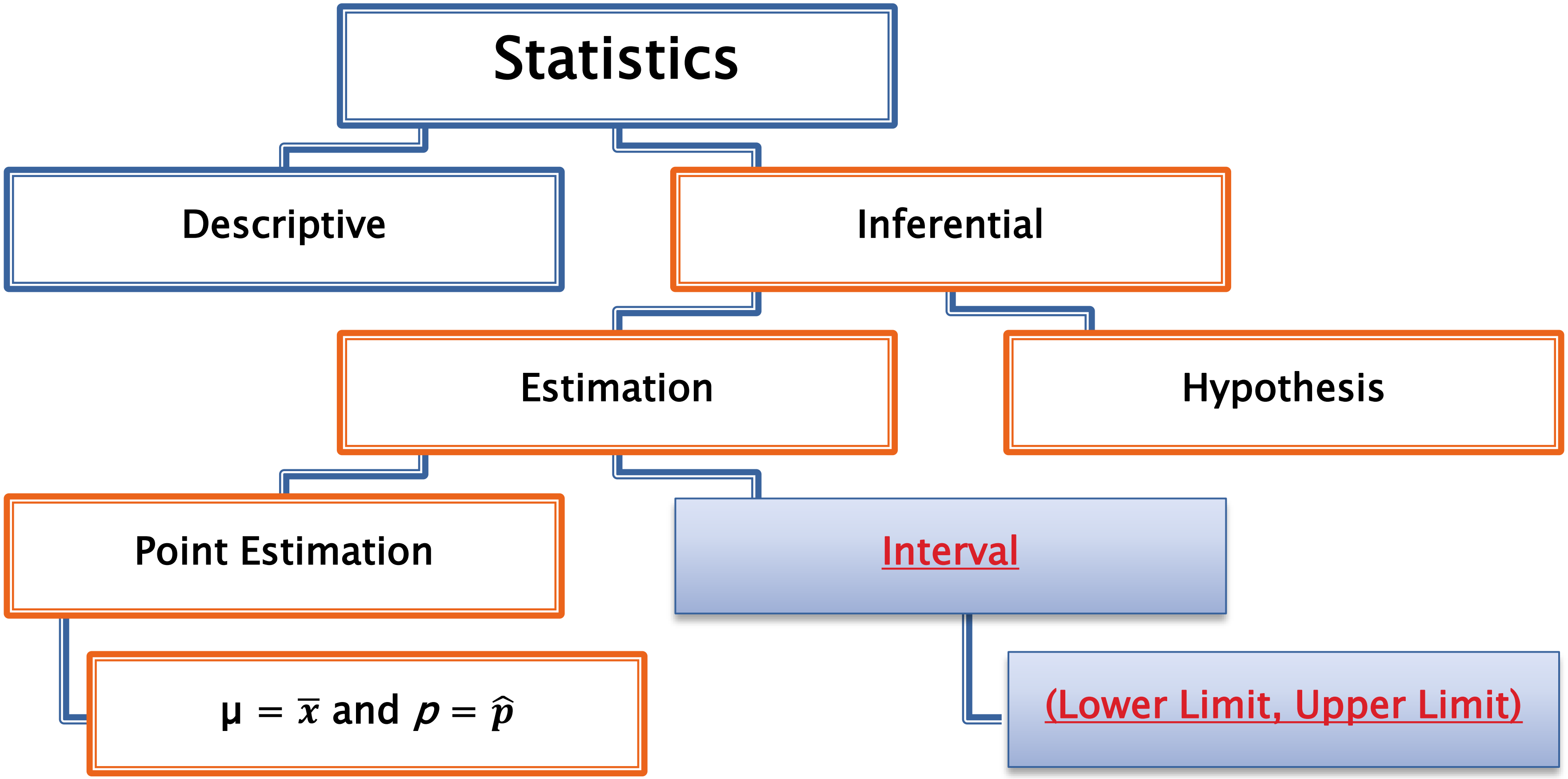
STATISTICAL ANALYSIS - LECTURE 10

Dr. Mahmoud Mounir

mahmoud.mounir@cis.asu.edu.eg

STATISTICAL INFERENCE

Recap


Statistics
Descriptive
Inferential
Estimation
Hypothesis
Point Estimation
Interval
$\mu = \overline{x}$ and $p = \widehat{p}$
(Lower Limit, Upper Limit)

Statistical Inference

- The statistical inference is the process of making judgment about a population based on the properties of a random sample from the population.

Estimators

Estimator (Sample Statistic)		Population Parameter
$\bar{x}$	ESTIMATES	μ
S^2	ESTIMATES	σ^2
s	ESTIMATES	σ
$\hat{p}$	ESTIMATES	p
$\bar{x}_1 - \bar{x}_2$	ESTIMATES	$\mu_1 - \mu_2$
$\hat{p}_1 - \hat{p}_2$	ESTIMATES	$p_1 - p_2$

Confidence Interval (CI)

Population Parameter = Estimator $\pm$ Margin of Error (d)

Where:

Margin of Error (d) = Critical Value of Z * Standard Error

Population Parameter $\in$

[Estimator − Margin of Error (d) , Estimator + Margin of Error (d)]

Confidence Interval (CI) for Single Mean

$$\mu = \bar{x} \pm z\left(\frac{\sigma}{\sqrt{n}}\right)$$

$$\mu \in \left[\bar{x} - z\left(\frac{\sigma}{\sqrt{n}}\right), \bar{x} + z\left(\frac{\sigma}{\sqrt{n}}\right)\right]$$

Confidence Interval (CI) for Single Proportion

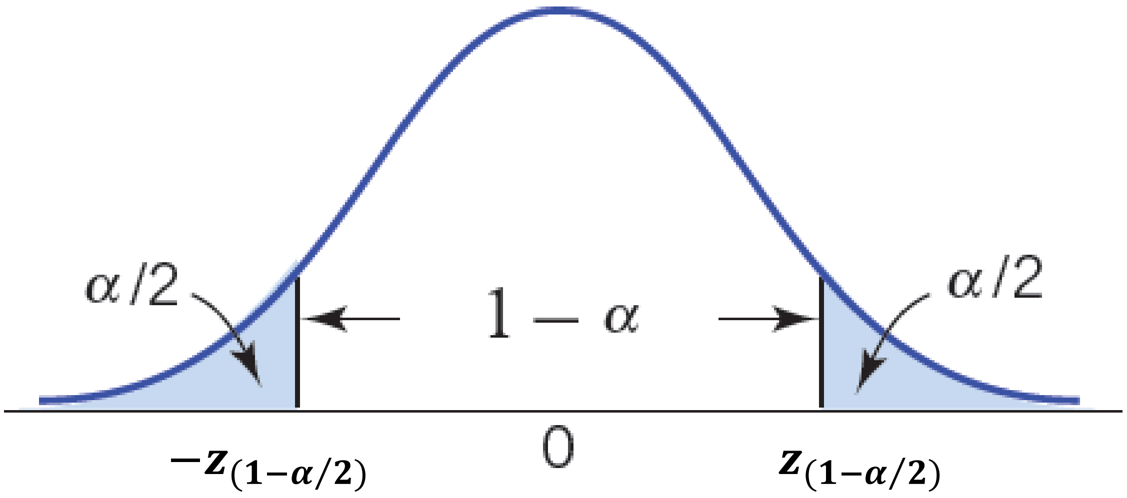
$$p = \hat{p} \pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$p \in \left[\,\hat{p} - z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\,,\ \hat{p} + z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\,\right]$$

Standard Error for Different Sample Statistics

Estimator (Sample Statistic)		Standard Error
$\bar{x}$		$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$
$\hat{p}$		$\sigma_{\hat{p}} = \sqrt{\frac{p*(1-p)}{n}}$
$\bar{x}_1 - \bar{x}_2$		$\sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
$\hat{p}_1 - \hat{p}_2$		$\sigma_{(\hat{p}_1-\hat{p}_2)} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$

Confidence Interval (CI)

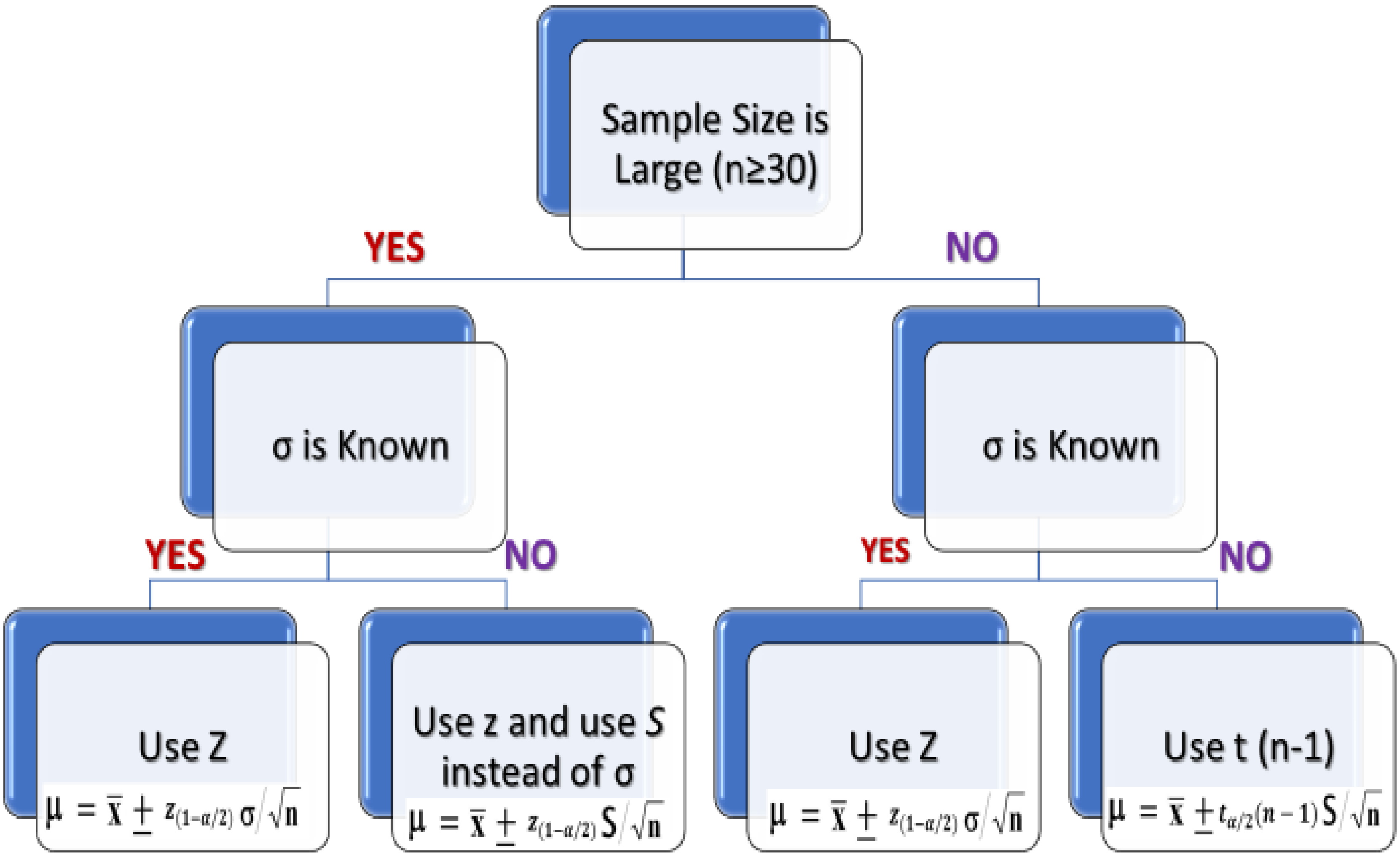


(1-α) **Confidence Level**
α **Confidence Coefficient**

Critical Value for Z:

$1-\alpha$	$100(1-\alpha)\%$	α	$1-\alpha/2$	$z_{1-(\alpha/2)}$
0.90	90%	0.10	0.95	1.65
0.95	95%	0.05	0.975	1.96
0.98	98%	0.02	0.990	2.33
0.99	99%	0.01	0.995	2.58

Confidence Interval for Population Mean



Confidence Interval (CI) for the Difference Between Two Population Means

$$(\mu_1 - \mu_2) = (\bar{x}_1 - \bar{x}_2) \pm z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

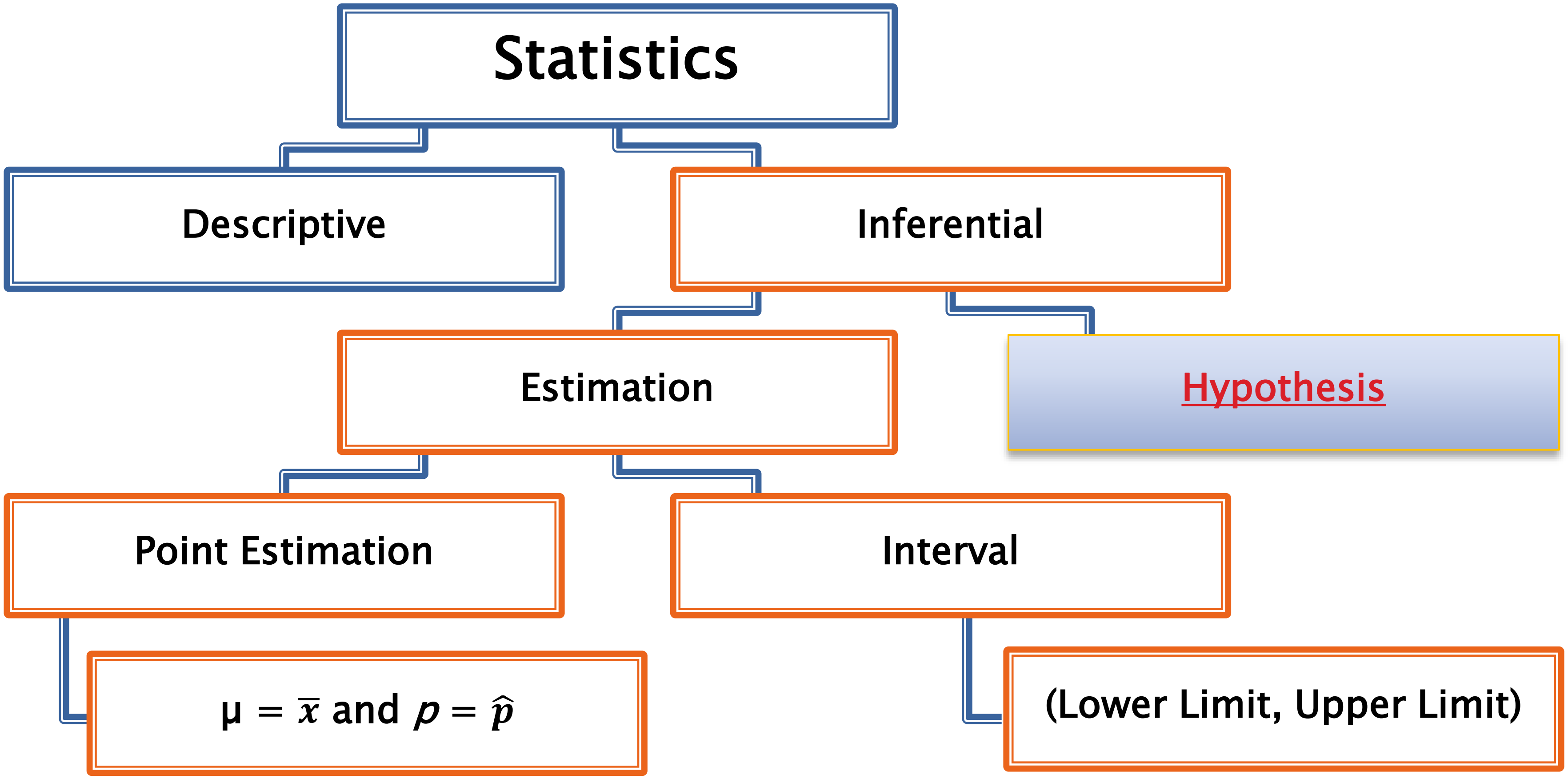
$$(\mu_1 - \mu_2) \in \left[(\bar{x}_1 - \bar{x}_2) - z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},\ (\bar{x}_1 - \bar{x}_2) + z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right]$$

Confidence Interval (CI) for Difference Between Two Population Proportions

$$(p_1 - p_2) = (\hat{p}_1 - \hat{p}_2) \pm z\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$$(p_1 - p_2) \in$$
$$\left[(\hat{p}_1 - \hat{p}_2) - z\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}, (\hat{p}_1 - \hat{p}_2) + z\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\right]$$

Hypothesis Tests



Hypothesis

DEFINITION 4.4.1 (Statistical Hypothesis)

Statistical hypothesis is an argument about a specific statistical question, and this argument can be true or wrong.

- It is a claim or statement about a population parameter.
- For example, the average age of college students equals 23 years.

Null Hypothesis (H_o)

DEFINITION 4.4.2 (Null Hypothesis)
The null hypothesis is a statement under investigation or testing.

- Usually, the null hypothesis represents a statement of "no effect," "no difference," or, put another way, "things haven't changed."
- ***For example, H_o: $\mu = 23$***

Alternate Hypothesis (H_a) or (H_1)

DEFINITION 4.4.3 (Alternate Hypothesis)

The **alternate hypothesis** is a statement we will adopt in the situation in which the evidence (data) is so strong that you reject the null hypothesis.

- This test is a statistical test is designed to assess the strength of the evidence (data) against the null hypothesis.
- ***For example,*** $H_1: \mu \neq 23$

Test Statistics

DEFINITION 4.4.4 (Test Statistic)

A test statistic is a quantity calculated from the given sample and can be used to make a decision in a test of hypotheses.

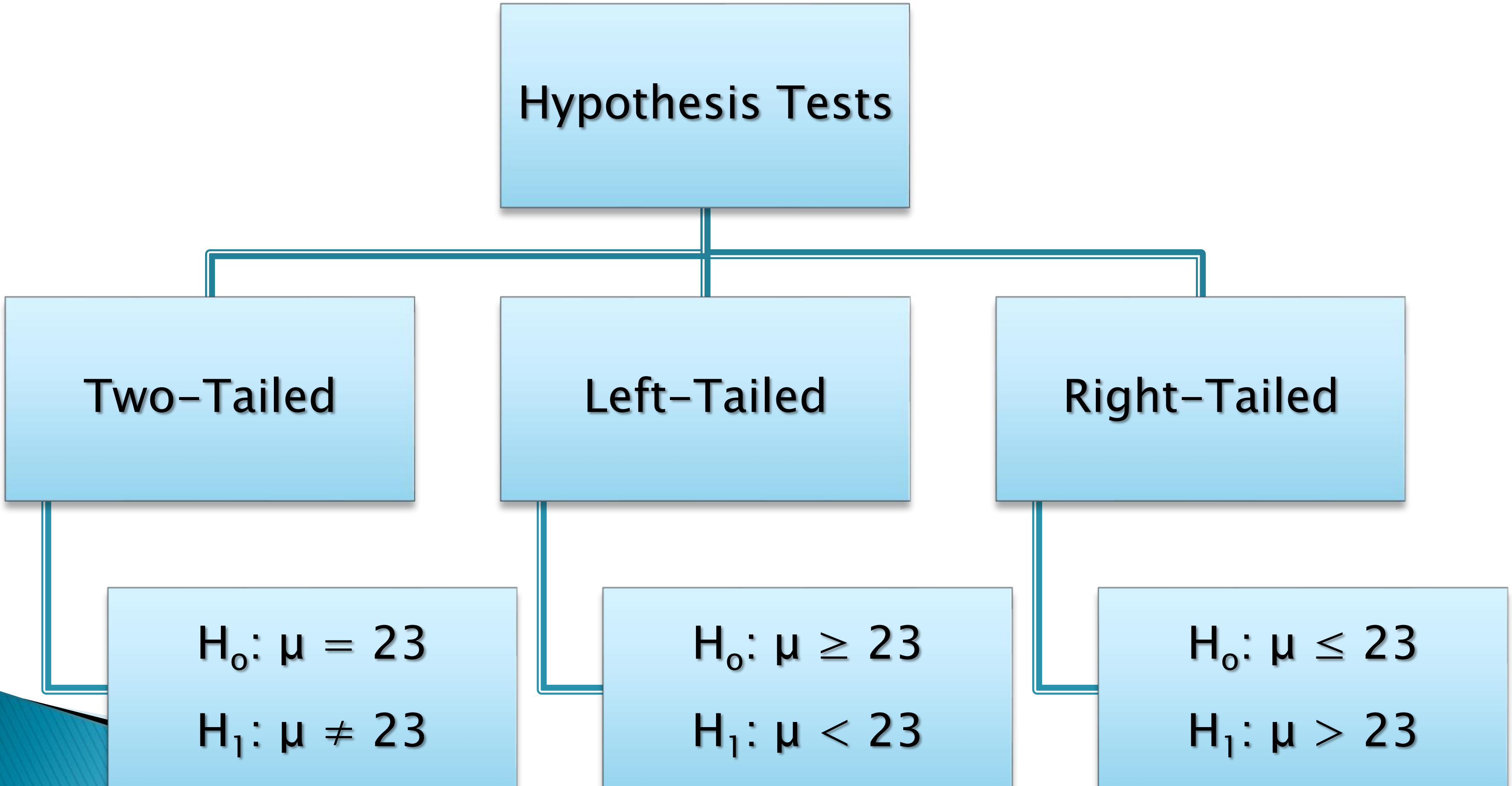
Different test statistics for population mean hypotheses testing

Case	Population	Sample size	Standard deviation	Statistic and distribution
1	normal	any sample size	known	$Z = \dfrac{\bar{x} - \mu}{\sigma \,/\, \sqrt{n}} \sim N(0,1)$
2	any population	large $(n \geq 30)$	known	$Z = \dfrac{\bar{x} - \mu}{\sigma \,/\, \sqrt{n}} \sim N(0,1)$
3	any population	large $(n \geq 30)$	unknown, use the sample standard deviation instead of	$Z = \dfrac{\bar{x} - \mu}{s \,/\, \sqrt{n}} \sim N(0,1)$

The Critical Region

DEFINITION 4.4.5 (The Critical Regions)

The critical region is a region that produced by the value(s) that corresponds to the rejection of the null hypothesis at some chosen level of significance.

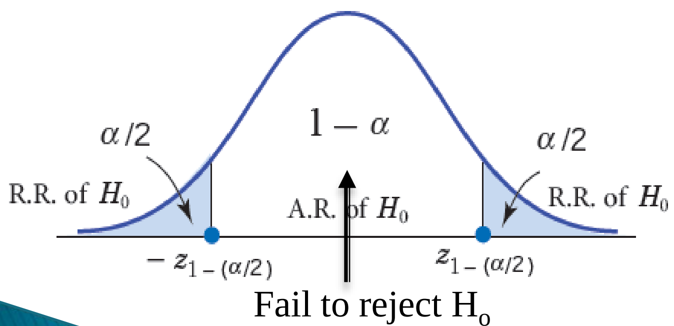


The Critical Region

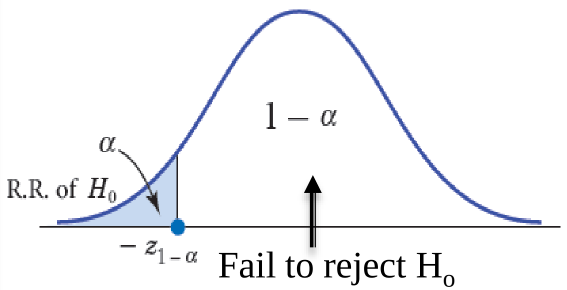
Hypothesis Tests

- **Two–Tailed**
  - $H_o: \mu = 23$
  - $H_1: \mu \neq 23$
- **Left–Tailed**
  - $H_o: \mu \geq 23$
  - $H_1: \mu < 23$
- **Right–Tailed**
  - $H_o: \mu \leq 23$
  - $H_1: \mu > 23$

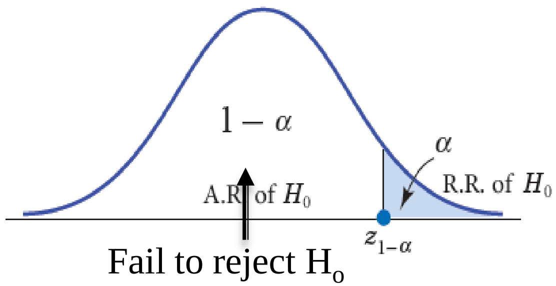
Two – tailed critical region



Left – tailed critical region



Right – tailed critical region



α: Significance Level

Steps for Hypothesis Test

- ***Step 1:*** State the null and alternative hypothesis (H_o and H_1).
- ***Step 2:*** Specify the suitable test statistic.
- ***Step 3:*** Compute the critical value for the statistic.
- ***Step 4:*** Make a decision.

  Reject H_o or Fail to reject H_o

- ***Step 5:*** Make a conclusion

  Accept H_1 or we don't have enough evidence to support H_1

Example (1)

In recent years, the mean age of all college students in a city has been 23. A random sample of 42 students revealed a mean age of 23.8 suppose their ages are normally distributed with a population standard deviation of σ=2.4 can we infer at α=0.05 that the population mean has changed?

Solution

$\mu = 23$	$n = 42$	$\sigma = 2.4$	$\bar{x} = 23.8$

***Step 1*: State the null and alternative hypothesis (H_o and H_1)**

Null hypothesis (H_o)	$\mu = 23$
alternative hypothesis (H_1)	$\mu \neq 23$

***Step 2*: Specify the suitable test statistic**

Since $n \geq 30$ and σ is known,
we will use Z-statistic

$\mu = 23$	$\sigma = 2.4$	$n = 42$	$\bar{x} = 23.8$
$\mu_{\bar{x}} = \mu = 23$	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2.4}{\sqrt{42}} = 0.37$	$z_{\bar{x}} = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{23.8 - 23}{0.37} = 2.16$	

Solution

Step 3**: *Compute the critical value for the statistic***

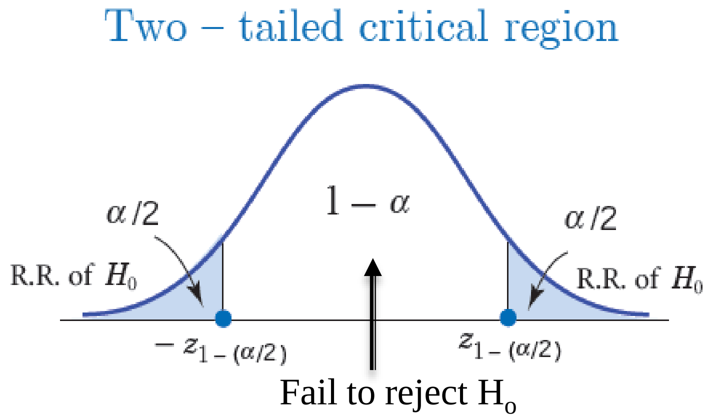
$(1-\alpha) = 0.95$	$\alpha = 0.05$	$\frac{\alpha}{2} = .025$
$1 - \frac{\alpha}{2} = 0.975$	$\pm z_{1-\alpha/2} = \pm 1.96$	

$1-\alpha$	$100(1-\alpha)\%$	α	$1-\alpha/2$	$z_{1-(\alpha/2)}$
0.90	90%	0.10	0.95	1.65
0.95	95%	0.05	0.975	1.96
0.98	98%	0.02	0.990	2.33
0.99	99%	0.01	0.995	2.58

Solution

***Step 4**: Make a decision*

$$z_{\bar{x}} = 2.16 \text{ which is greater than } 1.96$$
$$=> z_{\bar{x}} = 2.16 > 1.96, \text{ so we reject } H_0$$



***Step 5**: Make a conclusion*

We accept H_1, there is enough evidence that mean age has changed at $\alpha = 0.05$

Example (2)

In the previous example, if the confidence level (α) = 0.02, what can we infer about the population mean ?

Solution

$\mu = 23$	$n = 42$	$\sigma = 2.4$	$\bar{x} = 23.8$

***Step 1*: State the null and alternative hypothesis (H_o and H_1)**

Null hypothesis (H_o)	$\mu = 23$
alternative hypothesis (H_1)	$\mu \neq 23$

***Step 2*: Specify the suitable test statistic**

Since $n \geq 30$ and σ is known, we will use Z-statistic

$\mu = 23$	$\sigma = 2.4$	$n = 42$	$\bar{x} = 23.8$
$\mu_{\bar{x}} = \mu = 23$	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2.4}{\sqrt{42}} = 0.37$	$z_{\bar{x}} = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{23.8 - 23}{0.37} = 2.16$	

Solution

***Step 3**: Compute the critical value for the statistic*

$(1-\alpha) = 0.98$	$\alpha = 0.02$	$\frac{\alpha}{2} = .01$
$1 - \frac{\alpha}{2} = 0.99$	$\pm z_{1-\alpha/2} = \pm 2.33$	

1- α	$100(1-\alpha)\%$	α	1- $^{\alpha}/_{2}$	$z_{1-(\alpha/2)}$
0.90	90%	0.10	0.95	1.65
0.95	95%	0.05	0.975	1.96
0.98	98%	0.02	0.990	2.33
0.99	99%	0.01	0.995	2.58

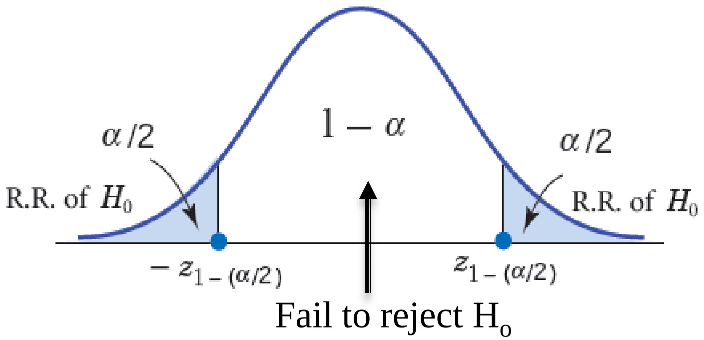
Solution

***Step 4**: Make a decision*

$z_{\bar{x}} = 2.16$ which is greater than 1.96

=> $z_{\bar{x}} = 2.16 < 1.96$, so we fail to reject H_0

Two – tailed critical region



***Step 5**: Make a conclusion*

We don't have enough evidence to accept H_1, or we don't have enough evidence that mean age has changed at $\alpha = 0.02$

Example (3)

A random sample of 27 observations from a large population has a mean of 22 and a standard deviation of 4.8. can we conclude at $\alpha = 0.01$ that a population mean is significantly below 24?

Solution

$\mu = 24$	$n = 27$	$s = 4.8$	$\bar{x} = 22$

***Step 1*: State the null and alternative hypothesis (H_o and H_1)**

Null hypothesis (H_o)	$\mu \geq 23$
alternative hypothesis (H_1)	$\mu < 23$

***Step 2*: Specify the suitable test statistic**

Since $n < 30$ and σ is unknown,
we will use t-statistic

$\mu = 24$	$n = 27$	$s = 4.8$	$\bar{x} = 22$
$\mu_{\bar{x}} = \mu = 24$	$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{4.8}{\sqrt{27}} = 0.924$	$t = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{22 - 24}{0.924} = -2.165$	

Solution

***Step 3**: Compute the critical value for the statistic*

$(1-\alpha)=0.98$	$\alpha=0.01$
	$-t_{\alpha}(27-1)=-t_{0.01}(26)=-2.479$

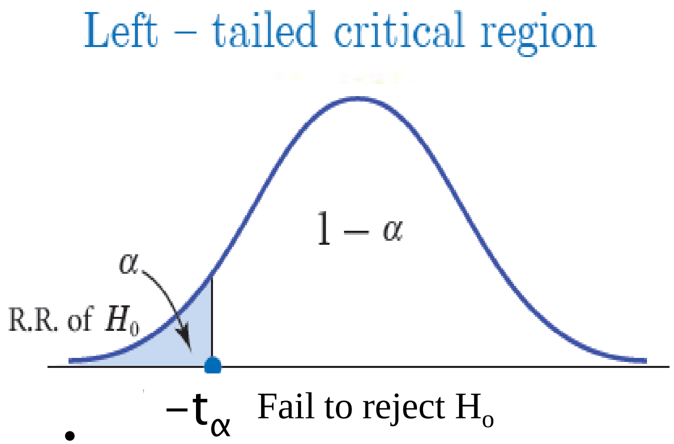
ν	α 0.02	0.015	0.01	0.0075	0.005	0.0025	0.0005
26	2.162	2.296	2.479	2.605	2.779	3.067	3.707

Solution

***Step 4**: Make a decision*

$t = 2.165$ which is greater than -2.479 ,

So we fail to reject H_0



***Step 5**: Make a conclusion*

We don't have enough evidence to accept H_1, or we fail to support H_1