SMT1201 ENGINEERING MATHEMATICS III

UNIT V

THEORY OF SAMPLING AND TESTING OF HYPOTHESIS

Test of Hypothesis - test of significance - Large samples - Z test - single proportion - difference of proportions - Single mean - difference of means - Small samples - Student‘s t test - single mean - difference of means -Test of variance - Fisher’s test - Chi square test - goodness of fit - independence of attributes.

THEORY OF SAMPLING AND TEST OF HYPOTHESIS

Population:
The group of individuals, under study is called is called population.

Sample:
A finite subset of statistical individuals in a population is called Sample.

Sample size:
The number of individuals in a sample is called the Sample size.

Parameters and Statistics:
The statistical constants of the population are referred as Parameters and the statistical constants of the Sample are referred as Statistics.

Standard Error :
The standard deviation of sampling distribution of a statistic is known as its standard error and is denoted by (S.E)

Test of Significance :
It enable us to decide on the basis of the sample results if the deviation between the observed sample statistic and the hypothetical parameter value is significant or the deviation between two sample statistics is significant.

Null Hypothesis:
A definite statement about the population parameter which is usually a hypothesis of no-difference and is denoted by $H_{o.}$

Alternative Hypothesis:
Any hypothesis which is complementary to the null hypothesis is called an Alternative Hypothesis and is denoted by $H_{1.}$

Errors in Sampling:
Type I and Type II errors.
Type I error: Rejection of H_0 when it is true.
Type II error: Acceptance of H_0 when it is false.

Two types of errors occur in practice when we decide to accept or reject a lot after examining a sample from it. They are Type 1 error occurs while rejecting H_o when it is true. Type 2 error occurs while accepting H_o when it is wrong.

Critical region:

A region corresponding to a statistic t in the sample space S which lead to the rejection of H_o is called Critical region or Rejection region. Those regions which lead to the acceptance of H_o are called Acceptance Region.

Level of Significance :

The probability α that a random value of the statistic "t" belongs to the critical region is known as the level of significance. In otherwords the level of significance is the size of the type I error. The levels of significance usually employed in testing of hypothesis are 5% and 1%.

One tail and two tailed test:

A test of any statistical hyposthesis where the alternate hypothesis is one tailed(right tailed/ left tailed) is called one tailed test.

For the null hypothesis H_0 if $\mu = \mu_0$ then.

$H_1 = \mu > \mu_0$ (Right tail)

$H_1 = \mu < \mu_0$ (Left tail)

$H_1 = \mu \# \mu_0$ (Two tail test)

Types of samples :

Small sample and Large sample

Small sample ($n \leq < 30$) : "Students t test, F test , Chi Square test

Large sample ($n>30$) : Z test.

95 % confidence limit for the population mean μ in a small test.

Let x be the sample mean and n be the sample size. Let s be the sample S.D.

Then $\bar{x} \pm t_{0.05}$ (s/√n-1)

Application of t – distribution

When the size of the sample is less than 30, 't' test is used in (a) single mean and (b) difference of two means.

Distinguish between parameters and statistics.

Statistical constant of the population are usually referred to as parameters. Statistical measures computed from sample observations alone are usually referred to as statistic.

In practice, parameter values are not known and their estimates based

Write short notes on critical value.

The critical or rejection region is the region which corresponds to a predetermined level of significanceα. Whenever the sample statistic falls in the critical region we reject the null hypothesis as it will be considered to be probably false. The value that separates the rejection region from the acceptance region is called the critical value.

Define level of significance explain.

The probability α that a random value of the statistic't' belongs to the critical region is known as the level of significance. In other words level of significance is the size of type I error. The levels of significance usually employed in testing of hypothesis are 5% and 1%.

Outline the assumptions made when the't' test us applied for difference of means.

(i) Degree of freedom is $n_1 + n_2 - 2$.

(ii) The two population variances are believed to be equal.

(iii) $S = \sqrt{\frac{(n_1 s_1^2 + n_2 s_2^2)}{(n_1 + n_2 - 2)}}$ is the standard error.

<u>Type I Student t test for single mean</u>

$$|t| = \frac{\bar{x} - \mu}{s / \sqrt{n-1}}$$

Where $\bar{x}$ the sample mean, μ is is the population mean, s is the SD and n is the number of observations.

Problems :

1. The mean weakly sales of soap bars in departmental stores were 146.3 bars per store. After an advertising campaign the mean weekly sales in 22 stores for a typical week increased to 153.7 and showed a SD of 17.2. Was the advertising campaign successful?

Solution:

Calculated t value = 1.97 and Tabulated Value = 1.72(at 5% level of significance with 21 degrees of freedom)

Calculated value > Tabulated value, Reject Ho (Null hypothesis)

2. A sample of 26 bulbs gives a mean life of 990 hours with SD of 20 hours. The manufacturer claims that the mean life of bulbs is 1000 hours. Is the sample not upto the standard.

Solution:
Calculated t value = 2.5
Tabulated Value = 1.708(at 5% level of significance with 25 degrees of freedom)
Calculated value > Tabulated value, Reject Ho (Null hypothesis)

3. The average breaking strength of steel rod is specified to be 18.5 thousand pounds. To test this sample of 14 rods was tested. The mean and SD obtained were 17.85 and 1.955 respectively. Is the result of the experiment significant?

Solution:
Calculated t value = 1.199
Tabulated Value = 2.16(at 5% level of significance with 13 degrees of freedom)
Calculated value < Tabulated value, Accept Ho (Null hypothesis)

4. Find the confidence limits of the mean of the population for a random sample of size 16 from a normal population with mean 53 and SD √10 with t value at 5% for 15 Degrees of freedom is 2.13.

Solution
(54.68, 51.31)

Type II Student t test when SD not given

$|t| = (\bar{x} - \mu) / (s/\sqrt{n})$
Where $\bar{x} = \Sigma(x)/n$ and $s^2 = 1/(n-1)\ \Sigma\ (x-\bar{x})^2$

PROBLEMS

Students t test where SD of the sample is not given directly)

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

Solution:
Calculated t value = 0.62
Tabulated Value = 2.26(at 5% level of significance with 9 degrees of freedom)
Calculated value < Tabulated value, Accept Ho (Null hypothesis)
95% confidence limits: (86.99, 107.4)

2. The heights of 10 males of a given locality are found to be 70,67,62,68,61,68,70,64,64,66 inches. Is it reasonable to believe that the average height is greater than 64 inches Test at 5%.

Solution:
Calculated t value = 2
Tabulated Value = 1.833(at 5% level of significance with 9 degrees of freedom)

Calculated value > Tabulated value, Reject Ho (Null hypothesis)

3. Certain pesticide is packed into bags by a machine. A random sample of 10 bags is drawn and their contents are found to be as follows: 50,49,52,44,45,48,46,45,49,45. Test if the average packing to be taken 50 grams

Solution:

Calculated t value = 3.19

Tabulated Value = 2.262 (at 5% level of significance with 9 degrees of Freedom)

Calculated value > Tabulated value, Reject Ho (Null hypothesis)

Type III Student t test for difference of means of two samples

To test the significant difference between two mea n $\bar{x}_1$ and $\bar{x}_2$ of sample sizes n_1 and n_2 use the statistic.

$$|t| = (\bar{x}_1 - \bar{x}_2) / s\sqrt{((1/n_1) + (1/n_2))}$$

Where $s^2 = (n_1 s_1^2 + n_2 s_2^2)/(n_1 + n_2 - 2)$

s_1 and s_2 being the sample standard deviations degree of freedom being $n_1 + n_2 - 2$.

PROBLEMS

1. Samples of two types of electric light bulbs were tested for length of life and following data were obtained.

Type I	Type II
Sample size $n_1 = 8$	$n_2 = 7$
Sample means $x_1 = 1234$ hrs	$x_2 = 1036$ hrs
Sample S.D. $s_1 = 36$ hrs	$s_2 = 40$ hrs

Is the difference in the means sufficient to warrant that type I is superior to type II regarding length of life.

Solution:

Calculated t value = 9.39

Tabulated Value = 1.77 (at 5% level of significance with 13 degrees of freedom)

Calculated value > Tabulated value, Reject Ho (Null hypothesis)

2. Below are given the gain in weights (in N) of pigs fed on two diets A and B.

Diet A	25	32	30	34	24	14	32	24	30	31	35	25		
Diet B	44	34	22	10	47	31	40	32	35	18	21	35	29	22

Test if the two diets differ significantly as regards their effect on increase in weight.

Solution:

Calculated t value = 0.609

Tabulated Value = 2.06 (at 5% level of significance with 25 degrees of freedom)
Calculated value < Tabulated value, Accept Ho (Null hypothesis)

3. The nicotine content in milligrams of two samples of tobacco were found to be as follows:

Sample A	24	27	26	21	25	
Sample B	27	30	28	31	22	36

Can it be said that two samples come from normal populations having the same mean.

Solution:

Calculated t value = 1.92
Tabulated Value = 2.262 (at 5% level of significance with 9 degrees of freedom)
Calculated value < Tabulated value, Accept Ho (Null hypothesis)

4. The means of two random samples of sizes 9 and 7 are given as 196.42 and 198.82. The sum of the squares of the deviations from mean is 26.94 and 18.73 respectively. Can the sample be considered to have been drawn from the same normal population?

Solution:

Calculated t value = 2.63
Tabulated Value = 2.15 (at 5% level of significance with 14 degrees of freedom)
Calculated value > Tabulated value, Reject Ho (Null hypothesis)

F - TEST

To test if the two samples have come from same population we use F test (OR) To test if there is any significant difference between two estimates of population variance.

F= GREATER VARIANCE/SMALLER VARIANCE

(OR)

$F = S_1^2 / S_2^2$

Where

$S_1^2 = \Sigma (x-\bar{x})^2 / n_1-1$

$S_2^2 = \Sigma (y-\bar{y})^2 / n_2-1$

Where n_1 is the first sample size and n_2 is the second sample size

1. Applications of F-test.

To test whether if 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.

2. Uses f test in sampling

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.

If the sample variance S^2 is not given we can obtain the population variance by using the relation

$S_1^2 = n_1 s_1^2/(n_1-1)$ and $S_2^2 = n_2 s_2^2/(n_2-1)$

If we have to test whether the samples come from the same normal population then the problem has to be solved by both the t test and the f tests.

(i) To test the equality of the variances by F test
(ii) To test the equality of means by t test

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 observation it was 102. 6. Test whether this difference is significant at 5 % level.

Solution:

Calculated F value = 1.057
Tabulated Value = 3.29 (at 5% level of significance with (7,9) degrees of freedom)
Calculated value < Tabulated value, Accept Ho (Null hypothesis)

2. Two random samples gave the following results.

Sample	Size	Sample mean	Sum of squares of deviations from the mean
1	10	15	90
2	12	14	108

Test whether the samples come from the same normal population.

Solution:

Calculated F value = 1.018

Tabulated Value = 2.9 (at 5% level of significance with (9,11) degrees of freedom)

By t test Calculated t value = 0.74

Tabulated Value = 2.086 (at 5% level of significance).

In both the tests of sampling

Calculated value < Tabulated value, Accept Ho (Null hypothesis)

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?

Solution:

Calculated F value = 1.37

Tabulated Value = 4.95 (at 5% level of significance with (6,5) degrees of freedom)

Calculated value < Tabulated value, Accept Ho (Null hypothesis)

4. The nicotine content in milligrams of two samples of tobacco were found to be as follows:

Sample A	24	27	26	21	25	
Sample B	27	30	28	31	22	36

Can it be said that two samples come from normal populations having the same variances.

Solution:

Calculated F value =4.07

Tabulated Value = 6.26 (at 5% level of significance with (5,4) degrees of freedom)

Calculated value < Tabulated value, Accept Ho (Null hypothesis)

CHI-SQUARE TEST

CHI-SQUARE TEST FORMULAE

$$\psi^2 = \Sigma \frac{(O-E)^2}{E}$$

Where O is the observed frequency and E is the Expected frequency

1. Define Chi square test of goodness of fit.

Under the test of goodness of fit we try to find out how far observed values of a given phenomenon are significantly different from the expected values. The Chi square statistic can be used to judge the difference between the observed and expected frequencies.

2. Give the main use of Chi-square test.

To test the significance of discrepancy between experimental values and the theoretical values, obtained under some theory or hypothesis.

3. Write the condition for the application of ψ^2 test.

ψ^2 test can be applied only for small samples.

4. How is the number of degrees of freedom of chi-square distribution fixed for testing the goodness of fit of a poisson distribution for the given data.

Degree of freedom = n – 1 where n is the no. of observations.

CHI-SQUARE TEST FOR INDEPENDENCE OF ATTRIBUTES

An attribute means a quality or characteristic. Eg. Drinking, smoking, blindness, honesty

2 X 2 CONTINGENCY TABLE

Consider any two attributes A and B. A and B are divided into two classes.

OBSERVED FREQUENCIES

A	a	b
B	c	d

EXPECTED FREQUENCIES

E(a)= (a+c)(a+b)/N	E(b)=(b+d)(a+b)/N	a+b
E(c) = (a+c)(c+d)/N	E(d)=(b+d)(c+d)/N	c+d
a+c	b+d	N(Total frequencies)

PROBLEMS

1. A die is thrown 264 times with the following results. Show that the die is biased

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

Solution:

Calculated ψ^2 value =17.6362

Tabulated Value = 11.07 (at 5% level of significance with 5 degrees of freedom)

Calculated value > Tabulated value, Reject Ho (Null hypothesis)

2. 200 digits were chosen at random from a set of tables. The frequencies of the digits were

Digits	0	1	2	3	4	5	6	7	8	9
Frequency	18	19	23	21	16	25	22	20	21	15

Use the ψ^2 test to assess the correctness of the hypothesis that the digits were distributed in the equal number in the tables from which these were chosen.

Solution:

Calculated ψ^2 value = 4.3

Tabulated Value = 16.919 (at 5% level of significance with 9 degrees of freedom)

Calculated value< Tabulated value, Accept Ho (Null hypothesis)

3. Two groups of 100 people each were taken for testing the use of a vaccine 15 persons contracted the disease out of the inoculated persons while 25 contracted the disease in the other group. Test the efficiency of the vaccine using chi square test.

Solution:

Calculated ψ^2 value = 3.125

Tabulated Value = 3.184 (at 5% level of significance with 1 degrees of freedom)

Calculated value< Tabulated value, Accept Ho (Null hypothesis)

4. In a certain sample of 2000 families 1400 families are consumers of tea. Out of 1800 Hindu families, 1236 families consume tea. Use Chi square test and state whether there is any significant difference between consumption of tea among Hindu and Non – Hindu families.

Solution:

Calculated ψ^2 value = 15.238

Tabulated Value = 3.841 (at 5% level of significance with 1 degrees of freedom)

Calculated value > Tabulated value, Reject Ho (Null hypothesis)

5. Given the following contingency table for hair colour and eye colour. Find the value of Chi-Square and is there any good association between the two

Hair colour / Eye colour	Fair	Brown	Black
Grey	20	10	20
Brown	25	15	20
Black	15	5	20

Solution:

Calculated ψ^2 value = 3.6458

Tabulated Value = 9.488 (at 5% level of significance with 4 degrees of freedom)

Calculated value < Tabulated value, Accept Ho (Null hypothesis)

LARGE SAMPLES

TEST OF SIGNIFICANCE OF LARGE SAMPLES

If the size of the sample n>30 then that sample is called large sample.

Type 1. Test of significance for single proportion

Let p be the sample proportion and P be the population proportion, we use the statistic Z= (p-P) / $\sqrt{(PQ/n)}$

Limits for population proportion P are given by p±3$\sqrt{(PQ/n)}$
Where q = 1-p

1. A manufacture claimed that at least 95% of the equipment which he supplied to a factory conformed to specifications. An examination of a sample of 200 pieces of equipment revealed that 18 were faulty. tEst his claim at 5% level of significance.

Solution:
Calculated Z value = 2.59
Tabulated Value = 1.96 (at 5% level of significance) Calculated value > Tabulated value, Reject Ho (Null hypothesis)

2. 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.

Solution:
Calculated Z value = 2.04
Tabulated Value = 1.645 (at 5% level of significance) Calculated value > Tabulated value, Reject Ho (Null hypothesis)

3. A die is thrown 9000 times and of these 3220 yielded 3 or 4. Is this consistent with the hypothesis that the die was unbiased?

Solution:
Calculated Z value = 4.94 since z>3
Calculated value > Tabulated value, Reject Ho (Null hypothesis)

4 A random sample of 500 apples were taken from the large consignment and 65 were found to be bad. Find the percentage of bad apples in the consignment.

Solution:
(0.175, 0.085) Hence percentage of bad apples in the consignment lies between 17.5% and 8.5%

Type II Test of significance for difference of proportions

Let n_1 and n_2 are the two sample sizes and sample proportions are p_1 and p_2

$$Z = \frac{(p_1 - p_2)}{\sqrt{pq(1/n_1 + 1/n_2)}}$$ where $p = (n_1p_1+n_2p_2)/n_1+n_2$ and $q=1-p$

Proplems

1. Before an increase in excise duty on tea, 800 persons out of a sample of 1000 persons were found to be tea drinkers. After an increase in duty 800 people were tea drinkers in the sample of 1200 people. Using standard error of proportions state whether there is a significant decrease in the consumption of tea after the increase in the excise duty.

Solution:

Calculated Z value = 6.972

Tabulated value at 5% (one tail) = 1.645

Calculated value > Tabulated value, Reject Ho (Null hypothesis)

2. In two large populations there are 30% and 25% respectively of fair haired people. Is this difference likely to be hidden in samples of 1200 and 900 respectively from the two populations.

Solution:

Calculated Z value = 2.55

Tabulated value at 5% = 1.96

Calculated value > Tabulated value, Reject Ho (Null hypothesis)

<u>Type III Test of significance for single Mean</u>

$z = \bar{x} - \mu / (\sigma/\sqrt{n})$ where $\bar{x}$ is the same mean

μ is the population mean, s is the population S.D.

n is the sample size.

The values of $\bar{x} \pm 1.96\ (\sigma/\sqrt{n})$ are called 95% confidence limits for the mean of the population corresponding to the given sample.

The values of $\bar{x} \pm 2.58\ (\sigma/\sqrt{n})$ are called 99% confidence limits for the mean of the population corresponding to the given sample.

PROBLEMS

1. A sample of 900 members has a mean of 3.4 cms and SD 2.61 cms. Is the sample from a large population of mean is 3.25 cm and SD 2.61 cms. If the population is normal and its mean is unknown find the 95% confidence limits of true mean.

Solution:

Calculated Z value = 1.724
Tabulated value at 5% = 1.96
Calculated value < Tabulated value, Accept Ho (Null hypothesis)
Limits (3.57, 3.2295)

2. An insurance agent has claimed that the average age of policy holders who issue through him is less than the average for all agents which is 30.5 years. A random sample of 100 policy holders who had issued through him gave the following age distribution.

Age	**16-20**	**21-25**	**26-30**	**31-35**	**36-40**
No of persons	**12**	**22**	**20**	**30**	**16**

Test the significant difference at 5% level of significance.

Solution:

Calculated Z value = 2.68
Tabulated value at 5% = 1.645
Calculated value > Tabulated value, Reject Ho (Null hypothesis)

3 Write down the test statistic for single mean for large samples.

$\mu = \bar{X} - \mu / (\sigma/\sqrt{n})$ where $\bar{X}$ is the same mean
μ is the population mean, s is the population S.D.
n is the sample size.

4. The mean score of a random sample of 60 students is 145 with a SD of 40. Fine the 95 % confidence limit for the population mean.

Solution $\bar{z} = X \pm 1.96 (\sigma/\sqrt{n})$
$= 145 \pm (1.96) (40/\sqrt{60})$
$= 145 \pm 10.12$
$= 155.12$ or 134.88

$\therefore$ The confidence limits are 155.12 and 134.88.

<u>Type IV Test of significance for Difference of means</u>

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

PROBLEMS

1. The means of 2 large samples of 1000 and 2000 members are 67.5 inches and 68 inches respectively. Can the samples be regarded as drawn from the same population of SD 2.5 inches.

Solution:

Calculated Z value = 5.16
Tabulated value at 5% = 1.96
Calculated value > Tabulated value, Reject Ho (Null hypothesis)

2. The mean yield of wheat from a district A was 210 pounds with SD 10 pounds per acre from a sample of 100 plots. In another district the mean yield was 220 pounds with sD 12 pounds from a sample of 150 plots. Assuming that the SD of yield in the entire state was 11 pounds test whether there is any significant difference between the mean yield of crops in the two districts.

Solution:

Calculated Z value = 7.041
Tabulated value at 5% = 1.96
Calculated value > Tabulated value, Reject Ho (Null hypothesis)

PRACTICE PROBLEMS

1. Ten cartoons are taken at random from an automatic filling machine. The mean net weight of 10 cartoons is 11.802 and SD is 0.15. Does the sample mean differ significantly from the weight of 12?

Solution:

Calculated t value = 4
Tabulated Value = 2.26(at 5% level of significance with 9 degrees of freedom)
Calculated value > Tabulated value, Reject Ho(Null hypothesis)

2. A random sample of size 20 from a normal population gives a sample mean of 42 and sample SD 6. Test if the population mean is 44?

Solution:

Calculated t value = 1.45
Tabulated Value = 2.09(at 5% level of significance with 19 degrees of freedom)
Calculated value < Tabulated value, Accept Ho(Null hypothesis)

3. A machine which produces mica insulating washers for using electric devices is said to turn out washers having a thickness of 10 mm. A sample of 10 washers has an average of 9.52 mm with SD of 0.6 mm. calculate student's t test.

Solution:

Calculated t value = 2.528
Tabulated Value = 2.26(at 5% level of significance with 9 degrees of freedom)

Calculated value > Tabulated value, Reject Ho(Null hypothesis)

4. The mean lifetime of 25 fans produced by a company is computed to be 1570 hours with SD 120 hrs. The company claims that the average life of fans produced by them is 1600 hours. Is the claim acceptable.

Solution:

Calculated t value = 1.22

Tabulated Value = 2.06(at 5% level of significance with 24 degrees of freedom)

Calculated value < Tabulated value, Accept Ho(Null hypothesis)

5. From a population of students 10 are selected. Their weekly packet money observed as 20,22,21,15,25,19,18,20,21,22. Test if the sample supports that on an average student get Rs.25 as packet money.

Solution:

Calculated t value = 1.89

Tabulated Value = 2.26 (at 5% level of significance with 24 degrees of freedom)

Calculated value < Tabulated value, Accept Ho(Null hypothesis).

6. Ten individuals are chosen from random and their heights are found to be in inches 63,63,64,65,66,69,69,70,70,71. Discuss the solution that the mean height of the universe is 65?

Solution :

Calculated t value = 2.02

Tabulated Value = 2.26 (at 5% level of significance with 9 degrees of freedom)

Calculated value < Tabulated value, Accept Ho(Null hypothesis).

7. An IQ test was given to 5 persons before and after they were trained. Results are given below.

IQ before training	110	120	123	132	125
IQ after training	120	118	125	136	121

Test if there is any change in the IQ after the training program.

Solution :

Calculated t value = 0.816

Tabulated Value = 2.78 (at 5% level of significance with 4 degrees of freedom)

Calculated value < Tabulated value, Accept Ho(Null hypothesis).

8. Memory capacity of 10 girls were tested before and after training. State if the training was effective or not

Before	12	14	11	8	7	10	3	0	5	6
After	15	16	10	7	5	12	10	2	3	8

Solution :
Calculated t value = 1.3646
Tabulated Value = 2.26 (at 5% level of significance with 9 degrees of freedom)
Calculated value < Tabulated value, Accept Ho(Null hypothesis).

9. 1.Two random samples gave the following results. Test whether the samples come from the same normal population.

Sample	Size	Sample Mean	Sum of squares of deviations from the mean
1	10	15	90
2	12	14	108

Solution:
Calculated F =1.018,Tabulated F for (9,11) d.f at 5% level=2.90. Since Calculated F < Tabulated F, the null hypothesis H_0 is accepted. Calculated t =0.74,Tabulated t for 20 d.f at 5% level=2.086. Since Calculated t < Tabulated t, the null hypothesis H_0 is accepted.

10. The fatality rate of typhoid patients is believed to be 17.26%. In a certain year 640 patients suffering from typhoid were treated in a metropolitan hospital and only 63 patients died. Can you consider the hospital efficient?
Ans:z=4.96, H_0 rejected.

11. A salesman in a departmental store claims that at most 60 percent of the shoppers entering the store leave without making a purchase. A random sample of 50 shoppers showed that 35 of them left without making a purchase. Are these sample results consistent with the claim of the salesman?
Ans:z=1.443, H_0 accepted

12. In a large city A, 20% of a random sample of 900 school boys had a slight physical defect. In another large city B, 18.5% of a random sample of 1600 school boys had the same defect. Is the difference between the proportions significant?
Ans:z=0.92, H_0 accepted.

13. Before and increase in excise duty on tea, 800 people out of a sample of 1000 were consumers of tea. After the increase in duty 800 out of a sample of 1200 persons. Find whether there is a significant decrease in the consumption of tea after the increase in duty.
Ans: z=6.82, H_0 is rejected.

14. A sample of 100 students is taken from a large population. The mean height of the students in this sample is 160cm. Can it be reasonably regarded that, in the population, the mean height is 165cm, and the SD is 10cm?

Ans:z=5, H_0 rejected.

15. A simple sample of heights of 6400 English men has a mean of 170cm and SD of 6.4cm, while a sample of heights of 1600 Americans has a mean of 172cm and a SD of 6.3cm. Do the data indicate that Americans, on the average taller than Englishmen?
Ans: z=11.32, H_0 rejected.

16. The average marks scored by 32 boys is 72 with SD of 8, while that for 36 girls is 70 with SD of 6. Test at 1% level whether boys perform better than girls.
Ans:z-1.15, H_0 accepted.

17. A random sample of 600 men chosen from a certain city contained 400 smokers. In another sample of 900 men chosen from another city, there were 450 smokers. Do the data indicate that (i)the cities are significantly different with respect to smoking habit among men? and (ii)the first city contains more smokers than the second?
Ans:z=6.49,(i)yes (ii)yes

18. In a college, 60 junior students are found to have a mean height of 171.5cm and 50 senior students are found to have a mean height of 173.8 cm. Can we conclude, based on these data, that the juniors are shorter than the seniors at 1% level assuming that the SD of students of that college is 6.2cm?
Ans:No, z=1.937

19. Tests made on the breaking strength of 10 pieces of a metal gave the following results: 578,572,570,568,572,570, 570,572,596 and 584kg. Test if the mean breaking strength of the wire can be assumed as 577 kg?
Ans:yes,t=0.65

20. A mechinist is expected to make engine parts with axle diameter of 1.75cm. A random sample of 10 parts shows a mean diameter of 1.85cm, with SD of 0.1cm. On the basis of this sample, would you say that the work of the machinist is inferior?
Ans: yes, t=3

21. A certain injection administered to each of the 12 patients resulted in the following increases of blood pressure: 5, 2, 8, -1, 3, 0, 6, -2, 1, 5, 0, 4. Can it be

concluded that the injection will be in general, accompanied by an increase in BP?
Ans: yes, t=2.89

22. The mean life time of a sample of 25 bulbs is found as 1550h, with SD of 120h. The company manufacturing the bulbs claims that the average life of their bulbs is 1600h. Is the claim acceptable?
Ans: yes, t=2.04

23. Two independent samples of sizes 8 and 7 contained the following values: Sample 1: 19, 17, 15, 21, 16, 18, 16, 14 and Sample 2: 15, 14, 15, 19, 15, 18, 16. Is the difference between the sample means significant?
Ans:No,t=0.93

24. The average production of 16 workers in a factory was 107 with SD of 9, while 12 workers in another comparable factory had an average production of 111 with SD of 10. Can we say that the production rate of workers in the latter factory is more than that in the former factory?
Ans: No, t=1.067

25. The following table gives the number of fatal road accidents that occurred during the 7 days of the week. Find whether the accidents are uniformly distributed over the week.
Ans:χ^2=4.17, accidents occur uniformly

Day	**Sun**	**Mon**	**Tue**	**Wed**	**Thu**	**Fri**	**Sat**
Number	8	14	16	12	11	14	9

26. 1000 families were selected at random in a city to test the belief that high income families usually send their children to public schools and the low income families often said their children to government schools. From the following results test whether income and type of schooling are independent.
Ans:χ^2=22.5, reject H_0

Income	**School**	
	Public	**Govt.**
Low	370	430
High	130	70

27. Three samples are taken comprising 120 doctors, 150 advocates and 130 university teachers. Each person chosen is asked to select one of the three categories that best represents his feeling toward a certain national policy. The three categories are in favour of the policy(F), against the policy(A), and indifferent toward the policy(I). The results of the interviews are given below. On

the basis of this data can it be concluded that the views Doctors, Advocates, and University teachers are homogeneous in so far as National policy under discussion is concerned.

Ans:χ^2=27.237, reject H_0

Occupation	Reaction		
	F	**A**	**I**
Doctors	80	30	10
Advocates	70	40	40
University teachers	50	50	30

28. A marketing agency gives you the following information about age groups of the sample informants and their liking for a particular model of scooter which a company plans to introduce. On the basis of the data can it be concluded that the model appeal id independent of the age group of the informants?

Ans:χ^2=42.788, reject H_0

	Age group of informants		
	Below 20	**20 – 39**	**40 – 59**
Liked	125	420	60
Disliked	75	220	100

29. A certain drug is claimed to be effective in curing cold. In an experiment on 500 persons with cold, half of them were given the drug and half of them were given the sugar pills. The patient's reaction to the treatment are recorded and given below. On the basis of this data, can it be concluded that the drug and sugar pills differ significantly in curing cold?

Ans:χ^2=3.52, do not differ significantly

	Helped	**Harmed**	**No effect**
Drug	150	30	70
Sugar Pills	130	40	80

01. **Test of significance of the difference between sample proportion and population proportion**

The test statistic $z = \dfrac{p - P}{\sqrt{\dfrac{PQ}{n}}}$

02. **Test of significance of the difference between two sample proportions:**

The test statistic $z = \dfrac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ where $P = \dfrac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$

$Q = 1 - P$

03 **Test of significance of the difference between sample mean and population mean:**

$z = \dfrac{\bar{x} - \mu}{S/\sqrt{n}}$; $H_0: \bar{x} = \mu$, $H_1: \bar{x} \neq \mu$; $H_1: \bar{x} > \mu$; $H_1: \bar{x} < \mu$

04. **Test of significance of the difference means of two samples:**

$z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$ Here $H_0: \bar{x}_1 = \bar{x}_2$; $H_1: \bar{x}_1 \neq \bar{x}_2$; $H_1: \bar{x}_1 > \bar{x}_2$, $H_1: \bar{x}_1 < \bar{x}_2$

NOTE:

01 The table value of z

5 % level : Two tailed test :- 1.96

one tailed test :- 1.645

02 Type II Error :- Accept H_0 when it is False

Type I Error :- Reject H_0 when it is True.

Page 1

PROBLEMS:-

01. The fatality rate of typhoid patients is believed to be 17.26%. In a certain year 640 patients suffering from typhoid were treated in a Metropolitan hospital and 63 patients died. Can you consider the hospital efficient

Solution:-

$H_0: p = P$ i.e. The hospital is not efficient.

$H_1: p < P$

Given: $P = 17.26\% = 0.1726$; $Q = 1 - P = 1 - 0.1726 = 0.8274$; $p = \frac{63}{640} = 0.0984$

$$z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.0984 - 0.1726}{\sqrt{\frac{0.1726 \times 0.8274}{640}}} = -4.96$$

$\therefore |z| = 4.96$

The table value of z at 5% level (one tailed test) = 1.645

The calculated value of z is greater than the table value. Hence H_0 is rejected i.e. The hospital is efficient.

---

02. In a large city A, 20% of a random sample of 900 school boys had a slight physical defect. In another city B, 18.5% of a random sample of 1600 school boys had the same defect. Is the difference between the proportions significant?

Solution:

Given: $p_1 = 20\% = 0.2$; $p_2 = 18.5\% = 0.185$, $n_1 = 900$, $n_2 = 1600$.

$H_0: P_1 = P_2$; $H_1: P_1 \neq P_2$

Now, $P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{180 + 296}{900 + 1600} = 0.1904$

$$z = \frac{P_1 - P_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$z = \frac{0.2 - 0.185}{\sqrt{0.1904 \times 0.8096 \times \left(\frac{1}{900}+\frac{1}{1600}\right)}} = 0.92$$

The table value of z at 5% level = 1.96

Since the calculated value of z is with in the table value Hence H_0 is accepted. i.e. The difference between P_1 and P_2 is not significant.

---

03. A Sample of 100 Students is taken from a large Population. The mean height of the Students in this Sample is 160 cm. Can it be reasonably regarded that, in the Population, the mean height is 165 cm, and the SD is 10 cm?

Solution:-

Given; $\bar{x} = 160$, $n = 100$, $\mu = 165$, $\sigma = 10$

$H_0: \bar{x} = \mu$; $H_1: \bar{x} \neq \mu$

$$z = \frac{\bar{x} - \mu}{S/\sqrt{n}} = \frac{160 - 165}{10/\sqrt{100}} = -5$$

$$|z| = 5$$

The table value of z at 5% level is = 1.96

Since the calculated value of z is greater than the table value.

Hence H_0 is accepted

i.e. It is not Statistically Correct to assume that $\mu = 165$.

04 A simple sample of heights of 6400 English men has a mean of 170cm and an standard deviation of 6.4 cm. While a simple sample of heights of 1600 Americans has a mean of 172 cm and an SD 6.3 cm. Do the data indicate the Americans are, on the average, taller than the Englishmen?

Solution:

Given $n_1 = 6400$, $\bar{x}_1 = 170$, $S_1 = 6.4$, $n_2 = 1600$, $\bar{x}_2 = 172$, $S_2 = 6.3$

$H_0: \bar{x}_1 = \bar{x}_2$; $H_0: \bar{x}_1 < \bar{x}_2$

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{170 - 172}{\sqrt{\frac{(6.4)^2}{6400} + \frac{(6.3)^2}{1600}}} = -11.32$$

$\therefore |z| = 11.32$ The table value of z at 5% level = 1.645.

Since the calculated value of z is greater than the table value. Hence H_0 is Rejected. (ie) Americans are on the average, taller than Englishmen.

01. **Students t-test:**

$$|t| = \frac{\bar{x} - \mu}{S/\sqrt{n}} \quad \text{where} \quad S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

Here $H_0: \bar{x} = \mu$; $H_1: \bar{x} \neq \mu$, $H_1: \bar{x} > \mu$, $H_1: \bar{x} < \mu$

The total no of degrees of freedom $v = n-1$

Note:- If standard deviation is given $|t| = \frac{\bar{x} - \mu}{S/\sqrt{n-1}}$

02. **Student's t test for difference means:**

$$|t| = \frac{\bar{x}_1 - \bar{x}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{where} \quad S^2 = \frac{\sum (x_i - \bar{x}_1)^2 + \sum (x_i - \bar{x}_2)^2}{n_1 + n_2 - 2}$$

Here $H_0: \bar{x}_1 = \bar{x}_2$; $H_1: \bar{x}_1 \neq \bar{x}_2$; $H_1: \bar{x}_1 > \bar{x}_2$, $H_1: \bar{x}_1 < \bar{x}_2$

The total no of degrees of freedom $v = n_1 + n_2 - 2$

Note:- If standard deviation is given, $S^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}$

03. **F test: [Variance Test]:**

$$F = \frac{S_1^2}{S_2^2} \text{ (or) } \frac{S_2^2}{S_1^2} \quad \text{where} \quad S_1^2 = \frac{\sum (x_i - \bar{x}_1)^2}{n_1 - 1}$$

$$S_2^2 = \frac{\sum (x_i - \bar{x}_2)^2}{n_2 - 1}$$

$H_0: \sigma_1^2 = \sigma_2^2$

$H_1: \sigma_1^2 \neq \sigma_2^2$

(or)

$$F = \frac{\text{larger variance}}{\text{Smaller variance}}$$

The total no of degrees of freedom $v = (n_1 - 1, n_2 - 1)$ (or) $v = (n_2 - 1, n_1 - 1)$

01. Tests made on the breaking strength of 10 pieces of metal wire gave the results 578, 572, 570, 568, 572, 570, 570, 572, 596 and 584 kg. Test if the mean breaking strength of the wire can be assumed as 577 kg.

<u>Solution:</u>

Let $H_0: \bar{x} = \mu \ (577)$; $H_1: \bar{x} \neq \mu$

Now, $\bar{x} = \dfrac{578 + 572 + 570 + 568 + 572 + 570 + 570 + 572 + 596 + 584}{10}$

$= \dfrac{5752}{10} = 575.2$

x	$(x_i - \bar{x})^2$
578	7.84
572	10.24
570	27.04
568	51.84
572	10.24
570	27.04
570	27.04
572	10.24
596	432.64
584	77.44
	681.6

Now, $S^2 = \dfrac{\sum (x_i - \bar{x})^2}{n-1}$

$= \dfrac{681.6}{10-1}$

$S^2 = 75.73$

$S = \sqrt{75.73}$

$= 8.702$

Now, $|t| = \dfrac{\bar{x} - \mu}{S/\sqrt{n}}$

$= \dfrac{575.2 - 577}{8.702/\sqrt{10}}$

$|t| = \dfrac{-1.8}{2.751}$

$t = 0.654$

The total no. of degrees of freedom $v = n-1$
$= 10-1 = 9$

For 9 degrees of freedom table value of t at 5% level is $= 2.26$

Since the calculated value is with in the table value. Hence H_0 is accepted. ... strength of the wire can be assumed as 577 kg.

02. A certain injection administered to each of 12 patients resulted in the following increases of blood pressure. 5, 2, 8, -1, 3, 0, 6, -2, 1, 5, 0, 4. Can it be concluded that the injection will be, in general accompanied by an increase in B.P.

Solution:

Let $H_0: \bar{x} = \mu\ (0)$; $H_1: \bar{x} > \mu$

Now, $\bar{x} = \dfrac{5+2+8-1+3+0+6-2+1+5+0+4}{12} = \dfrac{31}{12} = 2.58$

x	$(x-\bar{x})^2$
5	5.8564
2	0.3364
8	29.3764
-1	12.8164
3	0.1764
0	6.6564
6	11.6964
-2	20.9764
1	2.4964
5	5.8564
0	6.6564
4	2.0164
	104.9168

Now, $s^2 = \dfrac{\sum (x_i - \bar{x})^2}{n-1}$

$= \dfrac{104.9168}{12-1}$

$s^2 = 9.537$

$s = 3.088$

$|t| = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} = \dfrac{2.58 - 0}{3.088/\sqrt{12}}$

$= \dfrac{2.58}{0.8914}$

$= 2.89$

The total no. of degrees of fredom $v = n-1$
$= 12-1 = 11$

for 11 degrees of freedom the table value for t at 5% level is (one-tailed test) $= 1.80$

Calculated value of t is greater than table value. Hence H_0 is Rejected. (ie) we may conclude that the injection is accompained by an increase in B.P

03 A machinist is expected to make engine parts with axle diameter of 1.75 cm. A Random Sample of 10 parts shows a mean diameter 1.85 cm with Standard deviation of 0.1 cm. On the basis of this sample, would you say that the work of the machinist is inferior?

Solution:

Given $\bar{x} = 1.85$, $s = 0.1$, $n = 10$, $\mu = 1.75$

Here $H_0: \bar{x} = \mu$; $H_1: \bar{x} \neq \mu$

$$|t| = \frac{\bar{x} - \mu}{s/\sqrt{n-1}} = \frac{1.85 - 1.75}{0.1/\sqrt{10-1}} = \frac{0.1}{\frac{0.1}{3}} = 3$$

From the t-table for $v = 9$; $t_{0.05} = 2.26$

Since the calculated value of t is greater than the table, Hence H_0 is rejected. Hence the work of the machinist can be assumed to be inferior.

---

04. The nicotine contents in two random samples of tobacco are given below.

Sample 01 : 21 24 25 26 27

Sample 02 : 22 27 28 30 31 36

Can you say that the two samples came from the same Population?

Solution:

(i) **F - Test:-**

Here $H_0: \sigma_1^2 = \sigma_2^2$; $H_1: \sigma_1^2 \neq \sigma_2^2$

Now; $\bar{x}_1 = \frac{21+24+25+26+27}{5}$

$= \frac{123}{5} = 24.6$

$\bar{x}_2 = \frac{22+27+28+30+31+36}{6}$

$= \frac{174}{6} = 29$

x_1	$(x_1-\bar{x}_1)^2$ $(x_1-24.6)^2$	x_2	$(x_2-\bar{x}_2)^2 = (x_2-29)^2$
21	12.96	22	49
24	0.36	27	04
25	0.16	28	01
26	1.96	30	01
27	5.76	31	04
		36	49
	21.2		108

Now, $S_1^2 = \frac{\sum (x_1-\bar{x}_1)^2}{n_1-1} = \frac{21.2}{5-1} = 5.3$

$S_2^2 = \frac{\sum (x_1-\bar{x}_2)^2}{n_2-1} = \frac{108}{6-1} = 21.6$

$F = \frac{\text{Larger variance}}{\text{Smaller variance}} = \frac{S_2^2}{S_1^2}$

$= \frac{21.6}{5.3}$

$F_c = 4.07$

The table value of $F = (n_2-1, n_1-1)$ (ie) $(5, 4)$ degrees of freedom at 5% level is 6.26.

Since the calculated value of F is with in the table value. Hence H_0 is accepted

The variances of the two Populations can be regarded as equal

(ii) **Students t test:-** $H_0: \bar{x}_1 = \bar{x}_2$; $H_1: \bar{x}_1 \neq \bar{x}_2$

Now, $$|t| = \frac{\bar{x}_1 - \bar{x}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Now, $$S^2 = \frac{\sum (x_i - \bar{x}_1)^2 + \sum (x_i - \bar{x}_2)^2}{n_1 + n_2 - 2}$$

$$= \frac{21.2 + 108}{5 + 6 - 2}$$

$$S^2 = 14.35$$

$$S = \sqrt{14.35} = 3.78$$

$$|t| = \frac{24.6 - 29}{3.78\sqrt{\frac{1}{5} + \frac{1}{6}}}$$

$$|t| = \frac{-4.4}{3.78 \times 0.605}$$

$$|t| = -1.92$$

$$|t| = 1.92$$

The total no of degrees of freedom $v = n_1 + n_2 - 2$

$= 5 + 6 - 2$

$= 9$

The table value of t at 5% level is for 9 degrees of freedom $= 2.26$

Since the calculated value of t is not greater than the table value, hence H_0 is accepted.

(iii) The means of two samples do not differ significantly.

Conclusion:- The two samples could have been drawn from the same normal population.

15. Samples of two types of electric bulbs were tested for length of life and the following data were obtained.

	Size	mean	Standard deviation
Sample 01	8	1234 hours	36 hours
Sample 02	7	1036 hours	40 hours

Is the difference in the means sufficient to warrant type I bulbs are superior to type II bulbs?

Solution:

Given $H_0: \bar{x}_1 = \bar{x}_2$; $H_1: \bar{x}_1 > \bar{x}_2$

and $\bar{x}_1 = 1234$, $S_1 = 36$, $n_1 = 8$, $\bar{x}_2 = 1036$, $n_2 = 7$, $S_2 = 40$.

Now $S^2 = \dfrac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2} = \dfrac{8(36)^2 + 7(40)^2}{8+7-2} = \dfrac{21568}{13} = 1659.07$

$S = \sqrt{1659.07} = 40.73$

$|t| = \dfrac{\bar{x}_1 - \bar{x}_2}{S\sqrt{1/n_1 + 1/n_2}} = \dfrac{1234 - 1036}{40.73\sqrt{\frac{1}{8} + \frac{1}{7}}} = \dfrac{198}{40.73 \times 0.5175} = 9.39$

The total no. of degrees of freedom $\nu = n_1 + n_2 - 2$

$= 8 + 7 - 2$

$= 13$

For 13 degrees of freedom the table value of t at 5% level is (one-tailed test) $= 1.77$

Since the calculated value of t is greater than the table value

Hence H_0 is rejected.

(a) Type I bulbs may be regarded superior to type II bulbs.

PROBLEMS FOR PRACTICE

01. The following table gives the Biological values of Protein from cow's milk and buffalo's milk at a certain level. Examine if the average values of Protein in the two samples significantly differ.

Cow's milk :	1.82	2.02	1.88	1.61	1.81	1.54
Buffalo's milk :-	2.00	1.83	1.86	2.03	2.19	1.88

[Ans:- $\bar{x}_1 = 1.78$, $\bar{x}_2 = 1.965$, $|t| = 2.03$, table value at $v = 10 = 2.23$]

02. Two independent samples of eight and seven items respectively had the following values of the variable.

Sample 01 :-	9	11	13	11	15	9	12	14
Sample 02 :-	10	12	10	14	9	8	10	

Do the two estimates of Population variance differ significantly at 5% level of significance?

[$F = \frac{4.79}{3.96} = 1.21$, F (table $(7, 6)$) = 4.21, Ho is accepted]

Chi Square Test:- χ^2 Test:-

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$, O - observed frequency , E - Expected frequency.

The total no of degrees of freedom $v = n-1$

(or) $v = (r-1)(c-1)$

Uses of χ^2 distribution:-

(i) χ^2 distribution is used to test the goodness of fit.

(ii) It is used to test the independence of attributes.

Problems:

Q1 The following data give the number of aircraft accidents that occured during the various days of a week.

Day :-	Mon	Tues	Wed	Thu	Fri	Sat
No of accidents :-	15	19	13	12	16	15

Test whether the accidents are uniformly distributed over the week.

Solution:

H_0: Accidents occur uniformly over the week.

H_1: Accidents do not occur uniformly over the week.

$$E = \frac{15+19+13+12+16+15}{6} = 15.$$

O :	15	19	13	12	16	15
E :	15	15	15	15	15	15
$(O-E)^2$:	0	16	4	9	1	0

$$\therefore \chi^2 = \sum \frac{(O-E)^2}{E} = \frac{0}{15} + \frac{16}{15} + \frac{4}{15} + \frac{9}{15} + \frac{1}{15} + \frac{0}{15} = \frac{30}{15} = 2$$

The total no of degrees of freedom $v = n-1 = 6-1 = 5$

For 5 degrees of freedom the table value of χ^2 at 5% level = 11.07

Since the calculated of value of χ^2 is with in table value hence H_0 is accepted

02 Theory predicts that the proportion of beans in 4 groups A, B, C, D should be 9 : 3 : 3 : 1. In an experiment among 1600 beans, the numbers in the 4 groups were 882, 313, 287 and 118. Does the experiment support the theory?

Solution:-

H_0: The experiment support the theory.

H_1: The experiment do not support the theory.

O :	882	313	287	118
E :	$\frac{9}{16} \times 1600 = 900$	$\frac{3}{16} \times 1600 = 300$	$\frac{3}{16} \times 1600 = 300$	$\frac{1}{16} \times 1600 = 100$
$(O-E)^2$:	324	169	169	324

$$\psi^2 = \Sigma \frac{(O-E)^2}{E} = \frac{324}{900} + \frac{169}{300} + \frac{169}{300} + \frac{324}{100} = 4.72$$

The total no. of degrees of freedom $v = n-1 = 4-1 = 3$

For 3 degrees of freedom the table value at 5% level is = 7.82

Since the calculated value is within the table value, hence H_0 is accepted.

i.e. the experimental data support the theory.

---

03 The following data is collected on two characters. Based on this can you say that there is no relation between smoking and literacy?

	Smokers	Non Smokers
Literates	83	57
Illiterates	45	68

Solution:-

H_0 :- Literacy and smoking habit are Independent.

H_1 : Literacy and Smoking habit are dependent.

Given	Smokers	Non Smokers	Row total
Literates	83	57	140
Illiterates	45	68	113
Column Total	128	125	Grand Total = 253

O	E = $\frac{\text{Row total} \times \text{Column total}}{\text{Grand total}}$	$\frac{(O-E)^2}{E}$
83	$\frac{140 \times 128}{253} = 70.83 \simeq 71$	$12^2/71 = 2.03$
57	$\frac{140 \times 125}{253} = 69.17 \simeq 69$	$12^2/69 = 2.09$
45	$\frac{113 \times 128}{253} = 57.17 \simeq 57$	$12^2/57 = 2.53$
68	$\frac{113 \times 125}{253} = 55.83 \simeq 56$	$12^2/56 = 2.57$

$\chi^2 = \sum \frac{(O-E)^2}{E} = 2.03 + 2.09 + 2.53 + 2.57$

$= 9.22$

Total no. of degrees of freedom $v = (r-1)(c-1)$

$= (2-1)(2-1) = 1$

For 1 degrees of freedom the table value of χ^2 at 5% level is 3.84

Since the calculated value of χ^2 is greater than the table value

Hence H_0 is Rejected

i.e. There is some association between literacy and Smoking.