



# Introduction Into Probability Theory

**MTH 231**

**Lecture 8**

**Chapter 6**

**Fundamental of Statistical Analysis  
Parameter Estimation, Confidence Intervals & Test  
of Hypothesis**



# Today's lecture

---

- Parameter Estimation
- Confidence Intervals
- Sampling Error
- Test of Hypothesis

# Parameter Estimation

## Populations and Samples

The terms population and sample are defined in statistics as follows:

**Population:** It is a collection of all possible individuals, about which we require information. A population may be finite or infinite. For example the population consisting of all bolts produced in a factory on a given day is finite; the population consisting of all grains of sand in the world is infinite.

For a population of size **N**, its mean,  $\mu$ , is given by

$$\mu = \frac{\sum_{i=1}^N x_i}{N} = \sum_{i=1}^N x_i \frac{1}{N}$$

**Sample:** A sample is a portion of the population of interest.

For a sample of size **n**, the sample mean,  $\bar{X}$ , is given by

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$$

# Point and Interval Estimates

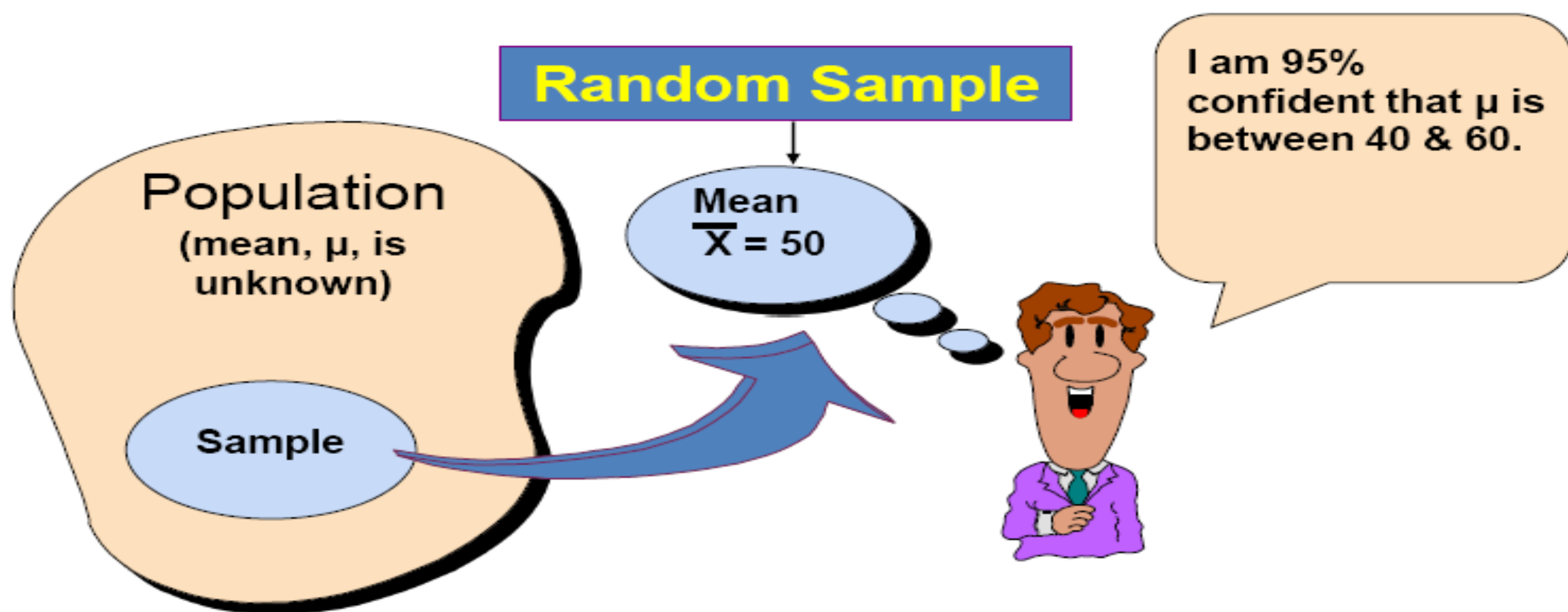
- A **parameter** is a population measure (e.g.  $\mu$ ,  $\sigma^2$ ).
- A **statistic** is a sample function (e.g.  $\bar{X}$ ,  $S^2$ ).
- Hence statistics may be regarded as random variables.
- Statistics are used to estimate parameters and are called **point estimators**.
- A **point estimate** of a parameter is a single numerical value of a respective estimator.
- The standard deviation of an estimator is called the **standard error**.

## Point Estimates

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

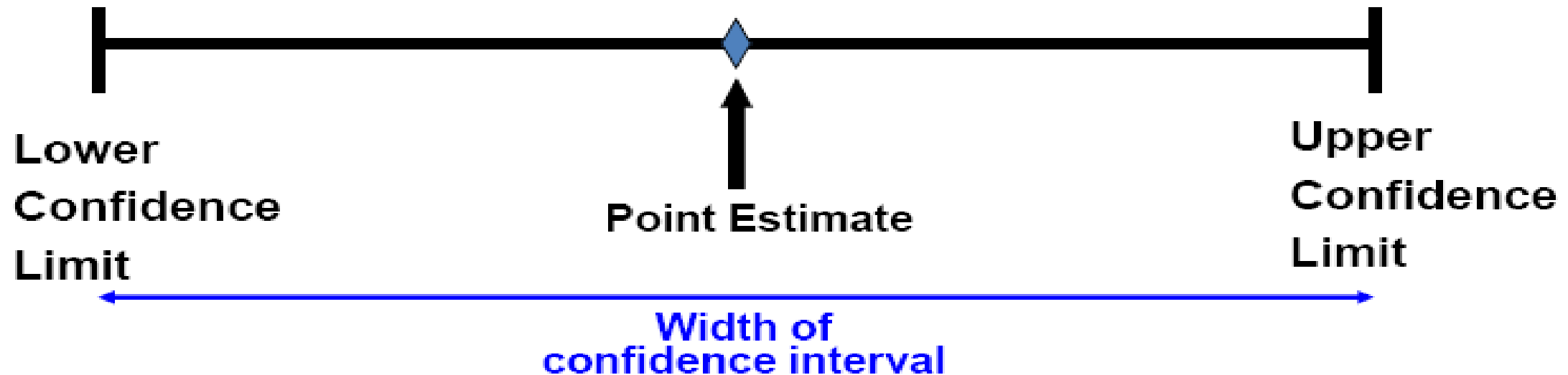
# Estimation Process

---



# Interval Estimates

- An **interval estimate of a parameter** is an interval within which the parameter is estimated to exist.
- The **confidence level of an interval estimate** is the probability that the interval contains the parameter.
- Notation: An interval estimate with a confidence level  $1 - \alpha$ , is referred to as a  **$1 - \alpha$  confidence interval**.



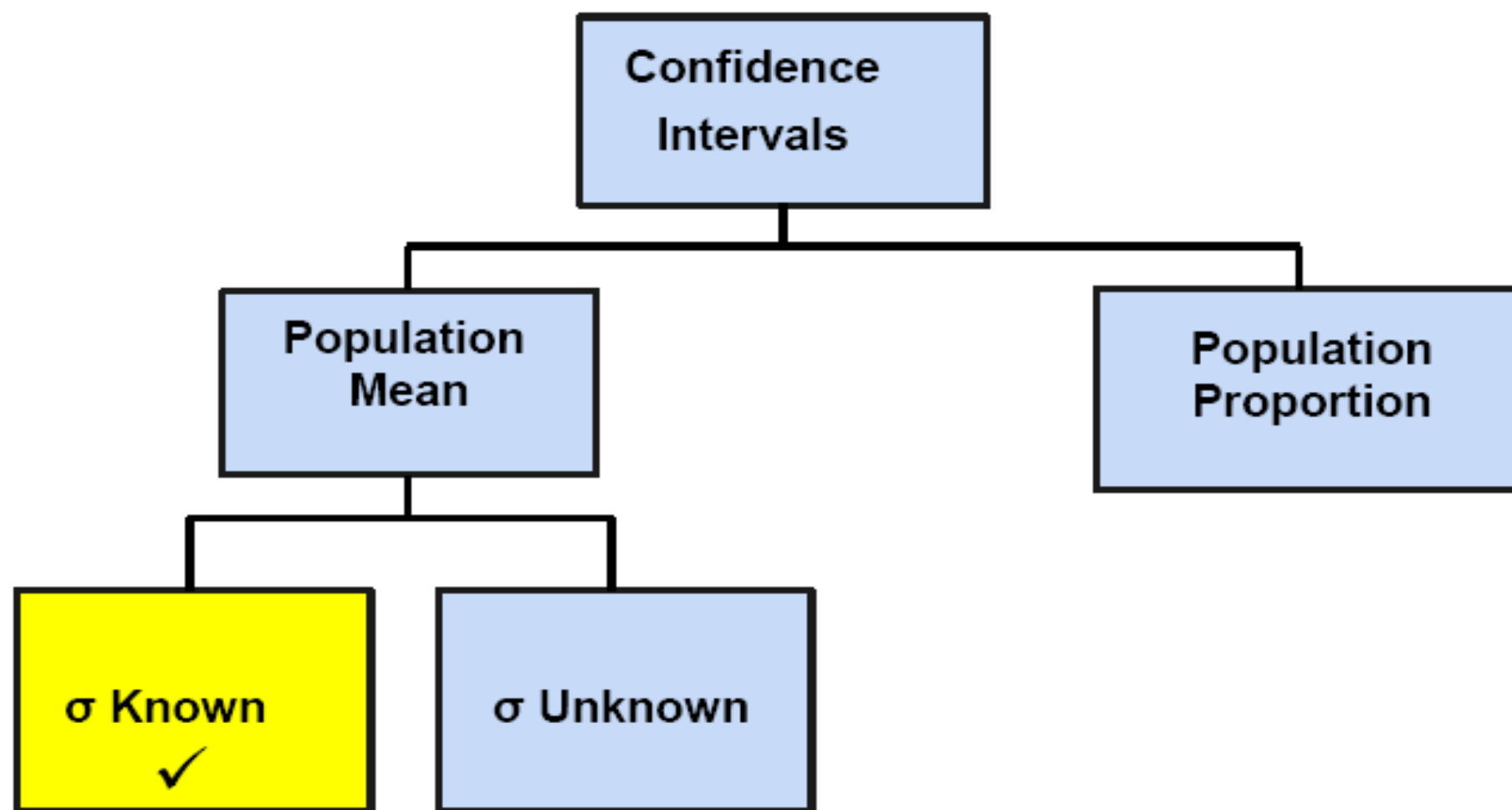
## Confidence Interval Estimate

---

- 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**
- An interval gives a **range** of values:
  - Based on observation from 1 sample
  - Gives information about closeness to unknown population parameters
  - Stated in terms of level of confidence
  - Can never be 100% confident

# Confidence Intervals

---

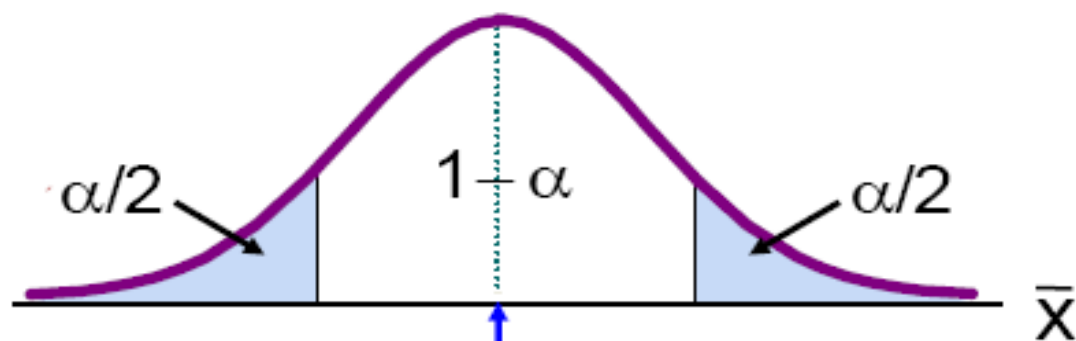




# Intervals and Level of Confidence

*(continued)*

Sampling Distribution of the Mean



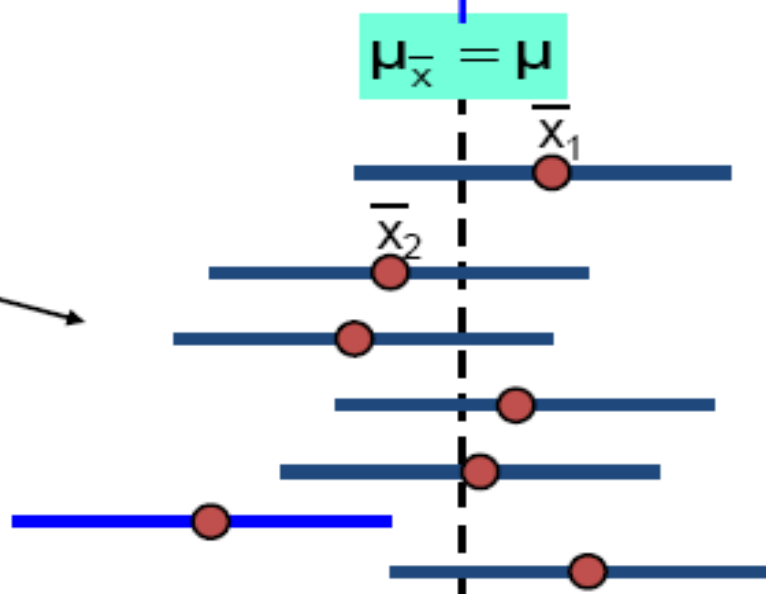
$$\mu_{\bar{x}} = \mu$$

Intervals  
extend from

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

to

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



Confidence Intervals

$(1 - \alpha) \times 100\%$   
of intervals  
constructed contain  
 $\mu$

# Interval Estimates of the Population Mean

- An interval estimate on the mean is an interval centered at the sample mean is:

$$(\bar{X} - \varepsilon, \bar{X} + \varepsilon)$$

- $\varepsilon$  is the **maximum error of estimation**.
- Saying that  $\mu \in (\bar{X} - \varepsilon, \bar{X} + \varepsilon)$  is equivalent to saying that  $\mu = \bar{X} \pm \varepsilon$
- How confident we are in this statement depends on  $(1 - \alpha)$  (the degree of confidence).

**The general formula for all confidence intervals is:**

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

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

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

## 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

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

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

# ① A $(1-\alpha)$ 100% two-sided confidence interval of $\mu$

---

Or,

$$\left( \bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \quad \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)$$

$$\bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

Where  $Z_{\frac{\alpha}{2}}$  is the Z- value above which we find an area of  $\frac{\alpha}{2}$  that

is  $P(Z > Z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$

## ▪ Assumptions

Population standard deviation  $\sigma$  is known

Population is normally distributed

## ② A $(1-\alpha)$ 100% one-sided confidence interval of $\mu$

---

(i) A  $(1 - \alpha)$  100% One-sided Lower interval for  $\mu$

$$(\bar{X} - Z_{\alpha} \frac{\sigma}{\sqrt{n}}, \quad \infty)$$

Or,

$$\bar{X} - Z_{\alpha} \frac{\sigma}{\sqrt{n}} < \mu < \infty$$

(ii) A  $(1 - \alpha)$  100% One-sided Upper interval for  $\mu$

$$(-\infty, \quad \bar{X} + Z_{\alpha} \frac{\sigma}{\sqrt{n}})$$

Or,

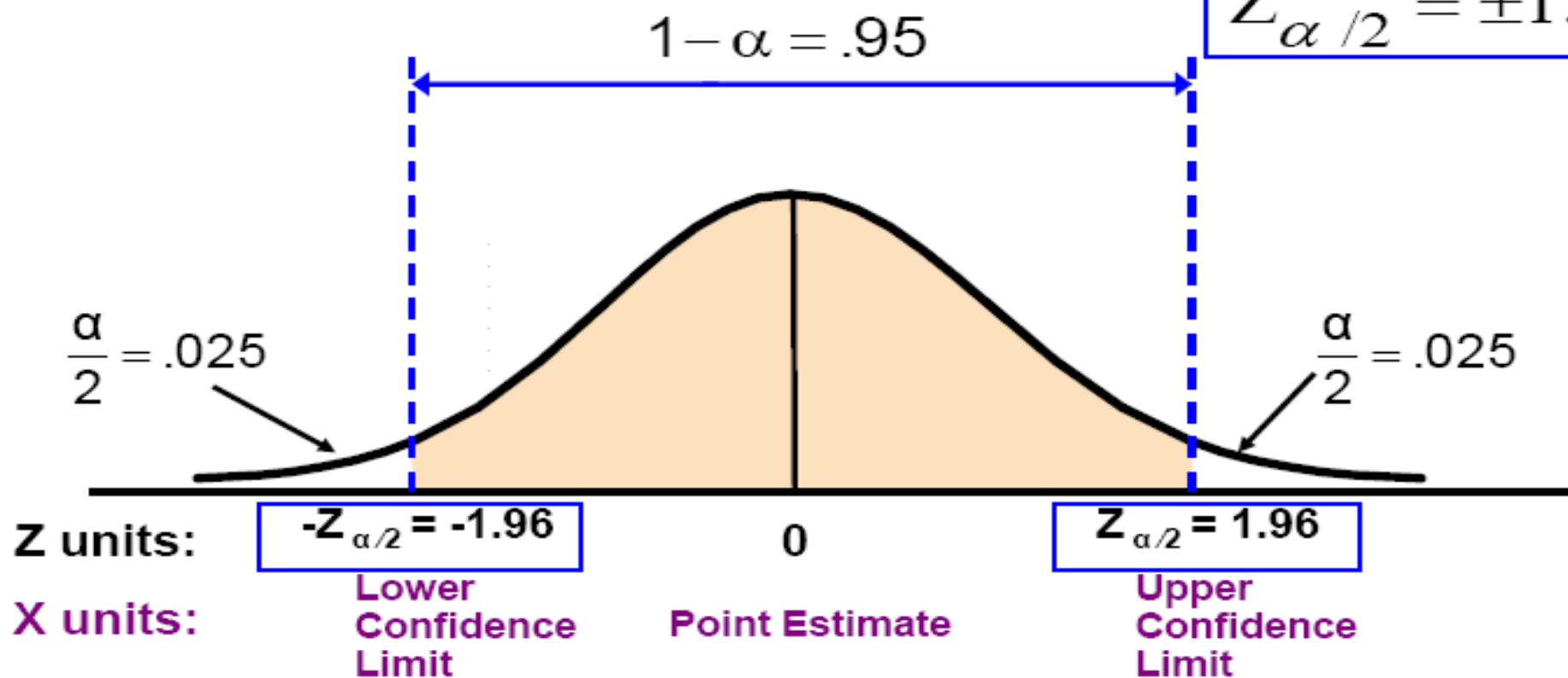
$$-\infty < \mu < \bar{X} + Z_{\alpha} \frac{\sigma}{\sqrt{n}}$$

## Finding the Critical Value, Z

- Consider a 95% confidence interval:

$$\Phi(z_{\alpha/2}) = 0.975$$

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



$$P(Z > Z_{\frac{\alpha}{2}}) = 1 - \Phi(Z_{\frac{\alpha}{2}}) = \frac{\alpha}{2} \quad \text{Then, } \Phi(Z_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2} \quad \text{Similarly, } \Phi(Z_{\alpha}) = 1 - \alpha$$

## Common Levels of Confidence

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

<i>Confidence Level</i>	<i>Confidence Coefficient, <math>1 - \alpha</math></i>	<i>Z value</i>
80%	.80	1.28
90%	.90	1.645
95%	.95	1.96
98%	.98	2.33
99%	.99	2.57
99.8%	.998	3.08
99.9%	.999	3.27

that is

- 90% (with  $\alpha = 0.1$ ) and critical value  $z_{\alpha/2} = 1.645$ .
- 95% (with  $\alpha = 0.05$ ) and critical value  $z_{\alpha/2} = 1.96$ .
- 99% (with  $\alpha = 0.01$ ) and critical value  $z_{\alpha/2} = 2.575$ .



[illegible]



➤ **Example (1):** Suppose that when a signal having value  $\mu$  is transmitted from location A the value received at location B is normally distributed with mean  $\mu$  and variance 4. That is, if  $\mu$  is sent, then the value is  $\mu + N$  where  $N$ , representing noise, is normal with mean 0 and variance 4. To reduce error, suppose the same value is sent 9 times. If the successive values received are 5, 8.5, 12, 15, 7, 9, 7.5, 6.5, and 10.5.

(a) Construct a 95% confidence interval for  $\mu$

(b) Determine the lower and upper 95 percent confidence interval estimates of  $\mu$ .

### Solution

$$\sigma^2 = 4, \text{ so } \sigma = 2$$

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n} = \frac{81}{9} = 9$$

**(a)** A 95% confidence interval for  $\mu$ ,  $\alpha = 5\% = 0.05$

$$(\bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \quad \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}})$$

$$Z_{\frac{\alpha}{2}} = Z_{0.025} = 1.96$$

Then ,95 percent confidence interval for  $\mu$  is ,

$$(9 - 1.96 \frac{2}{\sqrt{9}}, \quad 9 + 1.96 \frac{2}{\sqrt{9}}) = (7.69, 10.31)$$

**(b)**  $\alpha = 0.05$  , and  $Z_{\alpha} = Z_{0.05} = 1.645$

**(i)** the lower 95 percent confidence interval estimates of  $\mu$  is

$$(\bar{X} - Z_{\alpha} \frac{\sigma}{\sqrt{n}}, \quad \infty) = (7.903, \infty)$$

**(ii)** the upper 95 percent confidence interval estimates of  $\mu$  is

$$(-\infty, \quad \bar{X} + Z_{\alpha} \frac{\sigma}{\sqrt{n}}) = (-\infty, 10.097)$$

➤ **Example (2):** The average zinc concentration recovered from a sample of zinc measurements in 36 different locations is found to be 2.6 grams per milliliter. Find the 95% and 99% confidence interval for the mean zinc concentration in the river. Assume that the population standard deviation is 0.3.

**Solution:**  $\bar{X} = 2.6$  ,  $\sigma = 0.3$ ,  $n = 36$

(i) 95% confidence interval of  $\mu$ ,  $\alpha = 0.05$

$$\left( \bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \quad \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right), \quad Z_{\frac{\alpha}{2}} = Z_{0.025} = 1.96$$

Then, the 95 percent confidence interval for  $\mu$  is,

$$\left( 2.6 - 1.96 \frac{0.3}{\sqrt{36}}, \quad 2.6 + 1.96 \frac{0.3}{\sqrt{36}} \right) = (2.50, 2.70)$$

(ii) 99% confidence interval of  $\mu$ ,  $\alpha = 0.01$  ,  $Z_{\frac{\alpha}{2}} = Z_{0.005} = 2.575$

Then, the 99 percent confidence interval for  $\mu$  is,

$$\left( 2.6 - 2.575 \frac{0.3}{\sqrt{36}}, \quad 2.6 + 2.575 \frac{0.3}{\sqrt{36}} \right) = (2.47, 2.73)$$

➤ **Example (3):** An electric scale gives a reading equal to the true weight plus a random error that is normally distributed with mean 0 and standard deviation  $\sigma = 0.1$  mg. Suppose that the results of five successive weightings of the same object are as follows : 3.142, 3.163, 3.155, 3.150, and 3.141.

- (a) Determine a 95% confidence interval estimate of the true weight.
- (b) Determine a 99% confidence interval estimate of the true weight.
- (c) Determine the lower and upper 95% confidence interval estimates of  $\mu$ .

**Solution:**  $\sigma = 0.1$  ,  $n = 5$ ,  $\bar{X} = \sum_{i=1}^n x_i \div n = 3.1502$

(a) A 95% confidence interval for the true weight  $\mu$ ,  $\alpha = 0.05$ ,

$$\left( \bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \quad \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)$$

$$Z_{\frac{\alpha}{2}} = Z_{0.025} = 1.96$$

Then , the 95 percent C.I. for  $\mu$  is: **(3.0625, 3.2378)**

**(b)** A 99% confidence interval for  $\mu$ ,  $\alpha = 0.01$

$$(\bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \quad \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}})$$

$$Z_{\frac{\alpha}{2}} = Z_{0.005} = 2.58$$

Then , the 99 percent confidence interval for  $\mu$  is ,

$$(3.1502 - 2.58 \frac{0.1}{\sqrt{5}}, \quad 3.1502 + 2.58 \frac{0.1}{\sqrt{5}}) = (3.0348, 3.2656)$$

**(c)**  $\alpha=0.05$  , and  $Z_{\alpha} = Z_{0.05} = 1.645$

**(i)** the lower 95 percent confidence interval estimates of  $\mu$  is

$$(\bar{X} - Z_{\alpha} \frac{\sigma}{\sqrt{n}}, \quad \infty) = (3.0766, \infty)$$

**(ii)** the upper 95 percent confidence interval estimates of  $\mu$  is

$$(-\infty, \quad \bar{X} + Z_{\alpha} \frac{\sigma}{\sqrt{n}}) = (-\infty, 3.2238)$$

➤ **Example (4):** The standard deviation of test scores on a certain achievement test is 11.3. If a random sample of 81 students had a sample mean score of 74.6, find a 90 percent confidence interval estimate for the average score of all students.

### Solution

$$\sigma = 11.3, n = 81, \text{ and } \bar{X} = 74.6$$

A 90% C.I. for the average score of all students  $\mu$ ,  $\alpha = 0.1$

$$\left( \bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \quad \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)$$

$$Z_{\frac{\alpha}{2}} = Z_{0.05} = 1.645$$

$$\left( 74.6 - 1.645 \frac{11.3}{\sqrt{81}}, \quad 74.6 + 1.645 \frac{11.3}{\sqrt{81}} \right) = (72.5346, 76.6654)$$

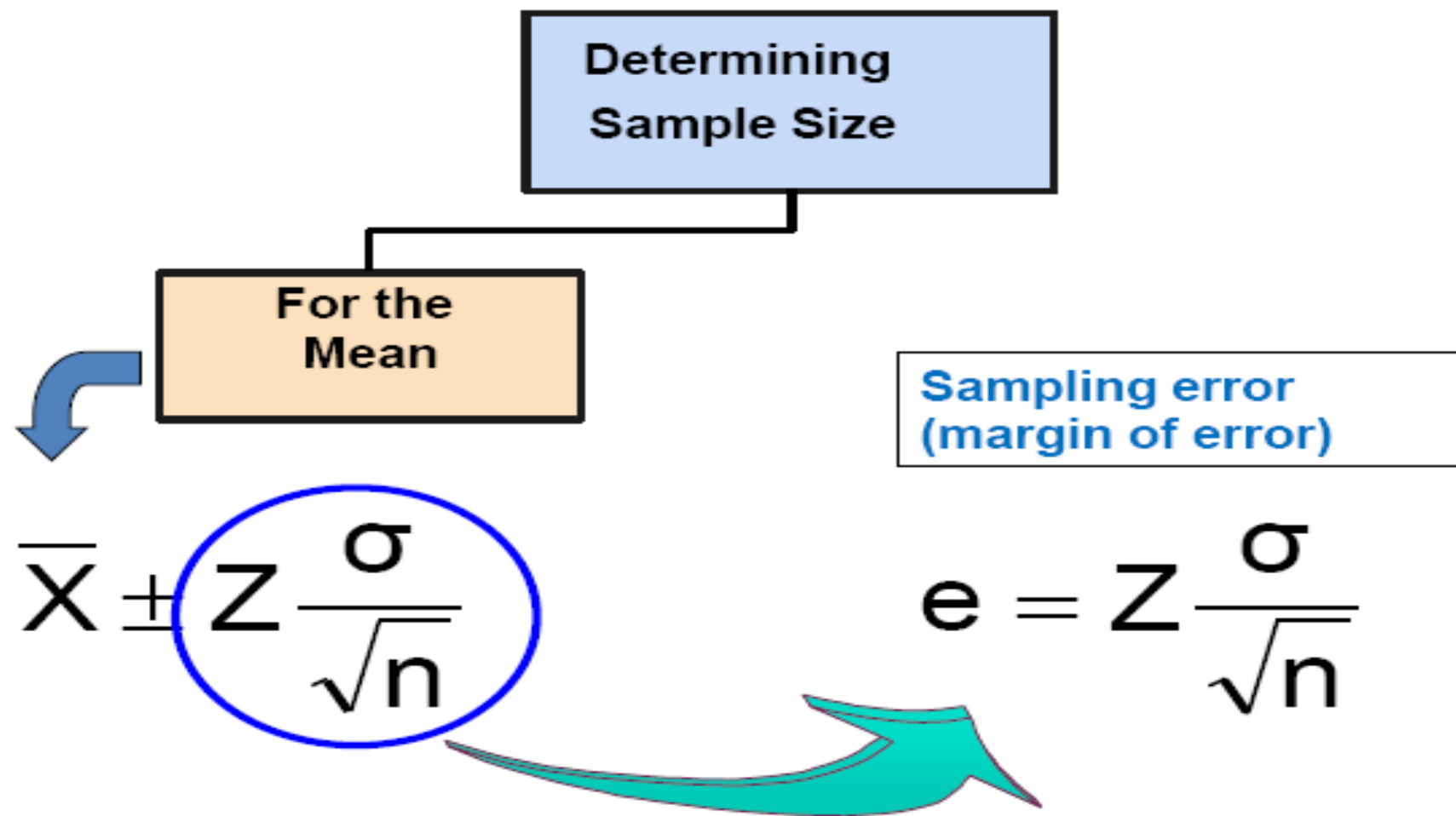
# Sampling Error

---

- The required sample size can be found to reach a desired **margin of error (e)** with a specified level of confidence ( $1 - \alpha$ )
- The margin of error is also called **sampling error**
  - the amount of imprecision in the estimate of the population parameter
  - the amount added and subtracted to the point estimate to form the confidence interval

# Determining Sample Size

---





# Determining Sample Size

*(continued)*

Determining  
Sample Size

For the  
Mean

$$e = Z \frac{\sigma}{\sqrt{n}}$$

Now solve for  
n to get

$$n = \frac{Z^2 \sigma^2}{e^2}$$

# Determining Sample Size

*(continued)*

---

- To determine the required sample size for the mean, you must know:
  - The desired level of confidence ( $1 - \alpha$ ), which determines the critical Z value
  - The acceptable sampling error (margin of error),  $e$
  - The standard deviation,  $\sigma$

## Required Sample Size Example

### ➤ Example (5):

If  $\sigma = 45$ , what sample size is needed to estimate the mean within  $\pm 5$  with 90% confidence?

$$n = \frac{Z^2 \sigma^2}{e^2} = \frac{(1.645)^2 (45)^2}{5^2} = 219.19$$

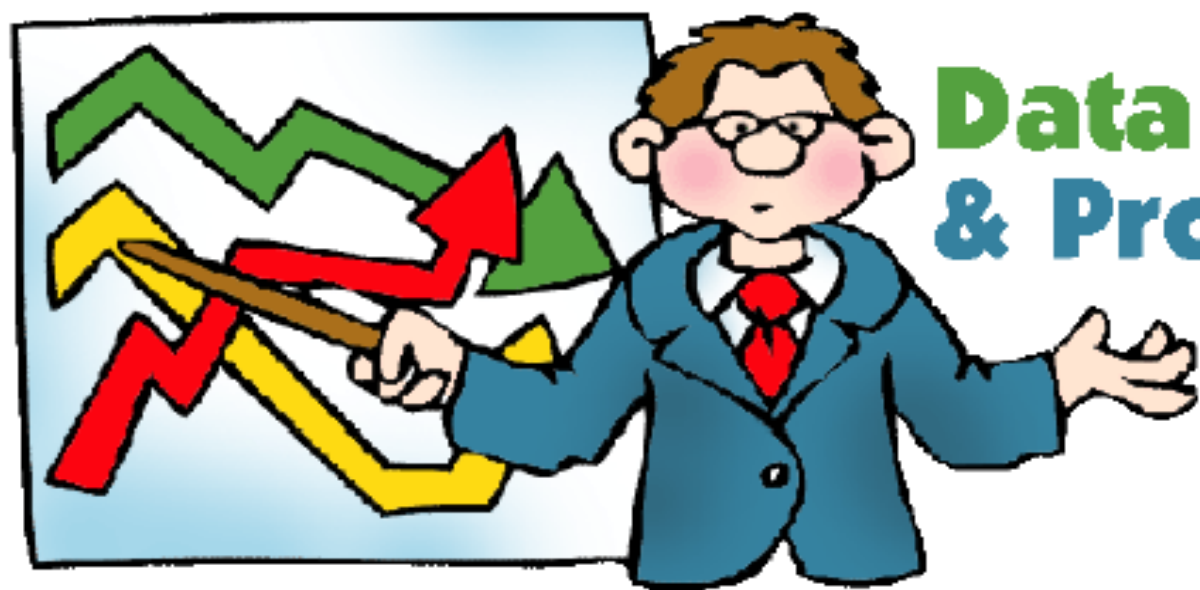
So the required sample size is  **$n = 220$**

(Always round up)

# Hypothesis Testing

---

- ❑ The purpose of hypothesis testing is to help the researcher or administrator in reaching a decision concerning a population by examining a sample from that population



**Data Analysis  
& Probability**



# What is a Hypothesis?

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

- For a population mean

**Example:** The mean monthly cell phone bill of this city is  $\mu = \$42$



## The Null Hypothesis, $H_0$

- States the assumption (numerical) to be tested

**Example:** The average number of TV sets in U.S. homes is equal to three ( $H_0: \mu = 3$ )

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

$$H_0 : \mu = 3$$

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



- 
- Begin with the assumption that the null hypothesis is true

*Similar to the notion of innocent until proven guilty!*

- Always contains “=”, “≤” or “≥” 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 above status
- Never contains the “=”, “≤” or “≥” sign
- May or may not be proven
- Is generally the hypothesis that the researcher is trying to prove

# Hypothesis Testing Process

**Claim:** the population mean age is 50.  
(Null Hypothesis:  
 $H_0: \mu = 50$  )



Population

Now select a random sample

Is  $\bar{X}=20$  likely if  $\mu = 50$ ?

If not likely,

**REJECT**  
Null Hypothesis

Suppose the sample mean age is 20:  $\bar{X} = 20$



Sample

# 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 .01, .05, or .10
- Is selected by the researcher at the beginning
- Provides the critical value(s) of the test

## Hypothesis-Testing Common Phrases

Is greater than  
Is above  
Is higher than  
Is longer than  
Is bigger than  
Is increased

>

Is less than  
Is below  
Is lower than  
Is shorter than  
Is smaller than  
Is decreased or  
reduced from

<

Is equal to  
Is the same as  
Has not changed from  
Is the same as

=

Is not equal to  
Is different from  
Has changed from  
Is not the same as

≠



# Level of Significance and the Rejection Region

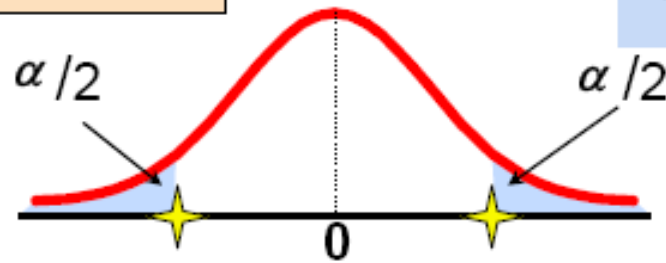
Level of significance =  $\alpha$

★ Represents critical value

$$H_0: \mu = 3$$

$$H_1: \mu \neq 3$$

Two-tailed test

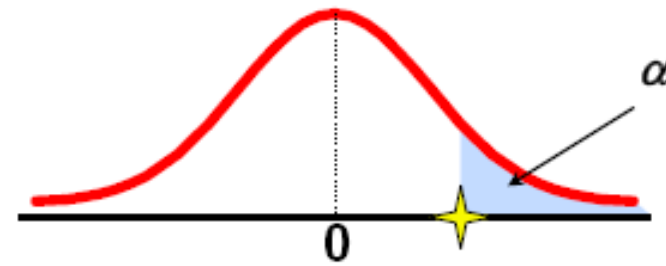


Rejection region is shaded

$$H_0: \mu \leq 3$$

$$H_1: \mu > 3$$

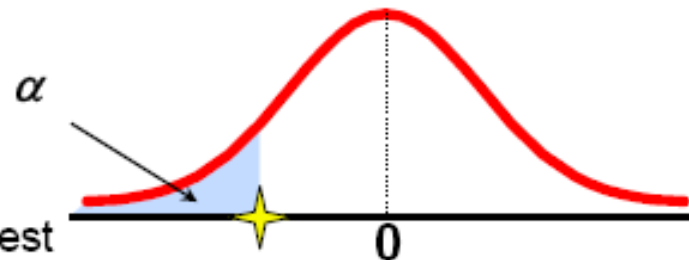
Upper-tailed test



$$H_0: \mu \geq 3$$

$$H_1: \mu < 3$$

Lower-tailed test



# Errors in Making Decisions

---

- **Type I Error**

- Rejecting the null hypothesis when it's true.
- Considered a serious type of error

The probability of Type I Error is  $\alpha$



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

- **Type II Error**

- Accepting the null hypothesis when it's false.

probability of Type II Error is  $\beta$

## Type I & II Error Relationship

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

# Outcomes and Probabilities

## Possible Hypothesis Test Outcomes

		Actual Situation	
		$H_0$ is true	$H_0$ is false
Decision	accept $H_0$	correct decision probability = $1 - \alpha$ = confidence level	incorrect decision (type II error) probability = $\beta$
	reject $H_0$	incorrect decision (type I error) probability = $\alpha$ = level of significance	correct decision probability = $1 - \beta$ = power of the test

## Z Test of Hypothesis for the Mean ( $\sigma$ Known)

- Convert sample statistic ( $\bar{X}$ ) to a Z **test statistic**

Hypothesis  
Tests for  $\mu$

$\sigma$  Known

$\sigma$  Unknown

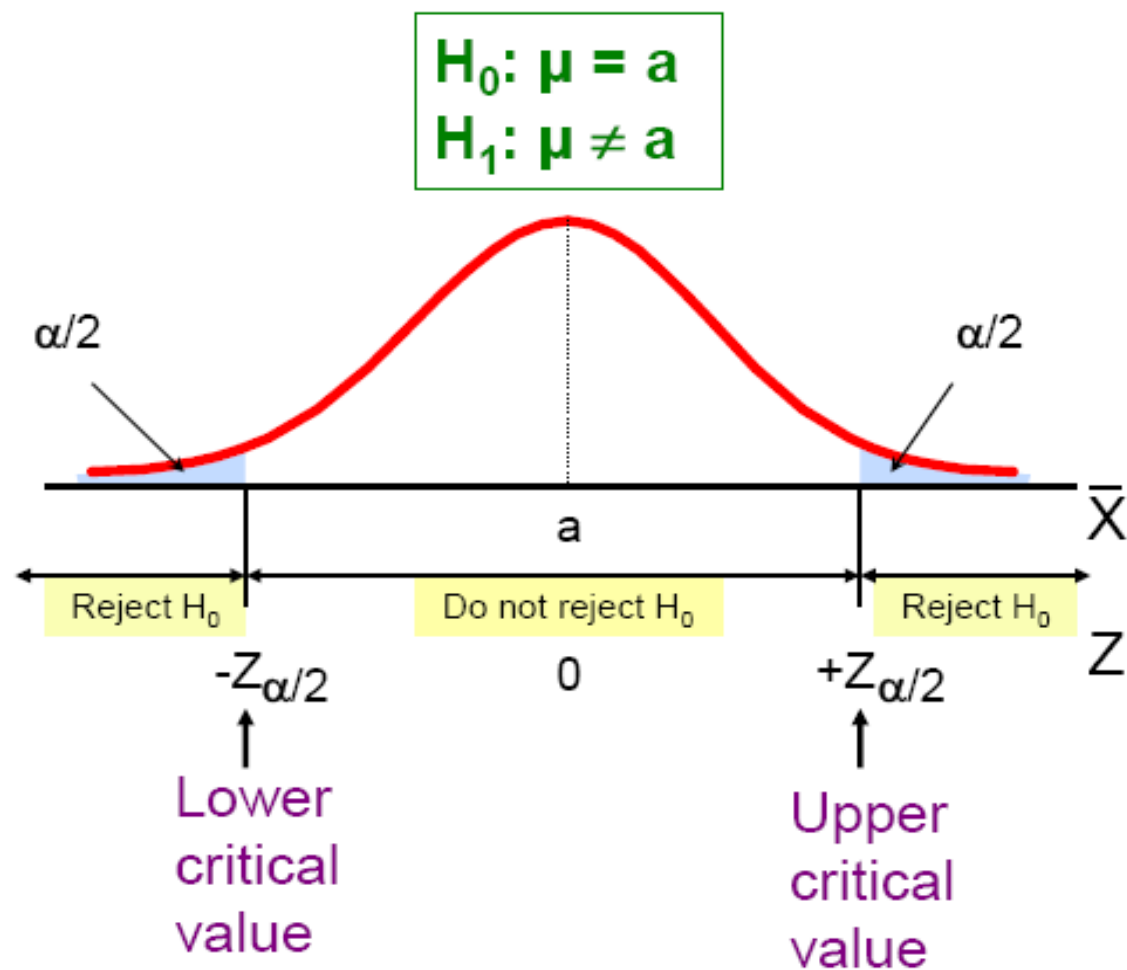
The test statistic is:

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

$$\text{Test value} = \frac{(\text{observed value}) - (\text{expected value})}{\text{standard error}}$$

# Two-Tailed Tests

- There are two cutoff values (critical values), defining the regions of rejection



# One-Tailed Tests

---

- In many cases, the alternative hypothesis focuses on a particular direction

$$H_0: \mu \geq 3$$

$$H_1: \mu < 3$$



This is a **lower**-tailed test since the alternative hypothesis is focused on the lower tail below the mean of 3

$$H_0: \mu \leq 3$$

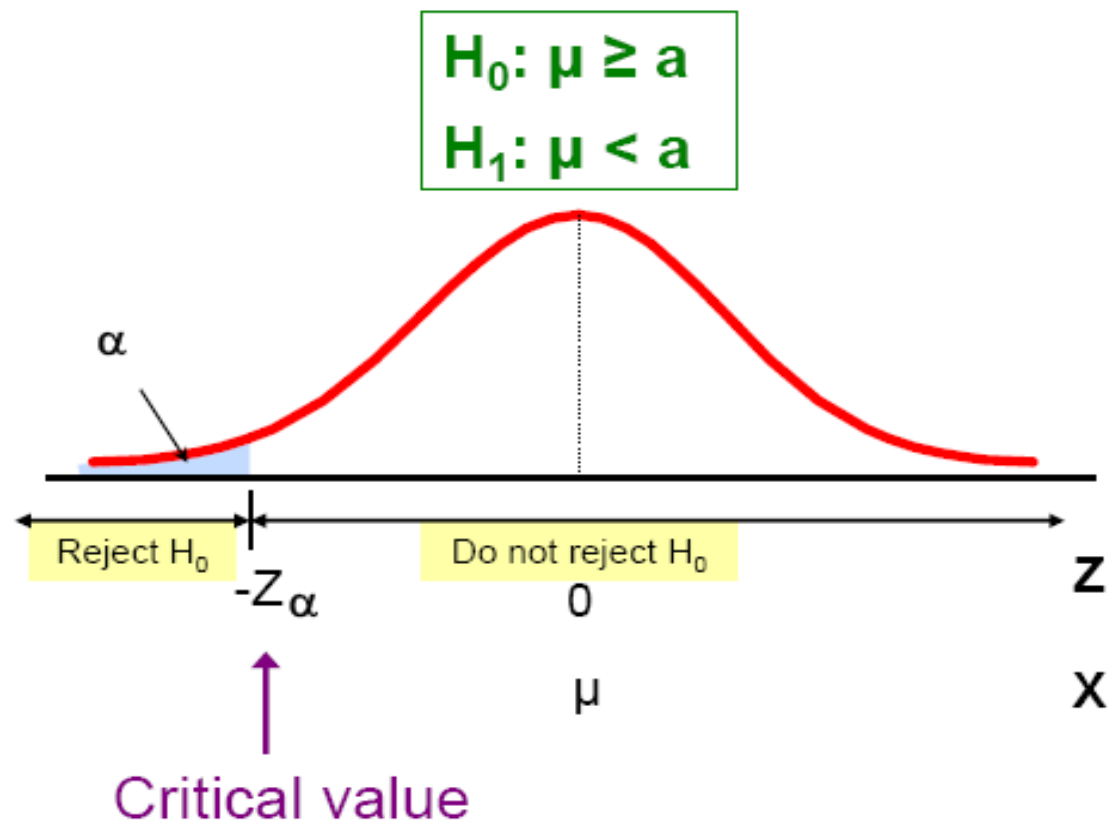
$$H_1: \mu > 3$$



This is an **upper**-tailed test since the alternative hypothesis is focused on the upper tail above the mean of 3

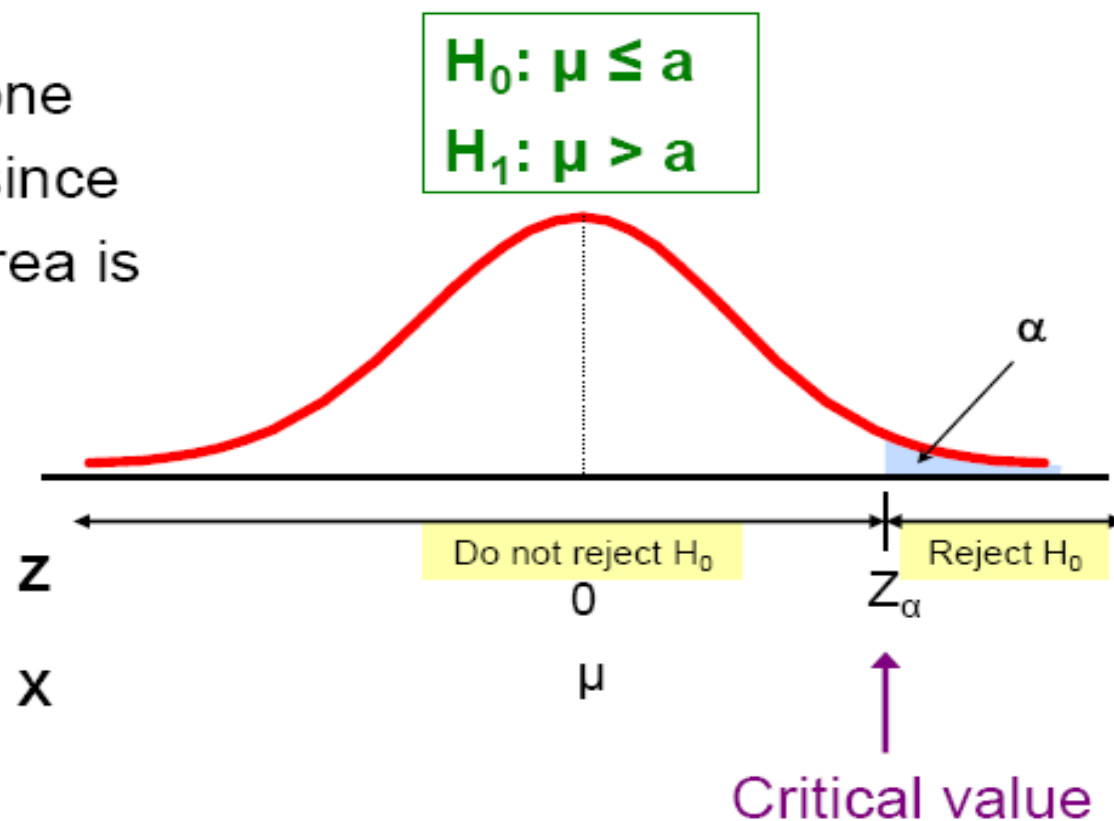
# Lower-Tailed Tests

- There is only one critical value, since the rejection area is in only one tail



# Upper-Tailed Tests

- There is only one critical value, since the rejection area is in only one tail





---

**Procedure Table**  
**Solving Hypothesis-Testing Problems**  
**(Traditional Method)**

- Step 1** State the hypotheses and identify the claim.
- Step 2** Find the critical value(s) from the appropriate table.
- Step 3** Compute the test statistic value.
- Step 4** Make the decision to reject or not reject the null hypothesis.
- Step 5** Summarize the results.

## ➤ Example(1): Two-tailed Z test ( $\sigma$ Known)

---

At a 5% significance level, test the claim that the true mean # of TV sets in U.S. homes is equal to 3. Consider a sample of size 100 with mean 2.84  
(Assume  $\sigma = 0.8$ )

- 1- State the appropriate null and alternative hypotheses
  - $H_0: \mu = 3$       $H_1: \mu \neq 3$      (This is a two-tailed test)
- 2- Specify the desired level of significance and set up the critical values
  - $\alpha = .05$  is chosen for this test
  - For  $\alpha = .05$ , the critical Z values are  $\pm 1.96$



## Hypothesis Testing Example

*(continued)*

3- Compute the test statistic

- The sample results are

$n = 100$ ,  $\bar{X} = 2.84$  ( $\sigma = 0.8$  is assumed known)

So the test statistic is:

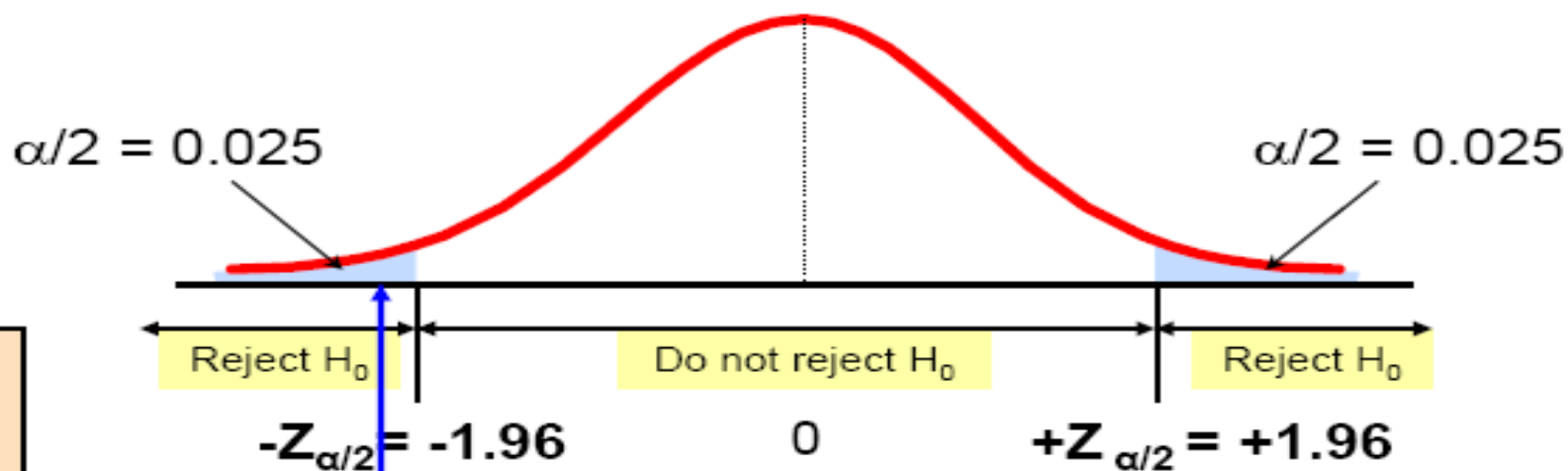
$$Z_c = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{2.84 - 3}{\frac{0.8}{\sqrt{100}}} = \frac{-.16}{.08} = -2.0$$



# Hypothesis Testing Example

(continued)

4- Is the test statistic in the rejection region?



Reject  $H_0$  if  
 $Z_c < -1.96$  or  
 $Z_c > 1.96$ ;  
otherwise do  
not reject  $H_0$

Here,  $Z_c = -2.0 < -1.96$ , so the  
test statistic is in the rejection  
region



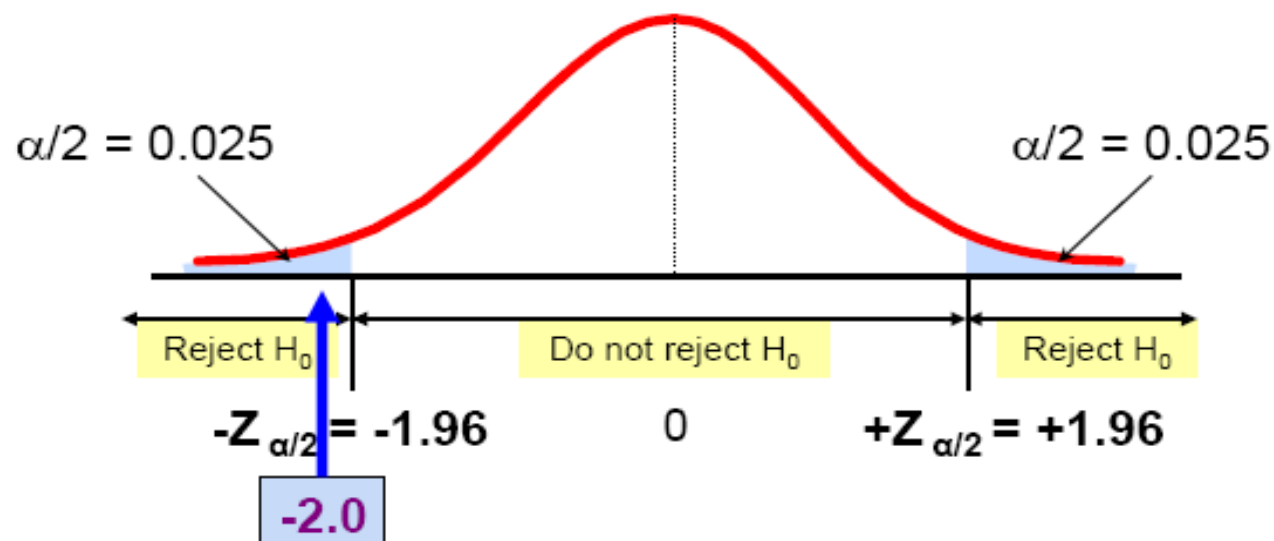
[illegible]



## Hypothesis Testing Example

*(continued)*

5- Reach a decision and interpret the result



Since  $Z_c = -2.0 < -1.96$ , we reject the null hypothesis and conclude that there is sufficient evidence that the mean number of TVs in US homes is not equal to 3.



## ➤ Example (2): Upper-Tail Z Test for Mean ( $\sigma$ Known)

A phone industry manager thinks that customer monthly cell phone bill have increased, and now average over \$52 per month. The company wishes to test this claim for a sample of 64 with mean 53.1. Use 0.10 significance level. (Assume  $\sigma = 10$  is known)

1- Form of hypothesis test:

$H_0: \mu \leq 52$	the average is not over \$52 per month
$H_1: \mu > 52$	the average is greater than \$52 per month (i.e., sufficient evidence exists to support the manager's claim)

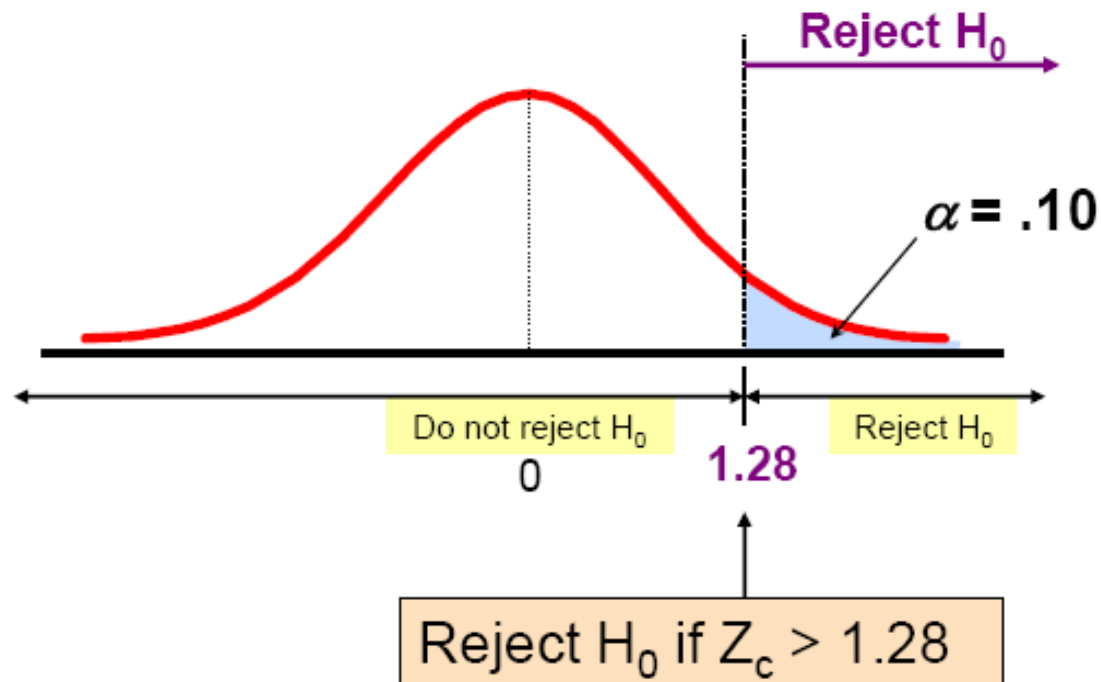


## Finding Rejection Region

*(continued)*

2- For the chosen  $\alpha = .10$  to this test, the critical value is 1.28 (why?)

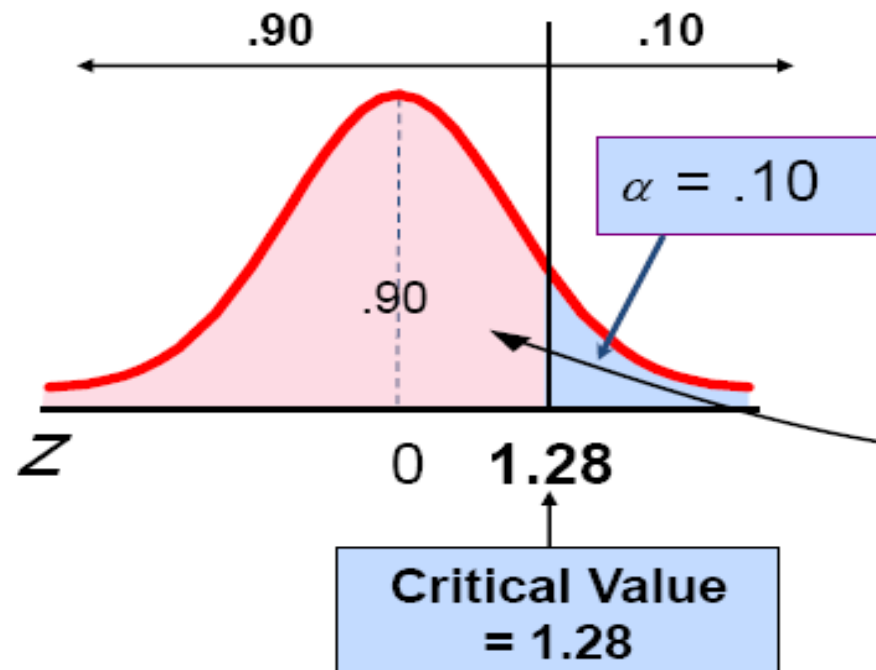
Find the rejection region:





# Review: One-Tail Critical Value

What is  $Z$  given  $\alpha = 0.10$ ?



Standard Normal  
Distribution Table (Portion)

$Z$	.07	<b>.08</b>	.09
1.1	.8790	.8810	.8830
<b>1.2</b>	.8980	<b>.8997</b>	.9015
1.3	.9147	.9162	.9177

[illegible]

## Test Statistic

*(continued)*

3- Compute the test statistic with  $n = 64$ ,  $\bar{X} = 53.1$ ,  
and  $\sigma = 10$

– Then the test statistic is:

$$Z_c = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{53.1 - 52}{\frac{10}{\sqrt{64}}} = 0.88$$

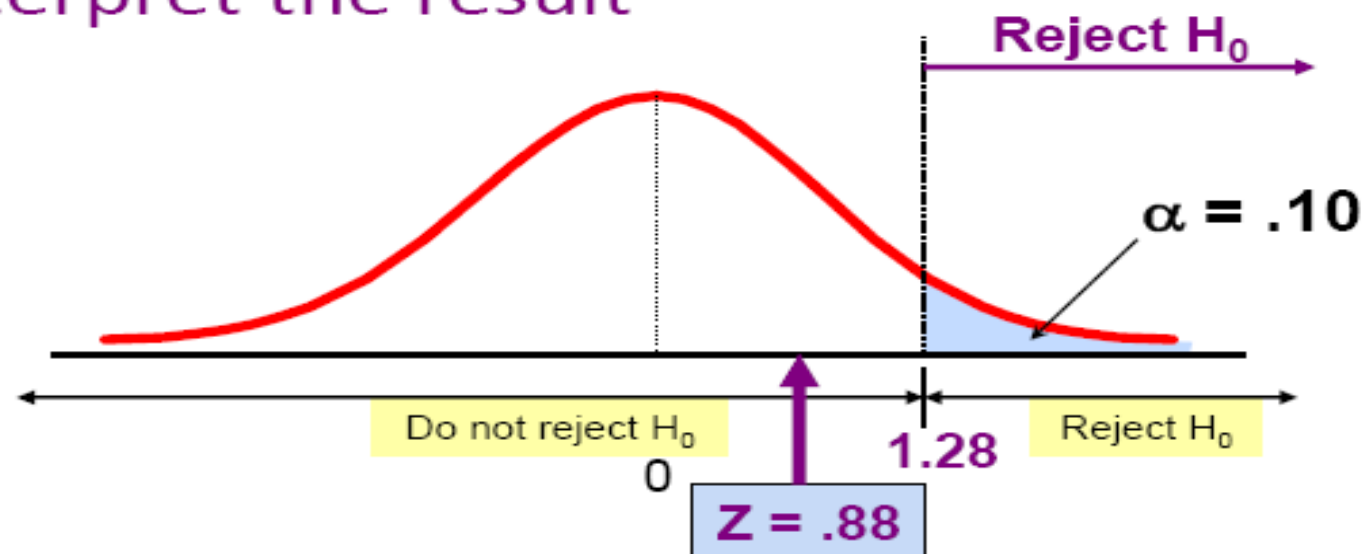


# Decision

*(continued)*

4- Reach a decision

5- interpret the result



**Do not reject  $H_0$  since  $Z = 0.88 \leq 1.28$**

i.e.: there is not sufficient evidence that the mean bill is over \$52



# Questions!

---

"Success doesn't come to you, you go to it."



**THANK YOU**

**PRACTICE!**  
**PRACTICE!**  
**PRACTICE!**

There is **NO**  
elevator to  
**SUCCESS.**

You have  
to take the  
**STAIRS.**