What is Hypothesis?

- Hypothesis is a predictive statement, capable of being tested by scientific methods, that relates an independent variables to some dependent variable.
- A hypothesis states what we are looking for and it is a proportion which can be put to a test to determine its validity

e.g.

Students who receive counseling will show a greater increase in creativity than students not receiving counseling

Characteristics of Hypothesis

- Clear and precise.
- Capable of being tested.
- Stated relationship between variables.
- limited in scope and must be specific.
- Stated as far as possible in most simple terms so that the same is easily understand by all concerned. But one must remember that simplicity of hypothesis has nothing to do with its significance.
- Consistent with most known facts.
- Responsive to testing with in a reasonable time. One can’t spend a life time collecting data to test it.
- Explain what it claims to explain; it should have empirical reference.

Null Hypothesis

- It is an assertion that we hold as true unless we have sufficient statistical evidence to conclude otherwise.
- Null Hypothesis is denoted by H_0
- If a population mean is equal to hypothesised mean then Null Hypothesis can be written as

$$H_0: \mu = \mu_0$$

Alternative Hypothesis

- The Alternative hypothesis is negation of null hypothesis and is denoted by H_a

If Null is given as $H_0: \mu = \mu_0$

Then alternative Hypothesis can be written as

$$H_a: \mu \neq \mu_0$$

$$H_a: \mu > \mu_0$$

$$H_a: \mu < \mu_0$$

Level of significance and confidence

- Significance means the percentage risk to reject a null hypothesis when it is true and it is denoted by α. Generally taken as 1%, 5%, 10%
- $(1-\alpha)$ is the confidence interval in which the null hypothesis will exist when it is true.

Risk of rejecting a Null Hypothesis when it is true

Designation	Risk α	Confidence $1-\alpha$	Description
Supercritical	0.001 0.1%	0.999 99.9%	More than $100 million (Large loss of life, e.g. nuclear disaster
Critical	0.01 1%	0.99 99%	Less than $100 million (A few lives lost)
Important	0.05 5%	0.95 95%	Less than $100 thousand (No lives lost, injuries occur)
Moderate	0.10 10%	0.90 90%	Less than $500 (No injuries occur)

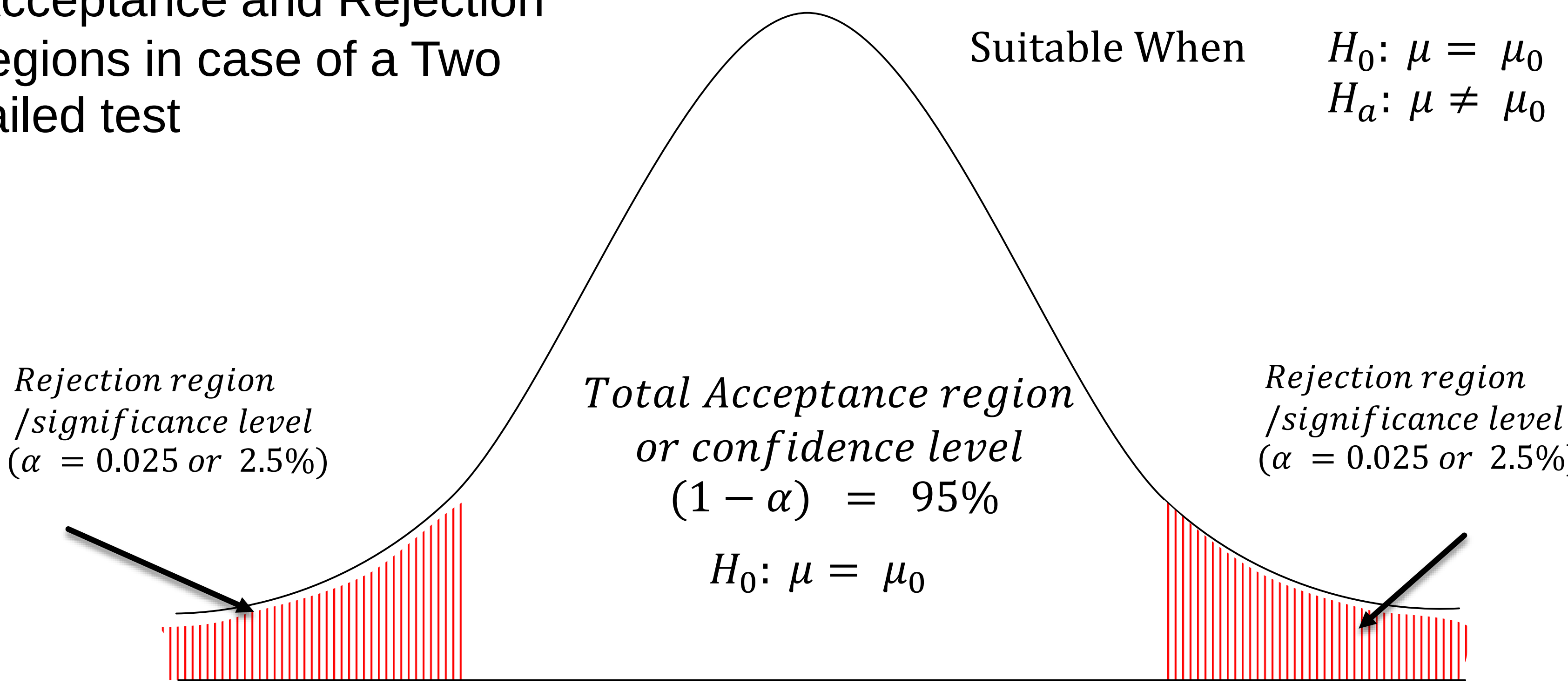
Type I and Type II Error

Situation	Decision	
	Accept Null	Reject Null
Null is true	Correct	Type I error (α *error*)
Null is false	Type II error (β *error*)	Correct

Two tailed test @ 5% Significance level

Acceptance and Rejection regions in case of a Two tailed test

Suitable When $H_0: \mu = \mu_0$
$H_a: \mu \neq \mu_0$

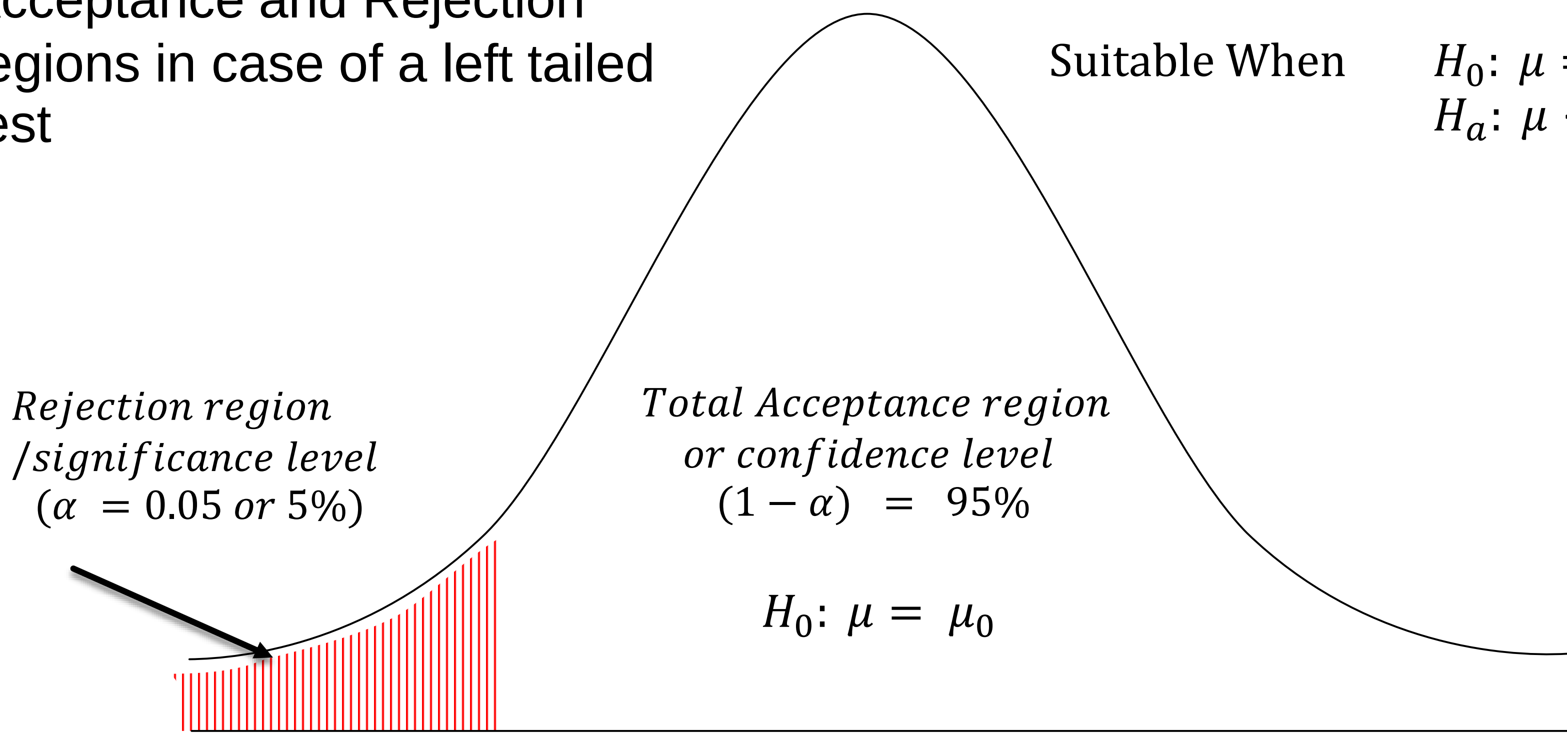


Left tailed test @ 5% Significance level

Acceptance and Rejection regions in case of a left tailed test

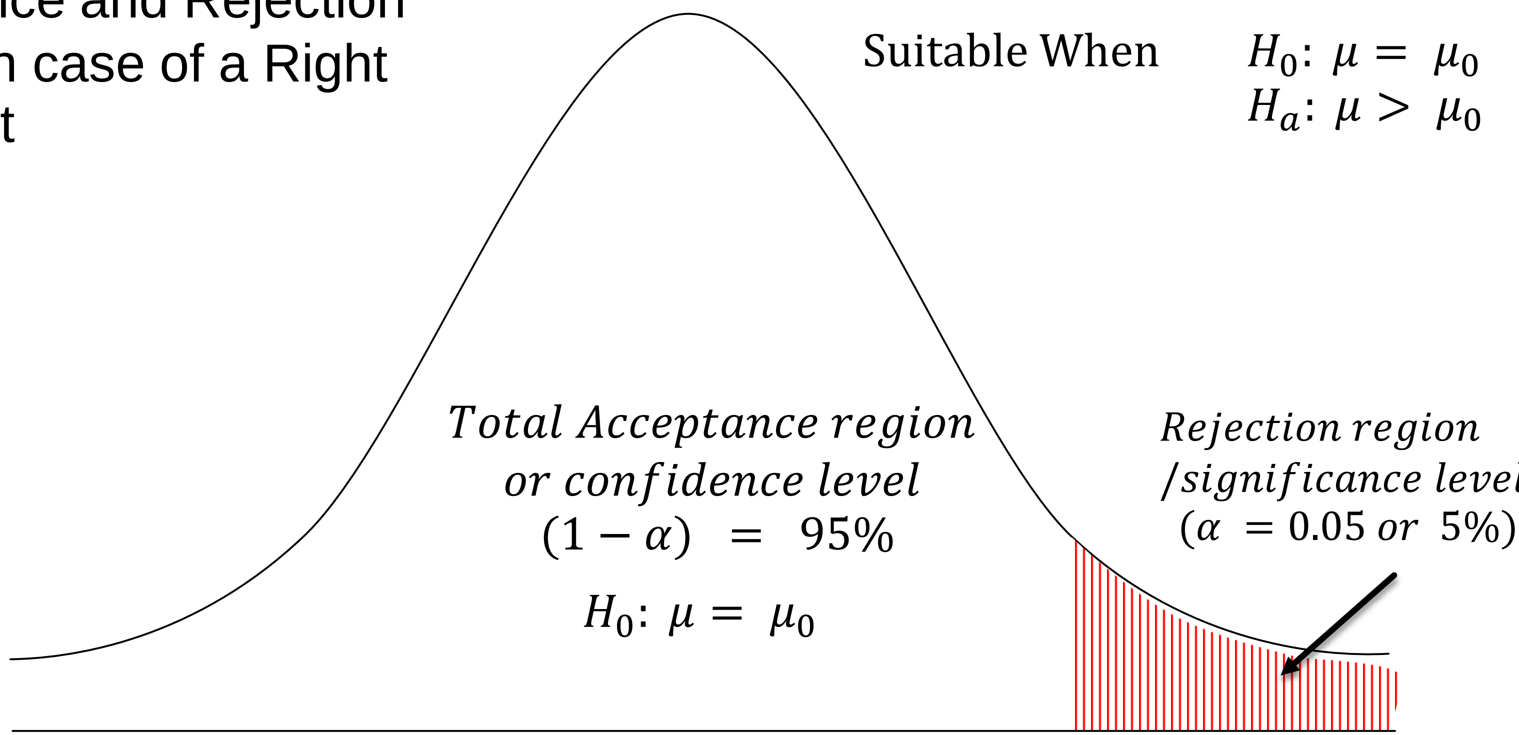
Suitable When $H_0: \mu = \mu_0$

$H_a: \mu < \mu_0$

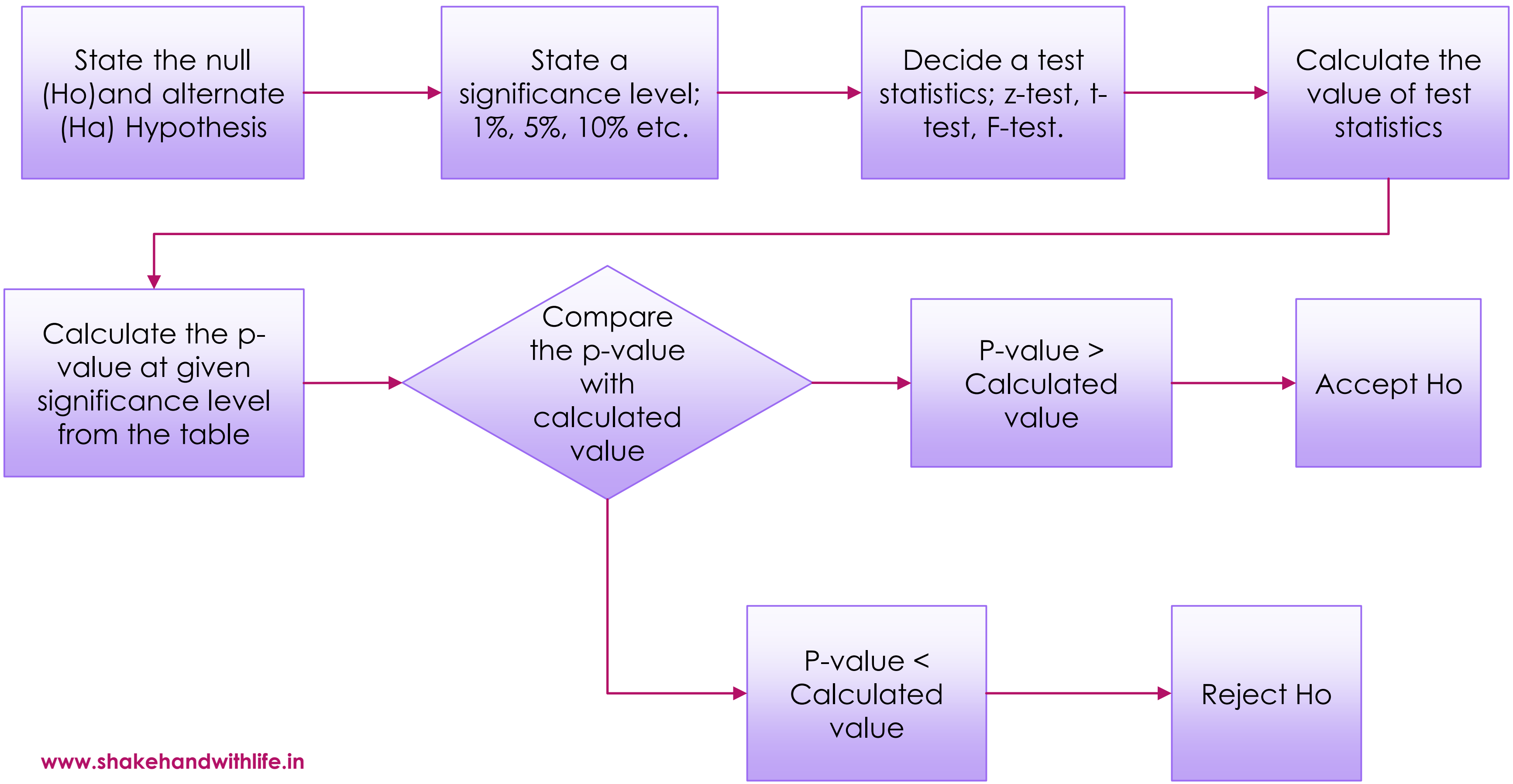


Right tailed test @ 5% Significance level

Acceptance and Rejection regions in case of a Right tailed test



Procedure for Hypothesis Testing



Hypothesis Testing of Means

Z-TEST AND T-TEST

Z-Test for testing means

Test Condition

- Population normal and infinite
- Sample size large or small,
- Population variance is known
- Ha may be one-sided or two sided

Test Statistics

$$Z = \frac{\bar{X} - \mu_{H_0}}{\sigma_p / \sqrt{n}}$$

Z-Test for testing means

Test Condition

- Population normal and finite,
- Sample size large or small,
- Population variance is known
- Ha may be one-sided or two sided

Test Statistics

$$z = \frac{\bar{X} - \mu_{H_0}}{\sigma_p / \sqrt{n} \times \left[\sqrt{(N-n)/(N-1)}\right]}$$

Z-Test for testing means

Test Condition

- Population is infinite and may not be normal,
- Sample size is large,
- Population variance is unknown
- Ha may be one-sided or two sided

Test Statistics

$$z = \frac{\bar{X} - \mu_{H_0}}{\sigma_s / \sqrt{n}}$$

Z-Test for testing means

Test Condition

- Population is finite and may not be normal,
- Sample size is large,
- Population variance is unknown
- Ha may be one-sided or two sided

Test Statistics

$$z = \frac{\bar{X} - \mu_{H_0}}{\sigma_s / \sqrt{n} \times \left[\sqrt{(N-n)/(N-1)}\right]}$$

T-Test for testing means

Test Condition

- Population is infinite and normal,
- Sample size is small,
- Population variance is unknown
- Ha may be one-sided or two sided

Test Statistics

$$t = \frac{\bar{X} - \mu_{H_0}}{\sigma_s / \sqrt{n}}$$

$$with\ d.f. = n - 1$$

$$\sigma_s = \sqrt{\frac{\sum (X_i - \bar{X})^2}{(n-1)}}$$

T-Test for testing means

Test Condition

- Population is finite and normal,
- Sample size is small,
- Population variance is unknown
- Ha may be one-sided or two sided

Test Statistics

$$t = \frac{\bar{X} - \mu_{H_0}}{\sigma_s/\sqrt{n} \times \left[\sqrt{(N-n)/(N-1)}\right]}$$

$$with\ d.f. = n - 1$$

$$\sigma_s = \sqrt{\frac{\sum(X_i - \bar{X})^2}{(n-1)}}$$

Hypothesis testing for difference between means

Z-TEST, T-TEST

Z-Test for testing difference between means

Test Condition

- Populations are normal
- Samples happen to be large,
- Population variances are known
- Ha may be one-sided or two sided

Test Statistics

$$z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_{p1}^2}{n_1} + \frac{\sigma_{p2}^2}{n_2}}}$$

Z-Test for testing difference between means

Test Condition

- Populations are normal
- Samples happen to be large,
- Presumed to have been drawn from the same population
- Population variances are known
- Ha may be one-sided or two sided

Test Statistics

$$z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\sigma_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

T-Test for testing difference between means

Test Condition

- Samples happen to be small,
- Presumed to have been drawn from the same population
- Population variances are unknown but assumed to be equal
- Ha may be one-sided or two sided

Test Statistics

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1 - 1)\sigma_{s1}^2 + (n_2 - 1)\sigma_{s2}^2}{n_1 + n_2 - 2}} \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$with\ d.f. = (n_1 + n_2 - 2)$$

Hypothesis Testing for comparing two related samples

PAIRED T-TEST

Paired T-Test for comparing two related samples

Test Condition

- Samples happens to be small
- Variances of the two populations need not be equal
- Populations are normal
- Ha may be one sided or two sided

Test Statistics

$$t = \frac{\overline{D} - 0}{\sigma_{diff.} / \sqrt{n}}$$

$$with\ (n-1)\ d.f.$$

$\overline{D}$ = Mean of differences

$\sigma_{diff.}$ = Standard deviation of differences

$n = Number\ of\ matched\ pairs$

Hypothesis Testing of proportions

Z-TEST

Z-test for testing of proportions

Test Condition

- Use in case of qualitative data
- Sampling distribution may take the form of binomial probability distribution
- Ha may be one sided or two sided
- $Mean = n.p$
- $Standard\ deviation = \sqrt{n.p.q}$

Test statistics

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p.q}{n}}}$$

$\hat{p} = proportion\ of\ sucess$

Hypothesis Testing for difference between proportions

Z-TEST

Z-test for testing difference between proportions

Test Condition

- Sample drawn from two different populations
- Test confirm, whether the difference between the proportion of success is significant
- Ha may be one sided or two sided

Test statistics

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}}$$

$\hat{p}_1 = \text{proportion of success in sample one}$

$\hat{p}_2 = \text{proportion of success in sample two}$

Hypothesis testing of equality of variances of two normal populations

F-TEST

F-Test for testing equality of variances of two normal populations

Test conditions

- The populations are normal
- Samples have been drawn randomly
- Observations are independent; and
- There is no measurement error
- Ha may be one sided or two sided

Test statistics

$$F = \frac{\sigma_{s1}^2}{\sigma_{s2}^2}$$

$with\ (n_1 - 1)$ and $(n_2 - 1)\ \mathrm{d.f.}$

σ_{s1}^2 is the sample estimate for σ_{p1}^2

σ_{s2}^2 is the sample estimate for σ_{p2}^2

Limitations of the test of Hypothesis

- Testing of hypothesis is not decision making itself; but help for decision making
- Test does not explain the reasons as why the difference exist, it only indicate that the difference is due to fluctuations of sampling or because of other reasons but the tests do not tell about the reason causing the difference.
- Tests are based on the probabilities and as such cannot be expressed with full certainty.
- Statistical inferences based on the significance tests cannot be said to be entirely correct evidences concerning the truth of the hypothesis.