UNIT-VI : TESTING OF HYPOTHESIS

Introduction :- The principal objective of "statistical inference" is to draw inferences about the population on the basis of data collected by sampling from the population. Statistical inferences consists of two major areas Estimation and Test of Hypothesis. Estimation was discussed in the Previous unit. In tests of hypothesis, a postulate or statement about a parameter of the population is tested for its validity (or) Truthfulness.

The main of "Testing of Hypothesis" is provide rules that lead to decision resulting in acceptance (or) rejection of statements about the population parameters.

Statistical Decisions :- These are the decisions (or) Conclusions about the population parameters on the basis of a random sample from the population.

Statistical Hypothesis :- It is an assumption (or) guess about the parameter(s) of the population distribution(s). And it is of two types.

(i) Null Hypothesis (ii) Alternative Hypothesis.

Null Hypothesis :- It is denoted by "H_0" is the statistical Hypothesis which is to be actually tested for the acceptance (or) rejection.

Alternative Hypothesis :- It is denoted by "H_1" is any hypothesis otherthan the Null Hypothesis.

Test of Hypothesis : It is the procedure to decide whether to accept or reject the null Hypothesis.

⟹ If Null Hypothesis is accepted then the result is said to be Non-significant. If it is rejected then the result is said to be significant.

⇒ Thus Test of hypothesis decides whether a statement concerning a parameter is true or false instead of estimating the value of the parameter.

⇒ Since the Test is based on sample observations the decision of acceptance (or) rejection of Null Hypothesis is always subjected to some error ((i.e) some amount of risk)

⇒ Type of Errors: 1. Type I error
2. Type II error.

Type I error :- The rejection of Null Hypothesis 'H_0' under H_0 is true.

Type II error :- The acceptance of Null Hypothesis 'H_0' under H_0 is false.

⇒

	Accept H_0	Reject H_0
H_0 is True	Correct Decision	Type-I Error
H_0 is False	Type-II Error	Correct Decision

Level of significance :- It is denoted by "α", is the probability of commiting type-I error. Thus, Level of significance measures the amount of risk (or) error associated in taking decisions.

⇒ (i.e) $\alpha = P(\text{type-I error})$, Generally '$\alpha$' takes 0.05 (or) 0.01, It is again called as size of the test.

NOTE:- If β = Probability of commiting type II error.

(i.e) $\beta = P(\text{Type II Error})$,

Here "$1-\beta$" is called Power of the test.

Simple Hypothesis :- If the hypothesis is of equality type then it is said to be simple Hypothesis. Generally Null Hypothesis is the simple Hypothesis.

Ex: $H_0 : \mu = \mu_0$ is Simple Hypothesis.

Composite Hypothesis :- If the hypothesis is stated interms of inequalities such as $<$, $>$ (or) $\neq$.

Ex: $H_1 : \mu \neq \mu_0$ (or) $H_1 : \mu < \mu_0$ (or) $H_1 : \mu > \mu_0$.

CRITICAL REGION :- It is the region of rejection of N.H. The area of critical region equals to level of significance

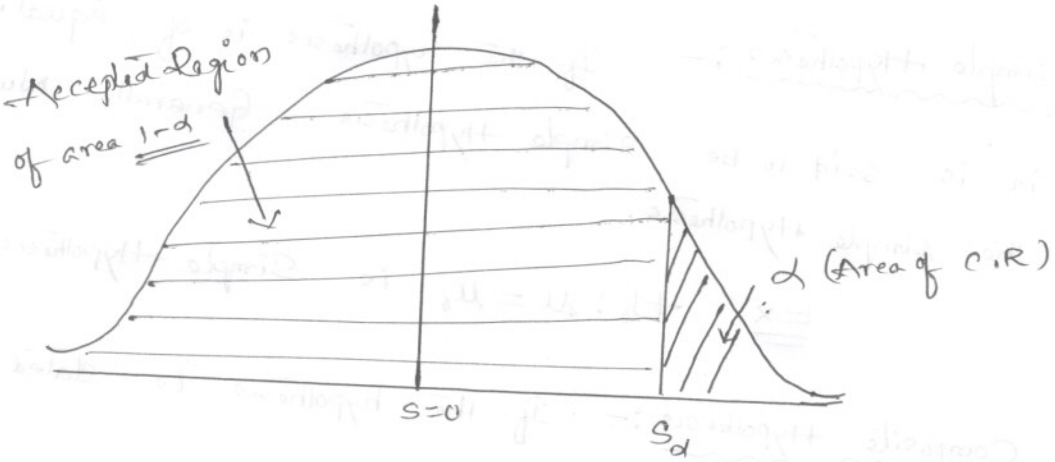
(i.e) Area of critical region $= \alpha =$ L.O.S.

NOTE :- Depending on the nature of Alternative Hypothesis, critical region may lie on one side (or) both side of the tails.

$\Rightarrow$ If C.R lies on one side then the corresponding test is said to be one tailed test. Here one tailed tests are two types (i) Right-tailed test (ii) Left-tailed test.

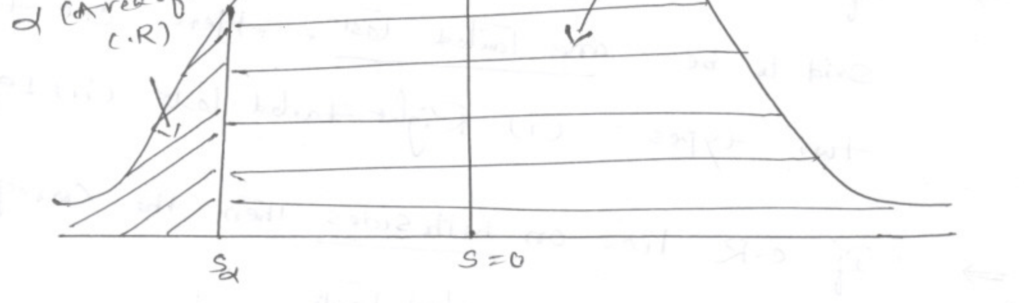
$\Rightarrow$ If C.R lies on both sides then the corresponding test is said to be two tailed test.

<u>Right tailed test</u> :- When the alternative hypothesis H_1 is of greaterthan type (i.e) $H_1 : \mu > \mu_0$ then the entire critical region of area 'α' lies on the <u>right side tail</u> of the probability density curve shown in following figure.



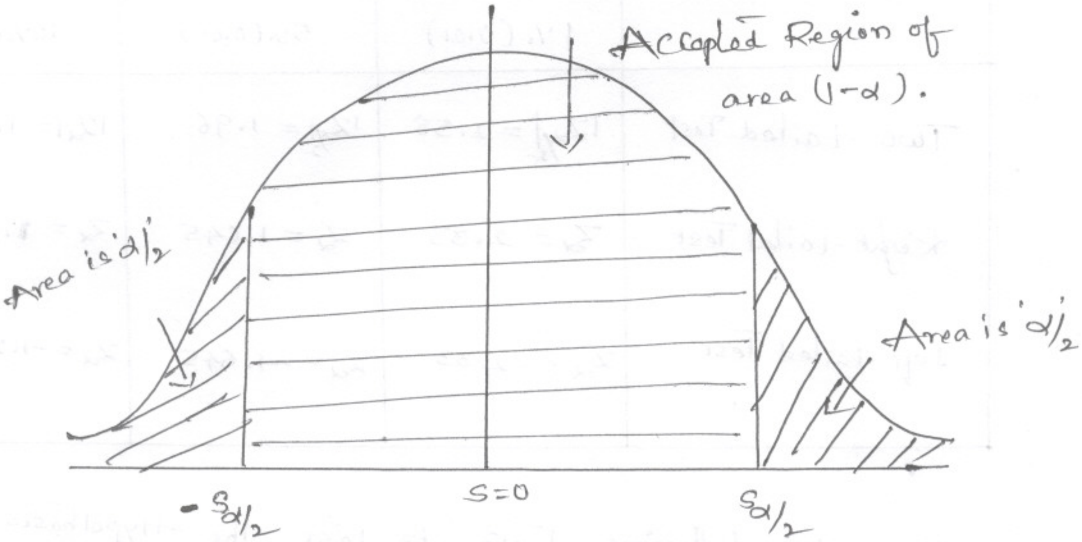
In such case the test of hypothesis is known as <u>Right one tailed test</u>.

<u>Left tailed test</u> :- When the alternative hypothesis H_1 is of lessthan type (i.e) $H_1 : \mu < \mu_0$ then the entire critical region of area 'α' lies on the <u>Left side tail</u> of the probability density curve shown in following figure



In such case the test of hypothesis is known as <u>Left one tailed test</u>.

Two tailed test :- When the alternative hypothesis H_1 is of not equal to type (i.e) $H_1: \mu \neq \mu_0$ then the entire critical region of area 'α' lies on the both sides of the probability density curve shown in following figure.



In such case the test of hypothesis is known as **Two tailed test.**

Working Rule for testing of Hypothesis :-

The following working rule or procedure may be adopted in testing of a statistical hypothesis.

step ①: Null Hypothesis : Set up a N.H (H_0) in clear terms.

step ②: Alternative Hypothesis : Set up the A.H (H_1) so that we could decide whether we should use one.(or) two-tailed test.

step ③: Level of Significance : select appropriate L.O.S 'α'.

step ④: Test statistic : $Z = \dfrac{t - E(t)}{S.E(t)}$

step ⑤: Conclusion : If $|Z_{cal}| < |Z_{table}|$, Accept H_0

If $|Z_{cal}| > |Z_{table}|$, Reject H_0.

Note: For the table values, the following table is very much useful.

⟹

Type of the test	Level of significance		
	1% (0.01)	5% (0.05)	10% (0.1)
Two tailed Test	$\lvert Z_{\alpha/2}\rvert = 2.58$	$\lvert Z_{\alpha/2}\rvert = 1.96$	$\lvert Z_{\alpha}\rvert = 1.645$
Right tailed Test	$Z_{\alpha} = 2.33$	$Z_{\alpha} = 1.645$	$Z_{\alpha} = 1.28$
Left tailed Test	$Z_{\alpha} = -2.33$	$Z_{\alpha} = -1.645$	$Z_{\alpha} = -1.28$

⟹ Here we have the following tests to test the Hypothesis.

(i) Hypothesis concerning one and two means

(ii) Hypothesis concerning Proportions

(I) HYPOTHESIS CONCERNING ONE & TWO MEANS :

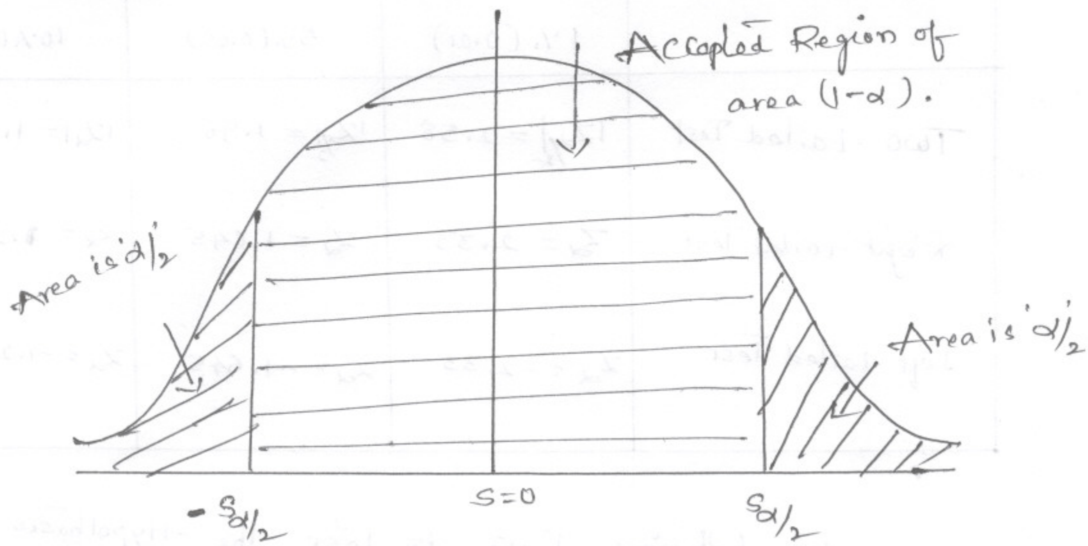
(1) TEST OF SIGNIFICANCE FOR SINGLE MEAN :-

Suppose we want to test whether the given sample of size 'n' has been drawn from a population with mean 'μ'. We setup null Hypothesis that there is no difference b/w $\bar{x}$ and μ, where $\bar{x}$ is the sample mean.

The test statistic is, $Z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

where 'σ' is the S.D of population.

Two tailed test :- When the alternative hypothesis H_1 is of not equal to type (i.e) $H_1: \mu \neq \mu_0$ then the entire critical region of area 'α' lies on the both sides of the probability density curve shown in following figure.



In such case the test of hypothesis is known as Two tailed test.

Working Rule for testing of Hypothesis :-

The following working rule or procedure may be adopted in testing of a statistical hypothesis.

step ①: Null Hypothesis : Set up a N.H (H_0) in clear terms.

step ②: Alternative Hypothesis : Set up the A.H (H_1) so that we could decide whether we should use one (or) two tailed test.

step ③: Level of significance : select appropriate L.O.S 'α'.

step ④: Test statistic : $Z = \dfrac{t - E(t)}{S.E(t)}$.

step ⑤: Conclusion : If $|Z_{cal}| < |Z_{tab}|$, Accept H_0.

If $|Z_{cal}| > |Z_{table}|$, Reject H_0.

Problems:

1. Mice with an average life span of 32 months will live upto 40 months when fed by a certain nutritious food. If 64 mice fed on this diet have an average lifespan of 38 months and S.D of 5.8 months, is there any reason to believe that average lifespan is lessthan 40 months.

Sol: Let μ = average lifespan of mice fed with nutritious food.

use 0.01 level of significance.

Null Hypothesis : H_0 : $\mu = 40$ months

Alternative Hypothesis H_1 : $\mu < 40$ months

Level of significance α : 0.01 (Left tailed)

Test Statistic : $Z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

Here $\bar{x} = 38$, $\mu = 40$, $\sigma = 5.8$, $n = 64$.

$$\therefore Z_{cal} = \frac{38 - 40}{5.8/\sqrt{64}} = -2.76$$

$$\Rightarrow |Z_{cal}| = 2.76$$

But the table value of Z at $\alpha = 0.01$ is $|Z_{table}| = 2.33$

Conclusion: Since $|Z_{cal}| > |Z_{table}|$ ($\because 2.76 > 2.33$)

Hence Reject H_0. $\Rightarrow$ Accept H_1.

Yes, there is reason to believe that the average lifespan of mice with nutritious food is lessthan 40 months.

2. According to the norms established for a mechanical aptitude test persons who are 18 years old have an average height of 73.2 with a standard deviation of 8.6. If 4 randomly selected persons of that age averaged 76.7, test the hypothesis $\mu = 73.2$ against the alternative hypothesis $\mu > 73.2$ at the 0.01 level of significance.

Sol: Given $n = 4$, $\mu = 73.2$, $\bar{x}$ = Mean of the sample = 76.7

$\sigma = 8.6$.

<u>Null Hypothesis</u> H_0 : $\mu = 73.2$

<u>Alternative Hypothesis</u> H_1 : $\mu > 73.2$ (Right tailed)

<u>Level of significance</u> α : 0.01

<u>Test statistic</u> : $Z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

$$\therefore Z_{cal} = \frac{76.7 - 73.2}{8.6/\sqrt{4}} = 0.814$$

$$\therefore |Z_{cal}| = 0.814$$

But the table value of 'Z' at $\alpha = 0.01$ is

$$|Z_{table}| = 2.33$$

<u>Conclusion</u> : Since $|Z_{cal}| < |Z_{table}|$ (Since $0.814 < 2.33$)

Accept H_0.

3. A sample of 64 students have a mean weight of 70 Kgs. Can this be regarded as a sample from a population with mean weight 65 Kgs and S.D 25 Kgs.

Sol: Here $n = 64$, $\bar{x} = 70$, $\mu = 65$, $\sigma = 25$

<u>Null Hypothesis</u> H_0 : $\mu = 65$

<u>Alternative Hypothesis</u> H_1 : $\mu > 65$ (Right tailed)

<u>L.O.S</u> α : 0.05

<u>Test statistic</u> $Z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} \Rightarrow Z_{cal} = \dfrac{70 - 65}{25/\sqrt{64}} = 1.6$

But Z_{table} $(\alpha = 0.05) = 1.645$

Since $|Z_{cal}| < |Z_{table}|$ $(\because 1.6 < 1.645)$

<u>Conclusion</u>: Accept H_0.

4. In a random sample of 60 workers, the average time taken by them to get to work is 33.8 minutes with a S.D of 6.1 minutes. Can we reject the null Hypothesis $\mu = 32.6$ minutes against the alternative hypothesis $\mu > 32.6$ at $\alpha = 0.025$.

(Hint: Z_{table} (at $\alpha = 0.025$) $= 2.58$)

Sol: Given $n = 60$, $\bar{x} = 33.8$, $\mu = 32.6$, $\sigma = 6.1$

Null Hypothesis $H_0 : \mu = 32.6$

Alternative Hypothesis $H_1 : \mu > 32.6$ (Right-tailed)

L.O.S : $\alpha = 0.025$

Test Statistic : $Z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

$$\therefore Z_{cal} = \frac{33.8 - 32.6}{6.1/\sqrt{60}} = \frac{1.2}{0.7875} = 1.5238$$

But Z_{table} ($\alpha = 0.025$) $= 2.58$

Conclusion: since $|Z_{cal}| < |Z_{table}|$, Accept H_0.

5. An ambulance service claims that it takes on the average lessthan 10 minutes to reach its destination in emergency calls. A sample of 36 calls has a mean of 11 minutes and the variance of 16 minutes. Test the significance at 0.05 L.o.s.

Sol:- (Here $\bar{x} = 11$, $\mu = 10$, $n = 36$, $\sigma^2 = 16 \Rightarrow \sigma = 4$.

$$Z_{cal} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{11 - 10}{4/\sqrt{36}} = 1.5$$

$Z_{table} = 1.645$

Conclusion: Accept H_0 (since $|Z_{cal}| < |Z_{table}|$).)

(2) TEST OF SIGNIFICANCE FOR DIFFERENCE OF MEANS:

Let $\bar{x}_1$ be the mean of a sample of size 'n_1' from a population with mean μ_1 and variance σ_1^2. Let $\bar{x}_2$ be the mean of a sample of size 'n_2' from a population with mean μ_2 and variance σ_2^2, then to test whether there is any significant difference b/w $\bar{x}_1$ and $\bar{x}_2$ we have to use the statistic

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

→ Note: If $\sigma_1^2 = \sigma_2^2 = \sigma^2$ (say) then $Z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$.

Problems:

1. Samples of students were drawn from two universities and from their weights in Kilograms, mean and S.D are calculated and shown below. Make a large sample test to test the significance of the difference b/w the means.

	Mean	S.D	Sample size
university-A	55	10	400
university-B	57	15	100

Sol: Given $\bar{x}_1 = 55$, $\bar{x}_2 = 57$, $\sigma_1 = 10$, $\sigma_2 = 15$, $n_1 = 400$, $n_2 = 100$

Null Hypothesis H_0: $\mu_1 = \mu_2$ (there is no difference)

Alternative Hypothesis H_1: $\mu_1 \neq \mu_2$ (two tailed)

Level of significance $\alpha = 0.05$ (assume)

Test statistic $Z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$.

$$\therefore\ Z_{cal} = \frac{55-57}{\sqrt{\frac{(10)^2}{400}+\frac{(15)^2}{100}}} = -1.26$$

$$\therefore\ |Z_{cal}| = 1.26$$

But Z_{table} (at $\alpha = 0.05$) $= 1.96$

Conclusion: Since $|Z_{cal}| < |Z_{table}|$, Hence Accept H_0

(i-e) There is no significant difference b/w the means.

2. The means of two large samples of sizes 1000 & 2000 members are 67.5 inches & 68.0 inches respectively. Can the samples be regarded as drawn from the same population of S.D 2.5 inches.

Sol: Given $n_1 = 1000$, $n_2 = 2000$

$\bar{x}_1 = 67.5$, $\bar{x}_2 = 68.0$ & $\sigma_1 = \sigma_2 = 2.5 = \sigma$ (say)

Null Hypothesis H_0: $\mu_1 = \mu_2$

Alternative Hypothesis H_1: $\mu \neq \mu_2$ (two tailed)

L.O.S : $\alpha = 0.05$ (assume)

Test statistic : $Z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$

$$\therefore\ Z_{cal} = \frac{67.5-68.0}{2.5\sqrt{\frac{1}{1000}+\frac{1}{2000}}} = -5.16$$

$$\therefore\ |Z_{cal}| = 5.16$$

But the table value of z (i.e) Z_{table} (at $\alpha = 0.05$) $= 1.96$

conclusion: Since $|Z_{cal}| > |Z_{table}|$, Reject the H_0.

3. To test the claim that men are taller than women, a survey was conducted resulting in the following data

Gender	Sample size	Mean Height (cm)	S.d (cm)
Men	1600	172	6.3
women	6400	170	6.4

Is the claim tenable at 0.01 L.O.S.

Sol: Here $n_1 = 1600$, $n_2 = 6400$

$\bar{x}_1 = 172$, $\bar{x}_2 = 170$

$\sigma_1 = 6.3$, $\sigma_2 = 6.4$

Null Hypothesis H_0: $\mu_1 = \mu_2$

Alternative Hypothesis H_1: $\mu_1 \neq \mu_2$ (two tailed)

L.O.S : $\alpha = 0.01$

Test Statistic : $Z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

$$\therefore \quad Z_{cal} = \frac{172 - 170}{\sqrt{\frac{(6.3)^2}{1600} + \frac{(6.4)^2}{6400}}} = 11.32$$

But Z_{table} (at $\alpha = 0.05$) $= 2.58$

Conclusion: Since $|Z_{cal}| > |Z_{table}|$, Reject H_0

Hence the claim is not tenable at $\alpha = 0.05$.

II) HYPOTHESIS CONCERNING PROPORTIONS:

(1) Test for significance for single proportion:

Suppose a large sample of size 'n' is taken from a normal population. To test the significant difference b/w the sample proportion 'p' and the population proportion 'P', we use the statistic

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}, \text{ where 'n' is the sample size}$$

$\Rightarrow$

> **Note:** Proportion means the probability of successes (i.e) If out of 'n' trails 'x' times we are succeded then Proportion $\Rightarrow P = \frac{x}{n}$.

Problems:

1. In a big city 325 men out of 600 men were found to be smokers. Does this information support the conclusion that the majority of men in this city are smokers.

Sol: Given $n = 600$, Number of smokers $= 325$

$\therefore$ p = proportion of Smokers $= \frac{325}{600} = 0.5417$

P = Population proportion of smokers in city $= 1/2 = 0.5$

$\therefore Q = 1 - P = 1 - 1/2 = 1/2$

Null Hypothesis H_0: $p = P$.

Alternative Hypothesis: $p > P$ (ie) $p > 0.5$ (Right Tailed).

L.O.S : $\alpha = 0.05$ (assume)

Test statistic: $Z = \frac{p - P}{\sqrt{PQ/n}} = \frac{0.5417 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{600}}}$

$= 2.04$

$\therefore |Z_{cal}| = 2.04$

But Z_{table} (at $\alpha = 0.05$) $= 1.645$

Conclusion: Since $|Z_{cal}| > |Z_{table}|$, Reject H_0.

(i.e) The majority of men in city are Smokers.

2. In a random sample of 125 cola drinkers, 68 said they prefer thumsup to pepsi. Test the null Hypothesis $P = 0.5$ against the alternative hypothesis $P > 0.5$.

Sol: Here, $n = 125$, $x = 68$

$\therefore\ p$ = Proportion of sample $= \frac{x}{n} = \frac{68}{125} = 0.544$

Null Hypothesis H_0: $P = 0.5$

Alternative Hypothesis H_1: $P > 0.5$ (Right tailed)

L.O.S : $\alpha = 0.05$ (assumed)

Test statistic : $Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.544 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{125}}}$

$= 0.9839$

$\therefore Z_{cal} = 0.9839$

But Z_{table} ($\alpha = 0.05$) $= 1.645$

Conclusion: Since $|Z_{cal}| < |Z_{table}|$

Hence Accept H_0.

3. Experience had shown that 20% of manufactured product is of the top quality in one day's production of 400 articles only 50 are of top quality. Test the hypothesis at $\alpha = 0.05$.

Sol: We have $n = 400$, $x = 50 \Rightarrow p = \frac{50}{400} = 0.125$

Null Hypothesis H_0: $P = 20\% \Rightarrow P = 0.2$

Alternative Hypothesis H_1: $P \neq 0.2$ (two tailed)

Level of significance: $\alpha = 0.05$

Test statistic: $Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.125 - 0.2}{\sqrt{\frac{(0.2)(0.8)}{400}}}$

$\Rightarrow \quad Z_{cal} = -3.75 \Rightarrow |Z_{cal}| = 3.75$

But Z_{table} (at $\alpha = 0.05$) $= 1.96$

Conclusion: Since $|Z_{cal}| > |Z_{table}|$, Reject H_0.

(i.e) $P = 20\%$ is not Correct.

4. A manufacturer claims that only 4% of his products are defective. A random sample of 500 were taken among which 100 were defective. Test the hypothesis at 0.05 level.

Sol: Here $x = 100$, $n = 500 \Rightarrow p = \frac{x}{n} = \frac{100}{500} = 0.2$

$P = 4\% = 0.04 \Rightarrow Q = 1 - P = 1 - 0.04 = 0.96$.

Null Hypothesis H_0: $P = 0.04$

Alternative Hypothesis H_1: $P > 0.04$ (Right tailed)

L.O.S : $\alpha = 0.05$

Test Statistic: $Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.2 - 0.04}{\sqrt{\frac{(0.04)(0.96)}{500}}} = \underline{18.26}$

$\therefore \quad Z_{cal} = 18.26$

But Z_{table} (at $\alpha = 0.05$) $= 1.645$

Conclusion: Since $|Z_{cal}| > |Z_{table}|$, Reject H_0.

5. In a random sample of 160 workers exposed to a certain amount of radiation 24 experienced some ill effects. Construct a 99% confidence interval for the corresponding true percentage.

Sol: Here $x = 24, \; n = 160 \Rightarrow p = \frac{x}{n} = \frac{24}{160} = 0.15 = P$ (say)

$\therefore \; Q = 1 - P = 1 - 0.15 = 0.85$

$\therefore \sqrt{\frac{PQ}{n}} = \sqrt{\frac{(0.15)(0.85)}{160}} = 0.028$

Confidence interval at 99% (i.e) $\alpha = 1\%$) is

$$\left(P - Z_{\alpha/2}\sqrt{\frac{PQ}{n}}, \; P + Z_{\alpha/2}\sqrt{\frac{PQ}{n}}\right)$$

$$\Rightarrow (0.15 - 1.96 \times 0.028, \quad 0.15 + 1.96 \times 0.028)$$

$$\Rightarrow (0.09512, \quad 0.20488).$$

6. 20 people were attacked by a disease and only 18 survived. Will you reject the hypothesis that the survival rate if attacked by this disease is 85%. infavour of the hypothesis that is more at 5% level.

Sol: Here $p = 18/20 = 0.9, \; P = 85\% = 0.85$

$\Rightarrow Q = 0.15$

$$Z_{cal} = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.9 - 0.85}{\sqrt{\frac{(0.85)(0.15)}{20}}} = 0.625$$

$Z_{table} = 1.645$ (at $\alpha = 0.05$)

Conclusion: Accept H_0.

(2) Test for significance of Difference of proportions:

Suppose two large samples of sizes n_1 & n_2 are taken respectively from two different populations. To test the significant difference b/w the proportions, the statistic is,

$$Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}, \quad \text{where } P = \frac{n_1 P_1 + n_2 P_2}{n_1+n_2} \quad (\text{or})$$

$$P = \frac{x_1 + x_2}{n_1 + n_2}$$

$$\therefore Q = 1 - P.$$

Problems:

1. A manufacturer of electronic equipment subjects samples of two completing brands of transistors to an accelerated performance test. If 45 of 180 transistors of the first kind and 34 of 120 transistors of the second kind fail the test, what can he conclude at the L.O.S $\alpha = 0.05$ about the difference b/w the corresponding sample proportions.?

Sol: Here, $n_1 = 180$, $n_2 = 120$

$x_1 = 45$, $x_2 = 34$

$$\therefore p_1 = \frac{x_1}{n_1} = \frac{45}{180} = 0.25, \quad p_2 = \frac{x_2}{n_2} = \frac{34}{120} = 0.283$$

$$\therefore P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{45+34}{180+120} = \frac{79}{300} = 0.263$$

$$Q = 1 - P = 1 - 0.263 = 0.737.$$

Null Hypothesis H_0: $p_1 = p_2$ (i.e) there is no difference

Alternative Hypothesis H_1: $p_1 \neq p_2$

Level of significance : $\alpha = 0.05$

Test statistic : $$Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$\therefore Z_{cal} = \frac{0.25 - 0.283}{\sqrt{(0.263)(0.737)\left(\frac{1}{180}+\frac{1}{120}\right)}} = -0.647$$

$$\therefore |Z_{cal}| = 0.647.$$

But Z_{table} (at $\alpha = 0.05$) $= 1.96$

Conclusion :- Since $|Z_{cal}| < |Z_{table}|$ ($\because 0.647 < 1.96$)

Accept H_0

(i.e) there no significant difference b/w two proportions.

2. On the basis of their total scores, 200 candidates of a civil service examination are divided into two groups, the upper 30% and the remaining 70%. Consider the first question of the examination. Among the first group, 40 had the correct answer, whereas among the second group, 80 had the correct answer. On the basis of these results, can one conclude that first question is not good at discriminating ability of the type being examined here?.

Sol: We have, $n_1 = 60$, $n_2 = 140$ ($\because n_1 = 30\%$ of $200 = 60$

$x_1 = 40$, $x_2 = 80$ $\quad$ $n_2 = 70\%$ of $200 = 140$)

$$\therefore p_1 = \frac{x_1}{n_1} = \frac{40}{60} = 0.667,$$

$$p_2 = \frac{x_2}{n_2} = \frac{80}{140} = 0.571$$

$$\therefore p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{40+80}{60+140} = \frac{120}{200} = 0.6$$

$$q = 1 - 0.6 = \underline{0.4}.$$

Null Hypothesis H_0: $p_1 = p_2$

Alternative Hypothesis H_1: $p_1 \neq p_2$ (two tailed)

Level of significance: $\alpha = 0.05$ (assumed)

Test Statistic: $Z = \dfrac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

$$\Rightarrow \quad Z_{cal} = \frac{0.667 - 0.571}{\sqrt{(0.6)(0.4)\left(\frac{1}{60} + \frac{1}{40}\right)}}$$

$$= \frac{0.096}{0.0756} = 1.27$$

But $\quad Z_{table}$ (at $\alpha = 0.05$) $= 1.96$

Conclusion: Since $|Z_{cal}| < |Z_{table}|$, Hence Accept the H_0.

(x)

UNIT : VII TEST OF SIGNIFICANCES

INTRODUCTION: A very important aspect of the sampling theory is the study of tests of significance which enable us to decide on the basis of the samples results if the deviation between the observed sample statistic and the hypothetical parameter value is significant, and the deviation b/w two samples statistics is significant.

⇒ NOTE: Here in this present unit we are going to discuss about the following distributions tests.

(i) Student's t - test

(ii) F - test

(iii) chisquared (χ^2) test

The above three tests are applied only for the samples of size $n \leq 30$ (small samples). So that the above tests are called Small Sample tests.

1. STUDENTS 't' - Test: Suppose we want to test

ⓐ If a random sample of size 'n' has been drawn from a Normal population with specified mean 'μ'.

ⓑ If the sample mean differs significantly from the hypothetical value 'μ' of the population mean

In this case the statistic is given by

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}},$$

where $\bar{x}$ = sample mean
μ = population mean
S = sample S.D
n = sample size.

NOTE:- Here At $(n-1)$ degrees of freedom the Confidence interval for 'μ' with $(1-\alpha)\%$ Confidence is givenby,

$$\bar{x} - t_\alpha \frac{s}{\sqrt{n}} \le \mu \le \bar{x} + t_\alpha \frac{s}{\sqrt{n}}.$$

Problems:

1. A random sample of six steel beams has a mean compressive strength of 58,392 p.s.i (pounds per square inch) with a S.D of 648 p.s.i. use this information and the L.O.S $\alpha = 0.05$ to test whether the true average compressive strength of the steel from which this sample came is 58,000 p.s.i. Assume normality.

Sol: Here $n = 6 < 30$ (Small Sample)

$\bar{x}$ = Sample mean = 58,392 p.s.i

s = S.D = 648 p.s.i

Degrees of freedom $= 6-1 = 5$.

Null Hypothesis H_0 : $\mu = 58\,000$

Alternative Hypothesis H_1 : $\mu \neq 58000$

Level of Significance $\alpha = 0.05$

Test Statistic : $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$

$$\Rightarrow t_{cal} = \frac{58,392 - 58000}{648/\sqrt{6}} = 1.482$$

Now the table value of 't' at $\alpha = 0.05$ with $6-1 = 5$ d.o.f is 3.365

Conclusion: Since $|t_{cal}| < |t_{table}|$ $(\because 1.482 < 3.365)$

Hence Accept H_0.

Hence the avg compressive strength of steel beem is equals to 58000 p.s.i.

2. Mean life-time (m|t) of Computers manufactured by a Company is 1120 hours with S.D of 125 hours. (a) Test the hypothesis that mean Lifetime of Computers has not changed if a sample of 8 computers has a m|t of 1070 hours

(b) Is there decrease in m|t? use (i) 0.05 (ii) 0.01

Sol:

Here $\mu = 1120, s = \sigma = 125$, $\bar{x} = 1070$, $n = 8$

Null Hypothesis H_0: $\mu = 1120$

Alternative Hypothesis H_1: $\mu \neq 1120$ (two tailed)

Level of significance: (i) $\alpha = 0.05$

(ii) $\alpha = 0.01$

Test Statistic: $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$

$$\Rightarrow t_{cal} = \frac{1070 - 1120}{125/\sqrt{8}} = -1.1313$$

$$\Rightarrow |t_{cal}| = 1.1313$$

(i) But At $\alpha = 0.05$ ($\Rightarrow \alpha/2 = 0.025$) $\}$ $|t_{table}|$ (at $8-1=7$ d.o.f) $= 2.365$

(ii) At $\Rightarrow \alpha = 0.01$ ($\Rightarrow \alpha/2 = 0.005$) $\}$ $|t_{table}|$ (at 7 d.o.f) $= 3.499$

Conclusion: In both cases ((i.e) $\alpha = 0.05$ & 0.01)

$|t_{cal}| < |t_{table}|$ $\therefore$ Accept H_0.

3. Producer of 'gutka' claims that the nicotine content in his 'gutka' on the average is 1.83 mg. Can this claim be accepted if a random sample of 8 'gutkas' of this type have the nicotine contents of 2.0, 1.7, 2.1, 1.9, 2.2, 2.1, 2.0, 1.6 mg?.

Sol:

Here $\bar{X} = (2.0+1.7+2.1+1.9+2.2+2.1+2.0+1.6)/8$

$= 15.6/8 = 1.95 \quad (\because n=8)$

$s^2 = \left[(2.0-1.95)^2 + (1.7-1.95)^2 + \cdots + (1.6-1.95)^2\right]/8-1$

$\Rightarrow s^2 = \sqrt{\frac{0.3}{7}} \Rightarrow s = 0.20702$

Null Hypothesis H_0: $\mu = 1.83$

Alternative Hypothesis H_1: $\mu > 1.83$

Test Statistic :- $t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$

Level of significance: $\alpha = 0.05$ (assumed)

$\therefore t_{cal} = \frac{1.95 - 1.83}{\frac{0.20702}{\sqrt{8}}} = 1.6395$

But t_{table} (at $\alpha = 0.05$) with $8-1=7$ d.o.f is 1.895

Conclusion: Since $|t_{cal}| < |t_{table}|$, Hence Accept H_0.

(i.e) Yes, the producer's claim can be accepted with 95% Confidence.

4. In 1950 in India the mean life expectancy was 50 years, if the life expectancies from a random sample of 11 persons are, 58.2, 56.6, 54.2, 50.4, 44.2, 61.9, 57.5, 53.4, 49.7, 55.4, 57.0 does it confirm the expected view?

(Hint: $H_0: \mu = 50$, $H_1: \mu \neq 50$, $\bar{x} = (58.2+56.6+\cdots+57.0)/11$

$= 54.41$

$s = 4.859$, $t_{cal} = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{-(50 - 54.41)}{4.859/\sqrt{11}} = 3.01$

$t_{table} = 2.228 = t_{0.025}$ with $11-1=10$ d.o.f

Accept H_1)

5. A random sample of 10 boys had the following I.Q's 70, 120, 110, 101, 88, 83, 95, 98, 107, 100. Do these data support the assumption of a population mean I.Q of 100? Find a reasonable range in which most of the mean I.Q values of samples of 10 boys lie?

Sol: Here $\bar{x} = (70 + 120 + \cdots + 107 + 100)/_{10} = 97.2$

$$s^2 = \left[(70-97.2)^2 + (120-97.2)^2 + \cdots + (100-97.2)^2\right]/(10-1)$$

$$\left(\because s^2 = \frac{1}{n-1}\sum_1^{'}(x_i-\bar{x})^2\right)$$

$$= \frac{1833.6}{9} = 203.73$$

$$\Rightarrow s = \sqrt{203.73} = 14.27.$$

Null Hypothesis H_0: $\mu = 100$

Alternative Hypothesis H_1: $\mu \neq 100$

L.O.S : $\alpha = 0.05$ (assume)

Test Statistic: $t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$

$$\Rightarrow t_{cal} = \frac{97.2-100}{14.27/\sqrt{10}} = -0.6204$$

$$\therefore |t_{cal}| = 0.6204$$

The table value of 't' at $\alpha = 0.05$ with $10-1 = 9$ d.o.f is

$$|t_{table}| = 2.26$$

Conclusion: Since $|t_{cal}| < |t_{table}|$, Accept H_0.

Also the 95% confidence limits are given by

$$\Rightarrow \bar{x} \pm t_{0.05}\frac{s}{\sqrt{n}}$$

$$= 97.2 \pm 2.26 \times \frac{14.27}{\sqrt{10}}$$

$$= 97.2 \pm 2.26 \times 4.514 = 97.2 \pm 10.2$$

The interval is $\Rightarrow (97.2 - 10.2,\ 97.2 + 10.2)$

$$= (87,\ 107.4).$$

5. A random sample of 10 boys had the following I.Q's 70, 120, 110, 101, 88, 83, 95, 98, 107, 100. Do these data support the assumption of a population mean I.Q of 100? Find a reasonable range in which most of the mean I.Q values of samples of 10 boys lie?

Sol: Here $\bar{x} = (70+120+ \cdots +107+100)/_{10} = 97.2$

$$s^2 = \left[(70-97.2)^2+(120-97.2)^2+ \cdots + (100-97.2)^2\right]/(10-1)$$

$$\left(\because s^2 = \frac{1}{n-1}\sum_1^n (x_i-\bar{x})^2\right)$$

$$= \frac{1833.6}{9} = 203.73$$

$$\Rightarrow s = \sqrt{203.73} = 14.27.$$

Null Hypothesis H_0: $\mu = 100$

Alternative Hypothesis H_1: $\mu \neq 100$

L.O.S : $\alpha = 0.05$ (assume)

Test Statistic: $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$

$$\Rightarrow t_{cal} = \frac{97.2-100}{14.27/\sqrt{10}} = -0.6204$$

$$\therefore |t_{cal}| = 0.6204$$

The table value of 't' at $\alpha=0.05$ with $10-1=9$ d.o.f is

$$|t_{table}| = 2.26$$

Conclusion: Since $|t_{cal}| < |t_{table}|$, Accept H_0.

Also the 95% confidence limits are given by

$$\Rightarrow \bar{x} \pm t_{0.05}\frac{s}{\sqrt{n}}$$

$$= 97.2 \pm 2.26 \times \frac{14.27}{\sqrt{10}}$$

$$= 97.2 \pm 2.26 \times 4.514 = 97.2 \pm 10.2$$

The interval is $\Rightarrow (97.2-10.2,\ 97.2+10.2)$

$$= (87,\ 107.4).$$

Student's 't' test for difference of means:

To test the significant difference b/w two means $\bar{x}_1$ & $\bar{x}_2$ of samples of sizes n_1 & n_2, we use statistic,

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \quad , \text{ where } \quad s^2 = \frac{\sum_1 (x_i-\bar{x})^2 + \sum_1 (y_i-\bar{y})^2}{n_1+n_2-2}$$

(or)

$$s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

Note: Here the table values are seen at (n_1+n_2-2) d.o.f.

Problems:

1. The means of two random samples of sizes 9 and 7 are 196.47 and 198.82 respectively. The sum of the squares of the deviations from the means are 26.94 & 18.73 respectively. Can the sample be considered to have been drawn from the same normal population.

Sol: Given $n_1 = 9$, $n_2 = 7$

$\bar{x}_1 = 196.42$, $\bar{y} = 198.82$

$\sum_1 (x-\bar{x})^2 = 26.94$, $\sum_1 (y_i-\bar{y})^2 = 18.73$

$$\therefore \quad s^2 = \frac{\sum_1 (x-\bar{x})^2 + \sum_1 (y_i-\bar{y})^2}{n_1+n_2-2} = \frac{26.94+18.73}{9+7-2}$$

$$= 3.26$$

$$\therefore \quad s = \sqrt{3.26} = 1.81$$

Null Hypothesis H_0: $\mu_1 = \mu_2$

Alternative Hypothesis H_1: $\mu_1 \neq \mu_2$

5. A random sample of 10 boys had the following I.Q's 70, 120, 110, 101, 88, 83, 95, 98, 107, 100. Do these data support the assumption of a population mean I.Q of 100? Find a reasonable range in which most of the mean I.Q values of samples of 10 boys lie?

Sol: Here $\bar{x} = (70 + 120 + \cdots + 107 + 100)/_{10} = 97.2$

$s^2 = \left[(70-97.2)^2 + (120-97.2)^2 + \cdots + (100-97.2)^2\right] / (10-1)$

$\left(\because s^2 = \frac{1}{n-1} \sum_i (x_i - \bar{x})^2\right)$

$= \frac{1833.6}{9} = 203.73$

$\Rightarrow s = \sqrt{203.73} = 14.27.$

Null Hypothesis H_0 : $\mu = 100$

Alternative Hypothesis H_1 : $\mu \neq 100$

L.O.S : $\alpha = 0.05$ (assume)

Test Statistic : $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$

$\Rightarrow t_{cal} = \frac{97.2 - 100}{14.27/\sqrt{10}} = -0.6204$

$\therefore |t_{cal}| = 0.6204$

The table value of 't' at $\alpha = 0.05$ with $10-1 = 9$ d.o.f is

$|t_{table}| = 2.26$

Conclusion: Since $|t_{cal}| < |t_{table}|$, Accept H_0.

Also the 95% confidence limits are given by

$\Rightarrow \bar{x} \pm t_{0.05} \frac{s}{\sqrt{n}}$

$= 97.2 \pm 2.26 \times \frac{14.27}{\sqrt{10}}$

$= 97.2 \pm 2.26 \times 4.514 = 97.2 \pm 10.2$

The interval is $\Rightarrow$ $(97.2 - 10.2, 97.2 + 10.2)$

$= (87, 107.4)$.

Level of significance: $\alpha = 0.05$ (5%)

Test statistic: $t = \dfrac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \dfrac{196.42 - 198.82}{1.81 \cdot \sqrt{\frac{1}{9} + \frac{1}{7}}}$

$\implies t_{cal} = -2.63$

$\therefore |t_{cal}| = 2.36$

But table value of 't' at $\alpha = 5\%$ with $\overset{(n_1+n_2-2)}{(9+7-2)} = 14$ d.o.f

is $\implies |t_{table}| = 2.15$

Conclusion: since $|t_{cal}| > |t_{table}|$

Hence Reject the H_0.

(i.e) The two samples are not drawn from the same population.

2. Two horses A and B were tested according to the time (in seconds) to run a particular track with the following results.

Hourse A (x_i)	28	30	32	33	33	29	24
Hourse B (y_i)	29	30	30	24	27	29	

Test whether the two horses have the same running capacity.

Sol:

Given $n_1 = 7$, $n_2 = 6$

Now $\bar{x} = \frac{1}{7}(28+30+32+33+33+29+24) = 31.286$

$\bar{y} = \frac{1}{6}(29+30+30+24+27+29) = 28.16$

$$s_1^2 = \frac{1}{n_1-1}\sum (x_i - \bar{x})^2 = \frac{1}{6}\left((28-31.286)^2 + (30-31.286)^2 + \dots + (24-31.286)^2\right]$$

$$\boxed{(n_1-1)s_1^2 = 31.4358}$$

$\therefore \quad (n_2-1)s_2^2 = \sum_1 (y_i-\bar{y})^2$

$$= (29-28.16)^2 + (30-28.16)^2 + \dots\dots + (29-28.16)^2$$

$$\boxed{(n_1-1)s_2^2 = 26.8336}$$

$$\therefore \quad S^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

$$\Rightarrow \quad S^2 = \frac{(31.4358 + 26.8336)}{7+6-2} = \frac{58.2694}{11} = \underline{\underline{5.23}}$$

$$\Rightarrow \quad S = \sqrt{5.23} = 2.3.$$

Null Hypothesis H_0: $\mu_1 = \mu_2$

Alternative Hypothesis H_1: $\mu_1 \neq \mu_2$

Level of significance: $\alpha = 0.05$ (assumed)

Test Statistic: $t = \dfrac{\bar{x}-\bar{y}}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \dfrac{31.286-28.16}{(2.3)\sqrt{\frac{1}{7}+\frac{1}{6}}}$

$$= 2.443$$

$\therefore \quad |t_{cal}| = 2.443$

But the table value of t at $\alpha = 0.05$ with $7+6-2 = 11$ d.o.f

is $|t_{table}| = 2.2$

Conclusion: Since $|t_{cal}| > |t_{table}|$ ($\because 2.443 > 2.2$)

$\therefore$ Reject $H_0 \Rightarrow$ Accept H_1.

(i.e) Both horses A & B do not have the same running capacity.

3. To examine the hypothesis that the husbands are more intelligent than the wives, an investigator took a sample of 10 couples and administered them a test which measures the I.Q. The results are as follows

Husbands	117	105	97	105	123	109	86	78	103	107
Wives	106	98	87	104	116	95	90	69	108	85

Test the hypothesis with a reasonable test at $\alpha = 0.05$.

Sol: We have, $n_1 = 10$, $n_2 = 10$

$$\bar{x} = (117+105+97+\dots+107)/_{10} = \frac{1030}{10} = 103$$

$$\bar{y} = (106+98+87+\dots+85)/_{10} = \frac{958}{10} = 95.8$$

Now $\sum_1 (x_i - \bar{x})^2 = [(117-103)^2 + (105-103)^2 + \dots + (107-103)^2]$

$= 1606$

$\sum_1 (y_i - \bar{y})^2 = [(106-95.8)^2 + (98-95.8)^2 + \dots + (85-95.8)^2]$

$= 1679.6$

$$\therefore s^2 = \frac{\sum_1 (x_i - \bar{x})^2 + \sum_1 (y_i - \bar{y})^2}{n_1 + n_2 - 2} = \frac{1606 + 1679.6}{18} = 182.53$$

$\Rightarrow \quad s = \sqrt{182.53} = 13.51$

Null Hypothesis H_0: $\mu_1 = \mu_2$ (i.e. No difference in I.Q)

Alternative Hypothesis H_1: $\mu_1 > \mu_2$ (Husbands are more intelligent than wives)

L.O.S : $\alpha = 0.05$

Test Statistic : $t = \dfrac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \dfrac{103 - 95.8}{(13.51)\sqrt{\frac{1}{10} + \frac{1}{10}}} = 1.19168$

$\therefore t_{cal} = 1.19168$

But $t_{table} = 1.734$ at $\alpha = 0.05$ with 18 d.o.f

Conclusion: Since $|t_{cal}| \ngtr |t_{table}| \Rightarrow$ Accept H_0.

4. Measuring specimens of nylon yarn, taken from two machines, It was found that '8' specimens from first machine had a mean denier of 9.67 with a S.D of 1.81 while 10 specimens from second machine had a mean denier of 7.43 with a S.D of 1.48. Assuming the proportions are normal, Test the hypothesis $H_0: \mu_1 - \mu_2 = 1.5$, against $H_1: \mu_1 - \mu_2 > 1.5$ at 0.05 L.O.S.

Note: The test statistic under the N.H $H_0: \mu_1 - \mu_2 = \delta$ is

$$t = \frac{\bar{x} - \bar{y} - \delta}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Sol:

Here $n_1 = 8$, $n_2 = 10$, $\bar{x} = 9.67$, $\bar{y} = 7.43$, $S_1 = 1.81$, $S_2 = 1.48$

Now $S^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2} = \frac{8(1.81)^2 + 10(1.48)^2}{8 + 10 - 2}$

$= 3.$

$\Rightarrow S = \sqrt{3} = 1.734$.

Null Hypothesis H_0: $\mu_1 - \mu_2 = 1.5 (= \delta)$

Alternative Hypothesis H_1: $\mu_1 - \mu_2 > 1.5$

L.O.S : $\alpha = 0.05$

Test Statistic: $t = \frac{\bar{x} - \bar{y} - \delta}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{9.67 - 7.43 - 1.5}{1.734\sqrt{\frac{1}{8} + \frac{1}{10}}}$

$= 0.9$

But t_{table} at $\alpha = 0.05$ with 16 d.o.f is 2.12

Conclusion: Since $|t_{cal}| < |t_{table}|$, Accept H_0.

NOTE: Sometimes we can take $S^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}$ instead of taking $S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$.

5. To compare two kinds of bumper guards, 6 of each kind were mounte on a car and then the car was run into a concrete wall. The following are the costs of repairs.

Guard 1	:	107	148	123	165	102	119
Guard 2	:	134	115	112	151	133	129

use the 0.01 L.O.S to test whether the difference b/w two sample means is significant.

Sol: We have, $n_1 = 6$, $n_2 = 6$, $\bar{x} = (107+148+123+165+102+119)/6$

$= 127.33$

$\bar{y} = (134+115+112+151+133+129)/6$

$= 129$

Now $\sum_1 (x_i - \bar{x})^2 = (107-127.33)^2 + (148-127.33)^2 + \cdots + (119-127.33)^2$

$= 2989.34$

$\sum_1 (y_i - \bar{y})^2 = (134-129)^2 + (115-129)^2 + \cdots + (129-129)^2$

$= 1010.$

$\therefore s^2 = \dfrac{\sum_1 (x_i - \bar{x})^2 + \sum_1 (y_i - \bar{y})^2}{n_1 + n_2 - 2} = \dfrac{2989.34 + 1010}{6+6-2}$

$= 399.934$

$\Rightarrow s = \sqrt{399.934} = 19.998$

Null Hypothesis $H_0: \mu_1 = \mu_2$

Alternative Hypothesis $H_1: \mu_1 \neq \mu_2$

L.O.S : $\alpha = 0.01$

Test statistic : $t = \dfrac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

$\Rightarrow t_{cal} = \dfrac{127.33 - 129}{19.99\sqrt{\frac{1}{6} + \frac{1}{6}}} = -0.1446$

$\Rightarrow |t_{cal}| = 0.1446$

but $|t_{table}| = 3.169$ (at $\alpha = 0.01$ with 10 d.o.f)

Conclusion:

Since $|t_{cal}| < |t_{table}|$

Accept H_0.

(ii) **F- Test :** To test whether there is any significant difference between two estimates of population variance.

To test if the two samples have come from the same population, we use F- test. In this case we set up null Hypothesis $H_0: \sigma_1^2 = \sigma_2^2$ (i.e) the population variances are same

under H_0 the test statistic is,

$$F = \frac{s_1^2}{s_2^2} \quad \text{where } s_1^2 > s_2^2$$

(or)

$$F = \frac{s_2^2}{s_1^2} \quad \text{where } s_2^2 > s_1^2$$

Note: We will take greater of the variances s_1^2 (or) s_2^2 in the numeral and adjust for the degrees of freedom accordingly.

(i.e) $$F = \frac{\text{Greater Variance}}{\text{Smaller variance}}$$

Problems:

1. In one sample of 8 observations the sum of the squares of deviations of the sample values from the sample mean was 84.4 and in the other sample of 10 observations it was 102.6. Test whether this difference is significant at 5% level.

Sol: Here $n_1 = 8$, $n_2 = 10$

Also $\sum_1 (x_i - \bar{x})^2 = 84.4$ (for first sample)

$\sum_1 (y_i - \bar{y})^2 = 102.6$ (for second sample)

$$\therefore \; S_1^2 = \frac{1}{n_1 - 1} \sum_1 (x_i^2 - \bar{x})^2 = \frac{1}{8-1}(84.4) = 12.057$$

$$S_2^2 = \frac{1}{n_2 - 1} \sum_1 (y_i - \bar{y})^2 = \frac{1}{10-1}(102.6) = 11.4$$

$$\therefore \; S_1^2 > S_2^2 .$$

Null Hypothesis H_0: $\sigma_1^2 = \sigma_2^2$ (or) $S_1^2 = S_2^2$

Alternative Hypothesis H_1: $\sigma_1^2 \neq \sigma_2^2$ (or) $S_1^2 \neq S_2^2$.

L.O.S : $\alpha = 0.05$

Test Statistic : $F = \frac{S_1^2}{S_2^2}$ $(\because S_1^2 > S_2^2)$

$\therefore F_{cal} = \frac{12.057}{11.4} = 1.057$

But table value of F at 5% L.O.S with $(\underset{=9}{8-1}, \underset{=9}{10-1})$ d.o.f

is $F_{table} = 3.29$

Conclusion: Since $F_{cal} < F_{table}$, ~~Hence~~ Accept Ho.

2. The measurements of the output of two units have given the following results. Assuming that both samples have been obtained from the normal populations at 10% significant level. Test whether the two populations have the same variance?

Sol:

UNIT-A :	14.1	10.1	14.7	13.7	14.0
UNIT-B :	14.0	14.5	13.7	12.7	14.1

Sol: Given $n_1 = 5$, $n_2 = 5$

$\bar{x} = (14.1 + 10.1 + 14.7 + 13.7 + 14.0)/5 = 13.32$

$\bar{y} = (14.0 + 14.5 + 13.7 + 12.7 + 14.1)/5 = 13.8$

$S_1^2 = \frac{1}{5-1}\left((14.1-13.32)^2 + (10.1-13.32)^2 + \dots + (14-13.32)^2\right)$

$= \frac{13.468}{4} = 3.372$

$S_2^2 = \frac{1}{5-1}\left[(14-13.8)^2 + (14.5-13.8)^2 + \dots + (14.1-13.8)^2\right]$

$= \frac{1.84}{4} = 0.46$

Null Hypothesis H_0 : $s_1^2 = s_2^2$

Alternative Hypothesis H_1: $s_1^2 \neq s_2^2$

Level of Significance : $\alpha = 0.01$

Test Statistic : $F = \dfrac{s_1^2}{s_2^2}$ $(\because s_1^2 > s_2^2)$

$$\therefore F_{cal} = \frac{3.372}{0.46} = 7.33$$

But The table value of F at $\alpha = 0.01$ with (4,4) d.o.f

is $F_{table} = 6.39$

Conclusion : Since $F_{cal} > F_{table}$

$\therefore$ Reject $H_0 \Rightarrow$ Accept H_1. $\Rightarrow s_1^2 \neq s_2^2$ is True

Hence there is significant different b/w the variances

3. The time taken by workers in performing a job by method - I and method II is given below

Method - I :	20	16	26	27	23	22	—
Method - II :	27	33	42	35	32	34	38

Do the data show that the variances of time distribution from population from which these samples are drawn do not differ significantly?

Sol: Here $n_1 = 6$, $n_2 = 7$,

$s_1^2 = 16.26$, $s_2^2 = 22.29$

$$F_{cal} = \frac{s_2^2}{s_1^2} = \frac{22.29}{16.26} = 1.3699 \simeq 1.37$$

$F_{table} = 4.95$ (at $\alpha = 0.05$ with (6,5) d.o.f)

Since $F_{cal} < F_{table}$ $\therefore$ Accept H_0

3. Chi Squared (χ^2) test :-

Chi squared (χ^2) test can be used to test the

(i) Goodness of fit (ii) Independence of attributes

(i) Chi Squared (χ^2) test for Goodness of fit :

Suppose we are given a set of observed frequencies obtained under some experiment and we want to test if the experimental results support a particular hypothesis or Theory. Karl Pearson developed a test for testing the significance of discrepancy b/w experimental values and the theoretical values obtained under some theorey. This test is known as χ^2-test of goodness of fit.

The statistic is : $$\chi^2 = \frac{(O-E)^2}{E}$$

Where O - observed frequency
E - Expected frequency

χ^2- test is used to test whether difference b/w observed and expected frequencies are significant.

NOTE: (i) If the data is given in a series of 'n' numbers then degress of freedom $= n-1$

(ii) In case of Binomial distribution, d.o.f $= n-1$

(iii) In case of Poisson distribution, d.o.f $= n-2$

(iv) In case of Normal distribution, d.o.f $= n-3$.

Problems:

1. The number of automobile accidents per week in a certain communi[ty] are as follows: 12, 8, 20, 2, 14, 10, 15, 6, 9, 4. Are these frequenci[es] in agreement with the belief that accident conditions were the same during this 10 week period.

Sol:

Expected frequency of accidents each week =

$$(12+8+20+2+14+10+15+6+9+4)/_{10} = \frac{100}{10} = 10$$

$\therefore$ Now,

O :	12	8	20	2	14	10	15	6	9	4
E :	10	10	10	10	10	10	10	10	10	10
$\frac{(O-E)^2}{E}$:	0.4	0.4	10	6.4	1.6	0	2.5	1.6	0.1	3.6

$$\therefore \sum \frac{(O-E)^2}{E} = 26.6$$

where O – observed

E – Expected

Null Hypothesis H_0: The accidents conditions were the same during the 10 week period

Here $\chi^2 = \sum \frac{(O-E)^2}{E} = 26.6$

But χ^2_{table} (at $\alpha = 0.05$ with $10-1=9$ d.of)

is 16.9

Since $\chi^2_{cal} > \chi^2_{table}$

Hence Reject the H_0.

(i.e) The accident conditions were not the same during the 10 week period.

2. A Sample analysis of examination results of 500 Students was made. It was found that 220 Students had failed. 170 had Secured a third class, 90 were placed in second class. and 20 got a first class. Do these figures commensurate with the general examination result which is in the ratio of 4:3:2:1 for the various categories respectively.

Sol: <u>Null Hypothesis H_0</u>: The observed results commensurate with the general examination results.

Expected frequencies are in the ratio of 4:3:2:1

Total frequency = 500

If we divide the total frequency 500 in the ratio of 4:3:2:1 we get the expected frequencies as 200, 150, 100, 50.

CLASS	OBSERVED FREQUENCY(O)	EXPECTED FREQUENCY(E)	$\frac{(O-E)^2}{E}$
Failed	220	200	2.00
Third	170	150	2.667
Second	90	100	1.000
First	20	50	18.000
			23.667

$$\therefore \chi^2_{cal} = \sum \frac{(O-E)^2}{E} = 23.667$$

But $\chi^2_{table} = 7.81$ at $\alpha = 0.05$ with $4-1=3$ d.o.f

<u>Conclusion</u>: Since $\chi^2_{cal} > \chi^2_{table}$ ($\because 23.667 > 7.81$)

H_0 is Rejected.

(i.e) The observed results are not Commensurate with the general examination results.

3. A die thrown 264 times with the following results. Show that the die is biased [Given $\chi^2_{0.05} = 11.07$ d.o.f]

No. appeared on the die :	1	2	3	4	5	6
Frequency :	40	32	28	58	54	60

Sol:

Null Hypothesis H_0: The die is unbiased

The Expected frequency of each of the numbers

1, 2, 3, 4, 5, 6 is $\frac{264}{6} = 44$.

O :	40	32	28	58	54	52
E :	44	44	44	44	44	44
$\frac{(O-E)^2}{E}$:	0.3636	3.2727	5.8181	4.4545	2.2727	1.4545

$$\therefore \chi^2_{cal} = \sum_1^1 \frac{(O-E)^2}{E} = 17.6362$$

But $\chi^2_{table} = 11.07$ (at $\alpha = 0.05$ with $6-1=5$ d.o.f)

Conclusion : Since $\chi^2_{cal} > \chi^2_{table}$

Hence Reject H_0

(i.e) The die is not unbiased.

5. By fitting the normal distribution, Test the goodness of fit for the following data:

x_i:	50	55	60	65	70	75	80	85	90	95	100
f_i:	2	3	5	9	10	12	7	2	3	1	0

Sol: According to Fitting of Normal Distribution, the expected frequencies are,

$E(X) = 1, 3, 6, 9, 11, 10, 7, 4, 2, 1, 0$

N.H H_0: Fitted Expected frequencies are best fitted one's

Here

O:	2	3	5	9	10	12	7	2	3	1	0
E:	1	3	6	9	11	10	7	4	2	1	0

$\frac{(O-E)^2}{E} = 1 \quad 0 \quad 0.166 \quad 0 \quad 0.0909 \quad 0.4 \quad 0 \quad 1 \quad 0.5 \quad 0 \quad 0$

$\therefore \chi^2_{cal} = \sum_1^{11} \frac{(O-E)^2}{E} = 3.1569$

But At $\alpha = 0.05$, with $11-1=10$ d.o.f

$\chi^2_{table} = 18.307$

Since $\chi^2_{cal} < \chi^2_{table}$

Accept H_0

4. By fitting poisson distribution test the goodness of fit for the following data?

x :	0	1	2	3	4
$f(x)$:	109	65	22	3	1

Sol:

According to <u>fitting of poisson distribution</u> the Expected frequencies are : 109, 66, 20, 4, 1 (Already did at poisson distribution

N.H H_0: Fitted expected frequencies are good.

$\therefore$

O :	109	65	22	3	1
E :	109	66	20	4	1
$\frac{(O-E)^2}{E}$:	0	0.0091	0.2	0.25	0

$$\therefore \quad \Sigma_1^1 \frac{(O-E)^2}{E} = 0.4591 = \chi^2_{cal}$$

But At $\alpha = 0.05$, with $5-1=4$ d.o.f

$$\chi^2_{table} = 9.488$$

Conclusion: Since $\chi^2_{cal} < \chi^2_{table}$

Accept H_0