Chapter 6(3)

Chi-Square Test

6(3).01 Introduction

In many statistical test, the samples came from normal populations. When this assumption cannot be justified, it is necessary to use procedures do not require that this conditions be met. These procedures are generally referred to as non-parametric methods. In this chapter we study only χ^2-test for small sample (n < 30). In most introductory statistics books, n=30 is treated as the break point between large and small samples.

The chi-square test is based on Chi-square distribution which was first used by Karl Pearson in the year 1900. It is denoted by χ^2 distribution.

6(3).02 Chi- square Distribution:

For large sample size, the sampling distribution of χ^2 can be closely approximated by a continuous curve is known as the chi-distribution.

The probability function of χ^2 distribution is given by

$$f(x^2) = c(\chi^2)^{\frac{1}{2}(v-2)} e^{\frac{\chi^2}{2}} \quad ; 0 \le \chi^2 < \infty$$

Where, v = Degrees of freedom = n-1, c= a constant depending only on the degree of freedom v, e= 2.71828.

The chi-square distribution has only one parameter v. This distribution is similar to the case of the t-distribution.

Properties or characteristics of the chi-square test or the chi-square distribution:

The important properties of chi-square distribution are given below:

(i) The mean of the chi-square distribution is given by the degree of freedom. That is, $E(\chi^2) = \nu$.

(ii) The variance of the chi-square distribution is twice the degree of freedom.

That is, $V(\chi^2) = 2\nu$.

(iii) the exact shape of the chi-square distribution depends on degree of freedom.

(iv) The sum of independent chi-square variates is also a chi-square variate.

(v) The chi-square distribution is a continuous probability distribution which has the value zero at its lower limit and extends to infinity in the positive direction negative value of the chi-square is not possible.

(vi) The chi-square curve is always positively skewed.

(vii) Chi-square values increases with the increase in degree of freedom.

(viii) The value of the chi-square lies between zero to infinite.

(ix) If the degree of freedom is increase, then the chi-square distribution tends to symmetrical distribution.

(x) The chi-square (χ^2) distribution is skewed when its degree of freedom is greater than 1.

(xi) The chi-square (χ^2) distribution tends to normal distribution when its degree of freedom is one (1).

(xii) The chi-square (χ^2) is only a parameter of a distribution. Its parameter is its degree of freedom.

All the Chi-square tests are one-tailed test and only the large values of χ^2 announce the rejection of the null hypothesis.

Uses of the chi-square distribution:

The chi-square distribution is a very powerful test for testing the hypothesis of a number of statistical problems. the important uses of the chi-square distribution are:

(i) Test of Association or Independence

(ii) Test of Homogeneity

(iii) Test of Goodness of Fit

All the Chi-square tests are one-tailed test and only the large values of χ^2 announce the rejection of the null hypothesis.

6(3).03 Chi-square (χ^2) test

The χ^2- test is one of the simplest and widely used non-parametric test. It was first used by Karl Pearson (1890). The quantity (χ^2) describes the magnitude of the difference between the observed set of observations and expected set of observations under an appropriate null hypothesis.

Conditions or assumptions of the chi-square (χ^2) test :

The following six basic conditions must be met in order for χ^2 analysis to be applied:

(i) The sample should be contain at least 50 observations.

(ii) The sample data must be drawn at random from the target population.

(iii) The sample observation must be independent of each other.

(iv) The data should be expressed in original units for convenience of comparison and not in percentage form.

(v) The expected frequency of any item or cell should not be less than 5. If it is less than 5, then frequencies from the adjacent items or cells are pooled together in order to make it 5 or more than 5.

(vi) The total of the observed frequencies and the expected frequencies are equal. That is, if the observed frequencies are $O_1, O_2, \ldots, O_k$ respectively and the expected frequencies are $E_1, E_2, \ldots, E_k$ respectively, then $\sum_{i=1}^{k} O_i = \sum_{i=1}^{k} E_i$.

Let $O_1, O_2, \ldots, O_k$ be the set of observed observations and $E_1, E_2, \ldots, E_k$ be the set of expected observations. Then the test statistic,

$$\chi^2_{(v)} = \frac{\sum_{i=1}^{k} (O_i - E_i)^2}{\sum_{i=1}^{k} E_i}, \text{ with degrees of freedom } v.$$

Where $v = k - r$ and r = The number of independent restrictions imposed.

In a χ^2-test the number of imposed restriction is only one and that is $\sum_{i=1}^{k} O_i = \sum_{i=1}^{k} E_i$.

Therefore, the degrees of freedom, $v = k-1$.

In a contingency table the degrees of freedom are calculated in a slightly different manner. The marginal total place the restriction in selecting cell frequencies.

For example, a 2×2 contingency table is given below :

	Red dress	Black dress	Marginal total
Boys	a	b	a + b
Girls	c	d	c + d
Marginal total	a+c	b+d	N=a+b+c+d (Grand total)

Thus for a 2×2 table, the number of cell frequencies assigned arbitrarily

$= (c - 1)(r - 1)$ [$\because$ c = Number of column, r = Number of row]

$= (2 - 2)(2 - 1) = 1 \times 1 = 1$

In this case, Chi-square statistic, $\chi^2 = \dfrac{N(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)} \sim \chi^2_{(r-1)(c-1)}$

Where, $(r-1)(c-1)$ is the degree of freedom.

Uses of χ^2-test :

Nowadays the χ^2-test is the most popular and widely used statistical inference procedure. It has a large number of applications as following :

(i) Test of Association or Independence

(ii) Test of Homogeneity

(iii) Test of Goodness of Fit

(iv) It is used to test the sample variance which is taken from the normal distribution.

(v) It is used to test the independence of the qualitative variables.

(vi) It is used to test the equality of the several correlation coefficients.

(vii) It is used to test the equality of the several proportions.

(viii) It is used to test the equality of the variances of the several distributions.

6(3).04 Yates' Corrections for continuity:

Yates correction is also known as Yates correction for continuity. It is used when at least one cell frequency is less than 5. The correction rule is to adjust the observed frequency in each cell of the 2×2 table. The adjustment is performed in such a way that the marginal total remains same. The operation will increase two cells by $\frac{1}{2}$ and will reduce two other cells by $\frac{1}{2}$, the marginal totals remain unchanged.

The corrected χ^2 value is computed as follows :

$$\chi^2 = \frac{N(|ad - bc| - \frac{N}{2})^2}{(a+b)(a+c)(c+d)(b+d)}$$

Example-1: By using the following information, find out whether there is any relationship between smoking and drinking

	Drinking	Not-drinking
Smoking	4	26
Not-smoking	10	30

Solution : We want to test, Ho : There is no relationship between smoking and drinking

Vs H_1 : There is some relation between smoking and drinking.

Since one cell frequency is less than 5. So the corrected χ^2 value is computed as follows :

$$\chi^2 = \frac{N(|ad\text{-}bc| - \frac{N}{2})^2}{(a+b)(a+c)(c+d)(b+d)}, \quad \text{with d.f.} = (2\text{-}1)(2\text{-}1) = 1$$

Here, a = 4, b = 26, c = 10, d = 30 and N = a+b+c+d = 4 + 26 + 10 + 30 = 70

$$\therefore \chi^2 = \frac{70(|4\times30 - 26\times10| - \frac{70}{2})^2}{(4+26)(4+10)(10+30)(26+30)}$$

$$= \frac{771750}{940800} = 0.82$$

The critical value of χ^2 with 1 d.f. at 5% level of significance is $\chi^2_{0.05,\ 1.} = 3.84$

Since $\chi^2 < \chi^2_{0.05,\ 1}$. So, the null hypothesis is not rejected. Therefore, there is no relationship between smoking and drinking.

6(3).05 Grouping when Frequencies are small:

If the value of frequencies is less than 10 and certainly not less than 5, then it is generally possible to overcome the difficulty by grouping two or more classes together. In other words, befor calculating the difference between observed and expected frequencies, one or more classes with theoritical frequencies less than 5 may be combined into a single category. In this case, the number of degree of freedom would be determined with number of observations after the regrouping.

For example:

Observed frequency	420	400	290	100	30	12	2	1
Expected frequency	410	390	300	90	27	8	1	0

The last three classes should be combined together. After grouping, the position would be as follows:

Observed frequency	420	400	290	100	30	15
Expected frequency	410	390	300	90	27	9

Here, degrees of freedom 8 - 2-1=5.
(N.B.: degrees of freedom are two less than the number of observations.)

Some important applications of chi-square test are discussed below:

6(3).05(i) Sampling Distribution of the Sample Variance:

The sampling distribution of the sample variance s^2 is concerned about the variability in a random sample. Since the value of sample variance (s^2) is cannot be negative, the distribution of the sample variance s^2 cannot be a normal distribution. In this case, the distribution the sample variance s^2 is a unimodal distribution which is positively skewed, is a chi-square distribution.
If the random sample of size n with the sample variance s^2 is drawn from the population with population variance σ^2, then we have the following relation:

$$s^2 = \frac{\chi^2\sigma^2}{v}$$

$$\Rightarrow s^2 = \frac{\chi^2\sigma^2}{n-1} \quad \text{[Where, d.f. v=n-1]}$$

$$\Rightarrow \chi^2 = \frac{(n-1)s^2}{\sigma^2}, \text{ which follows chi-square distribution.}$$

6(3).05(ii) Confidence interval for variance:

A 100(1- α) percent confidence interval for population variance σ^2 is constructed by first obtaining an interval about $\frac{(n-1)s^2}{\sigma^2}$. Two values of chi-square are selected from the table (Chi-square table in the appendix) such that $\alpha/2$ is to the left of the smaller value and $\alpha/2$ is to the right of the larger values.

Therefore, a 100(1- α)% confidence interval for $\frac{(n-1)s^2}{\sigma^2}$ is given by-

$$-\chi_{\alpha/2}^{2} < \frac{(n-1)s^2}{\sigma^2} < \chi_{(1-\alpha/2)}^{2}$$

By solving the above inequalities for σ^2, we get,

$$\frac{(n-1)s^2}{\chi_{1-\alpha/2}^{2}} < \sigma^2 < \frac{(n-1)s^2}{\chi_{\alpha/2}^{2}}$$

Which is the 100(1- α) percent confidence interval for population variance σ^2.

6(3).05(iii) Tests of Hypothesis Concerning Variance:

If the random sample of size n with the sample variance s^2 is drawn from the population with population variance σ^2, then the test statistic is $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$.

The calculated value of χ^2 lies in the rejection region if $\chi^2 < \chi_{1-\alpha/2}{}^2$ and $\chi^2 > \chi_{\alpha/2}{}^2$.

In this case, we reject the null hypothesis. Otherwise, we accept the null hypothesis.

Example-2: Weights in kg. of 10 students in a class are given below:

40, 45, 53, 38, 47, 43, 55, 48, 52, 49.

Are you agree that the variance of the distribution of weight of all students in that class from which the above sample of 10 students was drawn is equal to 20^2 kg. and obtain 95% confidence interval to estimate the true value of the populations variance.

Solution : Let, the population variance is σ^2.

We want to test the null hypothesis, $H_0 : \sigma^2 = 20$ against $H_1 : \sigma^2 \neq 20$

We know the test statistic, $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$, with d.f. = n-1

Table for calculation of s^2

Obtained marks (x_i)	x_i^2
40	1600
45	2025
53	2809
38	1444
47	2209
43	1849
55	3025
48	2304
52	2704
49	2401
$\Sigma x_i = 470$	$\Sigma x_i^2 = 22370$

Sample variance, $s^2 = \frac{1}{n-1}\{\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}\} = \frac{1}{10-1}\{22370 - \frac{(470)^2}{10}\} = \frac{1}{9}(22370 - 22090) = 31.11$

$\therefore \chi^2 = \frac{(10-1) \times 31.11}{20} = 13.99$

$\therefore$ The calculated value is $\chi^2 = 13.99$, with d.f. = n-1 = 10-1 = 9

The critical value of χ^2 with d.f. 9 at 5% level of significance is $\chi^2_{0.025, 9} = 19.022$

Since $\chi^2 < \chi^2_{0.025, 9}$.

So, the null hypothesis is accepted. Hence, the variance of the distribution of weight of all students is 20^2 kg.

To construct 95% confidence interval:

We obtain, $\chi^2_{(1-\alpha/2)} = \chi^2_{0.975} = 0.05063$ and $\chi^2_{\alpha/2} = \chi^2_{0.025} = 7.3776$.

So, the 95% C.I. for σ^2 is $\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{(1-\alpha/2)}}$

$$\Rightarrow \frac{(10-1)\times 31.11}{7.3776} < \sigma^2 < \frac{(10-1)\times 31.11}{0.05063}$$

$$\Rightarrow 37.95 < \sigma^2 < 5530.12$$

6(3).06 Test of Independence

In the test of independence, the population and sample are classified according to some attributes. The test will indicate only, whether or not any dependency relationship exists between the attributes. It will not indicate the degree of association or direction of the dependency. To conduct the test, a sample is drawn from the population of interest and observed frequencies are cross-classified according to the two criteria.

Test of two attributes in a contingency table :

The cross classification can be conveniently displayed means of a table is called contingency table.

Let, A and B be the two attributed characteristics. Attribute A having r mutually exclusive, exhaustive categories and attribute B having c mutually exclusive, exhaustive categories.

Suppose, O_{ij} (i = 1,2, -----, r ; j = 1, 2, ------, c) are the observed frequencies of i th row and j th column. Let, R_i (i = 1, 2, -----, r) are row total of i th row's observed frequencies and C_j (j = 1, 2, ------- ,c) are column total of j th column observed frequencies.

That is, $R_i = \sum_{j=1}^{c} O_{ij}$, $C_j = \sum_{i=1}^{r} O_{ij}$, i = 1, 2, ------, r and j = 1, 2, ----, c .

Hence the total observed frequencies, $N = \sum_{i=1}^{r} R_i = \sum_{j=1}^{c} C_j$

The following table showed of a standard contingency table of order r × c according to the number of rows and columns.

		Attribute B				Total
		B_1	B_2 ----------	B_j ----------	B_c	
Attribute A	A_1	O_{11}	O_{12} ----------	O_{1j} ----------	O_{1c}	R_1
	A_2	O_{21}	O_{22} ----------	O_{2j} ----------	O_{2c}	R_2
	-------	-------	-------	-------		-------
	A_i	O_{i1}	O_{i2} ----------	O_{ij} ----------	O_{ic}	R_i
	-------	-------	-------	-------		-------
	A_r	O_{r1}	O_{r2} ----------	O_{rj} ----------	O_{rc}	R_r
Total		C_1	C_2 ----------	C_j ----------	C_c	$N=\Sigma R_i = \Sigma C_j$

Now, let E_{ij} (i = 1, 2, --------, r; j = 1, 2, --------, c) are Expected frequencies of the observed frequencies O_{ij}, which can be determined in the following way :

$$E_{ij} = \frac{R_i \times C_j}{N} \quad (i = 1, 2, \text{--------}, r; \quad j = 1, 2, \text{--------}, c)$$

$$\text{Test statistic, } \chi^2 = \Sigma\Sigma \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2_{(r-1)(c-1)}$$

$$= \Sigma\Sigma \frac{O_{ij}^2}{E_{ij}} - N \quad \text{or, } \sum \frac{(O-E)^2}{E^2}$$

Here, df. = (r-1) (c-1)

Example-3 : The following table shows the number of recruits taking a preliminary and a final test in car driving.

		Preliminary	
		Pass	Fail
Final	Pass	605	135
	Fail	195	65

Use χ^2 test to discuss whether there is any association between the results of the preliminary and those of the final test.

Solution: We want to test,

Ho: There is no association between the results of preliminary and final test.

H_1 : There is association between results of preliminary and final test.

We know the test statistic,

$$\chi^2 = \Sigma\Sigma \frac{(O_{ij} - E_{ij})^2}{E_{ij}}, \text{ with d.f.} = (r - 1)(c - 1)$$

Where, O_{ij} = The observed frequency

and $E_{ij} = \frac{R_i \times C_j}{N}$ = The expected frequency

Table for the observed frequencies as follows :

		Preliminary		Total
		Pass	Fail	
Final	Pass	O_{11}= 605	O_{12}= 135	R_1= 740
	Fail	O_{21}= 195	O_{22}= 65	R_2= 260
Total		C = 800	C = 200	N = 1000

The expected frequencies are computed as follows :

$$E_{11} = \frac{R_1 \times C_1}{N} = \frac{740 \times 800}{1000} = 592 \; ; \qquad E_{12} = \frac{R_1 \times C_2}{N} = \frac{740 \times 200}{1000} = 148$$

$$E_{21} = \frac{R_2 \times C_1}{N} = \frac{260 \times 800}{1000} = 208 \; ; \qquad E_{22} = \frac{R_2 \times C_2}{N} = \frac{260 \times 200}{1000} = 52$$

$$\therefore \quad \chi^2 = \Sigma\Sigma \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \frac{(O_{11} - E_{11})^2}{E_{11}} + \frac{(O_{12} - E_{12})^2}{E_{12}} + \frac{(O_{21} - E_{21})^2}{E_{21}} + \frac{(O_{22} - E_{22})^2}{E_{22}}$$

$$= \frac{(605 - 592)^2}{592} + \frac{(135 - 148)^2}{148} + \frac{(195 - 208)^2}{208} + \frac{(65 - 52)^2}{52}$$

$$= \frac{169}{592} + \frac{169}{148} + \frac{169}{208} + \frac{169}{52} = 0.2855 + 1.1419 + 0.8125 + 3.2500 = 5.4899$$

$\therefore \chi^2_{cal.} = 5.4899$, with d.f.= (r - 1) (c - 1) = (2 - 1) (2 - 1) = 1 × 1 = 1

$\therefore$ The critical value of χ^2 with 1 d.f. at 5% level of significance is $\chi^2_{0.05,\,1.} = 3.84$

Since $\chi^2_{cal.} > \chi^2_{0.05,\,1}$, so the null hypothesis is rejected. That is, there is a significant association between the results of preliminary and final test.

6(3).07 Test of Goodness of Fit:

The chi-square (χ^2) test is a measure of probabilities of association between the attributes. It gives us an idea about the divergence between the observed and expected frequencies. Thus the test is also described as the test of "Goodness of fit". It is frequently of interest to explore the proposition that several populations are homogeneous with respect to some characteristic of interest. For example, we may be interested in knowing of some raw material available from several retailers is homogeneous.

Example- 4: The following table gives the number of aircraft accidents that occurred during the various days of the week. Find whether the accidents are uniformly distributed over the week.

Day	Sun.	Mon.	Tue.	Wed.	Thu.	Fri.	Sat.
No. of Accidents	14	16	8	20	11	9	14

Solution : We want to test, H_0 : Accidents are uniformly distributed over the week.

against, H_1 : Accidents are not uniformly distributed over the week.

We know, $\chi^2 = \Sigma \frac{(O-E)^2}{E}$, with d.f = n – 1

Here, O = The observed value, E = The expected value

Since the expected value (E) and the arithmetic mean ($\bar{x}$) are same in mathematically.

$$\therefore E = \frac{\text{Total number of accidents}}{\text{Total days}}$$

$$= \frac{14+16+8+20+11+9+14}{7} = \frac{92}{7} = 13.14$$

Table for calculation

Day	O	E	$(O-E)^2$	$(O-E)^2/E$
Sun.	14	13.14	0.7396	0.0563
Mon.	16	13.14	8.1796	0.6225
Tue.	8	13.14	26.4196	2.0106
Wed.	20	13.14	47.0596	3.5814
Thu.	11	13.14	4.5796	0.3485
Fri.	9	13.14	17.1396	1.3044
Sat.	14	13.14	0.7396	0.0563
				$\Sigma\frac{(O-E)^2}{E} = 7.98$

$\therefore \chi^2 = \Sigma\frac{(O-E)^2}{E} = 7.98$, $df = n - 1 = 7-1 = 6$

Comments : The calculated value of χ^2 is $\chi^2_{cal} = 7.98$ and the table value of χ^2 for dt. 6 at 5% level of significance is $\chi^2_{tab} = 12.59$.

Since, $\chi^2_{cal} < \chi^2_{tab}$. So, the null hypothesis is accepted. Hence, accidents are uniformly distributed over the week.

6(3).08 Test of Homogeneity :

It is frequently of interest to explore the proposition that several populations are homogeneous with respect to some characteristic of interest. We are interested in testing the null hypothesis that several populations are homogeneous with respect to the proportion of subject falling into several categories or some other criterion of classification. A random sample is drawn from each of the population and the number in each sample falling into each category is determined. The sample data is displayed in contingency table. The analytical procedure is same as that discussed for test of independence.

Example-5 : A random sample of 400 persons was selected from each of three age groups and each person asked to satisfy which of three types such as Drama, Talk show and Cinema of television programmes they preferred. The results are in the following table:

	Types of Programmes		
Age group	**Drama**	**Talk show**	**Cinema**
Under 30	120	30	50
30-40	10	75	15
45 and above	10	30	60

Test the hypothesis that the populations are homogeneous with respect to the types of television programmes they prefer.

Solution: We want to test, H_0 : The populations are homogeneous with respect to the types of television programmes they prefer.

We know, Test statistic, $\chi^2 = \Sigma \frac{(O-E)^2}{E}$, with df. = (r–1) (c-1)

Here, O = the observed frequency

$$E = \text{The expected frequency} = \frac{\text{Row total} \times \text{column total}}{\text{Total number of observation}}$$

Table for calculation

Age group	Drama	Talk show	Cinema	Total
Under 30	120	30	50	200
30-40	10	75	15	100
45 and above	10	30	60	100
Total	140	135	125	400

Arranging observed and expected frequencies in the following table:

O	E	$(O-E)^2$	$(O-E)^2/E$
120	70.00	2500.00	35.71
10	35.00	625.00	17.86
10	35.00	625.00	17.86
30	67.50	1406.25	20.83
75	33.75	1701.56	50.42
30	33.75	14.06	0.42
50	62.50	156.25	2.50
15	31.25	264.06	8.45
60	31.25	826.56	26.45
			$\Sigma\frac{(O-E)^2}{E} = 180.50$

$\therefore \chi^2 = \Sigma\frac{(O-E)^2}{E} = 180.50$, with df. = (3–1) (3-1)=4

Comments : The calculated value of χ^2 is $\chi^2_{cal} = 180.50$ and the table value of χ^2 for df 4 at 5% level of significance is $\chi^2_{tab} = 9.4877$.

Since, the calculated value of χ^2 is much greater than the table value. So, the null hypothesis is rejected. Hence, The populations are not homogeneous with respect to the types of television programmes they prefer.

6(3).09 Cautions while Applying chi-square test (χ^2 Test):

The chi-square test is very popularly used in practice. The chi-square test is unfortunate to find that the number of misuses of of the chi-square test has become surprisingly large. Some sources of error in the application of this test revealed are:

(i) Incorrect categorizing.

(ii) Application of non frequency data.

(iii) Small theoretical frequencies.

(iv) Failure to equalize the sum of the observed frequencies and the sum of the theoretical frequencies.

(v) Ignore of frequencies of non occurrence.

(vi) Indeterminate theoretical frequencies.

Mathematical Problems

The mathematical problems of this chapter can be divided into three formats.

Format (1) : Test about one variance value or Test for specified variance or population variance.

Format (2) : Test of Independence or, Test about the association (relationship or independence) between two attributes.

Format (3) : Test of Goodness of Fit

Tips to find the critical value or table value of χ^2- test:

Table-4 : Critical values of Chi-square (χ^2) .

Degree of	Upper Tails Areas (α)											
freedom	0.995	0.99	0.975	0.95	0.9	0.75	0.25	0.1	0.05	0.025	0.01	0.00
1			0.0010	0.0039	0.0158	0.1015	1.3233	2.7055	3.8415	5.0239	6.6349	7.8
2	0.0100	0.0201	0.0506	0.1026	0.2107	0.5754	2.7726	4.6052	5.9915	7.3778	9.2104	10.5
3	0.0717	0.1148	0.2158	0.3518	0.5844	1.2125	4.1083	6.2514	7.8147	9.3484	11.344	12.8
4	0.2070	0.2971	0.4844	0.7107	1.0636	1.9226	5.3853	7.7794	9.4877	11.143	13.276	14.8
5	0.4118	0.5543	0.8312	1.1455	1.6103	2.6746	6.6257	9.2363	11.070	12.832	15.086	16.7
6	0.6757	0.8721	1.2373	1.6354	2.2041	3,4546	7.8408	10.644	12.591	14.449	16.811	18.5
7	0.9893	1.2390	1.6899	2.1673	2.8331	4.2549	9.0371	12.017	14.067	16.012	18.475	20.2
8	1.3444	1.6465	2.1797	2.7326	3.4895	5.0706	10.218	13.361	15.507	17.534	20.090	21.9
9	1.7349	2.0879	2.7004	3.3251	4.1682	5.8988	11.388	14.683	16.919	19.022	21.666	23.5
10	2.1558	2.5583	3.2470	3.9403	4.8652	6.7372	12.548	15.987	18.307	20.483	23.209	25.1

For example:

The critical value of χ^2 with df 9 (which is the first coumn) at α = 5% = 0.05 level of significance

(a) $\chi^2_{cri.}$ or $\chi^2_{tab.}$ or $\chi^2_{\alpha, 9}$ = 16.919 (For one tailed test)

(b) $\chi^2_{cri.}$ or $\chi^2_{tab.}$ or $\chi^2_{\alpha, 9}$ = 19.022 (For two tailed test) .

Format (1) : Test for specified variance or population variance

Let, the population variance is σ^2.

We want to test the null hypothesis, $H_0 : \sigma^2 = \sigma_0^2$ against $H_1 : \sigma^2 \neq \sigma_0^2$

We know the test statistic, $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$, with d.f. = n-1

Problem- (1) : Heights in inches of 10 students are given below:

38, 40, 45, 53, 47, 43, 55, 48, 52, 49.

Test the variance of the distribution of height of all students from which the above sample of 10 students was drawn is equal to 20 square inches?

Solution : Let, the population variance is σ^2.

We want to test the null hypothesis, $H_0 : \sigma^2 = 20$ against $H_1 : \sigma^2 \neq 20$

We know the test statistic, $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$, with d.f. = n-1

Here, $s^2 = \frac{\sum(x-\bar{x})^2}{n-1}$ [Where, $\bar{x} = \frac{\sum x}{n}$]

Table for calculation of s^2

Height (in inch)	$x-\bar{x}$	$(x-\bar{x})^2$
38	-9	
40	-7	
45	-2	
53	6	
47	0	
43	-4	
55	8	
48	1	
52	5	
49	2	
$\sum x = 470$		$\sum(x-\bar{x})^2$=280

Sample mean, $\bar{x} = \frac{470}{10}$=47

Sample variance, $s^2 = \frac{280}{10-1} = 31.11$

$\therefore \chi^2 = \frac{(10-1) \times 31.11}{20} = 14$

$\therefore$ The calculated value is $\chi^2 = 14$, with d.f. = n-1 = 10-1 = 9

The critical value of χ^2 with d.f. 9 at 5% level of significance is 16.919

Since the calculated value of χ^2 is less than the critical value of χ^2. So, the null hypothesis is accepted. Hence, the variance of the distribution of height of all students in the population is 20 square inches.

Problem- (2) : Ten students are selected at random from a normal population and their obtained marks are found to be 60, 67, 71, 55, 66, 74, 81, 80, 77, 85. Do this data significantly satisfied that population variance is 27 and obtain 95% confidence interval to estimate the true value of the populations various.

Solution : Let, the population variance is σ^2.

We want to test the null hypothesis, $H_0 : \sigma^2 = 27$ against $H_1 : \sigma^2 \neq 27$

We know the test statistic, $\chi^2 = \dfrac{(n-1)s^2}{\sigma^2}$, with d.f. = n-1

Table for calculation of s^2

Obtained marks (x_i)	x_i^2
60	3600
67	4489
71	5041
55	3025
66	4356
74	5476
81	6561
80	6400
77	5929
85	7225
$\Sigma x_i = 716$	$\Sigma x_i^2 = 52102$

Sample variance, $s^2 = \frac{1}{n-1}\left\{\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}\right\} = \frac{1}{10-1}\left\{52102 - \frac{(716)^2}{10}\right\} = \frac{1}{9}(52102 - 51265.6) = 92.93$

$\therefore \chi^2 = \dfrac{(10-1) \times 92.93}{27} = 30.977$

$\therefore$ The calculated value is $\chi^2 = 30.977$, with d.f. = n-1 = 10-1 = 9

The critical value of χ^2 with d.f. 9 at 5% level of significance is 16.919

Since the calculated value of χ^2 is more than the critical value of χ^2. So, the null hypothesis is rejected. Hence, the given data is not significantly satisfied that the population varince is 27.

To construct 95% confidence interval:

We obtain, $\chi^2_{(1-\alpha/2)} = \chi^2_{0.975} = 0.05063$ and $\chi^2_{\alpha/2} = \chi^2_{0.025} = 7.3776$.

So, the 95% C.I. for σ^2 is $\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{(1-\alpha/2)}}$

$$\Rightarrow \frac{(10-1)\times 92.93}{7.3776} < \sigma^2 < \frac{(10-1)\times 92.93}{0.05063}$$

$$\Rightarrow 113.37 < \sigma^2 < 16519.26$$

Format (2) : Test of Independence or, Test about the association (relationship or independence) between two attributes.

Working structure :

We want to test, H_0 : There is no relationship / association (or, independence) between ....(1st variable)....and ....(2nd variable)......

Where H_1: There is a relationship / association (or; independence) between ....and ....

Test statistic, $\chi^2 = \Sigma\Sigma \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$, with d.f.= (r-1) (c-1)

Where, O_{ij} = The observed frequency, E_{ij} = The expected frequency $= \frac{R_i \times C_j}{N}$

Or, Test statistic, $\chi^2 = \Sigma \frac{(O-E)^2}{E}$, with df. = (r–1) (c-1)

Here, O = the observed frequency

E = The expected frequency $= \frac{\text{Row total} \times \text{column total}}{\text{Total number of observation}}$

From the given information, we have,

R_1= 1st row total, R_2 = 2nd row total, R_3 = 3rd row total etc.

C_1= 1st column total, C_2 = 2nd column total, C_3 = 3rd column total etc.

$E_{11} = \frac{R_1 \times C_1}{N}$, $E_{12} = \frac{R_1 \times C_2}{N}$, $E_{13} = \frac{R_1 \times C_3}{N}$,

$E_{21} = \frac{R_2 \times C_1}{N}$, $E_{22} = \frac{R_2 \times C_2}{N}$, $E_{23} = \frac{R_2 \times C_3}{N}$,

$E_{31} = \frac{R_3 \times C_1}{N}$, $E_{32} = \frac{R_3 \times C_2}{N}$, $E_{33} = \frac{R_3 \times C_3}{N}$,

By arranging the observed and expected frequencies in the table, $\Sigma\frac{(O-E)^2}{E} = *$

$\therefore \chi^2 = \Sigma\frac{(O-E)^2}{E} = *$, with df. = (r–1) (c-1) = (3–1) (3–1) = 4

The table value of χ^2 for 4 df at 5% level of significance is 9.4877.

Comment : If the calculated value of χ^2 is greater than the table value of χ^2. Then, the null hypothesis is rejected.

Problem - 3 : A residential hotel bought a total of 500 color television sets. Three different brands were purchased and their repair records were kept for each sets operation. The data is given below :

	Number of Repairs			
Brand	**0**	**1**	**2 or more**	**Total**
SONY	**143**	**70**	**37**	**250**
SINGER	**90**	**67**	**43**	**200**
RANGES	**17**	**13**	**20**	**50**
Total	**250**	**50**	**100**	**500**

Use an appropriate test at 5% level of significance to determine whether there is relationship between brand and number of repairs. [BBA, Third Sem., N. U. 2008]

Solution : We want to test, H_0 : There is no relationship between brand and number of repairs.

Where, H_1 : There is relationship between brand and number of repairs.

Test statistic, $\chi^2 = \Sigma \frac{(O-E)^2}{E}$, with df. = (r–1) (c-1)

Here, O = the observed frequency

E = The expected frequency $= \frac{\text{Row total} \times \text{column total}}{\text{Total number of observation}}$

From the given information, we have,

$R_1 = 250$, $R_2 = 200$, $R_3 = 50$

$C_1 = 250$, $C_2 = 150$, $C_3 = 100$, N= 500

$E_{11} = \frac{R_1C_1}{N} = \frac{250 \times 250}{500} = 125$, $E_{12} = \frac{R_1C_2}{N} = \frac{250 \times 150}{500} = 75$, $E_{13} = \frac{R_1C_3}{N} = \frac{250 \times 100}{500} = 50$,

$E_{21} = \frac{R_2C_1}{N} = \frac{200 \times 250}{500} = 100$, $E_{22} = \frac{R_2C_2}{N} = \frac{200 \times 150}{500} = 60$, $E_{23} = \frac{R_2C_3}{N} = \frac{200 \times 100}{500} = 40$

$E_{31} = \frac{R_3C_1}{N} = \frac{50 \times 250}{500} = 25$, $E_{32} = \frac{R_3C_2}{N} = \frac{50 \times 150}{500} = 15$, $E_{33} = \frac{R_3C_3}{N} = \frac{50 \times 100}{500} = 10$

Arranging the observed and expected frequencies in the following table :

O	E	$(O-E)^2$	$(O-E)^2/E$
143	125	324	2.592
90	100	100	1.000
17	25	64	2.560
70	75	25	0.333
67	60	49	0.817
13	15	4	0.267
37	50	169	3.380
43	40	9	0.225
20	10	100	10.000
			$\Sigma\frac{(O-E)^2}{E} = 21.174$

$\therefore \chi^2 = \Sigma\frac{(O-E)^2}{E} = 21.174$, df = (3–1) (3–1) = 4

The table value of χ^2 for 4 df at 5% level of significance is 9.4877.

Comment : Since the calculated value of χ^2 is greater than the table value of χ^2. So, the null hypothesis is rejected. Hence, there is relationship between brand and number of repairs.

Problem-(4) : From the following data find out whether there is any relationship between sex and preference for colour.

		Sex	
		Male	Female
Colour	Green	40	60
	White	35	25
	Yellow	25	15

Soluation : We want to test , H_0 : There is no relationship between sex and preference for colour.

Where H_1 : There is a relationship between sex and preference for colour.

We know, Test statistic, $\chi^2 = \Sigma \frac{(O-E)^2}{E}$, with df. = (r–1) (c-1)

Here, O = the observed frequency

E = The expected frequency $= \frac{\text{Row total} \times \text{column total}}{\text{Total number of observation}}$

From the given information, we have,

$R_1 = 100$, $R_2 = 60$, $R_3 = 40$

$C_1 = 100$, $C_2 = 100$, $N = 200$

$E_{11} = \frac{R_1 \times C_1}{N} = \frac{100 \times 100}{200} = 50$; $E_{12} = \frac{R_1 \times C_2}{N} = \frac{100 \times 100}{200} = 50$

$E_{21} = \frac{R_2 \times C_1}{N} = \frac{600 \times 100}{200} = 30$; $E_{22} = \frac{R_2 \times C_2}{N} = \frac{60 \times 100}{200} = 30$

$E_{31} = \frac{R_3 \times C_1}{N} = \frac{40 \times 100}{200} = 20$; $E_{32} = \frac{R_3 \times C_2}{N} = \frac{40 \times 100}{200} = 20$

Arranging the observed and expected frequencies in the following table :

O	E	$(O-E)^2$	$(O-E)^2/E$
40	50	100	2.0000
60	50	100	2.0000
35	30	25	0.8333
25	30	25	0.8333
25	20	25	1.2500
15	20	25	1.2500
			$\Sigma\frac{(O-E)^2}{E} = 8.1666$

$\therefore \chi^2_{cal.} = \Sigma\frac{(O-E)^2}{E} = 8.1666$, with d.f.= (r - 1) (c - 1) = (3 - 1) (2 - 1) = 2 × 1 = 2

∴ The critical value of χ^2 with 2 d.f. at 5% level of significance is $\chi^2_{0.05, 2} = 5.99$

Since $\chi^2_{cal.} > \chi^2_{0.05, 2}$. So, the null hypothesis is rejected. Hence, there is a significant relationship between sex and preference for colour.

Problem- (5) : By using the following information :

		Boys	
		Intelligent	**Unintelligent**
Father	**Skilled**	40	30
	Unskilled	70	54

Do these figures support the hypothesis that skilled fathers have intelligent boys?

Solution : We want to test,

H_0 : There is no relation between the skill of father and the intelligence of son.

H_1 : There is, somehow, a relation between the skill of father and son.

We know, Test statistic, $\chi^2 = \Sigma \frac{(O-E)^2}{E}$, with df. = (r–1) (c-1)

Here, O = the observed frequency

E = The expected frequency

$= \frac{\text{Row total} \times \text{column total}}{\text{Total number of observation}}$

From the given information, we have,

$R_1 = 70$, $R_2 = 124$, $C_1 = 110$, $C_2 = 84$, N= 194

The expected frequencies are computed as follows :

$E_{11} = \frac{R_1 \times C_1}{N} = \frac{70 \times 110}{194} = 39.69$; $E_{12} = \frac{R_1 \times C_2}{N} = \frac{70 \times 84}{194} = 30.31$

$E_{21} = \frac{R_2 \times C_1}{N} = \frac{124 \times 110}{194} = 70.31$; $E_{22} = \frac{R_2 \times C_2}{N} = \frac{124 \times 84}{194} = 53.69$

$$\therefore \chi^2 = \Sigma \frac{(O-E)^2}{E} = \frac{(40 - 39.69)^2}{39.69} + \frac{(30 - 30.31)^2}{30.31} + \frac{(70 - 70.31)^2}{70.31} + \frac{(54 - 53.69)^2}{53.69}$$

$$= \frac{0.0961}{39.69} + \frac{0.961}{30.31} + \frac{0.0961}{70.31} + \frac{0.961}{53.69}$$

$$= 0.00242 + 0.00317 + 0.00137 + 0.00179$$

$$= 0.00875$$

$\therefore \chi^2 = 0.00875$, with d.f.= (r - 1) (c - 1) = (2 - 1) (2 - 1) = 1 × 1 = 1

$\therefore$ The critical value of χ^2 with 1 d.f. at 5% level of significance is $\chi^2_{0.05,\,1.} = 3.84$

Since $\chi^2 < \chi^2_{0.05,\,1.}$

So, the null hypothesis is accepted. Hence, there is no relation between the skill of father and intelligence of son.

Problem- (6) : The following table gives for a sample of married women, the level of education and marrige adjustment score.

		Marriage adjustment score			
		Very low	Low	High	Very high
Level of eduction	College	24	97	62	58
	High school	22	28	30	41
	Middle school	32	10	11	20

Can you conclude from the above data that the higher the level of education, the greater is the degree of adjustment in marriage ?

Solution : We want to test,

Ho : There is no association between level of education and marriage adjustment score.

Where, H_1 : There is association between level of education and marriage adjustment score.

We know, Test statistic, $\chi^2 = \Sigma \frac{(O-E)^2}{E}$, with df. = (r–1) (c-1)

Here, O = the observed frequency

$$E = \text{The expected frequency} = \frac{\text{Row total} \times \text{column total}}{\text{Total number of observation}}$$

Table for observed frequencies

		Marriage adjustment score				Total
		Very low	Low	High	Very high	
Level of education	College	0_{11}= 24	0_{12}= 97	0_{13}= 62	0_{14}= 58	R_1= 241
	High school	0_{21}= 22	0_{22}= 28	0_{23}= 24	0_{24}= 41	R_2= 121
	Middle school	0_{31}= 32	0_{32}= 10	0_{33}= 11	0_{34}= 20	R_3= 73
		C_1= 78	C_2= 135	C_3= 103	C_4= 119	N = 435

The expected frequencies are computed as follows :

$$E_{11} = \frac{R_1 \times C_1}{N} = \frac{241 \times 78}{435} = 43.21, \quad E_{12} = \frac{R_1 \times C_2}{N} = \frac{241 \times 135}{435} = 74.79, \quad E_{13} = \frac{R_1 \times C_3}{N} = \frac{241 \times 103}{435} = 57.06, \quad E_{14} = \frac{R_1 \times C_4}{N} = \frac{241 \times 119}{435} = 65.93$$

$E_{21} = \frac{R_2 \times C_1}{N}$, $E_{22} = \frac{R_2 \times C_2}{N}$, $E_{23} = \frac{R_2 \times C_3}{N}$, $E_{24} = \frac{R_2 \times C_4}{N}$

$= \frac{121 \times 78}{435}$ $= \frac{121 \times 135}{435}$ $= \frac{121 \times 103}{435}$ $= \frac{121 \times 119}{435}$

$= 21.69$ $= 37.55$ $= 28.56$ $= 33.10$

$E_{31} = \frac{R_3 \times C_1}{N}$, $E_{32} = \frac{R_3 \times C_2}{N}$, $E_{33} = \frac{R_3 \times C_3}{N}$, $E_{34} = \frac{R_3 \times C_4}{N}$

$= \frac{73 \times 78}{435}$ $= \frac{73 \times 135}{435}$ $= \frac{73 \times 103}{435}$ $= \frac{73 \times 119}{435}$

$= 13.09$ $= 22.66$ $= 17.29$ $= 19.97$

$$\therefore \quad \chi^2 = \Sigma \frac{(O-E)^2}{E}$$

$$= \frac{(24 - 43.21)^2}{43.21} + \frac{(97 - 74.79)^2}{74.79} + \frac{(62 - 57.06)^2}{57.06} + \frac{(58 - 65.93)^2}{65.93}$$

$$+ \frac{(22 - 21.69)^2}{21.69} + \frac{(28 - 37.55)^2}{37.55} + \frac{(30 - 28.65)^2}{28.56} + \frac{(41 - 33.10)^2}{33.10}$$

$$= \frac{(32 - 13.09)^2}{13.09} + \frac{(10 - 22.66)^2}{22.66} + \frac{(11 - 17.29)^2}{17.29} + \frac{(20 - 19.97)^2}{19.97}$$

$$= 8.54 + 6.59 + 0.43 + 0.95 + 0.004 + 2.43 + 0.06 + 1.89 + 27.32$$

$$+ 7.07 + 2.29 + 0.00005 = 57.57$$

$\therefore \chi^2 = 57.57$, with d.f. = (r - 1) (c - 1) = (3 - 1) (4 - 1) = 2 × 3 = 6

$\therefore$ The critical value of χ^2 with df 6 at 5% level of significance is $\chi^2_{0.05,6.} = 12.59$

Since $\chi^2 > \chi^2_{0.05,6.}$ So, we can reject the null hypothesis. Hence, there is a significant association between the level of education and marriage adjustment score.

Problem - 7: The number of units sold by three salesmen over three month period are given below. By using 5% level of significance, test for the independence of salesmen and type of product. What is the conclusion?

		Product		
		Bread	**Biscuit**	**Chocolate**
Salesman	**Rahim**	**349**	**5**	**6**
	Karim	**19**	**35**	**9**
	Hakim	**7**	**45**	**25**

Solution : We want to test, Ho : Salesmen and type of product are independent.

We know, Test statistic, $\chi^2 = \Sigma\Sigma \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$, with d.f. = (r - 1) (c - 1)

Where, O_{ij} = The observed frequency

E_{ij} = The expected frequency $= \frac{R_i \times C_j}{N}$

Table for observed frequencies

		Product			Total
		Bread	Biscuit	Chocolate	
Salesman	Rahim	$0_{11} = 349$	$0_{12} = 5$	$0_{13} = 6$	$R_1 = 360$
	Karim	$0_{21} = 19$	$0_{22} = 35$	$0_{23} = 9$	$R_2 = 63$
	Hakim	$0_{31} = 7$	$0_{32} = 45$	$0_{33} = 25$	$R_3 = 77$
	Total	$C_1 = 375$	$C_2 = 85$	$C_3 = 40$	$N = 500$

The expected frequencies are computed as follows :

$E_{11} = \frac{R_1 \times C_1}{N} = \frac{360 \times 575}{500} = 270$ $\qquad E_{12} = \frac{R_1 \times C_2}{N} = \frac{360 \times 85}{500} = 61.2 \approx 61$

$E_{13} = \frac{R_1 \times C_3}{N} = \frac{360 \times 40}{500} = 28.8 \approx 29$ $\qquad E_{21} = \frac{R_2 \times C_1}{N} = \frac{63 \times 375}{500} = 47.25 \approx 47$

$E_{22} = \frac{R_2 \times C_2}{N} = \frac{63 \times 85}{500} = 10.71 = 11$ $\qquad E_{23} = \frac{R_2 \times C_3}{N} = \frac{63 \times 40}{500} = 5.04 \approx 5$

$E_{31} = \frac{R_3 \times C_1}{N} = \frac{77 \times 375}{500} = 57.75 \approx 58$ $\qquad E_{32} = \frac{R_3 \times C_2}{N} = \frac{77 \times 85}{500} = 13.09 \approx 13$

$E_{33} = \frac{R_3 \times C_3}{N} = \frac{77 \times 40}{500} = 6.16 \approx 6$

$$\therefore \chi^2 = \Sigma\Sigma \frac{(0_{ij} - E_{ij})^2}{E_{ij}} = \frac{(349-270)^2}{270} + \frac{(5-61)^2}{61} + \frac{(6-29)^2}{29} + \frac{(19-47)^2}{47} + \frac{(35-11)^2}{11}$$

$$+ \frac{(9-5)^2}{5} + \frac{(7-58)^2}{58} + \frac{(45-13)^2}{13} + \frac{(25-6)^2}{6}$$

$= 347.87$, with df $= (r-1)(c-1) = (3-1)(3-1) = 4$

Conclusion: For d.f.= 4, the value of χ^2 at 5% level of significance is 9.488. Since the computed value is greater than the table value. So, the null hypothsis is rejected. Hence, Salesmen and type of product are not independent.

Problem- (8) : By using χ^2 test find out weather there is any association between income level and type of colleging :

		College	
		Private	Government
Income	Low	200	400
	High	1000	400

Solution : We want to test, Ho : There is no association between income level and type of colleging. That is, the alternative hypothesis, H_1 : There is association between income level and type of colleging.

We know the test statistic, $\chi^2 = \Sigma\Sigma \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$, with d.f. $= (r - 1)(c - 1)$

Where, O_{ij} = The observed frequency

E_{ij} = The expected frequency $= \frac{R_i \times C_j}{N}$.

Table for observed frequencies

		College		Total
		Private	Government	
Income	Low	$0_{11}= 200$	$0_{12}= 400$	$R_1= 600$
	High	$0_{21}= 1000$	$0_{22}= 400$	$R_2= 1400$
Total		$C_1= 1200$	$C_2= 800$	$N = 2000$

The expected frequencies are computed as follows :

$$E_{11}= \frac{R_1 \times C_1}{N} = \frac{600 \times 1200}{2000} = 360 ;$$

$$E_{12} = \frac{R_1 \times C_2}{N} = \frac{600 \times 800}{2000} = 240$$

$$E_{21}= \frac{R_2 \times C_1}{N} = \frac{1400 \times 1200}{2000} = 840 ;$$

$$E_{22}= \frac{R_2 \times C_2}{N} = \frac{1400 \times 800}{2000} =560$$

$$\therefore \quad \chi^2 = \Sigma\Sigma \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \frac{(200 - 360)^2}{360} + \frac{(400 - 240)^2}{240} + \frac{(1000 - 840)^2}{840} + \frac{(400 - 560)^2}{560}$$

$$= \frac{25600}{360} + \frac{25600}{240} + \frac{25600}{840} + \frac{25600}{560} = 71.11 + 106.67 + 30.48 + 45.71 = 253.97$$

$\therefore \chi^2 = 253.97$, with d.f.= (r - 1) (c - 1) = (2 - 1) (2 - 1) = 1 × 1 = 1

$\therefore$ The critical value of χ^2 with 1 d.f. at 5% level of significance is $\chi^2_{0.05, 1.} = 3.84$

Since $\chi^2 > \chi^2_{0.05, 1}$, We can reject null hypothesis. Hence, there is a significant association between income level and type of colleging.

Problem - 9: A random sample of 518 employees was given a test to diagnose the of satisfication. Each employee's salary was also recorded in the table below. By using an appropriate level of significance, test to determine if salary and job satisfication are independent.

Salary / Satisfication	Under Tk. 25,000	Tk. 25,000- Tk. 35,000	Over Tk. 35,000	Total
High	126	61	38	225
Medium	71	93	69	233
Low	19	14	27	60
Total	216	168	134	518

Solution : We want to test, Ho : The salary and job satisfication are independent.

We know, Test statistic, $\chi^2 = \Sigma\Sigma \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$, with d.f. = (r - 1) (c - 1)

Where, O_{ij} = The observed frequency

E_{ij} = The expected frequency $= \frac{R_i \times C_j}{N}$.

Table for observed frequencies

Satisfication \ Salary	Under Tk. 25,000	Tk. 25,000-Tk. 35,000	Over Tk. 35,000	Total
High	$O_{11} = 126$	$O_{12} = 61$	$O_{13} = 38$	$R_1 = 225$
Medium	$O_{21} = 71$	$O_{22} = 93$	$O_{23} = 69$	$R_2 = 233$
Low	$O_{31} = 19$	$O_{32} = 14$	$O_{33} = 27$	$R_3 = 60$
Total	$C_1 = 216$	$C_2 = 168$	$C_3 = 134$	$N = 518$

The expected frequencies are computed as follows :

$$E_{11} = \frac{R_1 \times C_1}{N} = \frac{225\times216}{518} = 93.82,\ E_{12} = \frac{R_1\times C_2}{N} = \frac{225\times168}{518} = 72.97,\ E_{13} = \frac{R_1\times C_3}{N} = \frac{225\times134}{518} = 58.20,$$

$$E_{21} = \frac{R_2 \times C_1}{N} = \frac{233\times216}{518} = 97.16,\ E_{22} = \frac{R_2\times C_2}{N} = \frac{233\times168}{518} = 75.57, E_{23} = \frac{R_2\times C_3}{N} = \frac{233\times134}{518} = 60.27$$

$$E_{31} = \frac{R_3 \times C_1}{N} = \frac{60\times216}{518} = 25.02,\ E_{32} = \frac{R_3\times C_2}{N} = \frac{60\times168}{518} = 19.46,\ E_{33} = \frac{R_3\times C_3}{N} = \frac{60\times134}{518} = 15.52$$

$$\therefore \chi^2 = \frac{(126-93.82)^2}{93.82} + \frac{(61-72.97)^2}{72.97} + \frac{(38-58.20)^2}{58.20} + \frac{(71-97.16)^2}{97.16}$$

$$+ \frac{(93-75.57)^2}{75.57} + \frac{(69-60.27)^2}{60.27} + \frac{(19-25.02)^2}{25.02} + \frac{(14-19.46)^2}{19.46} + \frac{(27-15.52)^2}{15.52}$$

$$= 11.04 + 1.96 + 7.01 + 7.04 + 4.02 + 1.26 + 1.45 + 1.53 + 8.49 = 43.8$$

$\therefore$ The computed value, $\chi^2 = 43.8$, $df = (r-1)(c-1) = (3-1)(3-1) = 4$

Conclusion: For d.f.= 4, the value of χ^2 at 5% level of significance is 9.488. Since the computed value is greater than the table value. So, the null hypothsis is rejected. Hence, The salary and job satisfication are not independent.

Problem- (10) : A survey on the profession of 212 males was conducted to see whether their profession was associated with their father's profession.The following results were obtained:

		Son's profession			
		Agriculture	**Business**	**Clerk**	**Others**
Father's profession	**Agriculture**	23	4	15	9
	Business	8	19	16	20
	Clerk	4	3	28	6
	Others	3	15	12	27

Test the hypothesis that son's occupation is independent of father's occupation.

Solution : We want to test, Ho: Son's occupation is independent of father's occupation.

Against, H_1 : Sons' occupation is not independent of fathers' occupation.

We know the test statistic, $\chi^2 = \Sigma\Sigma \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$, with d.f. = (r - 1) (c - 1)

Where, O_{ij} = Observed frequency in the i th row and j the column.

E_{ij} = Expected frequency in the i th row and j the column = $\frac{R_i \times C_j}{N}$.

Table for the observed frequencies

		Son's occupation				
		Agriculture	Business	Clerk	Others	Total
Fathers occupa-tion	Agriculture	0_{11}= 23	0_{12}= 4	0_{13}= 15	0_{14}= 9	R_1= 51
	Business	0_{21}= 8	0_{22}= 19	0_{23}= 16	0_{24}= 20	R_2= 63
	Clerk	0_{31}= 4	0_{32}= 3	0_{33}= 28	0_{34}= 6	R_3= 41
	Others	0_{41}= 3	0_{42}= 15	0_{43}= 12	0_{44}= 27	R_4= 57
		C_1= 38	C_2= 41	C_3= 71	C_4= 62	N = 212

The expected frequencies are computed as follows :

$E_{11} = \frac{R_1 \times C_1}{N} = \frac{51 \times 38}{212} = 9.14$, $E_{12} = \frac{R_1 \times C_2}{N} = \frac{51 \times 41}{212} = 9.86$, $E_{13} = \frac{R_1 \times C_3}{N} = \frac{51 \times 71}{212} = 17.08$, $E_{14} = \frac{R_1 \times C_4}{N} = \frac{51 \times 62}{212} = 14.92$

$E_{21} = \frac{R_2 \times C_1}{N} = \frac{63 \times 38}{212} = 11.29$, $E_{22} = \frac{R_2 \times C_2}{N} = \frac{63 \times 41}{212} = 12.18$, $E_{23} = \frac{R_2 \times C_3}{N} = \frac{63 \times 71}{212} = 21.09$, $E_{24} = \frac{R_2 \times C_4}{N} = \frac{63 \times 62}{212} = 18.42$

$E_{31} = \frac{R_3 \times C_1}{N} = \frac{41 \times 38}{212} = 7.35$, $E_{32} = \frac{R_3 \times C_2}{N} = \frac{41 \times 41}{212} = 7.93$, $E_{33} = \frac{R_3 \times C_3}{N} = \frac{41 \times 71}{212} = 13.73$, $E_{34} = \frac{R_3 \times C_4}{N} = \frac{41 \times 62}{212} = 11.99$

$E_{41} = \frac{R_1 \times C_1}{N} = \frac{57 \times 38}{212} = 10.22$, $E_{42} = \frac{R_1 \times C_2}{N} = \frac{57 \times 41}{212} = 11.02$, $E_{43} = \frac{R_1 \times C_3}{N} = \frac{57 \times 71}{212} = 19.09$, $E_{44} = \frac{R_1 \times C_4}{N} = \frac{57 \times 62}{212} = 16.67$

$$\therefore \chi^2 = \Sigma\Sigma \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \frac{(23 - 9.14)^2}{9.14} + \frac{(4 - 9.86)^2}{9.86} + \frac{(15 - 17.08)^2}{17.08} + \frac{(9 - 14.92)^2}{14.82}$$

$$+\frac{(8-11.29)^2}{11.29}+\frac{(19-12.18)^2}{12.18}+\frac{(16-21.09)^2}{21.09}+\frac{(20-18.42)^2}{18.42}$$
$$+\frac{(4-7.35)^2}{7.35}+\frac{(3-7.93)^2}{7.93}+\frac{(28-13.73)^2}{13.73}+\frac{(6-11.99)^2}{11.99}$$
$$+\frac{(3-10.22)^2}{10.22}+\frac{(15-11.02)^2}{11.02}+\frac{(12-19.09)^2}{19.09}+\frac{(27-16.67)^2}{16.67}$$
$$= 21.02 + 3.48 + 0.25 + 2.36 + 0.96 + 3.82 + 1.23 + 0.134 + 1.53$$
$$+ 3.06 + 14.83 + 2.99 + 5.1 + 1.44 + 2.63 + 6.4 = 71.224$$

[**Alternative method :** $\chi^2 = \Sigma\Sigma \frac{O_{ij}^2}{E_{ij}} - N$

$$= \{\frac{(23)^2}{9.14}+\frac{(4)^2}{9.86}+\frac{(15)^2}{17.08}+\frac{(9)^2}{14.82}+\frac{(8)^2}{11.29}+\frac{(19)^2}{12.18}+\frac{(16)^2}{21.09}+\frac{(20)^2}{18.42}+\frac{(4)^2}{7.35}+\frac{(3)^2}{7.93}+\frac{(28)^2}{13.73}+$$
$$\frac{(6)^2}{11.99}+\frac{(3)^2}{10.22}+\frac{(15)^2}{11.02}+\frac{(12)^2}{19.09}+\frac{(27)^2}{16.67}\} - 212 = (57.88 + 1.62 + 13.17 + 5.47 + 5.67 + 29.64 +$$
$$12.14 + 21.72 + 2.18 + 1.13 + 57.10 + 3.00 + 0.88 + 20.42 + 7.54 + 43.73) - 212$$
$$= 283.25 - 212 = 71.295]$$

$\therefore \chi^2 = 71.224$, with d.f.= (r - 1) (c - 1) = (4 - 1) (4 - 1) = 9

$\therefore$ The critical value of χ^2 with 9 d.f. at 5% level of significance is $\chi^2_{0.05,9} = 16.9$

Since $\chi^2 > \chi^2_{0.05,9}$. So, We can reject the null hypothesis. Hence, sons' occupation is not independent of father's occupation.

Problem- (11) : A marketing agency provides you the following information about the age-groups of the sample informants and their likings for a particular model of sector which a company plans to intordduce :

	Age group of informants		
	Below 20	**20 - 39**	**40 - 59**
Liked	140	640	80
Disliked	65	320	120

On the basis of the above data, can it be concluded that the model appeal is independent of the age group of the informants ?

Solution : We want to test, Ho : The model is independent of age group of the informants.

Where as H_1 : The model is not independent of age group of the informants.

The test statistic, $\chi^2 = \Sigma\Sigma \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$, with d.f. = (r - 1) (c - 1)

Where, O_{ij} = The observed frequency

E_{ij} = The expected frequency $= \frac{R_i \times C_j}{N}$

Table for observed frequencies

	Age group of informants			Total
	Below 20	20 - 39	40 - 59	
Liked	$0_{11}= 140$	$0_{12}= 640$	$0_{13}= 80$	$R_1= 860$
Disliked	$0_{21}= 65$	$0_{22}= 320$	$0_{23}= 120$	$R_2= 505$
Total	$C_1= 205$	$C_2= 960$	$C_3= 200$	$N = 1365$

The expected frequencies are computed as follows :

$E_{11} = \frac{R_1 \times C_1}{N} = \frac{860 \times 205}{1365} = 129.16$, $E_{12} = \frac{R_1 \times C_2}{N} = \frac{860 \times 960}{1365} = 604.84$, $E_{13} = \frac{R_1 \times C_3}{N} = \frac{860 \times 200}{1365} = 126$

$E_{21} = \frac{R_2 \times C_1}{N} = \frac{505 \times 205}{1365} = 75.84$, $E_{22} = \frac{R_2 \times C_2}{N} = \frac{505 \times 960}{1365} = 355.16$, $E_{23} = \frac{R_2 \times C_3}{N} = \frac{505 \times 200}{1365} = 73.99$

$$\chi^2 = \Sigma\Sigma \frac{O_{ij}^2}{E_{ij}} - N = \frac{(140 - 129.16)^2}{129.16} + \frac{(640 - 604.46)^2}{604.46} + \frac{(80 - 126)^2}{126} + \frac{(65 - 75.84)^2}{75.84} + \frac{(320 - 355.16)^2}{355.16} + \frac{(120 - 73.99)^2}{73.99}$$

$$= 0.91 + 2.04 + 16.79 + 1.55 + 3.48 + 28.61 = 53.38$$

$\therefore \chi^2 = 53.38$, with d.f.= (r - 1) (c - 1) = (2 - 1) (3 - 1) = 2

$\therefore$ The critical value of χ^2 with 2 d.f. at 5% level of significance is, $\chi^2_{0.05,2.} = 5.99$

Since $\chi^2 > \chi^2_{0.05,2}$. We can reject the null hypothesis. Hence, from given data, it can be concluded that the model appeal is not independent of the age group of the informants.

Problem- (12) : A sample analysis of examination results of 200 BBA's was made. It was found that 46 students have failed, 68 secured a third class, 62 secured a second class and rest were placed in the first class. Are these figures commensurate with the general examination result which is in the ratio of 2 : 3 : 3 : 2, for various categories respectively?

Solution : We want to test,

Ho: There is no significant difference in the observed and expected results.

Vs H_1 : There is significant difference in the observed and expected results.

We know, Test statistic, $\chi^2 = \Sigma \frac{(O-E)^2}{E}$, with df. = (r–1) (c-1)

Here, O = the observed result

E = The expected result

From the given information, we have,

N = 200, ratio = 2 : 3 : 3 : 2 and Total = 2+3+3+2=10

Arranging observed and expected results in the following table and calculating χ^2 :

Category	Observed results (O)	Expected results (E)	$(O-E)^2$	$\frac{(O-E)^2}{E}$
Failed	46	$\frac{200 \times 2}{10}=40$	36	0.900
Third class	68	$\frac{200 \times 3}{10}=60$	64	1.067
Second class	62	$\frac{200 \times 3}{10}=60$	4	0.067
First class	24	$\frac{200 \times 2}{10}=40$	256	6.400
				$\Sigma\frac{(O-E)^2}{E}=8.434$

$\therefore \quad \chi^2 = \Sigma \frac{(O-E)^2}{E} = 8.434$, with d.f. v =4-1 = 3.

$\therefore$ The critical value of χ^2 with 8 d.f. at 5% level of significance is 7.81.

Since the calculated value of χ^2 is more than the table value. So, we can rejected the null hypothesis. Hence, we conclude that the given results are not commensurate with the general examination results.

Problem- (13) : Two treatments T_1 and T_2 were tried to control a certain types of plant disease. The following results were obtained-

Treatment T_1, 420 plants were examined and 80 were infected

Treatment T_2, 580 plants were examined and 120 were infected.

Can it conclude that treament T_2 is superior to T_1 ?

Solution : We want to test, Ho : Treatments T_1 and T_2 are alike.

Vs H_1 : Treatment T_2 is superior to T_1.

We know, Test statistic, $\chi^2 = \Sigma\Sigma \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$, with d.f. = (r - 1) (c - 1)

Where, O_{ij} = The observed frequency

and $E_{ij} = \frac{R_i \times C_j}{N}$ = The expected frequency

Table for the observed frequencies as follows:

		Infected	Not-infected	Total
Treatment	A	$0_{11}= 80$	$0_{12}= 340$	$R_1 = 420$
	B	$0_{21}= 120$	$0_{22}= 460$	$R_2= 580$
	Total	$C_1= 200$	$C_2= 800$	$N = 1000$

The expected frequencies are computed as follows :

$$E_{11}=\frac{R_1\times C_1}{N}=\frac{420\times 200}{1000}=84\ ;\qquad E_{12}=\frac{R_1\times C_2}{N}=\frac{420\times 800}{1000}=336$$

$$E_{21}=\frac{R_2\times C_1}{N}=\frac{580\times 200}{1000}=116\ ;\qquad E_{22}=\frac{R_2\times C_2}{N}=\frac{580\times 800}{1000}=464$$

$$\therefore\quad \chi^2=\Sigma\Sigma\frac{(O_{ij}-E_{ij})^2}{E_{ij}}=\frac{(80-84)^2}{84}+\frac{(340-336)^2}{336}+\frac{(120-116)^2}{116}+\frac{(460-464)^2}{464}$$

$$=\frac{16}{84}+\frac{16}{336}+\frac{16}{116}+\frac{16}{464}$$

$$=0.1905+0.0476+0.1379+0.0345=0.4105$$

$\therefore\ \chi^2=0.4105$, with d.f.$=(r-1)(c-1)=(2-1)(2-1)=1\times 1=1$

$\therefore$ The critical value of χ^2 with 1 d.f. at 5% level of significance is $\chi^2_{0.05,\ 2.}=3.84$

Since $\chi^2<\chi^2_{0.05,\ 2}$, which means that the observed p-value is greater than 0.05. So, the null hypothesis is not rejected. Hence, the treatments T_1 and T_2 are alike. That is, treatment T_2 is not superior to T_1.

Problem- (14) : Of the 500 workers in a factory exposed to an epidemic 350 in all were attacked, 200 had been inoculated and of these 100 were attacked. Set out the data in a tabular form and find out whether in the given case the association is or is not significant.

Solution : From the given problem, we have the tabular form as follows :

	Attacted	Not-attacted	Total
Inoculated	100	100	200
Not-inoculated	250	50	300
Total	350	150	500

We want to test the null hypothesis Ho: The inoculation and attack are independent.

Vs H_1 : The inoculation and attack are not independent.

We know the test statistic, $\chi^2=\Sigma\Sigma\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$, with d.f. = (r - 1) (c - 1)

Where, O_{ij} = The observed frequency

and $E_{ij}=\frac{R_i\times C_j}{N}$ = The expected frequency

Table for the observed frequencies

	Attacked	Not-attacked	Total
Inoculated	$0_{11}=100$	$0_{12}=100$	$R_1=200$
Not-inoculated	$0_{21}=250$	$0_{22}=50$	$R_2=300$
Total	$C_1=350$	$C_2=150$	$N=500$

The expected frequencies are computed as follows :

$$E_{11}=\frac{R_1 \times C_1}{N}=\frac{200 \times 350}{500}=140; \qquad E_{12}=\frac{R_1 \times C_2}{N}=\frac{200 \times 150}{500}=60$$

$$E_{21}=\frac{R_2 \times C_1}{N}=\frac{300 \times 350}{500}=210 \ ; \qquad E_{22}=\frac{R_2 \times C_2}{N}=\frac{300 \times 150}{500}=90$$

$$\therefore \chi^2=\Sigma\Sigma\frac{(O_{ij}-E_{ij})^2}{E_{ij}}=\frac{(100-140)^2}{140}+\frac{(100-60)^2}{60}+\frac{(50-210)^2}{210}+\frac{(50-90)^2}{90}$$

$$=\frac{1600}{140}+\frac{1600}{60}+\frac{1600}{210}+\frac{1600}{90}=11.4286+26.6667+7.6190+17.7778=63.4921$$

$\therefore \chi^2 = 63.4921,$ with d.f. $=(r-1)(c-1)=(2-1)(2-1)=1\times 1=1$

$\therefore$ The critical value of χ^2 with 1 d.f. at 5% level of significance is $\chi^2_{0.05,\,1.}=3.84$

Since $\chi^2 > \chi^2_{0.05,\,1}$, which means that the observed p-value is less than 0.05. So, the null hypothesis is rejected. Hence, the inoculation and attack are independent.

Format (3) : Test of Goodness of Fit.

The chi-square (χ^2) test is a measure of probabilities of association between the attributes. It gives us an idea about the divergence between the observed and expected frequencies. Thus the test is also described as the test of "Goodness of fit".

let, The observed value = O, The expected value =E

Then Test statistic, $\chi^2=\Sigma\frac{(O-E)^2}{E}=\Sigma\frac{O^2}{E}-N,$ Where $\Sigma O=\Sigma E=N$

Here, Degree of freedom, $df=n-1$.

Problem - 15: The number of parts for a particular spare part in a factory was found to very from day to day. In a sample study the following information was obtained :—

Day	Mon.	Tues.	Wed.	Thr.	Fri.	Sat.	Total
No. of parts demanded	1124	1125	1110	1120	1126	1115	6720

Test the hypothesis that the number of parts demanded does not depend on the day of the week. (The table value of χ^2 for df 5 at 5% level of significance is 11.07)

[BBA, Third Semester, N.U. 2009]

Solution: We want to test, H_0: The number of parts demanded does not depend on the day of the week.

Against, H_1 : The number of parts demanded depend on the day of the week.

We know, $\chi^2=\Sigma\frac{(O-E)^2}{E}$, $df=n-1$

Here, O = The observed value, E = The expected value

Since the expected value (E) and the arithmetic mean ($\bar{x}$) are same in mathematically.

$$\therefore E=\frac{\text{Total number of parts demanded}}{\text{Total days}}$$

$$=\frac{1124+1125+1110+1120+1126+1115}{6}=\frac{6720}{6}=1120$$

Table for calculation

Day	O	E	$(O-E)^2$	$(O-E)^2/E$
Mon	1124	1120	16	0.0143
Tues.	1125	1120	25	0.0223
Wed.	1110	1120	100	0.0893
Thr.	1120	1120	0	0
Fri.	1126	1120	36	0.0321
Sat.	1115	1120	25	0.0223
				$\Sigma\frac{(O-E)^2}{E} = 0.1803$

$\therefore \chi^2 = \Sigma\frac{(O-E)^2}{E} = 0.1803$, df = n – 1 = 6–1 = 5

Comments : We have, The calculated value of χ^2 is $\chi^2_{cal} = 0.1803$ and the table value of χ^2 for df. 5 at 5% level of significance is $\chi^2_{tab} = 11.07$

Since $\chi^2_{cal} < \chi^2_{tab}$. So, the null hypothesis is accepted. Hence, the number of parts demanded does not depend on the day of the week.

Problem - 16: The following figures show the distribution of digits in numbers chosen at random from a mobile numbers:

Digit	0	1	2	3	4	5	6	7	8	9	Total
Frequency	1026	1107	997	966	1075	933	1107	972	964	853	10000

Test whether the digits may be taken to occur equally frequently in the mobile numbers. **[BBA, MARKETING, R.U. 2014]**

Solution: We want to test,

H_0: The digits in numbers chosen occur equally frequently in the mobile numbers.

We know, $\chi^2 = \Sigma\frac{(O-E)^2}{E}$, df = n – 1

Here, O = The observed frequency, E = The expected frequency

Since the expected frequency (E) and the arithmetic mean ($\bar{x}$) are same in mathematically.

$\therefore E = \frac{10000}{10} = 1000$

Table for calculation

Digits	O	E	$(O-E)^2$	$(O-E)^2/E$
0	1026	1000	676	0.676
1	1107	1000	11449	11.449
2	997	1000	9	0.009
3	966	1000	1156	1.156
4	1075	1000	5625	5.625
5	933	1000	4489	4.489
6	1107	1000	11449	11.449
7	972	1000	784	0.784
8	964	1000	1296	1.296
9	853	1000	21609	21.609
				$\Sigma\frac{(O-E)^2}{E} = 58.542$

$\therefore \chi^2 = \Sigma\frac{(O-E)^2}{E} = 58.542$, $df = n-1 = 10-1 = 9$

Comments : We have, The calculated value of χ^2 is $\chi^2_{cal} = 58.542$ and the table value of χ^2 for dt. 5 at 5% level of significance is $\chi^2_{tab} = 16.919$.

Since the calculated value of χ^2 is more than the table value. So, we can reject the null hypothesis. Hence, the digits are not uniformly distributed in the mobile numbers.

Problem - 17: The following figures show the number of accidents per day in a certain city were as follows:

Day	1	2	3	4	5	6	7	8	9	10	Total
Frequency	12	8	20	2	14	10	15	6	9	4	100

Test whether these frequencies in agreement with the belief that accident conditions were the same during this 10 day period.

[BBA, MARKETING, J.U. 2015]

Solution: We want to test,

H_0: The number of accidents per day in a certain city are consistent with the belief that accident conditions were the same during this 10 day period.

We know, $\chi^2 = \Sigma\frac{(O-E)^2}{E}$, $df = n-1$

Here, O = The observed number , E = The expected number

Since the expected number (E) and the arithmetic mean ($\bar{x}$) are same in mathematically.

$\therefore E = \frac{100}{10} = 10$

Table for calculation

Digits	O	E	$(O-E)^2$	$(O-E)^2/E$
1	12	10	4	0.4
2	8	10	4	0.4
3	20	10	100	10.0
4	2	10	64	6.4
5	14	10	16	1.6
6	10	10	0	0.0
7	15	10	25	2.5
8	6	10	16	1.6
9	9	10	1	0.1
10	4	10	36	3.6
				$\Sigma\frac{(O-E)^2}{E} = 26.6$

$\therefore \chi^2 = \Sigma\frac{(O-E)^2}{E} = 26.6$, $df = n-1 = 10-1 = 9$

Comments : We have, The calculated value of χ^2 is $\chi^2_{cal} = 26.6$ and the table value of χ^2 for dt. 5 at 5% level of significance is $\chi^2_{tab} = 16.919$.

Since the calculated value of χ^2 is more than the table value. So, we can reject the null hypothesis. Hence, the accident conditions are not the same (uniform) during this 10 day period.

Problem-18: A survey of 1000 families with 5 children each, revealed the following distribution:

No. of boys	5	4	3	2	1	0
No. of girls	0	1	2	3	4	5
No. of families	38	144	342	287	164	25

Is result consistent with the hypothesis that male and female births are equally probable?

Solution:

We want to test the null hypothesis Ho: The male and female births are equally probable.

Vs H_1 : The male and female births are not equally probable.

Let us consider, x = The no. of male births.

We know, Probability function, $P(x) = {}^nC_x \, p^x \, q^{n-x}$; $x = 0, 1, 2, \text{--------}, n$ $[p + q = 1]$

According to null hypothesis, $p = q = \frac{1}{2} = 0.5$

$$\therefore \; P(x) = {}^5C_x \, (0.5)^x \, (0.5)^{5-x} ; \quad x = 0, 1, 2, \text{--------}, 5$$

$$\Rightarrow P(x) = {}^5C_x \, (0.5)^5 ; \quad x = 0, 1, 2, \text{-------}, 5$$

Table of fit of bionomial distribution is given below:

x	Observed frequency (O)	$P(x) = {}^5C_x(0.5)^5$	Expected frequency E=NP(x)
0	38	$P(0) = {}^5C_0 \, (0.5)^5 = 0.0312$	$31.2 \approx 31$
1	144	$P(1) = {}^5C_1 \, (0.5)^5 = 0.1562$	$156.2 \approx 156$
2	342	$P(2) = {}^5C_2 \, (0.5)^5 = 0.3125$	$312.5 \approx 313$
3	287	$P(3) = {}^5C_3 \, (0.5)^5 = 0.3125$	$312.5 \approx 313$
4	164	$P(4) = {}^5C_4 \, (0.5)^5 = 0.1562$	$156.2 \approx 156$
5	25	$P(5) = {}^5C_5 \, (0.5)^5 = 0.0312$	$31.2 \approx 31$
	$N = 1000$	$\Sigma P(x) = 1$ (app.)	$\Sigma E = 1000$ (app.)

Arranging observed and expected frequencies in the following table and calculating χ^2 :

Observed frequency (O)	Expected frequency E=NP(x)	$(O-E)^2$	$\frac{(O-E)^2}{E}$
38	31	49	1.58
144	156	144	0.92
342	313	841	2.69
287	313	676	2.16
164	156	64	0.41
25	31	36	1.16
N = 1000	$\Sigma E = 1000$ (app.)		$\Sigma \frac{(O-E)^2}{E} = 8.92$

$\therefore \quad \chi^2 = \Sigma \frac{(O-E)^2}{E} = 8.92,$ with d.f. $v = 6 - 1 = 5$.

$\therefore$ The critical value of χ^2 with 5 d.f. at 5% level of significance is 11.070.

Since the calculated value of χ^2 is less than the table value. So, we can accept the null hypothesis. Hence, we conclude that male and female births are equally probable.

Problem-19: Four coins are tossed 160 times and the number of heads appearing each time are noted. At the end, the following results were obtained:

Number of heads :	**0**	**1**	**2**	**3**	**4**
Frequency :	**17**	**52**	**54**	**31**	**6**

Use chi-square test of goodness of fit to determine whether the coins are unbiased.

Solution:

We want to test the null hypothesis Ho : The coins are unbiased.

Vs H_1 : The coins are biased.

Let us consider, x = The number of heads.

We know, Probability function, $P(x) = {}^nC_x \, p^x \, q^{n-x}$; $x = 0, 1, 2, \text{--------}, n$ $[p + q = 1]$

According to null hypothesis,

$$p = q = \frac{1}{2} = 0.5$$

$\therefore \; P(x) = {}^4C_x \, (0.5)^x \, (0.5)^{4-x}$; $x = 0, 1, 2, \text{--------}, 4$

$\Rightarrow P(x) = {}^4C_x \, (0.5)^4$; $x = 0, 1, 2, \text{-------}, 4$

Table of fit of bionomial distribution is given below:

x	Observed frequency (O)	$P(x) = {}^4C_x(0.5)^4$	Expected frequency E=NP(x)
0	17	$P(0) = {}^4C_0\,(0.5)^4 = 0.0625$	10
1	52	$P(1) = {}^4C_1\,(0.5)^4 = 0.2500$	40
2	54	$P(2) = {}^4C_2\,(0.5)^4 = 0.3750$	60
3	31	$P(3) = {}^4C_3\,(0.5)^4 = 0.2500$	40
4	6	$P(4) = {}^4C_4\,(0.5)^4 = 0.0625$	10
	N = 160	$\Sigma P(x) = 1$	$\Sigma E = 160$

Arranging observed and expected frequencies in the following table and calculating χ^2 :

Observed frequency (O)	Expected frequency E=NP(x)	$(O-E)^2$	$\frac{(O-E)^2}{E}$
17	10	49	4.90
52	40	144	3.60
54	60	36	0.60
31	40	81	2.03
6	10	16	1.60
N = 160	$\Sigma E =$		$\Sigma \frac{(O-E)^2}{E} = 12.73$

$$\therefore \quad \chi^2 = \Sigma \frac{(O-E)^2}{E} = 12.73$$

$\therefore \chi^2 = 12.73$, with d.f. $v = 5 - 1 = 4$.

$\therefore$ The critical value of χ^2 with 4 d.f. at 5% level of significance is 9.4877.

Since the calculated value of χ^2 is greater than the table value. So, we can reject the null hypothesis. Hence, we conclude that the coins are biased.

Problem- 20: At a level of significance of 1% or,0.10, can we conclude that the following 200 observations follow a poisson distribution?

No. of arrival (per hours)	0	1	2	3	4	5 or more
No. of hours	76	74	29	17	3	1

Solution:

We want to test the null hypothesis Ho: The given observations follow a poisson distribution.

Let, No. of arrival (per hours) = x and No. of hours = f.

We know, $\overline{x} = \frac{\Sigma fx}{N}$

Table for calculation

x	f	fx
0	76	0
1	74	74
2	29	58
3	17	51
4	3	12
5	1	5
	N = 200	$\Sigma fx = 200$

$\therefore\ \overline{x} = \frac{\Sigma fx}{N} = \frac{200}{200} = 1$

Let us consider that x be a poisson variate with parameter m.

Probability Function, $P(x) = \frac{e^{-m}\, m^x}{x!}$; $x = 0, 1, 2, \text{--------}, \infty$

Here, $E(x) = m$. $\therefore\ m = 1$

$\therefore\ P(x) = \frac{e^{-1}\,(1)^x}{x!}$; $x = 0, 1, 2, \text{--------}, \infty$

Table of fit of poisson distribution is given below:

x	Observed frequency (O)	$P(x) = \frac{e^{-1}\,(1)^x}{x!}$; $x = 0, 1, 2, \text{--------}, \infty$	Expected frequency E=NP(x)
0	76	$P(0) = \frac{e^{-1}.(1)^0}{0!} = \frac{0.3679\times1}{1} = 0.3679$	73.58 = 74 (app.)
1	74	$P(1) = \frac{e^{-1}.(1)^{-1}}{1!} = \frac{0.3679\times1}{1} = 0.3679$	73.58 = 73 (app.)
2	29	$P(2) = \frac{e^{-1}.(1)^2}{2!} = \frac{0.3679\times1}{2} = 0.18395$	36.79 = 37 (app.)
3	17	$P(3) = \frac{e^{-1}.(1)^3}{3!} = \frac{0.3679\times1}{6} = 0.0613$	12.26 = 12 (app.)
4	3	$P(4) = \frac{e^{-1}.(1)^4}{4!} = \frac{0.3679\times1}{24} = 0.0153$	3.06 = 3 (app.)
5	1	$P(5) = \frac{e^{-1}.(1)^5}{5!} = \frac{0.3679\times1}{120} = 0.0031$	0.62 = 1 (app.)
	N = 200	$\Sigma P(x) = 1$ (app.)	$\Sigma E = 200$ (app.)

Arranging observed and expected frequencies in the following table:

Observed frequency (O)	Expected frequency (E)	$(O-E)^2$	$\frac{(O-E)^2}{E}$
76	74	4	0.05
74	73	1	0.01
29	37	64	1.73
17	12	25	2.08
3	3	0	0.00
	1	0	0.00
N = 200	ΣE= 200 (app.)		$\Sigma\frac{(O-E)^2}{E} = 3.87$

$\therefore \quad \chi^2 = \Sigma\frac{(O-E)^2}{E} = 3.87$

$\therefore \chi^2 = 3.87$, with d.f. $v = 6 - 1 = 5$.

$\therefore$ The critical value of χ^2 with 5 d.f. at 1% level of significance is 9.2363.

Since the calculated value of χ^2 is less than the table value. So, we can accepted the null hypothesis.

Hence, we conclude that the given observations follow a poisson distribution.

Problem- 21: The figures given below are (a) the theoritical frequencies of a distribution and (b) the freqųencies of the distribution having the same mean, standard deviation and total frequency as in (a):

(a)	**1**	**12**	**66**	**220**	**495**	**792**	**924**	**792**	**495**	**220**	**66**	**12**	**1**
(b)	**2**	**15**	**6.6**	**210**	**484**	**799**	**943**	**799**	**484**	**210**	**66**	**15**	**2**

Do you think that the normal distribution provides a good fit to the data?

Solution: We want to test,

Ho: There is no difference between the observed frequencies and the expected frequencies as obtained by the normal distribution.

Vs H_1 : There is difference between the observed frequencies and the expected frequencies as obtained by the normal distribution.

We know, $\chi^2 = \Sigma\frac{(O-E)^2}{E}$, with df = n-1

Arranging observed and expected frequencies in the following table:

Observed frequency (O)	Expected frequency (E)	$(O-E)^2$	$\frac{(O-E)^2}{E}$
$\left.\begin{matrix}1\\12\end{matrix}\right\}=13$	$\left.\begin{matrix}2\\15\end{matrix}\right\}=17$	16	0.941
66	66	0	0.000
220	210	100	0.476
495	484	121	0.250
792	799	49	0.061
924	943	361	0.383
792	799	49	0.061
495	484	121	0.250
220	210	100	0.476
66	66	0	0.000
$\left.\begin{matrix}1\\12\end{matrix}\right\}=13$	$\left.\begin{matrix}2\\15\end{matrix}\right\}=17$	16	0.941
			$\Sigma\frac{(O-E)^2}{E}=3.839$

$\therefore\ \chi^2=\Sigma\frac{(O-E)^2}{E}=3.839$, with d.f. v = 13 - 2 - 3 = 8.

$\therefore$ The critical value of χ^2 with 8 d.f. at 5% level of significance is 15.507.

Since the calculated value of χ^2 is less than the table value. So, we can accepted the null hypothesis.

Hence, we conclude that there is no difference between the observed frequencies and the expected frequencies as obtained by the normal distribution.

Exercise- 6(3)

Part- A: Brief questions and answer

1. Define the chi-square distribution.

Ans. The chi-square test is based on Chi-square distribution which was first used by Karl Pearson in the year 1900. It is denoted by χ^2 distribution.

2. What is the probability function of the chi-square distribution?

Ans. The probability function of χ^2 distribution is given by