17

Chi-Square Test

INTRODUCTION

In the previous chapter on small sampling theory, it was necessary to make certain assumptions about the populations from which the samples were drawn. In many of the statistical tests, we had to assume that the samples came from normal populations. When this assumption cannot be justified, it is necessary to use procedures that do not require that these conditions be met. These procedures are generally referred to as non-parametric methods. In this chapter, we will discuss the χ^2 test which belongs to this category.

The χ^2 (pronounced as Chi-square) test is based on χ^2 distribution which was first used by Karl Pearson in the year 1900.

The Chi-square Distribution

For large sample size, the sampling (probability) distribution of χ^2 can be closely approximated by a continuous curve known as the Chi-square distribution. The probability function of χ^2 distribution is given by

$$F(\chi^2) = c(\chi^2)^{(v2)-1}\, e^{\chi^2/2}$$

where,

$e = 2.71828$

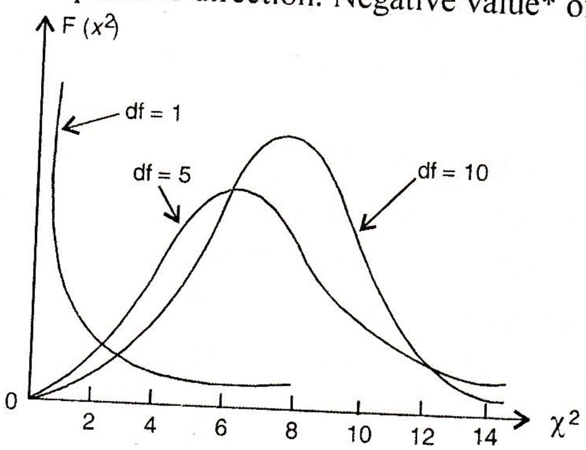
v = number of degrees of freedom

c = a constant depending only on v.

The Chi-square distribution has only one parameter v, the number of degrees of freedom. This is similar to the case of the t-distribution. Hence, $f(\chi^2)$ is a family of distributions, one for each value of v.

Important Properties of Chi-square Distribution

(1) χ^2 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 χ^2 is not possible.



*The value of χ^2 can never be negative, since the differences between the observed and expected frequencies are always squared.

(2) The exact shape of the distribution depends upon the number of degrees of freedom v. For different values of v, we shall have different shapes of the distribution. In general, when v is small, the shape of the curve is skewed to the right and as v gets larger, the distribution becomes more and more symmetrical and can be approximated by the normal distribution.

(3) The mean of the χ^2 distribution is given by the degrees of freedom, *i.e.*, $E(\chi^2) = v$ and variance is twice the degrees of freedom, *i.e.*, $V(\chi^2) = 2v$.

(4) As v gets larger, χ^2 approaches the normal distribution with mean v and standard deviation $\sqrt{2v}$. In practice, it has been determined that the quantity $\sqrt{2\chi^2}$ provides a better approximation to normality than χ^2 itself for values of 30 or more. The distribution of $\sqrt{2\chi^2}$ has a mean equal to $\sqrt{2v-1}$ and a standard deviation equal to one.

(5) The sum of independent χ^2 variates is also a χ^2 variate. Therefore, if χ_1^2 is a χ^2 variate with v_1 *d.f.* and χ_2^2 is another χ^2 variate with v_2 *d.f.* independent of χ_1^2, then their sum $\chi_1^2 + \chi_2^2$ is also a χ^2 variate with $v_1 + v_2$ *d.f.* This property is known as the additive property of χ^2.

Chi-square Test

The χ^2 test is one of the simplest and most widely used non-parametric tests in statistical work. It makes no assumptions about the population being sampled. The quantity χ^2 describes the magnitude of discrepancy between theory and observation, *i.e.*, with the help of χ^2 test we can know whether a given discrepancy between theory and observation can be attributed to chance or whether it results from the inadequacy of the theory to fit the observed facts. If χ^2 is zero, it means that the observed and expected frequencies completely coincide. The greater the value of χ^2, the greater would be the discrepancy between observed and expected frequencies. The formula for computing chi-square is :

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

where, O = observed frequency

E = expected or theoretical frequency.

The calculated value of χ^2 is compared with the table value of χ^2 for given degrees of freedom at specified level of significance. If the calculated value of χ^2 is greater than the table value, the difference between theory and observation is considered to be significant, *i.e.*, it could not have arisen due to fluctuations of simple sampling. On the other hand, if the calculated value of χ^2 is less than the table value, the difference between theory and observation is not considered significant, *i.e.*, it could have arisen due to fluctuations of sampling.

The number of degrees of freedom is described as the number of observations that are free to vary after certain restrictions have been imposed on the data. For a uniform distribution, we place one restriction on the expected distribution—the total of sample observations.

In a contingency table, the degrees of freedom are calculated in a slightly different manner. The marginal total or frequencies place the limit on our choice of selecting cell frequencies. The cell frequencies of all columns but one $(c-1)$ and of all rows but one $(r-1)$ can be assigned arbitrarily and so the number of degrees of freedom for all cell frequencies is $(c-1)(r-1)$ where, c refers to columns and r refers to rows. Thus, in a 2×2 table, the degrees of freedom would be $(2-1)(2-1) = 1$ and in a 3×3 table, the degrees of freedom would be $(3-1)(3-1) = 4$.

Conditions for the Application of χ^2 Test

The following five basic conditions must be met in order for chi-square analysis to be applied :

(1) The experimental data (sample observation) must be independent of each other.

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

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

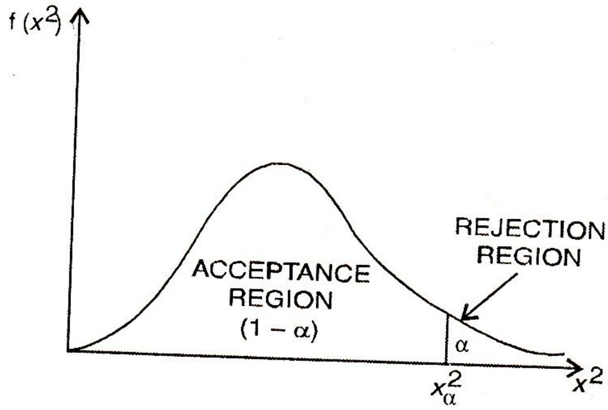
(4) The sample should contain at least 50 observations.

(5) There should not be less than five observations in any one cell (each data entry is known as a cell). For less than 5 observations, the value of χ^2 shall be over estimated and result in too many rejections of the null hypothesis.

Use of the Chi-square Table

To facilitate its many applications, the chi-square distribution has been extensively tabulated. The table of areas found in the Appendix gives value of χ^2 for various probabilities and various degrees of freedom. The value of α is given in the column headings, the degrees of freedom v are given in the rows and the body of the table gives the χ^2 values.

As depicted in the following figure, the value of χ^2 in the appendix table are given for various combinations of v and $1 - \alpha$.



Yates's Correction for Continuity

When using χ^2 analysis, it is important that a minimum of 80 per cent of the expected or theoretical frequencies in a cell be at least five and no cell have an expected frequency less than one. If the data results in expected frequencies less than five, wherever appropriate cell should be combined or the sample size should be increased until sufficient items fall into each cell.

The chi-square distribution is continuous distribution used with discrete data from a contingency table. When the expected frequencies are large, this approximate procedure is appropriate. In a 2 × 2 table, when expected frequencies are small, a correction was proposed by F. Yates in the year 1934 called "Yates's correction for continuity". The correction consists of :

$$\chi^2 \text{ (corrected)} = \frac{(|O_1 - E_1| - 0.5)^2}{E_1} + \frac{(|O_2 - E_2| - 0.5)^2}{E_2} + \ldots + \frac{(|O_k - E_k| - 0.5)^2}{E_k}$$

where $|O - E|$ means the absolute difference, ignoring plus and minus signs. Subtracting $\frac{1}{2}$ from the difference between O and E reduces the computed value of chi-square.

In general, the correction is made only when the number of degrees of freedom is $v = 1$. For large samples, this yields practically the same results as the uncorrected χ^2.

One problem with Yates's corrections should be noted. When the cells contain too few frequencies, ignoring Yates's correction might lead to excessive rejection of the hypothesis. On the other hand, Yates's correction tends to overcompensate for this and might result in excessive acceptance of the null hypothesis. The question is : what should be done by the analyst ? It may be reasonable to test the null hypothesis in the usual manner. If the hypothesis is accepted, we should be satisfied ; if rejection is indicated, then recalculate χ^2 using Yates's correction. Only if the null hypothesis is rejected without Yates's correction but accepted when the adjustment is used, should the analyst consider a more exact test than chi-square.

Grouping when Frequencies are Small

If small theoretical frequencies occur (less than 10 and certainly not less than 5), it is generally possible to overcome the difficulty by grouping two or more classes together. In other words, one or more classes with theoretical frequencies less than 5 may be combined into a single category before calculating the difference between observed and expected frequencies. The number of degrees of freedom would be determined with number of observations after the regrouping. This would be clear from the following :

Observed frequencies	: 364	376	218	89	33	13	2	1
Expected frequencies (based on Poisson dist.)	: 339	397	234	92	27	6	1	0

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

Observed frequencies	: 364	376	218	89	33	16
Expected frequencies	: 339	397	234	92	27	7

The degrees of freedom would now be $8 - 2 - 1 = 5$. (Note the degrees of freedom are two less than the number of observations.)

Some important applications of χ^2-test are discussed in detail below :

(1) Sampling Distribution of the Sample Variance. The sampling distribution of the sample variance s^2 is particularly important in problems where one is concerned about the variability in a random sample. Since s^2 must always be positive, the distribution of s^2 cannot be a normal distribution.

The distribution of s^2 is a unimodal distribution which is skewed to the right, is a chi-square distribution. When the parent population is normal, with variance σ^2 and if random samples of size n with sample variance s^2 is drawn, can be shown to be related as :

$$s^2 = \frac{\chi^2\sigma^2}{v} = \frac{\chi^2\sigma^2}{n-1}$$

Therefore, $$\chi^2 = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

i.e., follows χ^2 distribution with $v = n - 1$.

(2) Confidence Interval for Variance. Confidence interval for variance σ^2 is based on the sampling distribution of $(n - 1)\, s^2/\sigma^2$ which follows χ^2 distribution with $v = (n - 1)$. A $100(1 - \alpha)$ per cent confidence interval for σ^2 is constructed by first obtaining an interval about $(n - 1)s^2/\sigma^2$. Two values of χ^2 are selected from the table (given in appendix) such that $\alpha/2$ is to the left of the smaller value and $\alpha/2$ is to the right of the larger values. Since the chi-square distribution is not symmetrical, $-\chi^2_{\alpha/2}$ does not give the approximate value of the left side of the distribution. The point that does give the correct probability is that of χ^2 cutting off $1 - \alpha/2$ of the right tail.

Therefore, a $100\,(1 - \alpha)$ per cent confidence interval for $(n - 1)\, s^2/\sigma^2$ is given by

$$-\chi^2_{\alpha/2} < \frac{(n-1)s^2}{\alpha^2} < x^2_{1-\alpha/2}$$

Solving these inequalities for σ^2, we get

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

which is the required $100(1-\alpha)$ per cent confidence interval for σ^2.

(3) Tests of Hypothesis Concerning Variance. In testing hypothesis about the variance of a normally distributed population, the null hypothesis is $H_0 : \sigma^2 = \sigma_0^{\,2}$ where $\sigma_0^{\,2}$ is some specified value of the population variance.

We know that $\chi^2 = \dfrac{(n-1)s^2}{\sigma^2}$

where s^2 is computed from a random sample of size n.

If $\chi^2 < \chi^2_{1-\alpha/2}$ and $\chi^2 > \chi^2_{\alpha/2}$, *i.e.*, when the computed value of χ^2 lies in the rejection region, we reject the null hypothesis, otherwise we accept the null hypothesis. This is shown in the diagram given below :

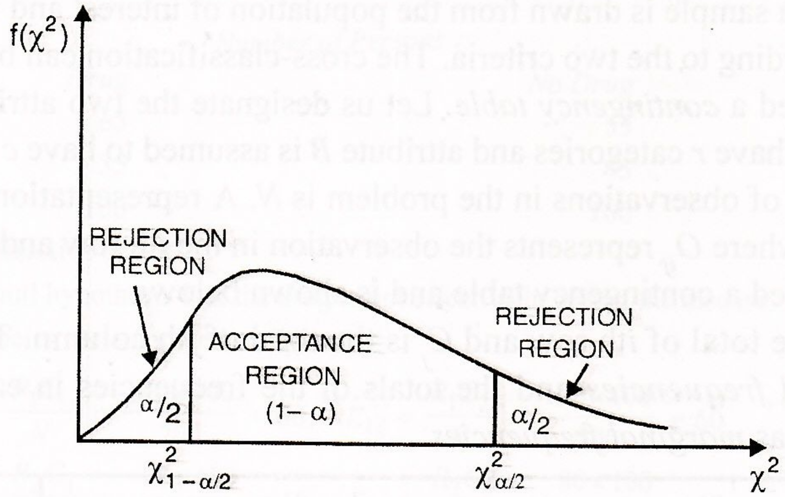


Illustration 1. Weights in kilograms of 10 shipments are given below :

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

Can we say that variance of the distribution of weight of all shipments from which the above sample of 10 shipments was drawn is equal to 20 square kilogram ?

Solution. Let the null hypothesis be that the variance of the distribution of shipments weight is 20 square kilogram, *i.e.*, $H_0 : \sigma^2 = 20$.

Weight (in kg.) X	$(X-\overline{X})$	$(X-\overline{X})^2$
38	-9	81
40	-7	49
45	-2	4
53	$+6$	36
47	0	0
43	-4	16
55	$+8$	64
48	$+1$	1
52	$+5$	25
49	$+2$	4
$\Sigma X = 470$	$\Sigma(X-\overline{X}) = 0$	$\Sigma(X-\overline{X})^2 = 280$

Sample mean $\bar{x} = \dfrac{\Sigma X}{n} = \dfrac{470}{10} = 47.$

Using the χ^2-test, statistic under the hypothesis $\sigma^2 = 20$

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{\Sigma(x-\bar{x})^2}{\sigma^2} = \frac{280}{20} = 14.$$

The table value of χ^2 for 9 *d.f.* at 5% level of significance is 16.919.

Since the calculated value of χ^2 is less than the tabulated value of χ^2, it is insignificant and the null hypothesis is accepted. Hence, we conclude that the data are consistent with the hypothesis that the variance of the distribution of weights of all shipments in the population is 20 kilograms.

(4) Test of Independence. One of the most frequent uses of χ^2 is for testing the null hypothesis that two criteria of classification are independent. They are independent if the distribution of one criterion in no way depends on the distribution of the other criterion. If they are not independent, there is an association between the two criteria. 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 the direction of the dependency.

To conduct the test, a sample is drawn from the population of interest and the observed frequencies are cross-classified according to the two criteria. The cross-classification can be conveniently displayed by means of a table called a *contingency table*. Let us designate the two attributes as A and B where, attribute A is assumed to have r categories and attribute B is assumed to have c categories. Furthermore, assume the total number of observations in the problem is N. A representation of these observations is shown below in a table where O_{ij} represents the observation in the ith row and jth column. Such a table in the matrix form is called a contingency table and is shown below.

In the table, R_i is the total of ith row and C_j is the total of jth column. The frequencies in these cells are termed as *cell frequencies* and the totals of the frequencies in each of the rows (R_i) and columns (C_j) are termed as *marginal frequencies.*

		Attribute B					Total
		B_1	B_2	B_3	B_j	B_c	
Attribute A	A_1	O_{11}	O_{12}	O_{13}	O_{1j}	O_{1c}	R_1
	A_2	O_{21}	O_{22}	O_{23}	O_{2j}	O_{2c}	R_2
	A_3	O_{31}	O_{32}	O_{33}	O_{3j}	O_{3c}	R_3
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	A_i	O_{i1}	O_{i2}	O_{i3}	O_{ij}	O_{ic}	R_i
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	A_r	O_{r1}	O_{r2}	O_{r3}	O_{rj}	O_{rc}	R_r
	Total	C_1	C_2	C_3	C_j	C_c	N

Expected cell frequencies are computed according to the multiplicative rule of probability. If two events are independent, the probability of their joint occurrence is equal to the product of their individual probabilities. Applying this rule to a contingency table, it is equivalent to say that, if two criteria of classification are independent, a joint probability is equal to the product of the two corresponding marginal probabilities. Thus, the expected cell frequencies are given by the formula :

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

To conduct the test, same χ^2 is employed as discussed earlier, *i.e.,*

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} \quad \text{or} \quad \Sigma \frac{(O - E)^2}{E}$$

will follow χ^2 distribution with $v = (r - 1)(c - 1)$ degrees of freedom.

While applying the test, the null hypothesis is that the two attributes are independent. If the calculated value of χ^2 is less than the table value at a specified level of significance, the null hypothesis holds true, *i.e.*, the two attributes are independent. If calculated value of χ^2 is greater than the table value, the null hypothesis is rejected, *i.e.,* the two attributes are associated.

Illustration 2. A sample of 200 persons with a particular disease was selected. Out of these, 100 were given a drug and the others were not given any drug. The results are as follows :

Number of Persons

	Drug	*No Drug*	*Total*
Cured	65	55	120
Not cured	35	45	80
Total	100	100	200

Test, whether the drug is effective or not.

Solution. Let us take the null hypothesis that the drug is not effective in curing the disease. Applying χ^2 test :

The expected* cell frequencies are computed as follows :

$$E_{11} = \frac{R_1 C_1}{N} = \frac{120 \times 100}{200} = 60; \quad E_{12} = \frac{R_1 C_2}{N} = \frac{120 \times 100}{200} = 60$$

$$E_{21} = \frac{R_2 C_1}{N} = \frac{80 \times 100}{200} = 40; \quad E_{22} = \frac{R_2 C_2}{N} = \frac{80 \times 100}{200} = 40$$

The table of expected frequencies is as follows :

60	60	120
40	40	80
100	100	200

O	E	$(O-E)^2$	$(O-E)^2/E$
65	60	25	0.417
35	40	25	0.625
55	60	25	0.417
45	40	25	0.625
			$\Sigma[(O-E)^2/E] = 2.084$

$$\chi^2 = \Sigma \frac{(O - E)^2}{E} = 2.084$$

*It may be noted that it is not necessary to calculate all the expected frequencies. It would be enough in a 2 × 2 table, if we calculate only one cell expected frequency. The others can be obtained by the process of deduction.

$$\chi^2 = \sum \frac{(O-E)^2}{E} = 2.084$$

$$v = (r-1)(c-1) = (2-1)(2-1) = 1$$

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

The calculated value of χ^2 is less than the table value. The null hypothesis is accepted. Hence, the drug is not effective in curing the disease.

Illustration 3. 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 patients' reactions to the treatment are recorded in the following table :

	Helped	*Harmed*	*No effect*	*Total*
Drug	150	30	70	250
Sugar pills	130	40	80	250
Total	280	70	150	500

On the basis of this data, can it be concluded that there is a significant difference in the effect of the drug and sugar pills ?

(MBA, Kumaun Univ., 2002; MBA, (HCA) DU, 2002)

Solution. Let us take the null hypothesis that there is no difference in the drug and sugar pills as far as their effect on curing cold is concerned.

Since it is a 2 × 3 table, the degrees of freedom would be $(2-1)(3-1) = 2$, *i.e.*, we will have to calculate only two expected frequencies and other four can be automatically determined.

Expected frequencies are computed as follows :

$$E_{11} = \frac{250}{500} \times 280 = 140; \quad E_{12} = \frac{250}{500} \times 70 = 35$$

The table of expected frequencies is :

140	35	75	250
140	35	75	250
280	70	150	500

Arranging the observed and expected frequencies in the following table :

O	E	$(O-E)^2$	$(O-E)^2/E$
150	140	100	0.714
130	140	100	0.714
30	35	25	0.714
40	35	25	0.714
70	75	25	0.333
80	75	25	0.333
			$\Sigma[(O-E)^2/E] = 3.522$

$$\chi^2 = \sum \frac{(O-E)^2}{E} = 3.522$$

The table of χ^2 for 2 *d.f.* at 5% level of significance is 5.99. The calculated value of χ^2 is less than the table value. Therefore the null hypothesis is accepted. Hence, we conclude that the drug and sugar pills do not differ significantly in curing cold.

(5) Test of Goodness of Fit. Tests of goodness of fit are used when we want to determine whether an actual sample distribution matches a known theoretical distribution. χ^2 test is popularly known as a test of goodness of fit for the reason, that it enables us to ascertain how well the theoretical distributions such as Binomial, Poisson, Normal, etc., fit empirical distribution, *i.e.*, those obtained from sample data. We hypothesize a theoretical distribution (Normal, for example) and then test to determine whether our sample came from or is comparable to the theoretical distribution. If there is a high degree of conformity between the two distributions, any slight difference may be assumed to be the result of sampling variation. On the other hand, any large discrepancy between the two distributions may lead to the conclusion that the sample was drawn from some theoretical distribution other than the one proposed.

While applying the chi-square test of goodness of fit, the null hypothesis usually states that the sample is drawn from the theoretical population distribution, and the alternate hypothesis usually states that it is not. The following illustrations would illustrate the use of χ^2 test of goodness of fit.

Illustration 4. The number of parts for a particular spare part in a factory was found to vary from day to day. In a sample study, the following information was obtained :

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

Test the hypothesis that the number of parts demanded does depend on the day of the week. *(MBA, Delhi Univ., 2000, 2005)*

Solution. Let us take the null hypothesis that the number of parts demanded does depend on the day of the week.

The number of spare parts demanded in a week are 6720 and if all days are same, we should expect 6720/6, *i.e.*, 1120 spare parts on a day of the week.

Day	*O*	*E*	$(O-E)^2$	$(O-E)^2/E$
Monday	1124	1120	16	0.014
Tuesday	1125	1120	25	0.022
Wednesday	1110	1120	100	0.089
Thursday	1120	1120	0	—
Friday	1126	1120	36	0.032
Saturday	1115	1120	25	0.022
				$\Sigma[(O-E)^2/E] = 0.179$

The table value of χ^2 for 5 *d.f.* at 5% level of significance is 11.07. The computed value of χ^2 is much less than the table value. The null hypothesis is accepted and we conclude that the demand for spare parts is dependent on the day of the week.

Illustration 5. A survey of 320 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	:	14	56	110	88	40	12

Is this result consistent with the hypothesis that male and female births are equally probable? *(MBA, IGNOU., 2002)*

Solution. Let us take the null hypothesis on the assumption that male and female births are equally probable, the probability of a male birth is $p = 1/2$. The expected number of families can be calculated by the use of binomial distribution. The probability of x male births in a family of 5 is given by

$$f(x) = {}^5C_x p^x q^{5-x} \qquad \text{[for } x = 0,1,2,3,4,5]$$

$$= {}^5C_x (½)^5 \qquad [\therefore p = q = ½]$$

To get the expected frequencies, multiply $f(x)$ by the total number $N = 320$. The calculations are shown below in the table :

x	$f(x)$	*Expected frequency* $= N f(x)$
0	${}^5C_0\left(\frac{1}{2}\right)^5 = 1/32$	$320 \times 1/32 = 10$
1	${}^5C_1\left(\frac{1}{2}\right)^5 = 5/32$	$320 \times 5/32 = 50$
2	${}^5C_2\left(\frac{1}{2}\right)^5 = 10/32$	$320 \times 10/32 = 100$
3	${}^5C_3(\left(\frac{1}{2}\right)^5 = 10/32$	$320 \times 10/32 = 100$
4	${}^5C_4\left(\frac{1}{2}\right)^5 = 5/32$	$320 \times 5/32 = 50$
5	${}^5C_5\left(\frac{1}{2}\right)^5 = 1/32$	$320 \times 1/32 = 10$

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

O	E	$(O-E)^2$	$(O-E)^2/E$
14	10	16	1.60
56	50	36	0.72
110	100	100	1.00
88	100	144	1.44
40	50	100	2.00
12	10	4	0.40
			$\Sigma[(O-E)^2/E] = 7.16$

$$\chi^2 = \Sigma \frac{(O-E)^2}{E} = 7.16$$

The table value of χ^2 for $v = 6 - 1 = 5$ at 5% level of significance is 11.07. The computed value of $\chi^2 = 7.16$ is less than the table value. Therefore, the null hypothesis is accepted. Thus, it can be concluded that male and female births are equally probable.

Illustration 6. The figures given below are (a) the theoretical frequencies of a distribution and (b) the frequencies 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	66	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. Let us take the null hypothesis that there is no difference in the observed frequencies and expected frequencies as obtained by the normal distribution.

Since the frequencies at the two corners are less than 5, they would be combined with the adjacent frequency.

O	E	$(O-E)^2$	$(O-E)^2/E$
1	2 }	16	0.941
12	15 }		
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
12	15 }		
1	2 }	16	0.941
			$\Sigma[(O-E)^2/E] = 3.839$

$v = 13 - 2 - 3 = 8$ (after grouping, 11 classes are left and for normal the degrees of freedom is less by 3 than the number of classes).

The table value of χ^2 for 8 $d.f.$ at 5% level of significance is 15.51. The calculated value of χ^2 is less than the table value and hence the fit is good.

(6) Test of Homogeneity. 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. Another way of stating the problem is to say that 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 a contingency table. The analytical procedure is same as that discussed for test of independence.

The main difference is that, in tests of independence, we are concerned with the problem whether the two attributes are independent or not while in tests of homogeneity, we are concerned whether the different samples come from the same population. Another difference is that test of independence involve a single sample but test of homogeneity involves two or more samples, one from each population. When there are two populations involved, and when the characteristics of interest consist of two categories, the test of homogeneity is the same as testing hypothesis about the difference between two population's proportions which was discussed in the chapter on tests of hypothesis.

Illustration 7. A random sample of 400 persons was selected from each of three age groups and each person was asked to specify which of three types of TV programmes be preferred. The results are shown in the following table :

TYPES OF PROGRAMME

Age group	*A*	*B*	*C*	*Total*
Under 30	120	30	50	200
30–44	10	75	15	100
45 and above	10	30	60	100
Total	140	135	125	400

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

Solution. Let us take the null hypothesis that the populations are homogeneous with respect to different types of television programmes they prefer.

O	E	$(O-E)^2$	$(O-E)^2/E$
120	70.00	2500.00	35.7143
10	35.00	625.00	17.8571
10	35.00	625.00	17.8571
30	67.50	1406.25	20.8333
75	33.75	1701.56	50.4166
30	33.75	14.06	0.4166
50	62.50	156.25	2.5000
15	31.25	264.06	8.4499
60	31.25	826.56	26.4499
			$\Sigma\,[(O-E)^2/E] = 180.4948$

$$\chi^2 = \Sigma \frac{(O-E)^2}{E} = 180.495$$

The table value of χ^2 for 4 *d.f.* at 5% level of significance is 9.488.

The calculated value of χ^2 is much greater than the table value. We reject the null hypothesis and conclude that the populations are not homogeneous with respect to the type of TV programmes preferred.

Cautions while Applying χ^2 Test

χ^2 test is very popularly used in practice. However, it is unfortunate to find that the number of misuses of χ^2 test has become surprisingly large. The test must be used with greater care, keeping in mind the assumptions on which it is based. Some sources of error in the application of this test revealed by a survey of all papers published in the journal of *Experiment Psychology* are :

(i) Small theoretical frequencies.
(ii) Neglect of frequencies of non-occurrence.
(iii) Indeterminate theoretical frequencies.
(iv) Incorrect or questionable categorizing.
(v) Failure to equalize the sum of the observed frequencies and the sum of the theoretical frequencies.
(vi) Use of non-frequency data.

It should also be noted that χ^2 test is not the only non-parametric test. There are many other non-parametric tests that can be used in business decisions.

MISCELLANEOUS ILLUSTRATIONS

Illustration 8. Of the 1,000 workers in a factory exposed to an epidemic, 700 in all were attacked, 400 had been inoculated and of these, 200 were attacked. On the basis of this information, can it be said that inoculation and attack are independent ?

Solution : The given information can be put in a tabular form as follows :

	Inoculated	*Not inoculated*	*Total*
Attacked	200	500	700
Not attacked	200	100	300
Total	400	600	1000

Let us take the null hypothesis that inoculation and attack are independent. Applying χ^2 test, the expected frequency corresponding to first row and first column is $E_{11} = \frac{700 \times 400}{1000} = 280$. Table of expected frequencies would be as follows :

280	420	700
120	180	300
400	600	1000

O	E	$(O-E)^2$	$(O-E)^2/E$
200	280	6400	22.857
200	120	6400	55.333
500	420	6400	15.238
100	180	6400	35.556
			$\Sigma[(O-E)^2/E] = 128.984$

$$\chi^2 = \Sigma \frac{(O-E)^2}{E} = 128.984$$

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

The calculated value of χ^2 is much greater than the table value. The null hypothesis is rejected. Hence, inoculation and attack are not independent.

Illustration 9. A sample analysis of examination results of 200 MBA's was made. It was found that 46 students have failed, 68 secured a third division, 62 secured a second division and rest were placed in the first division. Are these figures commensurate with the general examination result which is in the ratio of 2:3:3:2, for various categories respectively ? *(MBA, DU, 2002)*

Solution. Let us take the null hypothesis that there is no difference in the observed and expected results. On the basis of ratio 2 : 3 : 3 : 2, the expected number of students failing, getting third division, second division, and first division, should be $\frac{200 \times 2}{10} = 40, 60, 60, 40$ respectively.

Applying χ^2 test :

Category	O	E	$(O-E)^2$	$(O-E)^2/E$
Failed	46	40	36	0.900
Third Division	68	60	64	1.067
Second Division	62	60	4	0.067
First Division	24	40	256	6.400
				$\Sigma[(O-E)^2/E] = 8.434$

The table value of χ^2 for 3 *d.f.* at 5% level of significance is 7.81. The calculated value of χ^2 is greater than the table value. The null hypothesis does not hold true. Hence, the given results are not commensurate with the general examination results.

Illustration 10. An automobile manufacturing firm is bringing out a new model. In order to map out its advertising campaign, it wants to determine whether the model appeal depends on age group or not. The firm takes a random sample from persons attending a preview of the new model and obtained the results summarised below :

	AGE GROUPS				
Persons who	*Under 20*	*20–40*	*40–50*	*50 and over*	*Total*
Liked the car	146	78	48	28	300
Disliked the car	54	52	32	62	200
Total	200	130	80	90	500

Test, whether the model appeal and age groups are independent. *(MBA, DU, 2002)*

Solution. Let us take the null hypothesis that the model appeals equally to all the age groups, *i.e.*, model appeal does not depend on age groups.

$$E_{11} = \frac{300}{500} \times 200 = 120, \quad E_{12} = \frac{300}{500} \times 130 = 78, \quad E_{13} = \frac{300}{500} \times 80 = 48$$

The table of expected frequencies is :

120	78	48	54	300
80	52	32	36	200
200	130	80	90	500

O	E	$(O-E)^2$	$(O-E)^2/E$
146	120	676	5.633
54	80	676	8.450
78	78	0	–
52	52	0	–
48	48	0	–
32	32	0	–
28	54	676	12.519
62	36	676	18.778
			$\Sigma[(O-E)^2/E] = 45.38$

$$\chi^2 = \Sigma \frac{(O-E)^2}{E} = 45.38$$

The table value of χ^2 for 3 d.f. at 5% level of significance is 7.81. The calculated value of χ^2 is much greater than the table value. The null hypothesis is rejected. Hence, the model appeal depends on the age groups.

Illustration 11. The following figures show the distribution of digits in numbers chosen at random from a telephone directory :

Digit :	0	1	2	3	4	5	6	7	8	9	Total
Frequency :	1026	1107	997	966	1075	933	1107	972	964	853	10,000

Test, whether the digits may be taken to occur equally frequently in the directory.

[MBA, IIT, Roorkee, 2000; M.Com., Madras Univ., 2009]

Solution. The null hypothesis is that the digits occur equally frequently in the directory.

The expected frequency for each of the digits, 0, 1, 2, ...9 is 10,000/10 = 1,000

Arranging the observed and expected frequencies in the following table :

O	E	$(O-E)^2$	$(O-E)^2/E$
1,026	1,000	676	0.676
1,107	1,000	11,449	11.449
997	1,000	9	0.009
966	1,000	1,156	1.156
1,075	1,000	5,625	5.625
933	1,000	4,489	4.489
1,107	1,000	11,449	11.449
972	1,000	784	0.784
964	1,000	1,296	1.296
853	1,000	21,609	21.609
			$\Sigma\,[(O-E)^2/E] = 58.542$

$$\chi^2 = \sum \frac{(O-E)^2}{E} = 58.542$$

The table value of χ^2 for $\nu = 10 - 1 = 9$ *d.f.* at 5% level of significance is 16.919. The computed value of χ^2 is much greater than the table value. The null hypothesis is rejected. Thus, it can be concluded that the digits are not uniformly distributed in the directory.

Illustration 12. The number of automobile accidents per week in a certain city were as follows :

12, 8, 20, 2, 14, 10, 15, 6, 9, 4

Are these frequencies in agreement with the belief that accident conditions were the same during this 10 week period ?

(MBA, Delhi Univ., 1999)

Solution. Let the null hypothesis be that the number of accidents per week in a certain city are consistent with the belief that the accident conditions were same during the ten-week period.

As the total number of accidents over the 10 week period are 100, according to the statement of the null hypothesis, these accidents should be uniformly distributed over the 10 week period. Therefore, the expected number of accidents per week is equal to 100/10 = 10.

Week	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
				$\Sigma\,[(O-E)^2/E] = 26.6$

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

Table value of χ^2 at 5% level of significance for 9 *d.f.* is 16.819.

Since the calculated value of χ^2 is greater than the table value, therefore, the null hypothesis is rejected. Hence, we conclude that the accident conditions are not the same (uniform) over the 10 week period.

Illustration 13. 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 to test whether there is any significant difference between consumption of tea among Hindu and non-Hindu families. *(MBA, Madurai Kamaraj Univ., 2003)*

Solution. The above data can be conveniently arranged in the following table as :

	Hindus	*Non-Hindus*	*Total*
No. of families consuming tea	1236	164	1400
No. of families not consuming tea	564	36	600
Total	1800	200	2000

Let the null hypothesis be that the two attributes (consumption of tea and community) are independent.

The expected frequencies are computed as follows :

$$E_{11} = \frac{R_1C_1}{N} = \frac{1400 \times 1800}{2000} = 1260; \quad E_{12} = \frac{R_1C_2}{N} = \frac{1400 \times 200}{2000} = 140$$

The table of expected frequencies is :

1260	140	1400
540	60	600
1800	200	2000

Arranging the observed frequencies with the corresponding expected frequencies as given in the following table :

O	E	$(O-E)^2$	$(O-E)^2/E$
1236	1260	576	0.458
564	540	576	1.067
164	140	576	4.114
36	60	576	9.600
			$\Sigma\,[(O-E)^2/E] = 15.239$

$$\chi^2 = \Sigma\,\frac{(O-E)^2}{E} = 15.239$$

The table value of χ^2 for 1 *d.f.* at 5% level of significance is 3.841. Since the calculated value of χ^2 is much greater than the table value of χ^2, the null hypothesis is rejected. Hence, we conclude that the two communities differ significantly as regards to the consumption of tea.

Illustration 14. The following table gives the number of good and bad parts produced by each of three shifts in a factory :

Shift.	*Good*	*Bad*	*Total*
Day	900	130	1030
Evening	700	170	870
Night	400	200	600
Total	2000	500	2500

Is there any association between the shift and the quality of parts produced ?

(MBA, Kumaun Univ., 2000; MBA, Delhi Univ., 2005)

Solution. Let us take the null hypothesis that there is no association between the shift and quality of parts produced. The observed frequencies are :

Shift	*Good*	*Bad*	*Total*
Day	900	130	1030
Evening	700	170	870
Night	400	200	600
Total	2000	500	2500

The expected frequencies are computed as follows :

$$E_{11} = \frac{1030}{2500} \times 2000 = 824 \ ; \quad E_{21} = \frac{870}{2500} \times 2000 = 696$$

The table of expected frequencies is :

824	206	1030
696	174	870
480	120	600
2000	500	2500

O	E	$(O-E)^2$	$(O-E)^2/E$
900	824	5776	7.010
700	696	16	0.023
400	480	6400	13.333
130	206	5776	28.039
170	174	16	0.092
200	120	6400	53.333
			$\Sigma[(O-E)^2/E] = 101.83$

$$\chi^2 = \sum \frac{(O-E)^2}{E} = 101.83$$

For $\nu = 2$, $\chi^2_{0.05} = 5.991$.

Since the calculated value of χ^2 is much greater than the table value, the null hypothesis is rejected. On the basis of given data, we can, therefore, conclude that there is association between shift and the quality of parts produced.

Illustration 15. It has been stated that potential respondents are more likely to reply to questionnaire printed on light coloured paper than a dark coloured paper. Questionnaires were sent out on a random basis with the following results :

Colour used	*Response received*	*No response*	*Total*
Light	120	80	200
Dark	100	100	200
Total	220	180	400

Use an appropriate test at 5% level of significance to determine whether or not light colour paper yields better response.

Solution. Let us take the null hypothesis that the colour of the paper does not affect the response. Applying χ^2 test,

$$E_{11} = \frac{200}{400} \times 220 = 110; \quad E_{12} = \frac{200}{400} \times 180 = 90$$

The table of expected frequencies is :

110	90	200
110	90	200
220	180	400

O	E	$(O-E)^2$	$(O-E)^2/E$
120	110	100	0.909
100	110	100	0.909
80	90	100	1.111
100	90	100	1.111
			$\Sigma[(O-E)^2/E] = 4.04$

$$\chi^2 = \sum \frac{(O-E)^2}{E} = 4.04$$

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

The calculated value of χ^2 is greater than the table value. The null hypothesis is rejected. Hence, we may conclude that the colour of the paper may affect the response.

Illustration 16. A school bought a total of 500 colour television sets. Three different brands were purchased, and their repair records were kept for each set's operation. The data is given below :

Brand	*Number of Repairs* 0	1	2 or more	Total
A	143	70	37	250
B	90	67	43	200
C	17	13	20	50
Total	250	150	100	500

Is there a relationship between brand and number of repairs ?

Solution. Let us take the null hypothesis that there is no relationship between brand and number of repairs. Applying χ^2 test :

$$E_{11} = \frac{250}{500} \times 250 = 125 \; ; \quad E_{12} = \frac{250}{500} \times 150 = 75$$

$$E_{21} = \frac{200}{500} \times 250 = 100 \; ; \quad E_{22} = \frac{200}{500} \times 150 = 60.$$

The table of expected frequencies is :

125	75	50	250
100	60	40	200
25	15	10	50
250	150	100	500

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
			$\Sigma[(O-E)^2/E] = 21.174$

$$\chi^2 = \sum \frac{(O-E)^2}{E} = 21.174$$

For $\nu = 4$, $\chi^2_{0.05} = 9.488$.

The calculated value of χ^2 is more than the table value. The null hypothesis is rejected. Hence, there is a relationship between brand and number of repairs.

Illustration 17. The divisional manager of a retail chain believes the average number of customers entering each of the five stores in his division weekly is the same.

In a given week, a manager reports the following number of customers in their stores:

3000, 2960, 3100, 2780, 3160.

Test the divisional manager's belief at the 10 per cent level of significance. (*MBA, Delhi Univ., 2003*)

Solution. Let us take the null hypothesis that there is no significant difference in the number of customers entering each of the five stores. The number of customers entering each of the five stores is 15000, therefore, the expected frequency for each store is 15000/5 = 3000. Applying χ^2 test :

O	E	$(O-E)^2$	$(O-E)^2/E$
3000	3000	0	0
2960	3000	1600	0.533
3100	3000	10000	3.333
2780	3000	48400	16.133
3160	3000	25600	8.533
			$\Sigma[(O-E)^2/E] = 28.532$

$$\chi^2 = \sum \frac{(O-E)^2}{E} = 28.532$$

For $\nu = 4$, $\chi^2_{0.10} = 13.277$.

The calculated value of χ^2 is more than the table value. The null hypothesis is rejected. Hence, there is a significant difference in the number of customers entering each of the five stores.

Illustration 18. A die is thrown 150 times with the following results :

No. turned up :	1	2	3	4	5	6
Frequency :	19	23	28	17	32	31

Test the hypothesis that the die is unbiased.

Solution. Let us take the null hypothesis that there is no significant difference in the observed and expected frequencies in the throw of the die, *i.e.*, die is unbiased.

The expected frequencies for 1, 2, 3, etc. would be $\frac{150}{6} = 25$. Applying the χ^2 test :

O	E	$(O-E)^2$	$(O-E)^2/E$
19	25	36	1.44
23	25	4	0.16
28	25	9	0.36
17	25	64	2.56
32	25	49	1.96
31	25	36	1.44
			$\Sigma[(O-E)^2/E] = 7.92$

$$\chi^2 = \sum \frac{(O-E)^2}{E} = 7.92$$

For $\nu = 5$, $\chi^2_{0.05} = 11.07$.

The calculated value of χ^2 is smaller than the table value. The null hypothesis is accepted. Hence, the die seems to be unbiased.

Illustration 19. Use chi-square test to test if the two attributes in the following contingency table are independent :

Performance	Training: Intensive	Good	Average	Total
Above Average	100	150	40	290
Average	100	100	100	300
Poor	50	80	150	280
Total	250	330	290	870

(*MBA, Punjab Univ., 2005*)

Solution. Let us take the null hypothesis that the attributes performance and training are independent, *i.e.*, not associated. Applying χ^2 test :

$$E_{11} = \frac{290}{870} \times 250 = 83.33; \qquad E_{12} = \frac{290}{870} \times 330 = 110;$$

$$E_{21} = \frac{300}{870} \times 250 = 86.21\,; \qquad E_{22} = \frac{300}{870} \times 330 = 113.79.$$

The table of expected frequencies is :

83.33	110.00	96.67	290
86.21	113.79	100.0	300
80.46	106.21	93.33	280
250	330	290	870

O	E	$(O-E)^2$	$(O-E)^2/E$
100	83.33	277.89	3.335
100	86.21	190.16	2.21
50	80.46	927.81	11.53
150	110.00	1600.00	14.545
100	113.79	190.16	1.671
80	106.21	686.96	6.468
40	96.67	3211.49	33.221
100	100.00	0.00	0.000
150	93.33	3211.49	34.41
			$\Sigma\,[(O-E)^2/E] = 107.39$

$$\chi^2 = \sum \frac{(O-E)^2}{E} = 107.39$$

For $v = 4$, $\chi^2_{0.05} = 9.49$.

The calculated value of χ^2 is much greater than the table value. The null hypothesis is rejected. Hence, performance and training are associated.

Illustration 20. A cigarette company interested in the effect of sex on the type of cigarettes smoked and has collected the following data from a random sample of 150 persons :

Cigarette	Male	Female	Total
A	25	30	55
B	40	15	55
C	30	10	40
Total	95	55	150

Test, whether the type of cigarette smoked and sex are independent.

(*MBA, Osmania Univ., 2006*)

Solution. Let us take the null hypothesis that there is no association between the type of cigarettes smoked and the sex. Applying χ^2 test :

$E_{11} = \frac{55}{150} \times 95 = 34.83, E_{21} = \frac{55}{150} \times 95 = 34.83, E_{12} = \frac{55}{150} \times 55 = 20.17$

The table of expected frequencies is :

34.83	20.17	55
34.83	20.17	55
25.34	14.66	40
95	55	150

O	E	$(O-E)^2$	$(O-E)^2/E$
25	34.83	96.63	2.774
40	34.83	26.73	0.767
30	25.34	21.72	0.857
30	20.17	96.63	4.791
15	20.17	26.73	1.325
10	14.66	21.72	1.482
			$\Sigma\,[(O-E)^2/E\,] = 11.996$

$$\chi^2 = \Sigma \frac{(O-E)^2}{E} = 11.996$$

For $\nu = 2, \chi^2_{0.05} = 5.99$

The calculated value of χ^2 is greater than the table value. The null hypothesis is rejected. Hence, type of cigarette smoked and sex are not independent.

Illustration 21. A certain drug was administered to 456 males out of a total of 720 in a certain locality to test its efficacy against typhoid. Relevant data is given below :

	Infection	*No Infection*	*Total*
Administered the drug	144	312	456
Not Administered	192	72	264
Total	336	384	720

(*MBA, Sukhadia Univ.,* 2004)

Solution : Let us take the null hypothesis that there is no significant difference in the infection caused due to administration of drug or otherwise.

$$E_{11} = \frac{456}{720} \times 336 = 212.8$$

The table of expected frequencies is :

212.8	243.2	456
123.2	140.8	264
336	384	720

O	E	$(O-E)^2$	$(O-E)^2/E$
144	212.8	4733.44	22.24
192	123.2	4733.44	38.42
312	243.2	4733.44	19.46
72	140.8	4733.44	33.6
			$\Sigma\,[(0-E)^2/E] = 111.73$

$$\chi^2 = \Sigma \frac{(O-E)^2}{E} = 111.73$$

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

The calculated value of χ^2 is much higher than the table value. Hence, the hypothesis is rejected. There is a significant difference in the infection caused due to administration of drug.

Illustration 22. Five coins are tossed 3,200 times and the number of heads appearing each time are noted. At the end, the following results were obtained :

No. of heads :	0	1	2	3	4	5
Frequency :	80	570	1100	900	500	50

Use chi-square test of goodness of fit to determine whether the coins are unbiased. (*MBA, Hyderabad Univ., 2006*)

Solution. Let the null hypothesis be that the coins are unbiased. If the coins are unbiased, then the distribution of heads will follow binomial distribution. Calculating the expected frequencies by using the formula $f(x) = {}^nC_x\, p^x\, q^{n-x}$.

The table of expected frequencies is :

No. of heads	$f(x) = {}^nC_x\, p^x\, q^{n-x}$	*Expected frequency* $= N f(x)$
0	${}^5C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^5 = 1\left(\frac{1}{2}\right)^5$	$3200 \times \frac{1}{32} = 100$
1	${}^5C_1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^4 = 5\left(\frac{1}{2}\right)^5$	$3200 \times 5 \times \frac{1}{32} = 500$
2	${}^5C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^3 = 10\left(\frac{1}{2}\right)^5$	$3200 \times 10 \times \frac{1}{32} = 1000$
3	${}^5C_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^2 = 10\left(\frac{1}{2}\right)^5$	$3200 \times 10 \times \frac{1}{32} = 1000$
4	$\left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right) = 5\left(\frac{1}{2}\right)^5$	$3200 \times 5 \times \frac{1}{32} = 500$
5	${}^5C_5 \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^0 = 1\left(\frac{1}{2}\right)^5$	$3200 \times \frac{1}{32} = 100$

O	*E*	$(O-E)$	$(O-E)^2$	$(O-E)^2/E$
80	100	–20	400	4.0
570	500	+70	4,900	9.8
1,100	1,000	+100	10,000	10.0
900	1,000	–100	10,000	10.0
500	500	0	0	0.0
50	100	–50	2500	25.0
				$\Sigma(O-E)^2/E = 58.8$

$$\chi^2 = \sum \frac{(O-E)^2}{E} = 58.8.$$

For $v = 5$, $\chi^2_{0.05} = 11.07$.

Since the computed value of χ^2 is much greater than the table value of χ^2, therefore, we reject the null hypothesis. Hence, the coins are biased.

Illustration 23. A production supervisor is interested in knowing if the number of breakdowns on four machines is independent of the shift using the machines. Test this hypothesis based on the following sample information :

		Machine				
		A	*B*	*C*	*D*	*Total*
Shift	*Morning*	15	10	18	12	55
	Evening	12	8	15	10	45
	Total	27	18	33	22	100

(*MBA, Delhi Univ., 2004, 2007*)

Solution. Let us take the null hypothesis that the number of breakdowns is independent of the shift using the machines. The expected frequencies are :

$$E_{11} = \frac{55}{100} \times 27 = 14.85, E_{12} = \frac{55}{100} \times 18 = 9.90, \ E_{13} = \frac{55}{100} \times 33 = 18.15$$

The table of expected frequencies is :

14.85	9.90	18.15	12.10	55
12.15	8.10	14.85	9.90	45
27	18	33	22	100

O	E	$(O-E)^2$	$(O-E)^2/E$
15	14.85	0.0225	0.0015
12	12.15	0.0225	0.0019
10	9.90	0.0100	0.0010
8	8.10	0.0100	0.0012
18	18.15	0.0225	0.0012
15	14.85	0.0225	0.0015
12	12.10	0.0100	0.0008
10	9.90	0.0100	0.0010
			$\Sigma[(O-E)^2/E] = 0.0101$

$$\chi^2 = \Sigma \frac{(O-E)^2}{E} = 0.0101$$

For $\nu = 3$, $\chi^2_{0.05} = 7.81$

The calculated value of χ^2 is much less from the table value. The null hypothesis is accepted. Hence, the number of breakdowns is independent of the shift using the machines.

Illustration 24. Two sample polls of votes for two candidates A and B for a public office are taken, one from among residents of rural area and one from urban areas. The results are given below. Examine, whether the nature of the area is related to the voting preference in this election.

Votes for / *Area*	*A*	*B*	*Total*
Rural	620	380	1000
Urban	550	450	1000
Total	1170	830	2000

(*MBA., IGNOU, 2001; MBA, Anna Univ., 2003*)

Solution. Let us take the null hypothesis that the nature of area is not related to the voting preference in this election. Applying χ^2 test :

$$E_{11} = \frac{1000}{2000} \times 1170 = 585$$

The table of expected frequencies is :

585	415	1000
585	415	1000
1170	830	2000

O	E	$(O-E)^2$	$(O-E)^2/E$
620	585	1225	2.094
550	585	1225	2.094
380	415	1225	2.952
450	415	1225	2.952
			$\Sigma[(O-E)^2/E] = 9.992$

$$\chi^2 = \Sigma \frac{(O-E)^2}{E} = 9.992$$

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

The calculated value of χ^2 is greater than the table value. The null hypothesis is rejected. Hence, the nature of area is related to the voting preference.

Illustration 25. In setting sales targets, the marketing manager makes the assumption that order potentials are the same for each of the four sales territories. A sample of 200 sales data is given below :

Sales Territories

I	*II*	*III*	*IV*
60	45	59	36

Should the manager's assumption be rejected [Given : the chi-square value at 5% level of significance for 3 degrees of freedom is 7.81] *(MBA, Delhi Univ., 2003)*

Solution. Let us take the null hypothesis that order potentials are the same for each of the four sales territories. Applying χ^2 test : A sample of 200 sales for four territories is given. Therefore, the expected sale for each territory is 200/4 = 50.

O	E	$O-E$	$(O-E)^2$	$(O-E)^2/E$
60	50	10	100	2.00
45	50	–5	25	0.50
59	50	9	81	1.62
36	50	14	196	3.92
				$\Sigma[(O-E)^2/E] = 8.04$

The table value of χ^2 for 3 *d.f.* at 5% level of significance is 7.81 which is less than the calculated value of χ^2. The null hypothesis is rejected and we can conclude that marketing manager assumption is not justified. Hence, order potentials are not same for each of the four sales territories.

Illustration 26. A sample of parts provided the following table data on quality of parts by production shift :

Shift	*Number good*	*Number Defective*	*Total*
First	368	32	400
Second	285	15	300
Third	176	24	200
Total	829	71	900

Use five per cent level of significance to test the hypothesis that quality of parts is independent of the production shift. *(MBA, Delhi Univ. 2008)*

Solution. Let us take the null hypothesis that there is no significant difference between the quality of part produced and the production shift. Applying χ^2 test :

$$E_{11} = 368.44,\ E_{21} = 276.33$$

The table of expected frequencies is :

368.44	31.56	400
276.33	23.67	300
184.23	15.77	200
829	71	900

O	E	$(O-E)^2$	$(O-E)^2$/E
368	368.44	0.1936	0.0005
285	276.33	75.1869	0.2720
176	184.23	67.7329	0.3676
32	31.56	0.1936	0.0061
15	23.67	75.1689	3.1757
24	15.77	67.7329	4.2950
			$\Sigma[(0-E)^2/E] = 8.1169$

$$\chi^2 = \Sigma \frac{(O-E)^2}{E} = 8.1169$$

For $\nu = 2, \chi^2_{0.05} = 5.99$

The calculated value of χ^2 is greater than the table value. Hence, the null hypothesis is rejected. We, therefore, conclude that quality of part is not independent of the production shift.

Illustration 27. At a level of significance of 0.10, can we conclude that the following 400 observations follow a Poisson distribution ?

No. of arrivals (per hr.)	0	1	2	3	4	5 or more
No. of hours	20	57	98	85	78	62

(MBA, IGNOU, 2003)

Solution. To solve this question first, we have to calculate expected frequencies by applying Poisson distribution and then using χ^2 test of goodness of fit to conclude whether given observations follow the distributions or not.

Let us take the null hypothesis that the given data fits to Poisson distribution.

FITTING OF POISSON DISTRIBUTION

X	f	fX
0	20	0
1	57	57
2	98	196
3	85	255
4	78	312
5	62	310
	$N = 400$	$\Sigma fX = 1130$

$$\bar{X} = \frac{\Sigma fX}{N} = \frac{1130}{400} = 2.825$$

Expected Frequencies as per Poisson law

$$NP(0) = e^{-m} \times N = 0.06 \times 400 = 24 \qquad [e^{-m} = 0.06]$$

$$NP(1) = NP(0) \times m = 24 \times 2.825 = 67.8$$

$$NP(2) = NP(1) \times \frac{m}{2} = 67.8 \times \frac{2.825}{2} = 95.77$$

$$NP(3) = NP(2) \times \frac{m}{3} = 95.77 \times \frac{2.825}{3} = 90.18$$

$$NP(4) = NP(3) \times \frac{m}{4} = 90.18 \times \frac{2.825}{4} = 63.69$$

$$NP(5) = NP(4) \times \frac{m}{5} = 63.69 \times \frac{2.825}{5} = 35.99$$

Applying χ^2 test by rounding off the expected frequencies :

O	E	$(O-E)^2$	$(O-E)^2/E$
20	24	16	0.667
57	68	121	1.779
98	96	4	0.042
85	90	25	0.278
78	64	196	3.062
62	36	676	18.778
			$\Sigma[(O-E)^2/E] = 24.606$

$$\chi^2 = \Sigma \frac{(O-E)^2}{E} = 24.606$$

For $\nu = 5, \chi^2_{0.10} = 9.24$

The calculated value of χ^2 is greater than the table value. Therefore, we reject the null hypothesis. Hence, Poisson distribution does not provide good fit to the given data.

Illustration 28. In setting sales targets, the marketing manager makes the assumption that order potentials are the same for each of the four sales territories. A sample of 200 sales data is given below :

Sales Territories

I	*II*	*III*	*IV*
60	45	59	36

Should the manager's assumption be rejected. (*MBA, Delhi Univ., 2009*)

Solution. Let us take the null hypothesis that the order potentials are the same for each of the four sales territories. Hence, the expected sales target should be, *i.e.*, 50 in each sales territory. Applying χ^2 test :

O	E	$(O-E)^2$	$(O-E)^2/E$
60	50	100	2.00
45	50	25	0.50
59	50	81	1.62
36	50	196	3.92
			$\Sigma [(O-E)^2/E] = 8.04$

$$\chi^2 = \Sigma \frac{(O-E)^2}{E} = 8.04$$

For $\nu = 3, \chi^2_{0.05} = 7.81$

The calculated value of χ^2 is more than the table value. Hence, the null hypothesis is rejected. We, therefore, conclude that the order potential is not the same for each of the four sales territories.

Illustration 29. The following table gives the number of aircraft accidents that occurred during the various days of the week. Test, whether the accidents are uniformly distributed over the week.

Days	Mon.	Tue.	Wed.	Thurs.	Fri.	Sat.
No. of accidents	14	18	12	11	15	14

(*MBA, IGNOU, 2006*)

Solution. Let us take the null hypothesis that the accidents are uniformly distributed over the week. Applying χ^2 test :

Days	O	E	$(O-E)^2$	$(O-E)^2/E$
Mon.	14	14	0	0.000
Tue.	18	14	16	1.143
Wed.	12	14	4	0.286
Thurs.	11	14	9	0.643
Fri.	15	14	1	0.071
Sat.	14	14	0	0.000
	84			$\Sigma [(O-E)^2/E = 2.143$

$$\chi^2 = \Sigma \frac{(O-E)^2}{E} = 2.143$$

For $\nu = 5, \chi^2_{0.05} = 11.07$. The calculated value is much less than the table value. The null hypothesis is accepted. We, therefore, conclude that the accidents are uniformly distributed.

Illustration 30. One of the questions in a recent survey conducted by an Airline consultancy firm was "In the past 12 months, when travelling for business, what type of Airline ticket did you purchase most often?" The data obtained are shown in the following contingency table :

Type of ticket	*Type of Flight* Domestic	International	*Total*
First class	29	22	51
Business/executive class	95	121	216
Economy class	518	135	653
	642	278	920

Using 5% level of significance test for the independence of type of flight and type of ticket. (*MBA, Delhi Univ., 2006*)

Solution. Let us take the hypothesis that the type of ticket and type of flight are independent. Applying χ^2 test, let us calculate the expected frequencies :

$$E_{11} = \frac{51}{920} \times 642 = 35.59 \approx 36$$
$$E_{21} = \frac{216}{920} \times 642 = 150.73 \approx 151$$

36	15	51
151	65	216
455	198	653
642	278	920

O	E	$(O-E)^2$	$(O-E)^2/E$
29	36	49	1.361
95	151	3136	20.768
518	455	3969	8.723
22	15	49	3.267
121	65	3136	48.246
135	198	3969	20.045
			$\Sigma[(0-E)^2/E] = 102.410$

$$v = (r-1)(c-1) = (3-1)(2-1) = 2$$

For $\qquad v = 2, \chi^2_{0.05} = 5.99$

The calculated value of χ^2 is much greater than the table value. The hypothesis is rejected. Hence the type of ticket and the type of flight are not independent.

Illustration 31. In a study of brand loyalty in the automative industry, new car customers were asked whethr the make of their new car was the same as the make of their previous car. The break down of 600 responses shows the brand loyalty for domestic, European and American cars.

Purchaser	*Domestic*	*European*	*American*	*Total*
Same make	125	55	68	248
Different make	140	105	107	352
	265	160	175	600

Test the hypothesis to determine whether brand loyalty is independent of the manufacturer. Use level of significance 5%. What is your conclusion? If a significant difference is found, which manufacturer appears to have the greatest brand loyalty ? *(MBA, Delhi Univ., 2009)*

Solution. Let us take the hypothesis that the brand loyalty is independent of the manufacturer. Applying χ^2 test, calculation of expected values :

$$E_{11} = \frac{248}{600} \times 265 = 109.53 \approx 110$$
$$E_{21} = \frac{248}{600} \times 160 = 66.13 \approx 66$$

The table of expected frequencies shall be

110	66	72	248
155	94	103	352
265	160	175	600

O	E	$(O-E)^2$	$(O-E)^2/E$
125	110	225	2.045
140	155	225	1.452
55	66	121	1.833
105	94	121	1.287
68	72	256	3.556
107	103	16	0.155
			$\Sigma[(0-E)^2/E] = 10.328$

$$\chi^2 = \Sigma\frac{(O-E)^2}{E} = 10.328$$

$$v = (r-1)(c-1) = (3-1)(2-1) = 2$$

For $v = 2, \chi^2_{0.05} = 5.99$

The calculated value of χ^2 is much greater than the table value. Hence the hypothesis is rejected. Hence the brand loyalty is not independent of the manufacturer.

Illustration 32. The personnel department of IBM is doing a study about job satisfaction. A random sample of 375 employees was given a test designed to diagnose the level of job satisfaction. Each employee's salary was also recorded in the table below. Use an appropriate significance test to determine if salary and job satisfaction are independent at 5% level of significance.

Salary versus Job Satisfaction

Satisfaction	*Under \$ 50000*	*\$ 50000—\$ 75000*	*Over \$ 75000*	*Total*
High	30	30	20	80
Medium	100	85	30	215
Low	45	20	15	80
Total	175	135	65	375

(*MBA, Delhi Univ., 2009*)

Solution. The χ^2 test of significance would be appropriate in this case. Let us take the hypothesis that the salary and job satisfaction are independent. The expected frequencies are computed as follows :

$$E_{11} = \frac{80}{375}\times 175 = 37.33 \approx 37, \qquad E_{12} = \frac{80}{375}\times 135 = 28.8 \approx 29,$$

$$E_{21} = \frac{215}{375}\times 175 = 100.33 \approx 100, \qquad E_{22} = \frac{215}{375}\times 135 = 77.4 \approx 77.$$

The table of expected frequencies shall be :

37	29	14	80
100	77	38	215
38	29	13	80
175	135	65	375

O	E	$(O-E)^2$	$(O-E)^2/E$
30	37	49	1.324
100	100	0	0.000
45	38	49	1.289
30	29	1	0.034
85	77	64	0.831
20	29	81	2.793
20	14	36	2,571
30	38	64	1.684
15	13	4	0.308
			$\Sigma[(0-E)^2/E] = 10.834$

$$\chi^2 = \Sigma\frac{(O-E)^2}{E} = 10.834$$

For $v = 4, \chi^2_{0.05} = 9.488$

The calculated value of χ^2 is more than the table value. The hypothesis is rejected at 5% level and it is concluded that the salary and job satisfaction are not independent.

PROBLEMS

1-A : Answer the following questions, each question carries **one** mark:

(*i*) In contingency table, which of the following determines the degrees of freedom.

(*a*) $(r-1)(c-1)$, (*b*) $(r-1)(c+1)$, (*c*) $(r+1)(c+1)$, (*d*) $(r-1)(c+1)$ (*MBA, Madurai-Kamaraj Univ., 2003*)

(*ii*) Define χ^2 distribution.

(*iii*) What is χ^2 test of goodness of fit ? (*M. Com., M.K. Univ., 2002*)

(*iv*) Is χ^2 test a non-parametric test ?
(*v*) What are Yate's corrections ?
(*vi*) Explain how χ^2 test is used in the test of homogeneity.
(*vii*) How degrees of freedom are determined in testing of independence ?
(*viii*) What is additive property of χ^2 test ?
(*ix*) What precautions should be kept in mind while using the χ^2 test ?

1-B : Answer the following questions, each question carries **four** marks:
(*i*) Define χ^2 distribution. State the uses of χ^2 test.
(*ii*) Explain the characteristics of χ^2 test.
(*iii*) What is χ^2 distribution ? Describe the uses of χ^2 test ?
(*iv*) Explain χ^2 test of goodness of fit.
(*v*) Explain the significance of χ^2 distribution. *(MBA, Madras Univ., 2003)*

2. (*a*) What is χ^2 test ? What are its uses ? *(MBA, Sukhadia Univ., 2005)*
(*b*) What is χ^2 test ? Under what conditions is it applicable ? Point out its role in business decision-making.
3. What is chi-square test of independence ? What cautions are necessary while applying this test ?
4. Explain Yates's method of correction for small frequencies in contingency table.
5. What is χ^2 test of goodness of fit ? What cautions are necessary while applying this test ?
6. (*a*) What is χ^2 test ? Explain its important uses with the help of an example. *(MBA, UP Tech Univ., 2007)*
(*b*) What is χ^2 test ? Point out its applications. Under what conditions this test is applicable ?
7. Write short notes on :
(*i*) Yates's corrections for continuity, (*ii*) Degrees of freedom, (*iii*) Test of Goodness of fit.
8. Discuss the chi-square test of goodness of fit of theoretical distribution to an observed frequency distribution. State the conditions for the validity of chi-square test.
9. (*a*) Discuss the importance of χ^2 test. How is it used to test the association ?
(*b*) Describe the χ^2 test of significance and state the various uses to which it can be put.
(*c*) Show that the sum of two independent chi-square variables is also a chi-square variable. *(MBA, Anna Univ., 2003)*
10. 1000 families were selected at random in a city to test the belief that high income families usually send their children to public school and the low income families often send their children to government schools. The following results were obtained :

	School		
Income	*Public*	*Govt.*	*Total*
Low	370	430	800
High	130	70	200
Total	400	500	1000

Test, whether income and type of schooling are independent.
[$\chi^2 = 22.5$, No.]

11. A study is conducted of the volume of calls recieved on the switchboard of an insurance firm. A count is made of the number of incoming calls per minute for a sample of 120 minutes. The results of the study are shown below :

No. of calls per annum :	0	1	2	3	4	5
No. of minutes :	50	40	16	10	5	1

The statistician making the study believes that the incoming calls are distributed according to the Poisson distribution. Do you think his assumption is true ?

12. A survey was carried out in a state among the Doctors, belonging to the Kural Health Service cadre (500 doctors) and among the medical Education directorate cadre (300 teaching doctors). They were asked a question "would it be acceptable to you if the govt. proposes to hire all the doctors on fixed period contractual basis?" The doctors were to answer either as "Acceptable" or "Not acceptable". There was no third answer category "undecided". The following was the data compiled in a cross tabulated format :

Doctors	*Acceptable*	*Not acceptable*	*Total*
Rural Cadre	195	305	500
Teaching Cadre	140	160	300
Total	335	465	800

Apply χ^2 test and test the null hypothesis. *(MBA, HCA, Delhi Univ., 2008)*

13. Two factories using material from the same supplier and closely controlled to an agreed specification, produce output for a given period classified into three quality grades as follows :

		Quality grades (Output in tonnes)		
Factory	*A*	*B*	*C*	*Total*
X	42	13	33	88
Y	20	8	25	53
Total	62	21	58	141

Do this output figures show a significant difference at the 5% level ?

14. The theory predicts the production of beans in the four groups A, B, C and D should be 9 : 3 : 3 : 1. In an experiment among 1,600 beans, the numbers in the four groups were 882, 313, 287, and 118. Does the experimental result support the theory ?
[χ^2 = 4.72]

15. The following data relate to the sales in a time of trade depression, a certain article in great demand. Do the data suggest that the sales are significantly affected by depression ?

District where sales are :	*Districts*	
	Not hit by depression	*Hit by depression*
Satisfactory	140	60
Not satisfactory	40	60

[χ^2 = 25, yes]

(*M.Com., GND Univ., 2006*)

16. Based on information on 1,000 randomly selected fields about the tenancy status of the cultivators of these fields and use of fertilisers collected in an agroeconomic enquiry, the following classification was noted :

	Owned	*Rented*
Using fertilizer	416	184
Not using fertilizer	64	336

Would you conclude that owner-cultivators are more inclined towards the use of fertilizers?
[χ^2 = 273.5, yes]

(*M.Com., Osmania Univ., 2005*)

17. Two researchers adopted different sampling techniques while investigating the same group of students, to find the number of students falling in different intelligence levels. The results are as follows :

Researcher	*No. of students in each level*			
	Below average	*Average*	*Above average*	*Genius*
X	86	60	44	10
Y	40	33	25	2

Would you say that the sampling techniques adopted by the two researchers are significantly different ?
[χ^2 = 1.1971]

(*MBA, Delhi Univ., 2006*)

18. A random sample of size 20 from a normal population gives a sample mean of 42 and sample standard deviation of 6. Test the hypothesis that the population standard deviation is 9. Clearly state the alternative hypothesis you allow for and the level of significance adopted.
[χ^2 = 8.89]

19. A manufacturer of TV sets was trying to find out what variables influenced the purchase of a TV set. Level of income was suggested as possible variable influencing the purchase of TV set. A sample of 500 households was selected and the information obtained is classified as shown below :

	Have TV Set	*Do not have TV Set*
Low income group	0	250
Middle income group	50	100
High income group	80	20

Is there evidence from the above data of a relation between ownership of TV sets and level of income ?

20. A book has 700 pages. The number of pages with various numbers of misprints is recorded below. At the 5% significance level, are the misprints distributed according to Poisson law ?

No. of misprints :	0	1	2	3	4	5	Total
No. of pages with misprints :	616	70	10	2	1	1	700

[χ^2 = 12.81, No.]

(*M.Com., Delhi Univ., 2004*)

21. The following contingency table shows the classifications of 2,000 workers in a factory, according to the disciplinary action taken by the management and their promotional experience :

Disciplinary action	*Promotional Experience*	
	Promoted	*Not promoted*
Not-offenders	146	462
Offenders	54	1338

Test, whether the disciplinary action taken and promotional experience are independent.
[χ^2 = 1.227]

22. Four machines A, B, C and D are used to manufacture certain machine parts which are classified as first grade, second grade and third grade. The quality control engineer wants to test whether the quality of the product from the four machines is same. Data collected is as follows :

	Machines				
Grade	*A*	*B*	*C*	*D*	*Total*
First	620	750	400	530	2,300
Second	130	200	140	130	600
Third	50	50	60	40	200
Total	800	1,000	600	700	3,100

[$\chi^2 = 31.89$]

23. A certain drug is claimed to be effective in curing colds. In an experiment on 328 people with cold, half of them were given the drug and half of them were given sugar pills. The patients' reactions to the treatment are recorded in the following table :

	Helped	*Harmed*	*No effect*
Drug	104	20	40
Sugar pills	88	24	52

Test the hypothesis that the drug is no better than sugar pills for curing colds.

24. From the following data, find out whether there is any relationship between sex and preference of colour :

Colour	*Male*	*Female*	*Total*
Green	40	60	100
White	35	25	60
Yellow	25	15	40
Total	100	100	200

[$\chi^2 = 8.17$, yes] (*M.Com., Punjab Univ., 2005*)

25. In a survey of 200 boys, of whom 75 were intelligent, 40 had skilled fathers; while 85 of the unintelligent boys had unskilled fathers. Do these figures support the hypothesis that skilled fathers have intelligent boys ?

[$\chi^2 = 8.89$] (*MBA, Delhi Univ., 2003*)

26. The figures given below are (*i*) the theoretical frequencies of a distribution and (*ii*) the frequencies of the normal distribution having the same mean, standard deviation and the total frequency as in (*i*)

(*i*)	1	5	20	28	42	22	15	5	2
(*ii*)	1	6	18	25	40	25	18	6	1

Apply χ^2 test of goodness of fit.

27. Four different drugs have been developed for a certain disease. These drugs are used under three different environment (it is assumed that the environment might affect efficacy of drugs). The number of cases of recovery from the disease per 100 people who have taken the drugs is tabulated as follows :

	Drugs			
Environment	A_1	A_2	A_3	A_4
I	19	8	23	8
II	10	9	12	6
III	11	10	13	16

Test, whether the drugs differ in their efficacy to treat the disease, also whether there is any effect of environment on the efficacy of disease.

28. 2,000 digits were selected at random from a set of tables. The frequencies of the digits were given as below :

Digit :	0	1	2	3	4	5	6	7	8	9
Frequency :	180	200	190	230	210	160	250	220	210	150

Use the chi-square test to assess the correctness of the hypothesis that the digits were distributed in equal numbers in the tables from which these were chosen.

29. The result of a certain survey shows that out of 50 ordinary shops of small size, 35 are managed by men of which 17 are in cities, 12 shops in villages are run by women. Can it be inferred that shops run by women are relatively more in villages than in cities ? Use chi-square test.

[$\chi^2 = 3.572$]

30. For 2 × 2 contingency table :

	A	not A
B	a	b
not B	c	d

Prove that the chi-square test for independence of the two attributes A and B gives :

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

31. In a group of 100 persons, 56 were tall and 44 short. Of those who were tall 30 acted as leaders, 16 as followers and the rest were unclassifiable. Among those who were short, 14 acted as leaders, 22 as followers and the rest were unclassifiable. Tabulate the data and find out whether or not there is significant association between height and leadership.

32. In a study of market penetration, the marketing division of a company selected random samples of 200, 150 and 300 consumers from three cities and obtained the data given below. Do the data indicate that the extent of market penetration in the three cities is independent of the consumers knowledge of the product ?

City	*Never heard of product*	*Group heard but did not buy*	*Bought it at least once*	*Total*
1	36	55	109	200
2	45	56	49	150
3	54	78	168	300
Total	135	189	326	650

33. The number of machine malfunctions per shift at a factory is recorded for 180 shifts and the following data are obtained :

No. of malfunctions :	0	1	2	3	4	5	6	Total
No. of shifts :	82	42	31	12	8	3	2	180

What is a reasonable probability model for this type of data ?
Test, if this model describe the data adequately.

34. Study the effectiveness of three teaching methods (A), (B) and (C) from the following table :

Aptitude	*Age*			
	Young	*Middle*	*Old*	*Total*
Low	82(A)	87(B)	80(C)	249
Middle	92(B)	82(C)	81(A)	255
High	90(C)	83(A)	88(B)	261
Total	264	252	249	765

Do the teaching methods significantly differ in effectiveness on aptitude ?

35. An automobile company gives you the following information about age groups and the liking for particular model of car which it plans to introduce :

	Age Group			
	Below 25	*25–50*	*Above 50*	*Total*
Persons who liked the car	45	30	25	100
Disliked the car	55	20	25	100
Total	100	50	50	200

On the basis of above data, can it be concluded that the model appeal is independent of the age group ? *(MBA, DU, 2004)*
[$\chi^2 = 3$]

36. Boys and girls were sampled from a school and tested for their mathematical skills. Their classification into well skilled and poorly skilled categories was as below :

	Mathematical Skills		
	Good	*Poor*	*Total*
Boys	50	10	60
Girls	20	20	40
Total	70	30	100

Apply χ^2 test to find whether boys are better in mathematical skills to girls.
[$\chi^2 = 12.7$, yes]

37. L. Chandra, salesman for D. Paper Company, has 5 accounts to visit per day. It is suggested that the variable sales by Mr. Chandra may be described by the binomial distribution, with the probability of selling each account being 0.3. Given the following observed distribution of Chandra's number of sales per day, can we conclude that the distribution does in face follow the suggested distribution ? Use the .05 significance level.

No. of sales per day :	0	1	2	3	4	5
Frequency of no. of sales :	20	65	42	14	6	3

(*MFC, Delhi Univ., 2005*)

38. You are given the distribution of the number of defective units produced in a single shift in a factory over 100 shifts. Would you say that the defective units follow a Poisson distribution?

No. of defective units :	0	1	2	3	4	5	6
No. of shifts :	4	14	23	23	18	9	9

39. Price of a basket of goods and services showed the following trend in up-country and mid-town markets :

	Increasing	*Not increasing*
Mid-town	56	31
Up-country	18	6

Show of the trends in up-country prices and in mid-town prices has any significant association.

40. "A sample of 300 students of Under-Graduate and 300 students of Post-Graduate classes of a University were asked to give their opinion towards the autonomous colleges. 190 of the Under-Graduate and 210 of the Post-Graduate students favoured the autonomous status."

Present the above fact in the form of a frequency table and test at 5% level, that opinions of Under-Graduate and Post-Graduate students on autonomous status of colleges are independent.

41. Calculate the expected frequencies for the following data presuming the two attributes, *viz.*, condition of home and condition of child as independent :

	Condition of Home	
Condition of Child	*Clean*	*Dirty*
Clean	70	50
Fairly clean	80	20
Dirty	35	45

Use chi-square test at 5% level to state whether the two attributes are independent.
[$\chi^2 = 24.64$] (*M.Com., Madurai-Kamaraj Univ., 2005*)

42. 1000 students at college level were graded according to their IQ and economic conditions of their home. Use χ^2 test to find out, whether there is any association between economic condition at home and I.Q.

	I.Q.		
Economic Condition	*High*	*Low*	*Total*
Rich	460	140	600
Poor	240	160	400
Total	700	300	1000

[$\chi^2 = 31.75$] (*MBA, Osmania, Univ., MBA, Kumaun Univ., 2008*)

43. The following table gives the number of car accidents that occurred during the various days of the week. Find, whether the accidents are uniformly distributed over the week.

Day :	Sun.	Mon.	Tues.	Wed.	Thurs.	Fri.	Sat.
No. of accidents :	14	16	8	12	11	9	14

[$\chi^2 = 4.165$] (*M.Com., M.D. Univ., 2005; MBA, Delhi Univ., 2005*)

44. What are the assumptions in carrying out test of independence of attributes through chi-square? Set up an appropriate hypothesis for the data given below and draw your conclusions through some suitable test of significance method.

	Level of Intelligence		
Family Status	*Dull*	*Average*	*Brilliant*
Lower Middle	20	35	25
Middle	40	70	30
Upper Middle	40	30	30

45. (*a*) A marketing agency gives following information about the age groups and their liking for a particular model which the company plans to introduce :

	Age group			
	Below 20	*20–39*	*40–59*	*Total*
Liked	125	420	60	605
Disliked	75	220	100	395
Total	200	640	160	1000

On the basis of the above data, can it be concluded that the model appeal is independent of the age group.
[$\chi^2 = 42.79$] (*MBA, Kumaun Univ., 2002*)

(*b*) A die was thrown 9,000 times and of these 3,220 yielded a 3 or 4. Is this consistent with the hypothesis that the die was unbiased ? (*MBA, Bharathidasan Univ., 2001*)

46. A random sample of 400 persons was selected from each of three age groups and each person was asked to specify which of three types of TV programmes be preferred. The results are shown in the following table :

	Table Programme			
Age Group	*A*	*B*	*C*	*Total*
Under 30	120	30	50	200
30–44	10	75	15	100
45 and above	10	30	60	100
Total	140	135	125	400

Test the hypothesis that the populations are homogeneous with respect to the types of television programme they prefer. (*MBA, Guru Jambheshwar Univ., 2007*)

47. The following information is obtained concerning an investigation of 50 ordinary shops of small size :

	No. of Shops		
	in towns	*in villages*	*Total*
Run by Men	17	18	35
Run by Women	3	12	15
Total	20	30	50

Can it be inferred that shops run by women are relatively more in villages than in towns ? Use chi-square test.
[$\chi^2 = 0.121$] (*MBA, Madurai Kamaraj Univ., 2006*)

48. The number of analysis sum by three operators during different shifts is given below. Test the hypothesis that the performance of the operators is independent of shifts

	Operator		
	1	*2*	*3*
I	97	58	32
II	78	46	39

49. Fit a Poisson distribution to the following data and test for goodness of fit.

X :	0	1	2	3	4	5	6
f :	275	72	30	7	5	2	1

(*MBA, Anna Univ., 2007*)

*****