
where:  $\bar{X}$  = sample mean  
 $\mu$  = population mean  
 $\sigma$  = population standard deviation  
 $n$  = sample size

# Sampling Distribution Properties

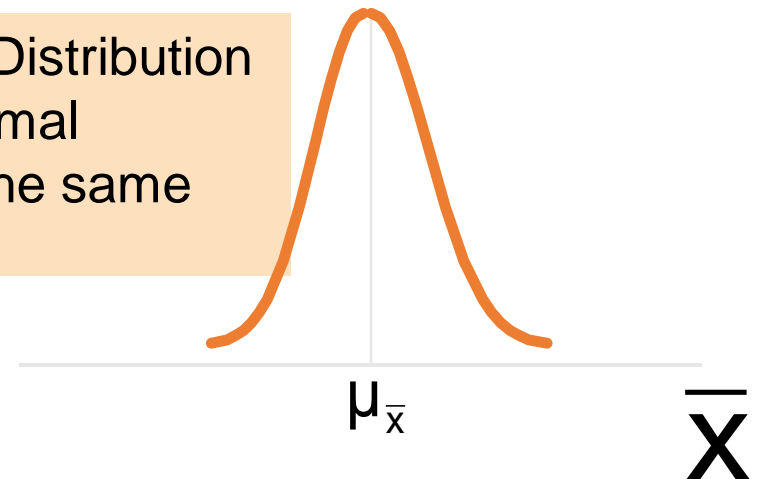
$$\mu_{\bar{X}} = \mu$$

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

Normal Population  
Distribution



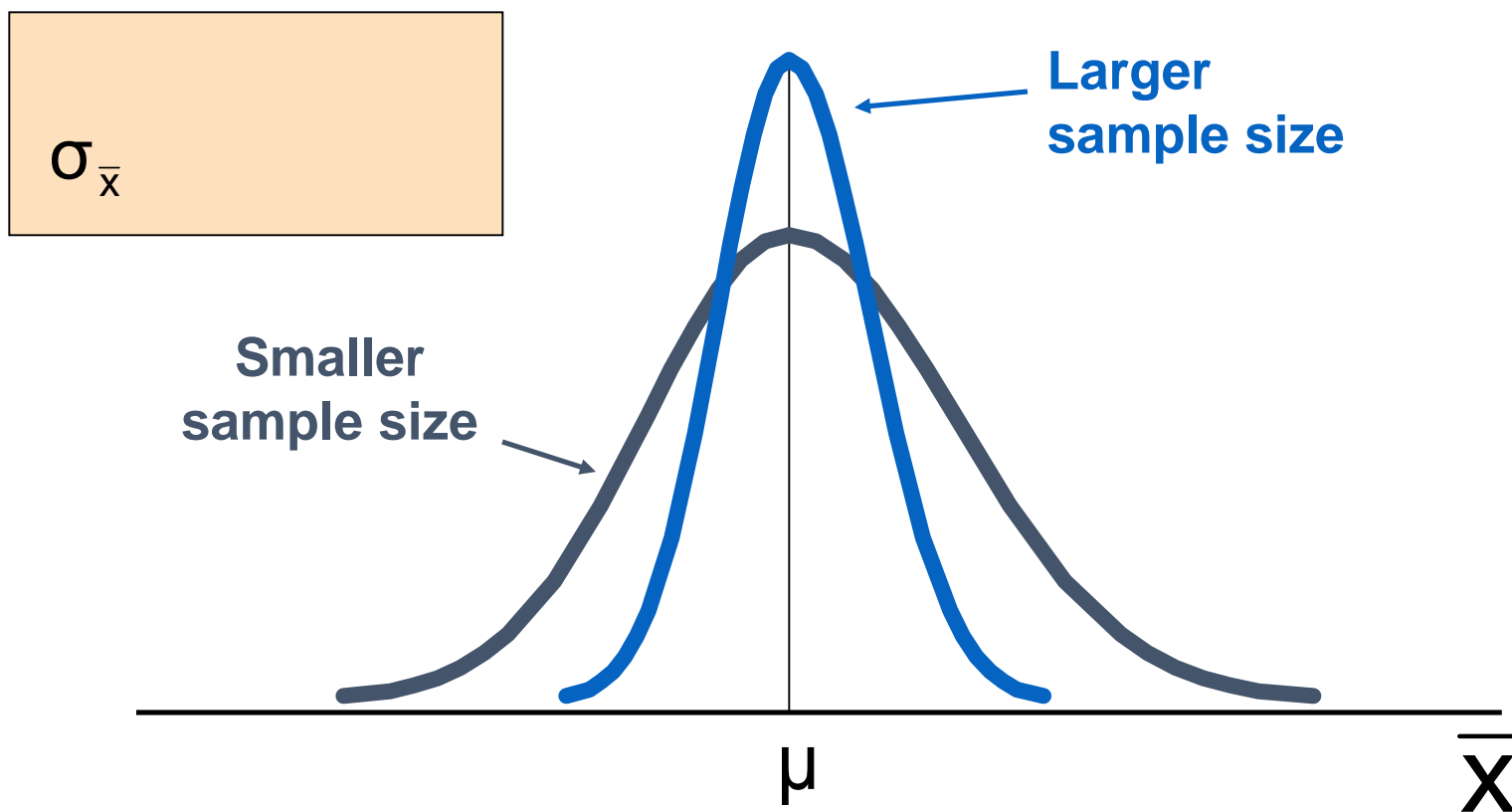
Sampling Distribution  
is also normal  
(and has the same  
mean)





# Sampling Distribution Properties

*(continued)*



# 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:

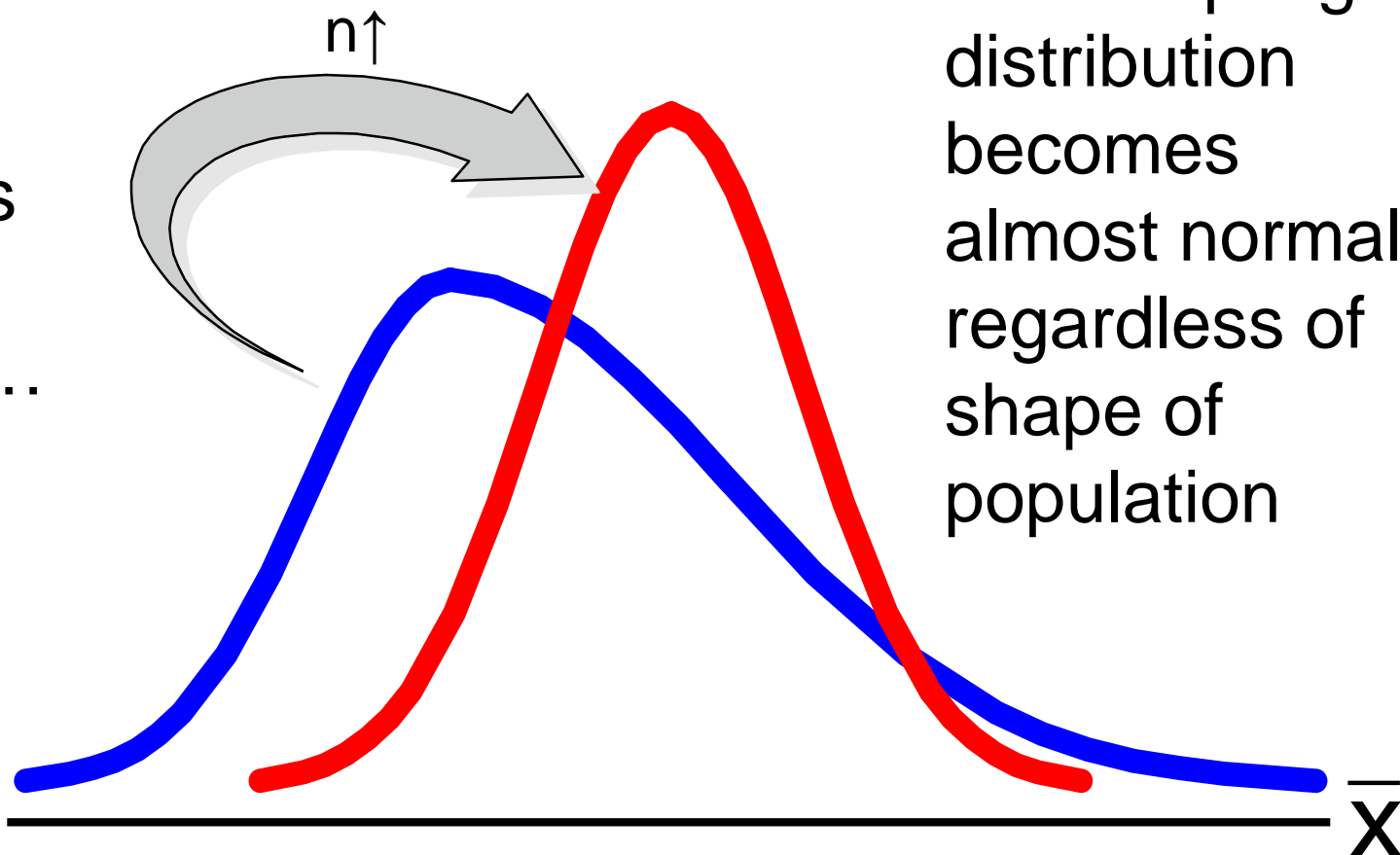
$$\mu_{\bar{x}} = \mu \text{ and}$$

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



# Central Limit Theorem

As the  
sample  
size gets  
large  
enough...



the sampling  
distribution  
becomes  
almost normal  
regardless of  
shape of  
population

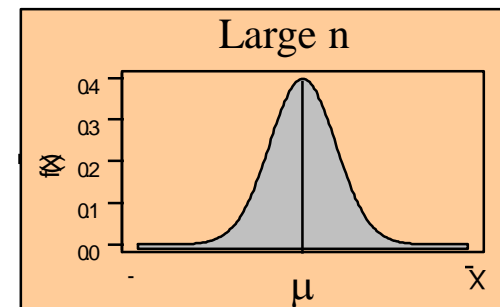
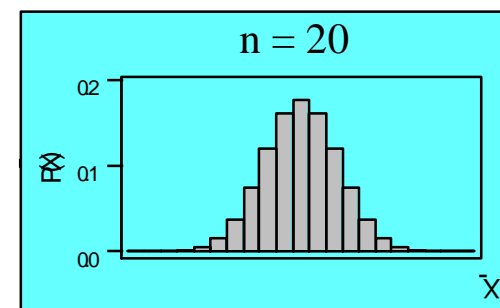
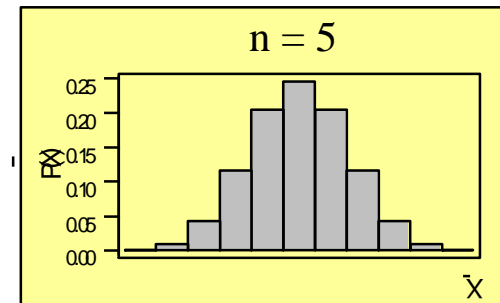




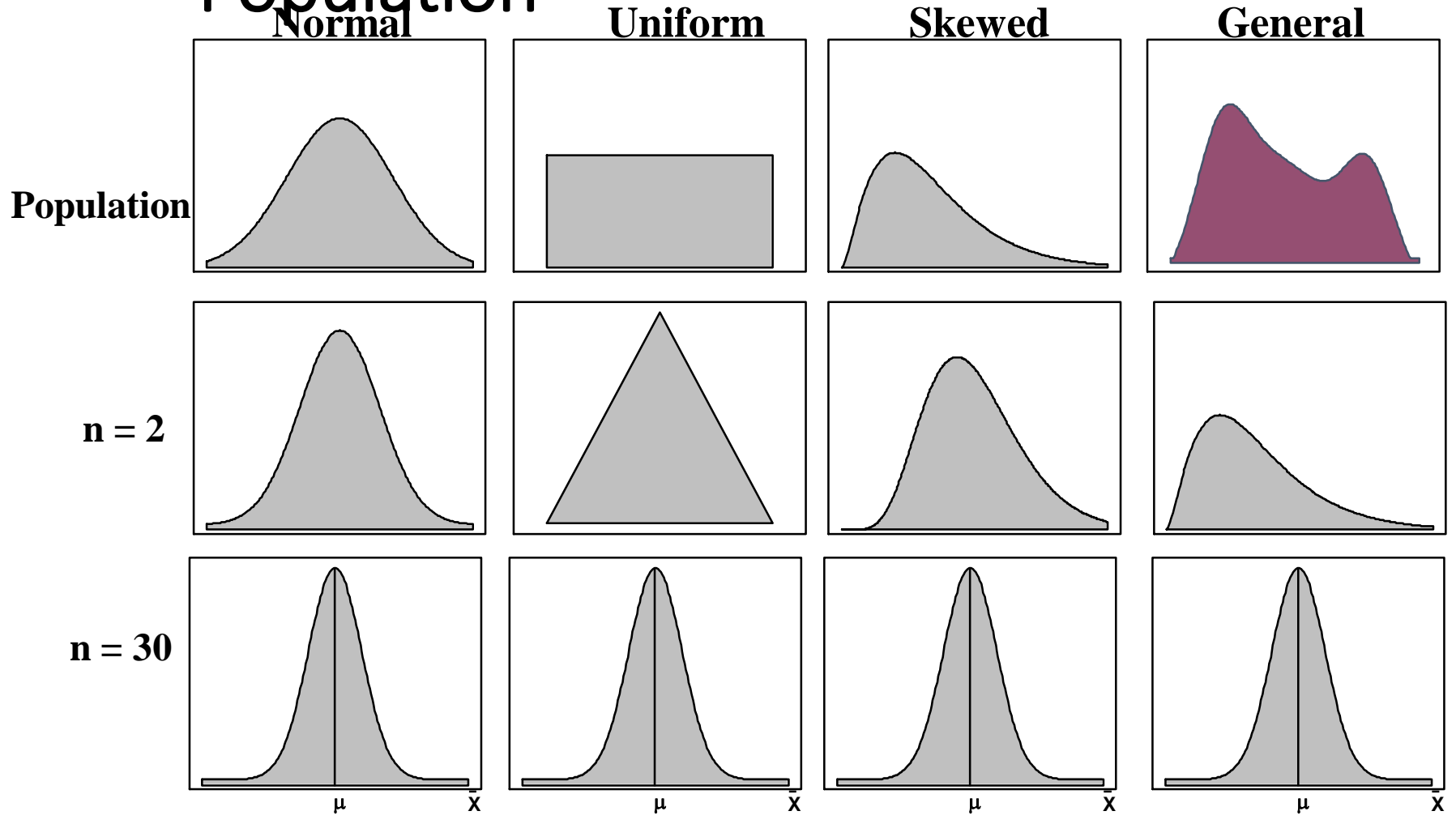
# The Central Limit Theorem

When sampling from a population with mean  $\mu$  and finite standard deviation  $\sigma$ , the sampling distribution of the sample mean will tend to a normal distribution with mean  $\mu$  and standard deviation  $\frac{\sigma}{\sqrt{n}}$  as the sample size becomes large ( $n > 30$ ).

For “large enough”  $n$ :  $\bar{X} \sim N(\mu, \sigma^2 / n)$



# The Central Limit Theorem Applies to Sampling Distributions from **Any** Population



# If the Population is **not** Normal

*(continued)*

Sampling distribution properties:

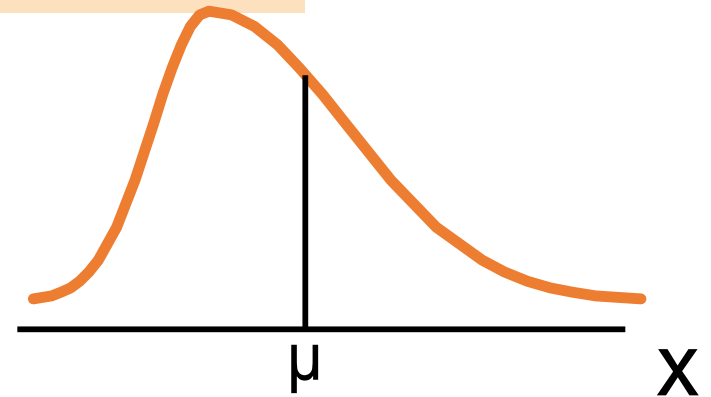
Central Tendency

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

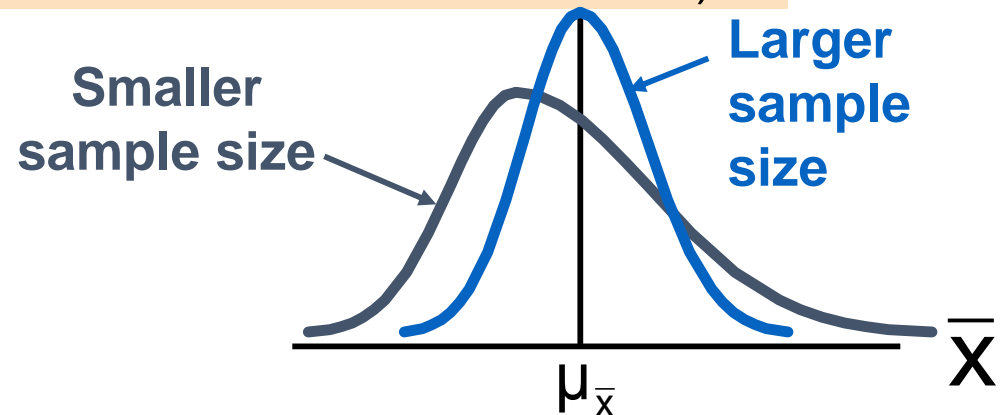
Variation

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

Population Distribution



Sampling Distribution  
(becomes normal as n increases)





# 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 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 ( $n > 30$ )
- ... so the sampling distribution of  $\bar{X}$  is approximately normal
- ... with mean  $\mu_{\bar{X}} = 8$
- ...and standard deviation

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{36}} = 0.5$$



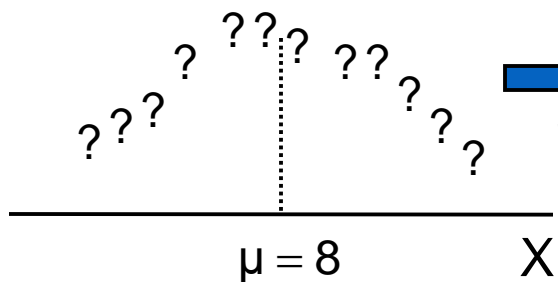
# Example

(continued)

Solution (continued):

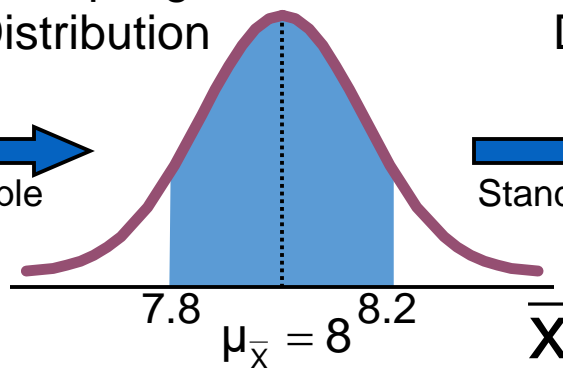
$$P(7.8 < \bar{X} < 8.2) = P\left(\frac{7.8 - 8}{\frac{3}{\sqrt{36}}} < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < \frac{8.2 - 8}{\frac{3}{\sqrt{36}}}\right)$$
$$= P(-0.4 < Z < 0.4) = \boxed{0.3108}$$

Population  
Distribution



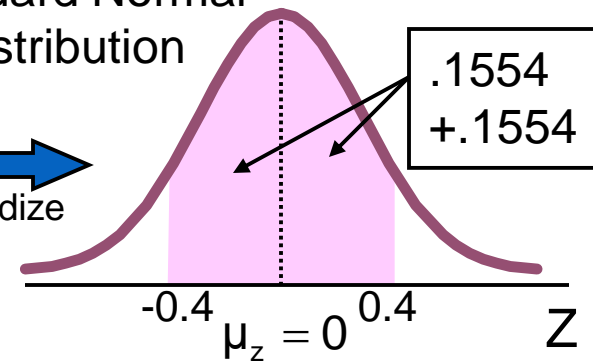
Sampling  
Distribution

Sample



Standard Normal  
Distribution

Standardize





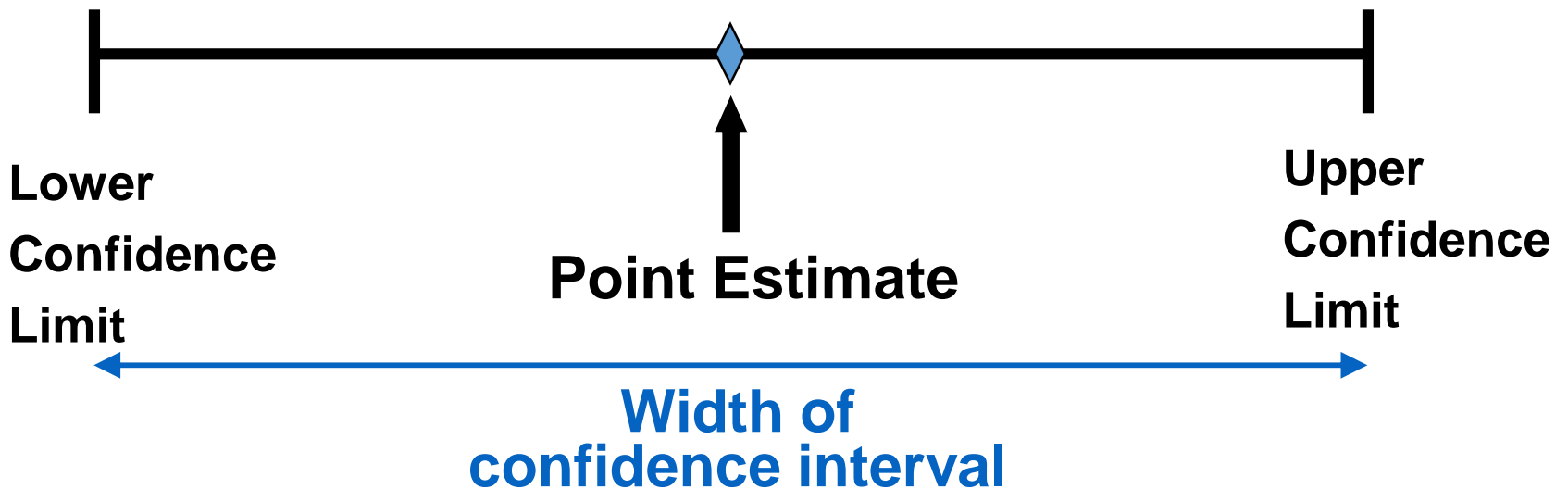
**Centurion**  
**UNIVERSITY**  
*Shaping Lives...*  
*Empowering Communities...*

# ESTIMATES



# Point and Interval Estimates

- A **point estimate** is a single number,
- a **confidence interval** provides additional information about variability





# Point Estimates

We can estimate a Population Parameter ...		with a Sample Statistic (a Point Estimate)
Mean	$\mu$	$\bar{X}$
Proportion	$p$	$p_s$



# Confidence Intervals

- 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**

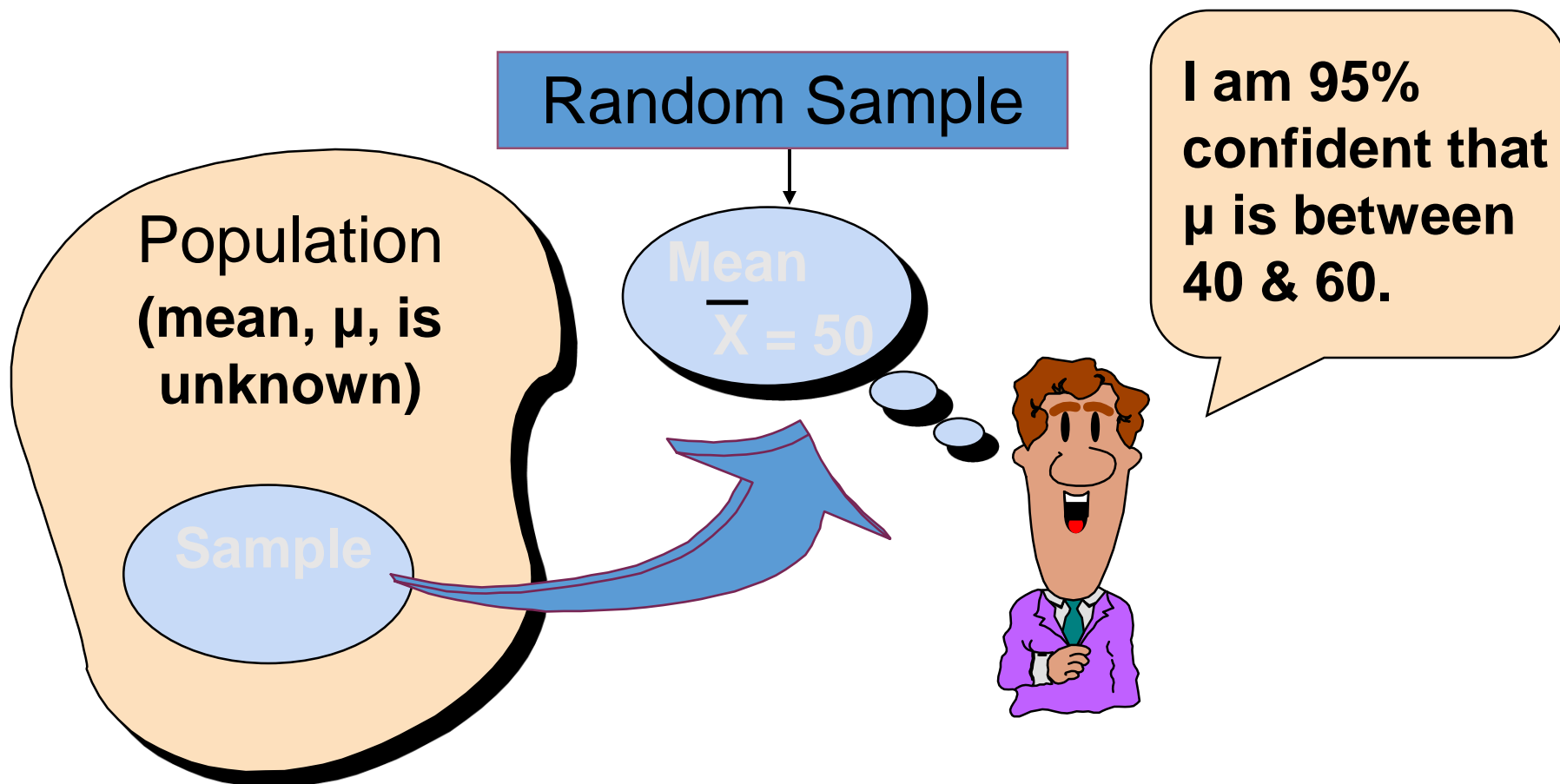


# Confidence Interval Estimate

- 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
    - Can never be 100% confident



# Estimation Process





# General Formula

- The general formula for all confidence intervals is:

**Point Estimate  $\pm$  (Critical Value)(Standard Error)**

# Confidence Level

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



# Confidence Level, $(1-\alpha)$

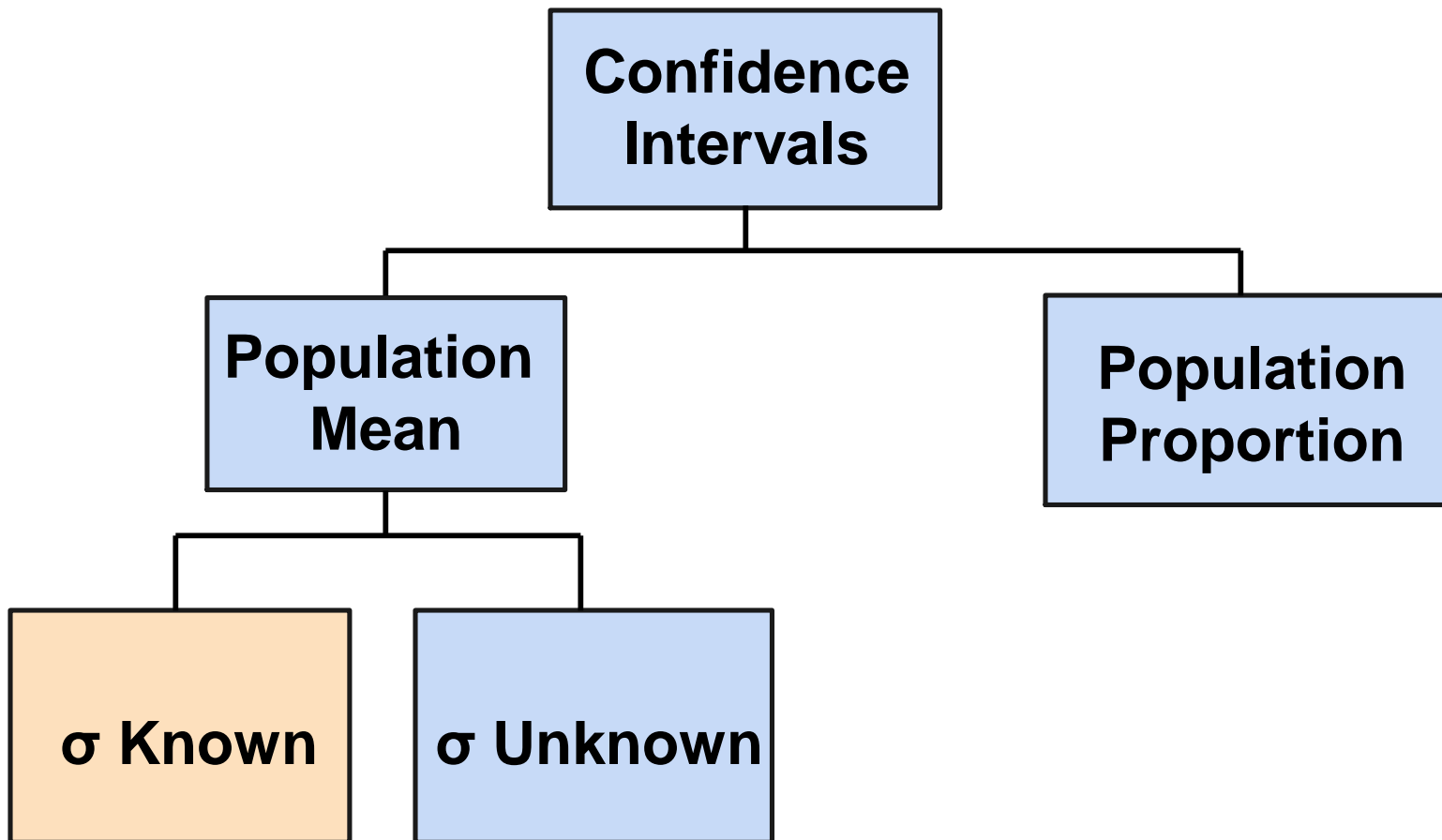
*(continued)*

- Suppose confidence level = 95%
- Also written  $(1 - \alpha) = 0.95$
- A relative frequency interpretation:
  - In the long run, 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



# Confidence Interval for $\mu$ ( $\sigma$ Known)

- Assumptions
  - Population standard deviation  $\sigma$  is known
  - Population is normally distributed
  - If population is not normal, use large sample
- Confidence interval estimate:

$$\bar{X} \pm Z \frac{\sigma}{\sqrt{n}}$$

- where  $\bar{X}$  is the point estimate

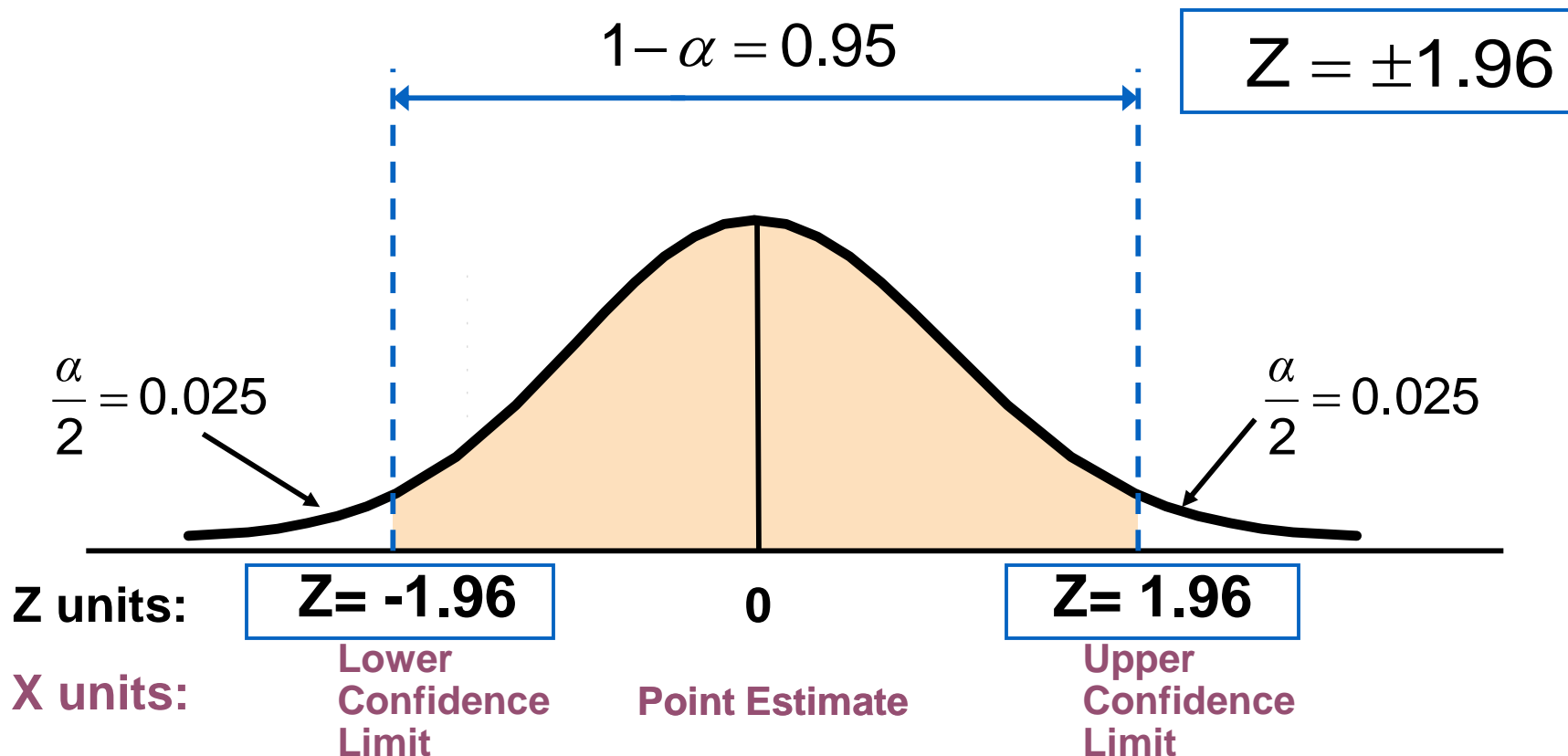
$Z$  is the normal distribution critical value for a probability of  $\alpha/2$  in each tail

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



# Finding the Critical Value, Z

- Consider a 95% confidence interval:



# Common Levels of Confidence

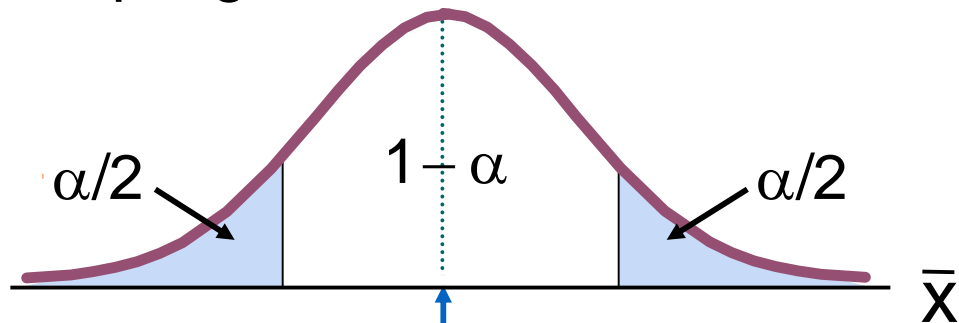
- Commonly used confidence levels are 90%, 95%, and 99%

<b>Confidence Level</b>	<b>Confidence Coefficient, <math>1 - \alpha</math></b>	<b>Z value</b>
<b>80%</b>	<b>0.80</b>	<b>1.28</b>
<b>90%</b>	<b>0.90</b>	<b>1.645</b>
<b>95%</b>	<b>0.95</b>	<b>1.96</b>
<b>98%</b>	<b>0.98</b>	<b>2.33</b>
<b>99%</b>	<b>0.99</b>	<b>2.58</b>
<b>99.8%</b>	<b>0.998</b>	<b>3.08</b>
<b>99.9%</b>	<b>0.999</b>	<b>3.27</b>



# Intervals and Level of Confidence

## Sampling Distribution of the Mean

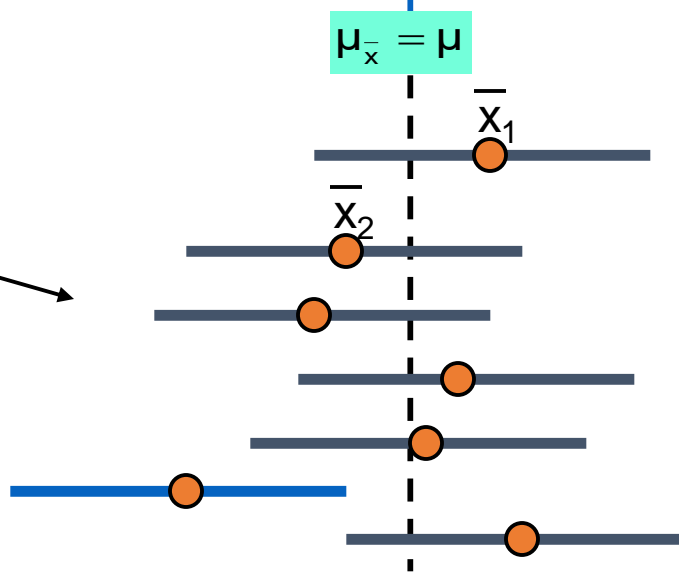


Intervals  
extend from

$$\bar{X} + Z \frac{\sigma}{\sqrt{n}}$$

to

$$\bar{X} - Z \frac{\sigma}{\sqrt{n}}$$



$(1 - \alpha) \times 100\%$   
of intervals  
constructed  
contain  $\mu$ ;  
 $(\alpha) \times 100\%$  do  
not.



# What is a Hypothesis?

- A hypothesis is a claim (assumption) about a population parameter:
  - population mean



**Example: The mean monthly cell phone bill of this city is  $\mu = \text{Rs.}42$**

- population proportion

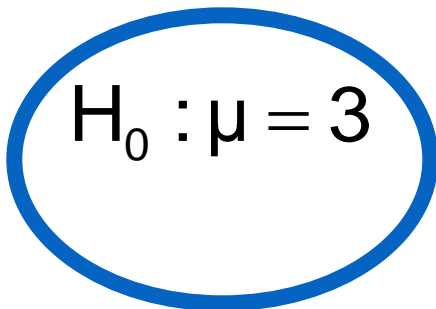
**Example: The proportion of adults in this city with cell phones is  $p = 0.68$**

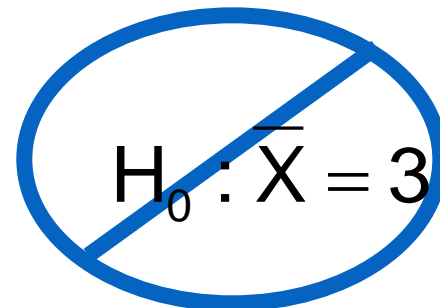
# The Null Hypothesis, $H_0$

- States the claim or assertion to be tested


$$H_0 : \mu = 3$$

- Is always about a population parameter, not about a sample statistic

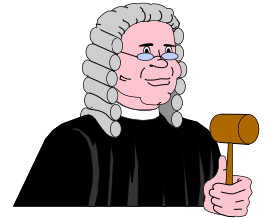

$$H_0 : \mu = 3$$


$$\cancel{H_0 : \bar{X} = 3}$$

# The Null Hypothesis, $H_0$

*(continued)*

- Begin with the assumption that the null hypothesis is true
  - Similar to the notion of innocent until proven guilty
- Refers to the status quo
- Always contains “=”, “ $\leq$ ” or “ $\geq$ ” sign
- May or may not be rejected





# The Alternative Hypothesis, $H_1$

- Is the opposite of the null hypothesis
  - e.g., The average number of TV sets in U.S. homes is not equal to 3 (  $H_1: \mu \neq 3$  )
- Challenges the status quo
- Never contains the “=” , “≤” or “≥” sign
- May or may not be proven
- Is generally the hypothesis that the researcher is trying to prove



# Level of Significance, $\alpha$

- **Defines the unlikely values of the sample statistic if the null hypothesis is true**
  - Defines **rejection region** of the sampling distribution
- Is designated by  **$\alpha$** , (level of significance)
  - Typical values are 0.01, 0.05, or 0.10
- Is selected by the researcher at the beginning
- Provides the critical value(s) of the test

# Level of Significance and the Rejection Region

Level of significance =  $\alpha$

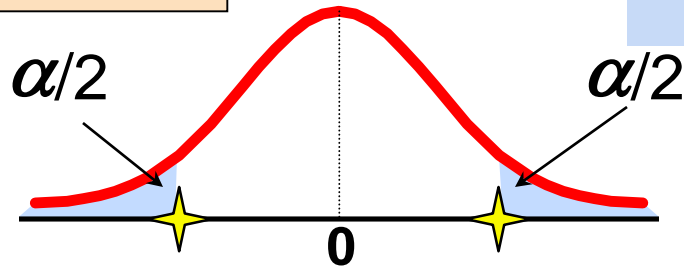
✦ Represents critical value

Rejection region is shaded

$H_0: \mu = 3$

$H_1: \mu \neq 3$

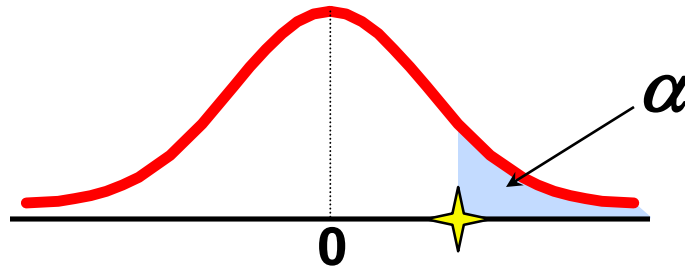
Two-tail test



$H_0: \mu \leq 3$

$H_1: \mu > 3$

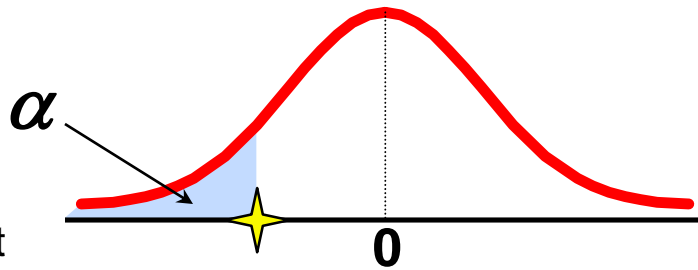
Upper-tail test



$H_0: \mu \geq 3$

$H_1: \mu < 3$

Lower-tail test





# Errors in Making Decisions

- **Type I Error**

- Reject a true null hypothesis
- Considered a serious type of error

The probability of Type I Error is  $\alpha$

- Called **level of significance** of the test
- Set by the researcher in advance



# Errors in Making Decisions

*(continued)*

- **Type II Error**
  - Fail to reject a false null hypothesis

The probability of Type II Error is  $\beta$



# Outcomes and Probabilities



## Possible Hypothesis Test Outcomes

	Actual Situation	
Decision	$H_0$ True	$H_0$ False
Do Not Reject $H_0$	No error ( $1 - \alpha$ )	Type II Error ( $\beta$ )
Reject $H_0$	Type I Error ( $\alpha$ )	No Error ( $1 - \beta$ )






**Key:**  
**Outcome**  
**(Probability)**

# Type I & II Error Relationship

- Type I and Type II errors cannot happen at the same time
  - Type I error can only occur if  $H_0$  is **true**
  - Type II error can only occur if  $H_0$  is **false**

If Type I error probability (  $\alpha$  ) , then  
Type II error probability (  $\beta$  ) 

# Factors Affecting Type II Error

- All else equal,
  - $\beta$   when the difference between hypothesized parameter and its true value
  - $\beta$   when  $\alpha$
  - $\beta$   when  $\sigma$  
  - $\beta$   when  $n$  