

Sampling Theory

Population OR Universe : →

Population is a group of items, units, or subjects which is under study.

Population is divided into four categories -

(i) Finite Population →

A Universe containing a finite number of members is called finite universe. For example, The Universe of the weight of students in particular class.

(ii) Infinite Population →

A universe with infinite number of members is known as infinite universe. For example, The Universe of pressures at various points in the atmosphere.

(iii) Real Population

(iv) Hypothetical Population →

The collection of all possible ways in which a specified event can happen is called a hypothetical Universe. e.g. The universe of heads and tails obtained by tossing a coin an infinite number of times is a hypothetical one.

Sample → A finite subset of a population or universe is called a sample. A sample is thus a small portion of the universe. The number of individuals in a sample is called sample size. The process of selecting a sample from a universe is called Sampling.

Note : → A sample should be chosen such that it is a true representative of the population.

Large Sample → A sample consists of more than 30 items is called Large Sample. Z-test is used for test of significance of Large Samples. $n \geq 30$

Small Samples → A sample consisting up to 30 items is called small samples. Small samples tests are, t-test, F test etc. ~~chi-square (χ^2) test.~~

$$n \leq 30$$

* Parameters And Statistics : →

The statistical constants of the ~~populati~~ population such as mean (μ), standard deviation (σ) are called parameters. The mean (\bar{x}), standard deviation s of a sample are known as statistic.

Population	Sample
Population size = N	Sample size = n
Population mean = μ	Sample mean = \bar{x}
Population standard deviation = σ	Sample standard deviation = s
Population proportion = p	Sample proportion = \hat{p}

Hypothesis : →

A Hypothesis is a statement about the population parameter. In other words, a hypothesis is a conclusion which is tentatively drawn on a logical basis.

A Statistical Test of Hypothesis consists of five parts :-

- ① The Null Hypothesis, denoted by H_0
- ② The alternative hypothesis denoted by H_a
- ③ The test statistic and its value
- ④ The rejection region
- ⑤ The Conclusion.

Note :- The Alternative Hypothesis $\neq H_a$ is contradiction of null hypothesis.

Null Hypothesis →

For applying the test of significance, we first set up a hypothesis which is a definite statement about the parameter called Null Hypothesis. It is denoted by H_0 .

Alternative Hypothesis →

Any hypothesis which is complementary to the null hypothesis is called an alternative hypothesis. It is denoted by H_a .

Large Sample Test OR Z-Test

① Testing of Significance for single proportion

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} \quad \text{or} \quad Z = \frac{x - np}{\sqrt{npq}} \quad p = \frac{x}{n}$$

- Note: → ① set the Null Hypothesis H_0
 ② set the Alternative hypothesis H_a

③ Test Statistic $Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$

P] Population
 Q]
 p - sample

④ Conclusion: —

Note-1 We know that $|Z| = 1.96$ at 5% level of significance

$|Z| = 2.58$ at 1% level of significance

In some dice throwing experiment, we do not throw dice 49152 times and of these 25145 yields 4, 5, 6, is this consistent with the hypothesis that the

dice were unbiased.

H_0 :- dice are unbiased.
 $P = \text{prob. of } 4, 5, 6 \text{ in a throw} = \frac{3}{6} = \frac{1}{2}$

$$q = 1 - p = \frac{1}{2}, \quad n = 49152, \quad x = 25145, \quad np = 24576$$

$$p = \frac{25145}{49152} = 0.511 \quad \sqrt{npq} = \sqrt{12288} = 110.88$$

$$Z = \frac{x - np}{\sqrt{npq}} = \frac{0.511 - 0.5}{110.88} = \frac{0.011}{110.88} = \frac{569}{110.88} = \approx 5.13$$

$|Z| = 1.96$ at 5% of level of signi.

$|Z| > 1.96$ H_0 is rejected.

Q.1 A manufacturer claims that only 4% of his products supplied by him are defective. A random sample of 600 products contains 36 defectives. Test the claim of the manufacturer at 5% level of significance.

Soln Here sample $n = 600$ i.e. $n > 30$ so we use Large sample test

$$n = 600$$

p = proportion of defective in the population

$$= 4\% = \frac{4}{100} = .04$$

$$Q = 1 - P = 1 - .04 = .96$$

\hat{p} = proportion of defective in the sample

$$\hat{p} = \frac{36}{600} = .06$$

Null Hypothesis (H_0): $P = .04$ is true i.e. the claim of manufacturer is accepted.

Alternative Hypothesis (H_a): $P \neq .04$

Test Statistic $Z = \left| \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} \right| = \frac{|.06 - .04|}{\sqrt{\frac{(.04)(.96)}{600}}} = 2.5$

Here $|Z| = 2.5$ ~~< 2.5~~

$$|Z| = 2.5 > 1.96$$

So Null Hypothesis is not accepted. Hence the claim of manufacturer is not accepted.

Q.2 A coin was tossed 400 times and the head turned up 216 times. Test the hypothesis that the coin is unbiased.

Soln $n = 400$, $P = \frac{1}{2}$, $Q = \frac{1}{2}$, $\hat{p} = \frac{216}{400} = .54$

$$\therefore Z = \left| \frac{\hat{p} - P}{\sqrt{\frac{pq}{n}}} \right| = 1.6, |Z| < 1.96, \text{ at } 5\% \text{ level of significance,}$$

so Hypothesis is accepted

Hence coin is unbiased.

A

$$0.025$$

A bag contains defective items also, the exact number being not known. A large sample 100 items from the bag has 10% defective items find 95% confidence limit for the proportion.

$$\text{Q: } p = \frac{10}{100} = 0.1 \quad Q = 1 - p = 0.9$$

$$\begin{aligned}\text{confidence limit will be } &= p \pm 2.05 \sqrt{\frac{pq}{n}} \\ &= 0.1 \pm 1.96 \sqrt{\frac{0.1 \times 0.9}{100}} \\ &= 0.1 \pm 1.96 (0.03) \\ &= 0.01588, 0.0412.\end{aligned}$$

Limit for proportion

p → population proportion is given for success. limit are. [also said to be 3 σ limit]

$$p - 3 \sqrt{\frac{pq}{n}} < p < p + 3 \sqrt{\frac{pq}{n}}$$

limit of proportion on the basis of sample are

given by $p \pm 3 \sqrt{\frac{pq}{n}}$ [also said to be 3 σ limit]

at confidence limit or at level of significance for population proportion.

$$p \pm 2.05 \sqrt{\frac{pq}{n}} \quad \left[\text{at } 5\% \text{ of level of significance or } 95\% \text{ of confidence.} \right]$$

limit for the proportion (on the basis of sample
are taken as $p \pm 3\sqrt{\frac{pq}{n}}$ → ①
also said to be 3σ limit.)

If A die is thrown 9000 times and a throw of 3 or 4 observed 3240 times. Show that the die can not be regarded as an unbiased one and find the limits between which probability of throw of 3 or 4 lies.

Given $n = 9000$, $x = \text{number of success} = 3240$

H0 die is unbiased.

$p = \text{probability of getting 3 or 4} = \frac{2}{6} = \frac{1}{3}$, $q = \frac{2}{3}$

$$Z = \frac{x - np}{\sqrt{npq}} = \frac{3240 - 9000 \times \frac{1}{3}}{\sqrt{9000 \times \frac{1}{3} \times \frac{2}{3}}} = \frac{-240}{\sqrt{2000}} = -5.36$$

$|Z| > 3$ H_0 is rejected. Hence coin is biased.

since p die is biased $p \neq \frac{1}{3}$ limit are given by

$$p \pm 3\sqrt{\frac{pq}{n}} \quad p = \frac{3240}{9000} = .36 \quad q = .64$$

$$= .36 \pm \sqrt{\frac{.36 \times .64}{9000}} = .36 \pm 0.15 \\ = .375 \text{ and } .345$$

Remark the probable limits for the observed proportion of success are given by

$$p - 3\sqrt{\frac{pq}{n}} < p < p + 3\sqrt{\frac{pq}{n}}$$

If p is not known then we will use ① formula for limit for the proportion.

Testing of Significance for Difference of Proportions

Q.2

Consider two samples X_1 and X_2 of sizes n_1 and n_2 respectively taken from two different populations. To test the significance of the difference between the sample proportions p_1 and p_2 , the test statistic is

$$Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}, \text{ where } P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2}$$

where H_0 : there is no significant difference b/w sample proportion.

Q.1 Before an increase in excise duty on tea, 800 people out of a sample of 1000 persons were found to be tea drinkers. After an increase in the duty, 800 persons were known to be tea drinkers in a sample of 1200 people. Do you think that there has been a significant decrease in the consumption of tea after an increase in the excise duty?

Sol'

$$n_1 = 1000, n_2 = 1200$$

$$p_1 = \frac{x_1}{n_1} = \frac{800}{1000} = \frac{4}{5}, p_2 = \frac{x_2}{n_2} = \frac{800}{1200} = \frac{2}{3}$$

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{1000\left(\frac{800}{1000}\right) + 1200\left(\frac{800}{1200}\right)}{1000 + 1200} = \frac{8}{11}$$

$$\therefore Q = 1 - P = 1 - \frac{8}{11} = \frac{3}{11}$$

Null Hypothesis (H_0): $p_1 = p_2$ i.e. there is no significant difference in the consumption of tea before and after increase of excise duty.

$$Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{\frac{4}{5} - \frac{2}{3}}{\sqrt{\frac{8}{11} \times \frac{3}{11} \left(\frac{1}{1000} + \frac{1}{1200}\right)}} = 6.842$$

$|Z| > 1.96$, and also $|Z| > 2.58$ both the significant value at 5% and 1% level of significance. So Null Hypothesis is rejected. i.e. there is a significant decrease in the consumption of tea due to increase in excise duty.

Q.2. A machine produced 16 defective articles in a batch of 500.
 After overhauling it produced 3 defectives in a batch of 100.
 Has the machine improved?

$$\text{Sol}^b \quad p_1 = \frac{16}{500} \quad \because n_1 = 500, \quad n_2 = 100$$

$$p_2 = \frac{3}{100}$$

No

Null Hypothesis (H_0): Machine has not improved due to overhauling
 i.e. $p_1 = p_2$

$$\therefore Z = \frac{p_1 - p_2}{\sqrt{PQ \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad \text{Here } P = \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2} = .032$$

$$Q = 1 - P$$

$$Z = .104$$

$|Z| < 1.96$ at 5% level of significance.

Hence Null Hypothesis H_0 accepted

i.e. The machine has not improved due to overhauling.

Q.3. 500 articles from a factory are examined and found to be 2% defective. 800 similar articles from a second factory are found to have only 1.5% defective. Can it reasonably be concluded that the products of the first factory are inferior to those of second?

$$\text{Sol}^c \quad \text{Here } n_1 = 500, \quad n_2 = 800$$

$p_1 = \text{proportion of defective from first factory} = .02$

$p_2 = \text{proportion of defective from second factory} = 1.5\% = .015$

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = .01692, \quad Q = 1 - P = .9830$$

Null Hypothesis (H_0): There is no significant difference between the two products. i.e. the products do not differ in quality

$$\therefore Z = \frac{p_1 - p_2}{\sqrt{PQ \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = .68 \quad \text{i.e. } |Z| < 1.96, \text{ the significant value of } Z \text{ at } 5\% \text{ level of significance.}$$

Hence Null Hypothesis is accepted i.e. products do not differ in quality.

(3) Testing of Significance for Single Mean : →

$$z = \frac{\text{Difference of mean}}{\text{standard error}} = \frac{\bar{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}$$

H_0 : there is no significant diff. b/w sample mean and population mean

- Q.1. The average marks in mathematics of a sample of 100 students was 51 with a s.d. of 6 marks. Could this have been a random sample from a population with average marks 50?

Solⁿ: Here $n = 100$, $\bar{x} = 51$, $\mu = 50$, $\sigma = 6$.

Null Hypothesis (H_0): $\mu = 50$

$$\therefore z = \frac{\text{Diff. of mean}}{\left(\frac{\sigma}{\sqrt{n}}\right)} = \frac{1}{\frac{6}{\sqrt{100}}} = 1.66 < 1.96 \text{ is the}$$

significant value of z at 5% level of significance.

Hence Null Hypothesis is accepted. i.e. the ~~sample~~ sample is drawn from the population with mean 50.

- Q.2. It is claimed that a random sample of 100 tires with a mean life of 15269 km is drawn from a population of tires which has a mean life of 15200 km and a s.d. of 1248 km. Test the validity of this claim.

Solⁿ: Null Hypothesis (H_0): Let us consider the null hypothesis that the sample of 100 tires has come from a universe with mean 15200 km and s.d 1248 km

$$z = \frac{\text{Diff. of mean}}{\left(\frac{\sigma}{\sqrt{n}}\right)} = \frac{15269 - 15200}{\frac{1248}{\sqrt{100}}} = .55 < 1.96$$

at 5% level of significance. Hence Null Hypothesis accepted.

Hence the claim is justified.

Q.3. A sample of 1000 students from Delhi University was taken and average weight was found to be 50.80 kg with S.D of 9.07 kg. Could the mean weight of students in the population be 54.43 kg.

Sol Null Hypothesis (H_0): Mean weight of student in the population be 54.43 kg

$$Z = \frac{\text{Mean of Difference}}{\left(\frac{\sigma}{\sqrt{n}}\right)} = \frac{|50.80 - 54.43|}{\frac{9.07}{\sqrt{1000}}} = 3.9168$$

Here $|Z| > 1.96$ at 5% level of significance.

So Null Hypothesis is rejected hence mean weight of student in the population can not be 54.43 kg.

Q4) A sample of 1000 members is found to have a mean of 3.42 cm. Could it be reasonably regarded as a simple sample from large population whose mean is 3.30 cm and standard deviation is 2.6?

Ans we have $\bar{x} = 3.42$, $n = 1000$, $\mu = 3.30$, $\sigma = 2.6$
Ho i.e suppose the sample is drawn from the population with mean $\mu = 3.30$ and $\sigma = 2.6$

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{(3.42 - 3.30)}{2.6 / \sqrt{1000}} = \frac{0.12}{0.082} = 1.46$$

$|Z| < 1.96$ at 5% level of significance.

H_0 is accepted

Test of Significance for Difference of Means of Two Samples

$$Z = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{\checkmark \text{Diff. of two mean}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

where \bar{x}_1, \bar{x}_2 be the means of two independent sample sizes n_1, n_2 (both large) and so σ_1 and σ_2 respectively

Q.1. A random sample of 150 villages was found to be 440 and the average populations per village was 32, and the S.D 32. Another random sample of 250 villages from the same district gave an average population 480 per village with S.D of 56. Is the difference between the averages of two samples statistically significant?

Sol's Here $n_1 = 150, \bar{x}_1 = 440, s_1 = 32, n_2 = 250, \bar{x}_2 = 480, s_2 = 56$

Here population S.D σ_1 and σ_2 are not given, so we use s_1 in place of σ_1 and s_2 in place of σ_2 $\underline{H_0}$ there is no significance diff. b/w mean (Average).

Thus

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\cancel{H_0})}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{440 - 480}{\sqrt{\frac{32^2}{150} + \frac{56^2}{250}}} = 9.09$$

$Z = 9.09 > 1.96$ at 5% level of significance

So difference between sample means

Hence Null Hypothesis rejected.

Remark H_0 :- there is no diff. b/w the population mean.

If the mean of two single large sample of 1000 and 2000 members are 67.5 inches and 68.0 inches respectively can the sample be regarded as drawn from the same population of standard deviation 2.5 inches?

$$\text{Ans} \quad n_1 = 1000, n_2 = 2000, \bar{x}_1 = 67.5, \bar{x}_2 = 68 \\ \sigma_1 = \sigma_2 = 2.5 = \sigma$$

H₀ null hypothesis the samples have been drawn from the same population of std. Deviation is 2.5 inches.

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{67.5 - 68.0}{2.5 \sqrt{\frac{1}{1000} + \frac{1}{2000}}} = -5.1$$

|Z| = 5.1 > 1.96 at the level of 5% level of Significance

H₀ is rejected and conclude that the samples are not from the same population with S.D 2.5

If mean and S.D calculated from the weights in kg of students of two groups taken from two universities are given below

	mean	S.D	sample size
University A	55	10	400
University B	57	15	100

Test the significance of the difference between the mean?

H₀ there is no significance diff. between the means

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

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{55 - 57}{\sqrt{\frac{10^2}{400} + \frac{(15)^2}{100}}} = -1.2648$$

|Z| = 1.26 < 1.96 H₀ is accepted there is no significance diff. between the means

Test for significance for single standard deviation

Let σ be S.D. of sample size 'n' drawn from a population with S.D. σ

Test statistic

$$Z = \frac{\bar{x} - \sigma}{\sigma / \sqrt{2n}}$$

Test of significance for the difference of S.D.

Let σ_1, σ_2 be the S.D. of two independent sample sizes n_1, n_2 drawn from populations with S.D. of population σ_1 and σ_2 respectively

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

If σ_1, σ_2 are not given
 $\sigma_1 \sim \bar{x}_1, \sigma_2 \sim \bar{x}_2$ for large sample

H_0 : S.D. do not differ significantly
Q: In a certain sample random sample of 72 items the S.D. is found to be 3.2". Is it reasonable to suppose that it is from a parent population whose S.D. is 3.1"? ?

Ans H_0 sample is from parent population with S.D. 3.1"

$$\bar{x} = 3.2, \sigma = 3.1, n = 72$$

$$Z = \frac{\bar{x} - \sigma}{\sigma / \sqrt{2n}} = \frac{3.2 - 3.1}{3.1 / \sqrt{144}} = 0.4$$

$$|Z| < 1.96$$

H_0 is accepted.

A random sample drawn from two countries gave (3) the following data relating to heights of adult males

mean height in inches

S.D. in inches.

number of sample

country A

67.42

country B

67.25

2.58

2.50

1000

1200

(i) Is the difference b/w the mean significant at 5% level of significance?

(ii) Is the difference b/w the S.D. significant.

$$\text{By (i)} \quad n_1 = 1000, \quad n_2 = 1200, \quad \bar{x}_1 = 67.42 \quad \bar{x}_2 = 67.25 \\ s_1 = 2.58, \quad s_2 = 2.50$$

Since sample sizes are large we can take H_0 there is no significant difference b/w mean of samples.

$$s_1 = s_2 = 2.58, \quad s_2 = s_2 = 2.50 \quad H_0 \text{ there is no diff. b/w mean of samples.}$$

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{67.42 - 67.25}{\sqrt{\frac{(2.58)^2}{1000} + \frac{(2.50)^2}{1200}}} = 1.56$$

$|Z| < 1.96 \therefore H_0$ is accepted at 5% level of significance

(iii) H_0 there is no significant difference b/w

$$\text{sample S.D.} \\ Z = \frac{s_1 - s_2}{\sqrt{\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}}} = \frac{2.58 - 2.50}{\sqrt{\frac{(2.58)^2}{2 \times 1000} + \frac{(2.50)^2}{2 \times 1200}}} = 0.077$$

$Z = 0.077 < 1.96$ at Hence H_0 is accepted at 5% level of significance.

A the mean yields of two set of plot and variability are as given
set of 40 plots set of 60 plots

mean yield per plot 1258

1243

S.D. / plot 34

28

whether the difference in the variability in yield significant

$$\frac{80}{\underline{80}}^h \quad n_1 = 40, \quad n_2 = 60, \quad \bar{x}_1 = 1258, \quad \bar{x}_2 = 1243 \\ \sigma_1 = 34, \quad \sigma_2 = 28$$

H_0 there is no significance difference in variability in yields

$$Z = \frac{\sigma_1 - \sigma_2}{\sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}} = \frac{34 - 28}{\sqrt{\frac{(34)^2}{80} + \frac{(28)^2}{120}}} = 1.3$$

$$|Z| = 1.3 < 1.96$$

Hence H_0 is true at the level of significance.

Student's t-test : →

The statistic t was introduced by W.S. Gosset in 1908. The quantity t is defined as

$$t = \frac{\bar{x} - \mu}{(\sigma/\sqrt{n})} = \frac{\text{Difference in Two Values}}{\text{Standard error}}$$

where n = number of observations in the sample

\bar{x} = $\frac{1}{n} \sum_{i=1}^n x_i$ is the sample mean

μ = The mean of population from which the sample has been drawn.

σ = $\sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$ is the standard deviation of the sample.

Application of the t-distribution →

(i) To test the significance of the mean of the sample.
(ii) To test the difference between two means or to compare

(iii) Two Sample.
Paired-t-test.

(iv) To test the sample coefficient of correlation.
To test the sample coefficient of regression.

W.B.
Yellamby

Testing the significance of Sample mean. \rightarrow

There is no significant difference between the sample mean and population mean or the sample which has been drawn from the population whose mean is μ_0 . Thus.

Null hypothesis $H_0 = \mu = \mu_0$

Alternative hypothesis $H_a = \mu \neq \mu_0$

Test statistic $t = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = t_{cal}$

t -tabulated: Generally at 1%, 5%, and 10% conclusion, tabulated values are taken from the table at $(n-1)$ degrees of freedom at the given level of significance.

Conclusion: If $|t_{cal}| \leq |t_{tab}|$, then ^{null} Hypothesis is accepted.

Q1 → A manufacturer of dry cells claimed that the life of cells is 24 hours. The mean life of a sample of 10 bad cells is 22.5 hours with a standard deviation of 3.0 hours. On the basis of available information, test whether the claim of manufacturer is correct at 5% level of significance.

Sol' Here. $\mu = 24$ hours, $\bar{x} = 22.5$, $s = 3.0$ hours.

Null hypothesis (H_0): The claim of manufacturer is correct i.e $\mu = 24$ hours.

Alternative Hypothesis (H_a): $\mu \neq 24$ hours.

Test statistic,

$$t = \frac{|\bar{x} - \mu|}{s/\sqrt{n}} = \frac{22.5 - 24}{\left(\frac{3}{\sqrt{10}}\right)} = 1.58$$

t -tabulated at 5% level of significance with $(10-1) = 9$ degrees of freedom = 2.26

Conclusion: Here $|t_{cal}| < |t_{tab}|$,

Null hypothesis can be accepted at 5% level of significance.

i.e manufacturer's claim is correct.

Q.2 - A machine is designed to produce insulating washers for electrical devices of an average thickness of .025 cm. A random sample of 10 washers was found to have an average thickness of .024 cm with a standard deviation of .002 cm. Test the significance of deviation. Value of t for 9 degree of freedom at 5% level is 2.262.

Sol' Here $\mu = .025$, $\bar{x} = .024$, $s = .002$

Null Hypothesis (H_0): $\mu = .025$, i.e The difference between \bar{x} and μ is not significant
Alt... i.e. Null hypothesis (H_a) " $\mu \neq .025$

$$\therefore \text{Test statistic } t = \frac{|\bar{x} - \mu|}{s/\sqrt{n}} = \frac{|1.024 - 1.025|}{(1.002)} \times \sqrt{10} \\ = 1.58$$

t -tabulated at 5% level of significance with $(10-1) = 9$ degrees of freedom = 2.262

Conclusion:- Now $|t_{\text{cal}}| < |t_{\text{tab}}|$

Therefore Null Hypothesis can be accepted at 5% level of significance, i.e. the product of machine (Washer) is passed. Hence Null Hypothesis is correct.

Q.3. In a random sampling, 10 individuals were chosen from a population and their heights in inches are found to be 63, 63, 64, 65, 66, 69, 69, 70, 70, 71. Discuss the suggestion that the mean height in the Universe is 65 inches, given that, for 9 degrees of freedom the value of student's t at 5% level of significance is 2.62

Sol^b. We know that the test-statistic is $t = \frac{|\bar{x} - \mu|}{s/\sqrt{n}}$ → ①

Now Sample mean $\bar{x} = \frac{\sum x}{n} = \frac{670}{10} = 67$, $\mu = 65$

Sample standard deviation $s = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}}$

x	$x - \bar{x}$	$(x - \bar{x})^2$
63	$63 - 67 = -4$	16
63	-4	16
64	-3	9
65	-2	4
66	-1	1
69	2	4
69	2	4
70	3	9
70	3	9
71	4	16

$$\therefore s = \sqrt{\frac{88}{9}} = 3.126$$

$$\therefore t = \frac{(\bar{M} - \mu)}{\sigma/\sqrt{n}} = \frac{(65 - 67)}{3.126} \times \sqrt{10} = 2.02$$

$$\therefore t_{\text{cal}} = 2.02$$

$$\text{Given } t_{\text{tab}} = 2.262$$

i.e. $|t_{\text{cal}}| < |t_{\text{tab}}|$ at 5% level of significance.

\Rightarrow This error could have arisen due to fluctuations and we may conclude that the data are consistent with the assumption of mean height in the universe of 65 inches. \therefore

Q.4 A drug was administered to 10 patients and the increments in their blood pressure were recorded to be 6, 3, -2, 4, -3, 4, 6, 0, 0, 2. Is it reasonable to believe that the drug has no effect on the change of blood pressure? Use 5% significance level and assume that for 9 degrees of freedom $t_{0.5, 9} = 2.26$.

Sol:

Let x denote the increment of Blood pressure.

Now $\bar{x} = \frac{6+3-2+4-3+4+1+2}{10} = 2$

x	$x - \bar{x}$	$(x - \bar{x})^2$
6	$6-2=4$	16
3	$3-2=1$	1
-2	$-2-2=-4$	16
4	$4-2=2$	4
-3	$-3-2=-5$	25
4	$4-2=2$	4
6	$6-2=4$	16
0	$0-2=-2$	4
0	$0-2=-2$	4
2	$2-2=0$	0
		90

$$\therefore s = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}} = \sqrt{\frac{90}{9}} = 3.16$$

\Rightarrow Null Hypothesis (H_0): $\mu_1 = \mu_2$ i.e., there is no change in blood pressure due to effect of drug.

Alternative Hypothesis (H_A): $\mu_1 \neq \mu_2$

test statistic:

$$t = \frac{|\bar{x} - M|}{\sigma/\sqrt{n}} = \frac{2}{3.16} \times \sqrt{10} = 2$$

$$t_{\text{cal}} = 2$$

And $t_{\text{tab}} = 2.26$ (given)

Conclusion \rightarrow Here $|t_{\text{cal}}| < |t_{\text{tab}}|$

Therefore Null Hypothesis must be accepted i.e there is no change in blood pressure due to the effect of drug.

Testing the Significance of the Difference between the Sample Mean

Let the two independent sample be x_1, x_2, \dots, x_{n_1} and y_1, y_2, \dots, y_{n_2} with mean \bar{x}, \bar{y} , from two normal population with means M_1 and M_2 and common variance σ^2

$$\text{Let } s_1^2 = \frac{1}{n_1-1} \sum (x - \bar{x})^2 \quad \text{If } n_1 \neq n_2$$

$$s_2^2 = \frac{1}{n_2-1} \sum (y - \bar{y})^2$$

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

$$\text{where } s^2 = \frac{\sum (x - \bar{x})^2 + \sum (y - \bar{y})^2}{n_1 + n_2 - 2} = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$$

$$\bar{x} = \frac{1}{n_1} \sum x_i ; \quad \bar{y} = \frac{1}{n_2} \sum y_i$$

The significance of t for $(n_1 + n_2 - 2)$ d.o.f
is tested in the same way

Q.1 The marks obtained in the same test by two groups of student are as follows.

First group	18	20	36	50	49	36	34	43	41
Second group	29	28	26	35	30	44	46		

Test the significance of difference between the arithmetic mean marks secured by the students of the above two groups.
 [given the value of t at 5% level of significance for 14 degrees of freedom is 2.14]

Solution.

Null Hypothesis (H_0): There is no significant difference between the mean of the above two groups i.e. $\mu_x = \mu_y$

Alternative Hypothesis (H_a): $\mu_x \neq \mu_y$

x	$x - \bar{x}$	$(x - \bar{x})^2$	First group			Second group		
			y	$y - \bar{y}$	$(y - \bar{y})^2$			
18	-19	361	29	-5	25			
20	-17	289	28	-6	36			
36	-1	1	26	-8	64			
50	13	169	35	1	1			
49	12	144	30	-4	16			
36	-1	1	44	10	100			
34	-3	9	46	12	144			
49	12	144						
41	4	16						
		1134						
			$n_1 = 9$					
				$n_2 = 7$				
					386			

$$\therefore s^2 = \frac{1}{n_1 + n_2 - 2} \left[\sum (x_i - \bar{x})^2 + \sum (y_j - \bar{y})^2 \right]$$

$$= \frac{1}{9+7-2} [1134 + 386] = 108.57$$

test statistic

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{s^2 \left[\left(\frac{1}{n_1} \right) + \left(\frac{1}{n_2} \right) \right]}} = \frac{\bar{x} - \bar{y}}{s} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} = \frac{\bar{x} - \bar{y}}{s}$$

$$= \frac{37 - 34}{\sqrt{108.57 \left\{ \frac{1}{9} + \frac{1}{7} \right\}}} = 0.5713.$$

Given $t_{tab} = 2.14$

$|t_{cal}| < |t_{tab}|$ at 5% level of significance

So H_0 may be ~~not~~ accepted.

i.e. there is no difference between the performance of the two groups of students.

Q Two types of batteries, A and B, are tested for their length of life and following results were obtained.

	No. of samples	mean	Variance
A	10	500	100
B	10	560	121

Is there a significant diff. in two mean at 5% of level.
Now $\bar{x} = 500$, $\bar{y} = 560$, $s_x^2 = 100$, $s_y^2 = 121$ H_0 no sig. diff. in means.

$$\text{var } s_x^2 = \frac{\sum (x_i - \bar{x})^2}{n} \Rightarrow 100 = \frac{\sum (x_i - \bar{x})^2}{10} \Rightarrow \sum (x_i - \bar{x})^2 = 1000$$

$$\text{similarly } \sum (y_i - \bar{y})^2 = 10 \times 121 = 1210$$

$$s^2 = \frac{1}{(n_1 + n_2 - 2)} \left\{ (\sum (x_i - \bar{x})^2 + \sum (y_i - \bar{y})^2) \right\}$$

$$= \frac{1000 + 1210}{10 + 10 - 2} = 122.77 \Rightarrow s = 11.08$$

$$t = \frac{\bar{x} - \bar{y}}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{60}{11.08 \sqrt{\frac{1}{10} + \frac{1}{10}}} = 11.36$$

$$|t| = 11.36 \quad \text{the tabulated value for } (10+10-2) \text{ d.f.} \\ = 2.010$$

$|t_{cal}| > |t_{tab}|$ at 5% level of significance.
 H_0 is rejected. i.e. the diff. b/w the means is highly significant

Paired t-test Difference of Means

In this case, we have n paired observation $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. Thus the size of samples is equal.

Let $d_i = x_i - y_i$ be the difference between x_i 's and y_i 's; ($i=1, 2, 3, \dots, n$)

Then Null Hypothesis (H_0): $\mu_d = \mu_y - \mu_x$ $n_1 = n_2$

Alternative Hypothesis (H_a): $\mu_d \neq \mu_y - \mu_x$

Test statistic: $t = \frac{\bar{d}}{\left(\frac{s}{\sqrt{n}}\right)}$ or $\frac{\bar{d}}{s/\sqrt{n}}$

Where $\bar{d} = \frac{1}{n} \sum_{i=1}^n d_i$

and $s^2 = \frac{1}{n-1} \sum_{i=1}^n (d_i - \bar{d})^2 = \frac{1}{n-1} \left[\sum d^2 - \frac{(\sum d)^2}{n} \right]$

Conclusion:— If $|t_{cal}| > |t_{tab}|$, then null hypothesis H_0 is rejected at $\alpha\%$ level of significance at $(n-1)$ degree of freedom

If a drug was administered to 10 patients and the increments in their blood pressure were recorded to be: - 6, 3, -2, 4, -3, 4, 6, 9, 0, 2
Is it reasonable to believe that drug has no effect on change of BP given $\alpha = 0.05$ for $t_{0.05} = 2.26$

Now let d denote the increment in BP

$$\bar{d} = \frac{1}{10} (6 + 3 - 2 + 4 - 3 + 4 + 6 + 0 + 0 + 2) = 2$$

$$d - \bar{d} = 4, 1, -4, 2, -5, 2, 4, -2, -2, 0$$

$$\therefore \sum (d - \bar{d})^2 = 16 + 1 + 16 + 4 + 25 + 4 + 16 + 4 + 4 + 0 = 90$$

$$s^2 = \frac{1}{n-1} \sum (d - \bar{d})^2 = \frac{90}{9} = 10 \Rightarrow s = \sqrt{10}$$

H_0 the drug has no effect change of BP

$$t = \frac{\bar{d}}{\left(\frac{s}{\sqrt{n}}\right)} = \frac{2}{\left(\frac{\sqrt{10}}{\sqrt{10}}\right)} = 2$$

$$|t_{tab}| = 2.26, |t_{cal}| < |t_{tab}| \text{ at } \alpha = 0.05$$

H_0 is accepted i.e. drug has no effect on change of BP

Q1: 10 soldiers visit the rifles range for two consecutive weeks. Their scores on the first week are:

67, 24, 57, 55, 63, 54, 56, 68, 33, 43. For the second week their scores in the same order are:

70, 38, 58, 58, 56, 67, 68, 75, 42, 38. Examine if there is any significant difference in their performance.

Sol': Let x and y denote the scores in the first and second week. Null Hypothesis (H_0): There is no significant difference between their performance i.e. $H_0: \mu_x = \mu_y$

Alternative Hypothesis (H_a): $\mu_x \neq \mu_y$

$$d_i = x_i - y_i \quad \text{Here } \bar{d} = \frac{\sum d}{n} = \frac{-50}{10} = -5$$

x	y	$d = x - y$	$d - \bar{d}$	$(d - \bar{d})^2$
67	70	-3	2	4
24	38	-14	-9	81
57	58	-1	4	16
55	58	-3	2	4
63	56	7	12	144
54	67	-13	-8	64
56	68	-12	-7	49
68	75	-7	-2	4
33	42	-9	-4	16
43	38	5	10	100
$\sum d = -50$				$\sum (d - \bar{d})^2 = 482$

$$\therefore s^2 = \frac{1}{(n-1)} \sum_{i=1}^n (d_i - \bar{d})^2 = \frac{1}{10-1} (482) = 53.55$$

$$s = \sqrt{53.55} = 7.3182$$

$$\text{Test statistic: } t = \frac{\bar{d}}{(s/\sqrt{n})} = \frac{-5}{\sqrt{53.55}} = -2.1606$$

$\therefore t_{\text{tab}} \text{ for } (10-1) = 9 \text{ degree of freedom at } 5\% \text{ level of significance}$
 $t_{\text{cal}} = -2.1606 < 2.26$

CHI-SQUARE TEST OF GOODNESS OF FIT

χ^2 -test [Chi-Square Test] > 50

~~(non-Parametric)~~

If $O_1, O_2, O_3, \dots, O_n$ is a set of observed (experimentally) frequencies and E_1, E_2, \dots, E_n is the corresponding set of expected (theoretical or hypothetical) frequencies, then:

Null Hypothesis (H_0): There is no significant difference between observed and expected frequencies.

Alternative Hypothesis (H_a): H_0 is not true

Test statistic :

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

Where χ^2 follows χ^2 -distribution with $(n-1)$ degrees of freedom

Conclusion: If $|\chi^2_{\text{cal}}| \geq |\chi^2_{\text{tab}}|$, then null hypothesis H_0 is rejected at $\alpha\%$ of significance, otherwise H_0 may be accepted.

- (i) If data is given in series (n) dot $= n-1$
- (ii) In the case of Binomial distribution dot $= n-1$
- (iii) In the case of poisson distribution dot $= n-2$
- (iv) In the case of normal distribution dot $= n-3$

Occurred - incidental (प्राकृतिक)

- Q.1 The following table gives the number of aircraft accidents occurred during the various days of the week. Find whether the accidents are uniformly distributed over the week.

Days	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Total
No. of accidents	14	16	8	12	11	9	14	(84)

(χ^2 for 4 degrees of freedom at 5% level of significance is 9.41)

Sol^b Null Hypothesis (H_0): The accidents are uniformly distributed over the week.

Alternative Hypothesis (H_a): H_0 is not true

Here Expected frequencies of accidents on any day $E = \frac{84}{7} = 12$

Since frequencies in some classes are less than 10, we regroup the data

Days	Sun	Mon	Tue+Wed	Thu	Fri+Sat	Total
Observed No. of accidents (O)	14	16	20	11	23	84
Expected no. of accidents (E)	12	12	24	12	24	84

Test statistic:

$$\begin{aligned}\chi^2 &= \sum_{i=1}^5 \frac{(O_i - E_i)^2}{E_i} \\ &= \frac{(14-12)^2}{12} + \frac{(16-12)^2}{12} + \frac{(20-24)^2}{24} + \frac{(11-12)^2}{12} + \frac{(23-24)^2}{24} \\ &= \frac{4}{12} + \frac{16}{12} + \frac{16}{24} + \frac{1}{12} + \frac{1}{24} = \frac{1}{24} [8 + 32 + 16 + 2 + 1] \\ &= \frac{59}{24} = 2.46\end{aligned}$$

Here degree of freedom = $5-1=4$

The tabulated value of χ^2 for 4 degrees of freedom at 5% level of significance = 9.41 i.e $\chi^2_{tab} = 9.41$

Conclusion: Here $|X_{\text{cal}}^2| < |X_{\text{tab}}^2|$ at 5% level of significance

i.e. H_0 Null Hypothesis is accepted.

i.e. The accidents are uniformly distributed uniformly distributed over the week.

Q.2 In a sample survey of public opinion, answer to the following questions are tabulated below.

- (i) Do you drink? option
(ii) Are you in favor of local option sale of liquor?

		QUESTION		Total
		Yes	No	
Yes	Yes	56	31	87
	No	18	6	24
Total		74	37	111

Can you infer whether or not the local option on the sale of liquor is dependent on individual drink?

Solution: The Null Hypothesis (H_0): The option on the sale of liquor is independent or not associated with individual drinking

Alternative Hypothesis (H_A): H_0 is not true.

$$\text{Expected frequency} = \frac{R \times C}{T}$$

Expected Frequency are tabulated as →

		Yes	No	
Yes	Yes	$\frac{87 \times 74}{111} = 58$	$\frac{87 \times 37}{111} = 29$	
	No	$\frac{74 \times 24}{111} = 16$	$\frac{24 \times 37}{111} = 8$	
Total		74	37	

Test statistic:

$$X^2 = \sum \frac{(O_i - E_i)^2}{E} = \frac{(56 - 58)^2}{58} + \frac{(18 - 16)^2}{16} + \frac{(31 - 29)^2}{29} + \frac{(6 - 8)^2}{8} = .957$$

$$\text{Degree of Freedom} = (R-1)(C-1) = (2-1)(2-1) = 1$$

The tabulated value of χ^2 at 5% and one degree of freedom is

$$\chi^2_{0.05,1} = 3.84$$

Conclusion: Here $|\chi^2_{cal}| \leq |\chi^2_{tab}|$

therefore the null hypothesis is accepted.

i.e. sale of liquor is independent of individual drinking.

Q.3 → A die is thrown 276 times and the result of these throws are given below:

No. of appeared on the die	1	2	3	4	5	6
Frequency	40	32	29	59	57	59

Test whether the die is biased or not.

Sol' Null Hypothesis (H_0): Die is unbiased

$$\text{Expected frequencies } E = \frac{276}{6} = 46$$

Observed frequency O	40	32	29	59	57	59
Expected frequency E	46	46	46	46	46	46

O	E	O-E	$(O-E)^2$
40	46	-6	36
32	46	-14	196
29	46	-17	289
59	46	13	169
57	46	11	121
59	46	13	169
			980

$$\therefore \chi^2 = \sum \frac{(O_i - E_i)^2}{E} = \frac{980}{46} = 21.304$$

Here degree of freedom = $(6-1) = 5$

$$\therefore \chi^2_{0.05,5} = 11.07$$

$\chi^2 > \chi^2_{0.05,5}$ so H_0 is rejected i.e. Die is biased R

4. Records taken of the number of male and female births in 800 families having four children are as follows:

No. of male births.	0	1	2	3	4
No. of Female births	4	3	2	1	0
No. of Families	32	178	290	236	94

Test whether the data are consistent with the hypothesis that the Binomial law holds and the chance of male birth is equal to that of female birth, namely $P = q = \frac{1}{2}$

Sol: Null Hypothesis (H_0): The data are consistent with the hypothesis of equal probability for male and female birth i.e $P = q = \frac{1}{2}$

We use Binomial distribution to calculate theoretical frequency given by

$$N(r) = N \times {}^n C_r p^r q^{n-r}$$

Where N is total frequency, $N(r)$ is the number of families with r male children, p and q are prob. of male and female birth, n is the number of children

$$N(0) = 800 \times {}^4 C_0 \left(\frac{1}{2}\right)^4 = 800 \times 1 \times \frac{1}{16} = 50$$

~~$$N(1) = 800 \times {}^4 C_1 \left(\frac{1}{2}\right)^4 = 800 \times 4 \times \frac{1}{8}$$~~

$$N(1) = {}^4 C_1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^3 = 800 \times \frac{4}{3} \left(\frac{1}{2}\right)^4 = 200$$

$$N(2) = 800 \times {}^4 C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^2 = 800 \times 6 \times \frac{1}{16} = 300$$

$$N(3) = 800 \times {}^4 C_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^1 = 800 \times 4 \times \frac{1}{16} = 200$$

$$N(4) = 800 \times {}^4 C_4 \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^0 = 800 \times 1 \times \frac{1}{16} = 50$$

$\therefore r$	E	$O-E$	$(O-E)^2$
32	50	-18	324
178	200	-22	484
290	300	-10	100
236	200	36	1296
94	50	44	1936

$$\therefore \chi^2 = \sum \frac{(O-E)^2}{E} = \frac{324}{50} + \frac{484}{200} + \frac{100}{300} + \frac{1296}{200} + \frac{1936}{50} \\ = 6.48 + 2.42 + 1.33 + 6.48 + 38.72 = 53$$

Conclusion → Table Value of χ^2 at 5% level of significance for 5-1=4 degree of freedom is 9.48

Here $|X_{cal}^2| > |X_{tab}^2|$, so Null Hypothesis is rejected
i.e. the chance of a male birth is not equal to that of a female ~~births~~ birth.

Q.5. A survey of 320 families with 5 children shows the following distribution:

No. of boys	5 boys	4 boys	3 boys	2 boys	1 boy	0 boys	Total
No. of girls	0 girls	1 girl	2 girls	3 girls	4 girls	5 girls	
No. of families	18	56	110	88	40	8	320

Given that Values of χ^2 for 5 degree of freedom are 11.0 and 15.1 at 0.5 and 0.1 significance level respectively, test the hypothesis that male and female births are equally probable.

Solution: → Null Hypothesis (H_0): Male and Female births are equally probable.

The prob. of. male births $P = \frac{1}{2}$, and female births $q = \frac{1}{2}$

so the expected frequency of r male births in a family of 5 out of 320 families = $N \cdot {}^5C_r P^r q^{5-r}$

$$\therefore \text{so, Expected frequency of 5 boys and 0 girl } (E_1) = 320 \times {}^5C_5 \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^0 \\ = 320 \times \frac{1}{32} = 10$$

$$\text{Expected frequency of 4 boys and 1 girl } (E_2) = 320 \times {}^5C_4 \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^1 \\ = 320 \times 5 \times \frac{1}{32} = 50$$

$$\text{Expected frequency of 3 boys and 2 girls } (E_3) = 320 \times {}^5C_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^2 \\ = 320 \times 10 \times \frac{1}{32} = 100$$

$$\text{Expected frequency of 2 boys and 3 girls } (E_4) = 320 \times {}^5C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^3 = 100$$

$$\text{Expected frequency of 1 boys and 4 girls } E_5 = 320 \times {}^5C_1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^4 = 50$$

$$\text{Expected frequency of 0 boy and 5 girls } (E_6) = 320 \times {}^5C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^5 = 10$$

table

Observed frequency (O_i)	Expected frequency (E_i)
18	10
56	50
110	100
88	100
40	50
8	10

Test statistic.

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

$$= \frac{(18-10)^2}{10} + \frac{(56-50)^2}{50} + \frac{(110-100)^2}{100} + \frac{(88-100)^2}{100} + \frac{(40-50)^2}{50} + \frac{(8-10)^2}{10}$$

$$= 6.4 + 1.72 + 1 + 1.44 + 2 + 1 = 11.96$$

$$\text{Degree of freedom} = n-1 = 6-1 = 5$$

Given χ^2_{tab} at 5% level of significance = 11

(i) ~~Given~~ Conclusion:- (i) since $|\chi^2_{\text{cal}}| > |\chi^2_{\text{tab}}|$ at 5% level of significance
so null hypothesis is rejected.

(ii) and χ^2_{tab} at 1% level of significance = 15.1 given

i.e $|\chi^2_{\text{cal}}| < |\chi^2_{\text{tab}}|$ at 1% level of significance so null hypothesis is accepted at 1% level of significance.

Q. The following table gives the number of accidents that took place in an industry during various days of the week. Test if accidents are uniformly distributed over the week.

Day	Mon	Tue	Wed	Thur	Fri	Sat
No. of accidents	14	18	12	11	15	14

Sol". Null hypothesis H_0 : The accidents are uniformly distributed over the week

Under this H_0 , the expected frequencies of the accidents on each of these days = $\frac{84}{6} = 14$

Observed frequency	14	18	12	11	15	14
Expected frequency	14	14	14	14	14	14
$(O_i - E_i)^2$	0	16	4	9	1	0

$$\chi^2 = \frac{\sum (O_i - E_i)^2}{E_i} = \frac{30}{14} = 2.1428$$

Conclusion - Calculated value of χ^2 is less than the tabulated value, H_0 is accepted i.e the accidents are uniformly distributed over the week

Table value of χ^2 at 5% level for (6-1= 5 d.f) is 11.09

~~0.33~~

~~4.743~~