Introduction

In the preceding lesson some tests of significance were discussed which is based on the assumption that the samples were drawn from normally distributed population. These tests are known as parametric tests as the testing procedure involves the assumption about the type of population of parameters. There are, however, many situations where it is not possible to assume a particular type of population distribution from which the samples are drawn. This leads to development of alternative techniques known as non-parametric or distribution free methods. Chi-square (χ^2) (pronounced as Ki-square) test is one of the important non-parametric test and one of the most commonly used test of significance. The chi-square test dates back to 1900, when Prof. Karl Pearson used it for frequency data classified into k-mutually exclusive categories. It is also a frequently used test in genetics, where one tests whether the observed frequencies in different crosses agree with the expected frequencies or not. The chi-square test is applicable to test the hypothesis of the variance of a normal population, goodness of fit of the theoretical distribution to the observed frequency distribution, in a one way classification having k-categories. It is also applied for the test of independence of attributes, when the frequencies are presented in a two-way classification called the contingency table. In this lesson we give chi-square test of various hypotheses.

19.2 Chi-Square Distribution

If X is a normal variate with mean and standard deviation σ viz., $X \sim N(\mu, \sigma^2)$ then $Z = \frac{X-\mu}{\sigma}$ is a standard normal variate .The square of a standard normal variate i.e., $\left(\frac{X-\mu}{\sigma}\right)^2$ is known as Chi-square (χ^2) variate with one degree of freedom (d.f.). If $X_1, X_2, ---, X_n$ are n independent random variate following normal distribution with means $\mu_1, \mu_2, ---, \mu_n$ and standard deviations $\sigma_1, \sigma_2, ---, \sigma_n$ respectively then the variate

$$\chi^2 = \left(\frac{X_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{X_2-\mu_2}{\sigma_2}\right)^2 + ---+ \left(\frac{X_n-\mu_n}{\sigma_n}\right)^2 = \sum_{i=1}^{n}\left(\frac{X_i-\mu_i}{\sigma_i}\right)^2$$

which is sum of the squares of n independent standard normal variates , follows Chi-square distribution with n d.f.

19.3 Applications of the χ^2–distribution

Chi-square distribution has a number of applications which are enumerated below:

a) To test if the population has a specified value of the variance σ^2.

b) Chi-square test of goodness of fit.

c) Chi-square test for independence of attributes.

19.3.1 Chi-Square Test for Population Variance:

Suppose on the basis of previous knowledge, we have a preconceived value of population variance σ_0^2.Suppose we draw a random sample of size n from this population. On the basis of n sample observations (X_1, X_2,.....,X_n), the population variance value σ_0^2 of the population variance σ^2 is to either be substantiated or refuted with the help of a statistical test. In this case we use Chi-square test. For this the null hypothesis is taken as

$$H_0: \sigma^2 = \sigma_0^2$$

$$H_1: \sigma^2 \neq \sigma_0^2$$

and is tested by the statistic

$$\chi^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma_0^2} = \frac{n\,s^2}{\sigma_0^2} \quad i = 1,2,3, \ldots., n$$

Follows a χ^2 distribution with (n-1) degree of freedom. Where

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

is the variance of the sample? If calculated value of chi-square is more than or equal to tabulated chi-square value at α% level of significance then H_0 is rejected at α level of significance otherwise if calculated χ^2 is less than tabulated χ^2 at α level of significance then H_0 is not rejected at α level of significance.

Example 1: An owner of a big firm agrees to purchase the product of a factory, if the produced items do not have variance of more than 0.5mm^2 in their length. To make sure of the specifications, the buyer selects a sample of 18 items from his lot. The length of each item was measured in mm which is given as under:

18.57	18.10	18.61	18.32	18.33	18.46	18.37	18.64	18.58
18.12	18.34	18.57	18.22	18.63	18.43	18.34	18.43	18.63

Test whether the sample has been drawn from the population having specified variance not more than value of 0.5

Solution : On the basis of the sample data, the hypothesis

$$H_0; \sigma^2 = 0.5 \text{ aganist } H_1: \sigma^2 > 0.5$$

can be tested by the statistic

$$\chi^2 = \frac{\sum_i (X_i - \overline{X})^2}{\sigma_0^2} \quad i = 1,2 \dots ,18$$

Calculate, $\sum_i (X_i - \overline{X})^2 = \sum_i X_i^2 - \frac{(\sum_i X_i)^2}{n}$

For the given data, $\sum_i X_i^2 = 6112.64; \sum_i X_i = 331.69$

$$\therefore \sum_i (X_i - \overline{X})^2 = 6112.64 - \frac{(331.69)^2}{18} = 6112.640 - 6112.125 = 0.515$$

Thus, χ^2 = 0.515/0.5 = 1.03. Tabulated value of χ^2 at 5% level of significance and 17 d.f. viz., $\chi^2_{0.0517}$ is 27.587. Since the calculated value of χ^2 (1.03) is less than 27.857, we donot reject the null hypothesis at α = 0.05. i.e., $\sigma^2 = 0.5$, It means that the buyer should purchase the lot having specified variance not more than value of 0.5 mm^2 length **.**

19.3.2 Chi- square test of goodness of fit

A very powerful test for testing the significance of the discrepancy between theory and experiment was given by Prof. Karl Pearson in 1900 and is known as chi-square test of goodness of fit. This test is used for testing the significance of discrepancy between experimental values and the theoretical values obtained under some theory or hypothesis. It enables us to find if the deviation of the experiment from theory is just by chance or is it really due to the inadequacy of the theory to fit the observed data.

Under the null hypothesis that there is no significant difference between the observed (experimental) values and the theoretical values i.e., there is good compatibility between theory and experiment , Karl Pearson proved that the statistic

$$\chi^2 = \frac{(O_1 - E_1)^2}{E_i} + \frac{(O_2 - E_2)^2}{E_2} + --- + \frac{(O_n - E_n)^2}{E_n} = \sum_{i=1}^{n} \left[\frac{(O_i - E_i)^2}{E_i} \right]$$

Follows a χ^2- distribution with (n-1) degree of freedom where O_i (i=1,2,....,n) is a set of observed (experimental) frequencies and E_i (i=1,2,..,n) be the corresponding set of expected (theoretical) frequencies obtained under some theory or hypothesis .

If calculated value of χ^2 is less than the corresponding tabulated value at (n-1) d.f. then it is said to be non-significant at the required level of significance. This implies that the discrepancy between experimental (observed) values and the theoretical (expected) values obtained under some theory or hypothesis may be attributed to chance. In other words, data do not provide us any evidence against the null hypothesis and we may conclude that there is good correspondence (fit) between theory and experiment. On the other hand if calculated value of χ^2 is greater than the corresponding tabulated value at (n-1) d.f. then it is said to be significant at the required level of significance. This implies that the discrepancy between experimental (observed) values and the theoretical (expected) values obtained under some theory or hypothesis cannot be attributed to chance and we reject the null hypothesis. In other words we conclude that the experiment does not support the theory.

Remarks:

a) The observed and expected frequencies are subjected to a linear constraint $\sum_{i=1} O_i = \sum_{i=1} E_i = N$, where N is the total frequency since it does not involve squares and higher powers of the frequencies. $\sum_{i=1}(O_i - E_i) = \sum_{i=1} O_i - \sum_{i=1} E_i = N - N = 0$ i.e. the sum of deviations of the observed and expected frequencies is always zero.

b) Sometimes the following formula is useful for computation of χ^2

$$\chi^2 = \sum_{i=1}^{n}\left[\frac{(O_i - E_i)^2}{E_i}\right] = \sum_{i=1}^{n}\frac{O_i^2 + E_i^2 - 2O_iE_i}{E_i}\chi^2 = \sum_{i=1}^{n}\frac{O_i^2}{E_i} + \sum_{i=1}^{n}E_i - 2\sum_{i=1}^{n}O_i$$

$$\because \sum_{i=1}^{n}O_i = \sum_{i=1}^{n}E_i = N \qquad \chi^2 = \sum_{i=1}^{n}\frac{O_i^2}{E_i} - N$$

where N is the total frequency.

c) χ^2-test depends only on the observed and expected frequencies and on degree of freedom (n-1) .It does not make any assumption regarding the parent population from which the observations care taken .Since χ^2 does not involve any population parameter it is known as statistic and the test is known as Non-parametric test or Distribution Free test.

19.3.2.1 Degrees of freedom

The number of independent variates which makes up the statistic (e.g., χ^2) is known as the degrees of freedom (d.f.). The no. of degrees of freedom in general is the total no. of observations minus the no. of independent linear constraints imposed on the observations. e.g., if k is the no. of independent constraints imposed on a set of data of n observations then d.f. = (n−k).Thus in a set of n observations usually, the degrees of freedom for χ^2 are (n 1), one d.f. being lost because of the linear constraint. $\sum_{i=1} O_i = \sum_{i=1} E_i = N,$. If r independent linear constraints are imposed on the cell frequencies, then the d.f. are reduced by r.

In addition if any of the population parameters (s) is (are) calculated from the given data and used for computing the expected frequencies then in applying χ^2 test of goodness of fit, we have to subtract one d.f. for each parameter estimated.

19.3.2.2 Conditions for the validity of χ^2 test

Following are the conditions which should be satisfied before χ^2 test can be applied.

(i) N the total number of frequencies must be large, greater than50.

(ii) The sample observations should be independent.

(iii) No theoretical cell frequency should be small. Five should be regarded as the very minimum and 10 is better. The chi-square distribution is essentially a continuous distribution but it cannot maintain its character of continuity if the cell frequency is less than 5). If any theoretical cell frequency is less than 5, then for the application of χ^2 test, it is pooled with the preceding or succeeding frequency so that the pooled frequency is more than 5 and finally adjusts for the d.f. lost in the pooling.

(iv) The constraints on the cell frequencies, if any, should be linear. Constraints which involve linear especially in the cell frequencies are called linear constraints such as $\sum_i O_i = \sum_i E_i = N$.

The above procedure is explained through following examples:

Example 2 : The following table gives the number of coli forms per ml in thousand bottles of sterilized milk:

No of coli forms (X_i)	0	1	2	3	4	5	6	7	8	9	10
No. of bottles (f_i)	2	8	46	116	211	243	208	119	40	7	0

Fit a binomial distribution to the above data and test the goodness of fit.

Solution

The fitting of this problem is already explained in example 10.6 of lesson 10 (module 3) .In the usual notations we have: n=10, N=1000,

$$\sum f_i X_i = 4971,\ \bar{X} = 4.971, p = \frac{4.971}{10} = 0.4971, q = 0.5029, \frac{p}{q} = \frac{0.4971}{0.5029} = 0.9985$$

putting r=0,1,2,3,---,10 in $f(r) = N \text{ x } {}^{n}C_{r}\, P^{r} q^{n-r}$ we get the expected frequency as given in the following table:

Table 19.1

No. of coliforms	**No. of bottles (f$_i$) (O$_i$)**		**Expected Frequency E $_i$**		**(O$_i$ - E$_i$)**	**(O$_i$ - E$_i$)^2**	$\frac{(O_i - Ei)^2}{E_i}$
0	2	10*	1	11*	-1	1	0.0909
1	8		10				
2	46		46		0	0	0.0000
3	116		120		-4	16	0.1333
4	211		207		4	16	0.0773
5	243		246		-3	9	0.0366
6	208		203		5	25	0.1232
7	119		115		4	16	0.1391
8	40		42		-2	4	0.0952
9	7	7*	9	10*	-3	9	0.9000
10	0		1				
Total	1000		1000				1.5956

* In the above table since some of expected cell frequencies being less than 5 , therefore, they have been merged with either preceding or succeeding frequencies. Accordingly the corresponding observed frequencies have also been merged. Thereafter the value of χ^2 is calculated as follows

$$\chi^2 = \sum_{i=1}^{n} \left[\frac{(O_i - E_i)^2}{E_i}\right] = 1.5956$$

Required Degrees of freedom : 11-1-1-2=7 (n=11,one d.f. is lost due to linear constraint $\sum_{i=1} O_i = \sum_{i=1} E_i = N,$; one d.f. is lost because the parameter p of binomial distribution is estimated from the given data;2 d.f. are lost due to pooling first and last two frequencies). Tabulated value of χ^2 for 7 d.f. and at 5% level of significance is 14.067.Since calculated value of χ^2 is less than tabulated value it is not significant. Thus we conclude that

Binomial distribution is a good fit to the given data or Binomial distribution fits well to the given data.

Example 3: The following table gives the number of lactations completed by 1000 cows of Tharparker breed:

No of lactations (X_i)	0	1	2	3	4	5	6	7	8	9	10
No. of females (f_i)	300	205	155	126	90	47	35	18	13	8	3

Fit a Poisson distribution to the above data and test its goodness of fit.

Solution :

The fitting of this problem is already explained in example 11.6 of lesson 11 (module 3) . In the usual notations we have : N=1000, $\Sigma f_i X_i = 2030, \bar{X} = 2.03 = m$

putting r=0,1,2,3,---,10 in

$$f(r) = nx\frac{e^{-m} m^r}{r\,!}$$

we get the expected frequency as given in the following table

Table 19.2

No. of lactation (X_i)	No. of Females (f_i) (Oi)	Expected Frequency (E_i)	(O_i - E_i)	$(O_i - E_i)^2$	$\frac{(O_i - Ei)^2}{E_i}$
0	300	131	169	28447.71	216.6034
1	205	267	-62	3795.93	14.2377
2	155	271	-116	13365.74	49.3911
3	126	183	-57	3261.89	17.81356
4	90	93	-3	8.58	0.092368
5	47	38	9	85.94	2.277851
6	35	13	59	3481.00	193.3889
7	18	4			

8	13	77*	1	18*			
9	8		0				
10	3		0				
Total	1000		1000				493.8049

* In the above table since some of expected cell frequencies being less than 5, therefore, they have been merged with either preceding or succeeding frequencies. Accordingly the corresponding observed frequencies have also been merged, Thereafter the value of χ^2 is calculated as follows

$$\chi^2 = \sum_{i=1}^{n} \left[\frac{(O_i - E_i)^2}{E_i}\right] = 611.4467$$

Required Degrees of freedom: 11-1-1-4=5 (n=11,one d.f. is lost due to linear constraint $\sum_{i=1} O_i = \sum_{i=1} E_i = N$;one d.f. is lost because the parameter m of Poisson distribution is estimated from the given data; 4 d.f. are lost due to pooling last five frequencies). Tabulated value of χ^2 for 5 d.f. and at 5% level of significance is 11. 07. Since calculated value of χ^2 is more than tabulated value it is significant. Thus, we conclude that Poisson distribution is not a good fit to the given data or Poisson distribution does not fit well to the given data.

19.3.3 Independence of attributes

Let us consider two attributes A and B, A divided into r classes $A_1,A_2,..,A_r$ and B divided into s classes $B_1,B_2.B_s$. (Such a classification in which attributes are divided into more than two classes is known as manifold classification). The various cell frequencies can be expressed in table known as rs (manifold) contingency table where (A_i) is the number of persons possessing the attributes A_i (i=1,2,,r), (B_j) is the number of persons possessing the attributes B_j, (j=1,2,.,s) and $(A_i\,B_j)$ is the number of persons possessing both the attributes A_i and B_j, (i=1,2,,r; j=1,2,,s). Also

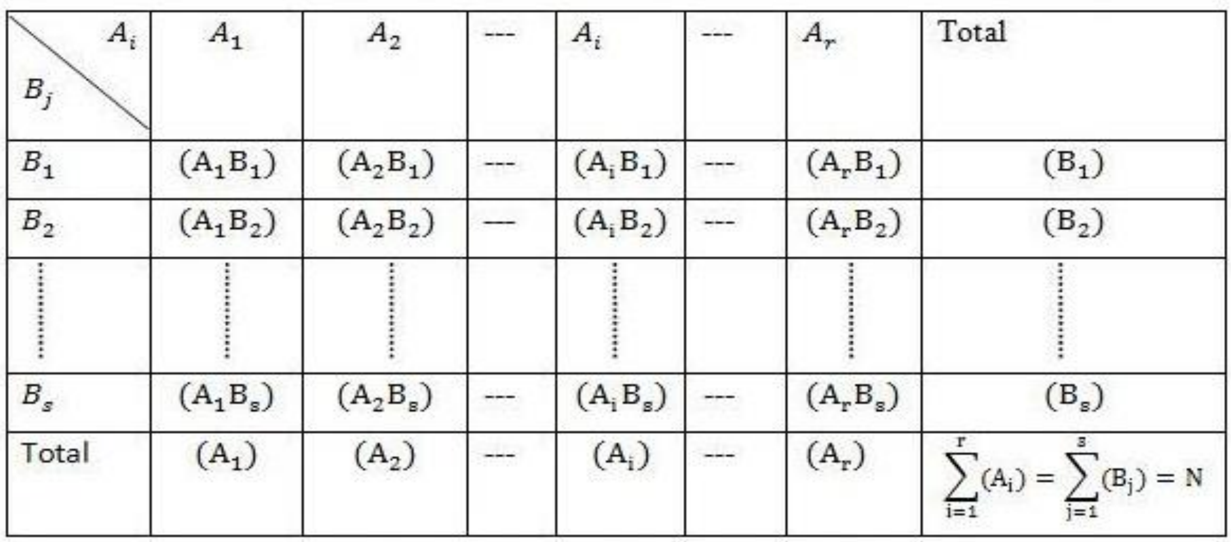
$$\sum_{i=1}^{r}(A_i) = \sum_{j=1}^{s}(B_j) = \text{N is the total frequency}$$

The problem is to test if two attributes A and B under consideration are independent or not. Under the null hypothesis that the attributes are independent, the theoretical cell frequencies are calculated as

$$E(A_1B_1) = \frac{(A_1)(B_1)}{N} \ldots\ldots\ldots E(A_iB_j) = \frac{(A_i)(B_j)}{N} \qquad \begin{matrix} i = 1,2,\ldots\ldots,r \\ j = 1,2,\ldots\ldots,s \end{matrix}$$

Hence expected frequency for any of the cell frequencies can be obtained by multiplying the row totals and column totals in which the frequency occurs and then dividing the product by the total frequency N.

Table 19.3 rxs Manifold Contingency Table



A_i B_j	A_1	A_2	---	A_i	---	A_r	Total
B_1	(A_1B_1)	(A_2B_1)	---	(A_iB_1)	---	(A_rB_1)	(B_1)
B_2	(A_1B_2)	(A_2B_2)	---	(A_iB_2)	---	(A_rB_2)	(B_2)
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
B_s	(A_1B_s)	(A_2B_s)	---	(A_iB_s)	---	(A_rB_s)	(B_s)
Total	(A_1)	(A_2)	---	(A_i)	---	(A_r)	$\sum_{i=1}^{r}(A_i) = \sum_{j=1}^{s}(B_j) = N$

Now for rxs observed frequencies (A_i B_j) and the corresponding expected frequencies E(A_i B_j) . Applying χ^2 test of goodness of fit, the chi-square statistic is given by

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = \sum_{i=1}^{r}\sum_{j=1}^{s}\left[\frac{\{A_iB_j - E(A_iB_j)\}^2}{E(A_iB_j)}\right] = \sum_{i=1}^{r}\sum_{j=1}^{s}\left[\frac{\left\{A_iB_j - \frac{(A_i)(B_j)}{N}\right\}^2}{\frac{(A_i)(B_j)}{N}}\right]$$

Follows χ^2 distribution with (r =1) (s = 1) d.f. Comparing this calculated value with the tabulated value for (r-1)(s-1) d.f. and at certain level of significance , we reject or accept the null hypothesis of independence of attributes at that level of significance.

19.3.3.1 Degrees of Freedom for rxs contingency table

In (r=s) contingency table in calculation of the expected frequencies, the row totals, the column totals and the grand total remain fixed. Further $\sum A_i = \sum B_j = N$. Further since the total number of cell frequencies is (rs), the required number of degrees of freedom is d.f. = rs (r + s 1) =(r1)(s1)

contingency table

Under the null hypothesis of independence of attributes, the value of χ^2 for the 2x2 contingency table

			Total
	a	b	a+b
	c	d	c+d
Total	a+c	b+d	N=a+b+c+d

is given by

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

which follows Chi-square distribution with (2-1)(2-1)= 1 degree of freedom .

19.3.3.3 Yates correction for continuity for 22 contingency table

In 22 contingency table, the no. of d.f. is (2−1) (2−1) =1. If any one of theoretical cell frequency is less than 5, then the using of pooling method for χ^2 results in χ^2 with 0 d.f. (since 1 d.f. is lost in pooling) which is meaningless. In this case we apply Yates correction for continuity which says that add 1/2 to cell frequency which is less than 5 and then adjust for remaining cell frequencies accordingly. The modified formula for χ^2 is as follows:

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

If N is large then Yates correction will make a very little difference and this can be applied in 2x2 contingency table only.

19.3.3.4 2x r Contingency table:

Under the hypothesis of independence of attributes, the value of χ^2 for 2xr contingency table:

	2xr Contingency Table					Total
	a_1	a_2	a_3	----	a_r	$n_1 = \sum_{i=1}^{r} a_i$
	b_1	b_2	b_3	----	b_r	$n_2 = \sum_{i=1}^{r} b_i$
Total	m_1	m_2	m_3	----	m_r	$N = n_1 + n_2 \; or \sum_{i=1}^{r} m_i$

can be computed from Brandt and Snedecor formula :

$$\chi^2 = \frac{N^2}{n_1 x n_2}\left[\frac{a_1^2}{m_1} + \frac{a_2^2}{m_2} + \frac{a_3^2}{m_3} + - - + \frac{a_r^2}{m_r} - \frac{n_1^2}{N}\right] = \frac{N^2}{n_1 x n_2}\left[\sum_{i=1}^{r} \frac{a_i^2}{m_i} - \frac{n_1^2}{N}\right]$$

Other form of this formula is

$$= \frac{1}{pq}\left[\sum a_i p_i - n_1 p\right] \text{ where } p = \frac{n_1}{N}, q = \frac{n_2}{N} \text{ and } p_i = \frac{a_i}{m_i}$$

which is χ^2 distribution with (2-1)(k-1)=k-1 d.f. The above procedure is explained through following examples:

Example 4: A milk producers union wishes to test whether the preference pattern of consumers for its product is dependent on income levels. A random sample of 500 individuals gives the following data:

Table 19.4

Income	**Product Preferred**			
	Product A	**Product B**	**Product C**	**Total**
Low	170	30	80	280
Medium	50	25	60	135
High	20	10	55	85
Total	240	65	195	500

Can you conclude that the preference patterns are independent of income levels?

Solution: Let us take the hypothesis that preference patterns are independent of income levels. On the basis of this hypothesis, the expected frequencies corresponding to different rows and columns shall be:

$$E_{11} = \frac{240 \times 280}{500} = 134.4 \qquad E_{12} = \frac{65 \times 280}{500} = 36.4$$

$$E_{21} = \frac{240 \times 135}{500} = 64.8 \qquad E_{22} = \frac{65 \times 135}{500} = 17.55$$

and so on. The expected frequencies would be as follows:

Table 19.5

134.40	36.40	109.20	280
64.80	17.55	52.65	135
40.80	11.05	33.15	85
240	65.00	195.00	500

Applying χ^2-test:

Table 19.6

O_i	E_i	$(O_i\text{-}E_i)^2$	$(O_i\text{-}E_i)^2/E_i$
170	134.60	1267.36	9.430
50	64.80	219.04	3.380
20	40.80	432.64	10.604
30	36.40	40.96	1.125
25	17.55	55.50	3.162
10	11.05	1.10	0.099
80	109.20	852.64	7.808
60	52.65	54.02	1.026
55	33.15	477.42	14.402
			$\sum (O_i - E_i)^2/E_i = 51.036$

$$\therefore \chi^2 = \sum \left[\frac{(O_i - E_i)^2}{E_i} \right] = 51.036$$

Degree of freedom = v =(r - 1) (c - 1) = (3 - 1) (3 - 1) = 4

The tabulated value of χ^2 for 4 d.f. at 5% level of significance i.e., $\chi^2_{0.05,4}$=9.488

Since the calculated value of χ^2 is greater than the table value, therefore, we reject the null hypothesis and hence conclude that preference patterns are not independent of income levels.

Example 5: In an experiment of cattle from tuberculosis, the following were obtained:

	Affected	Not Affected	Total
Inoculated	4	20	24
Not inoculated	6	50	56
Total	10	70	80

Calculate χ^2 and discuss the effect of vaccine in controlling susceptibility to tuberculosis.

Solution : N=80

H_0: The vaccine is not effective in controlling susceptibility to tuberculosis.

H_1: The vaccine is effective in controlling susceptibility to tuberculosis.

Since one of the observed frequency is less than 5. Applying Yates correction, we increase the value of that observed frequency by 0.5 and adjust other frequencies. The adjusted observed frequencies after Yates correction will be as follows:

	Affected	Not Affected	Total
Inoculated	4+0.5=4.5	20-0.5=19.5	24
Not inoculated	6-0.5=5.5	50+0.5=50.5	56
Total	10	70	80

$$\chi^2 = \frac{N\left[|ad - bc| - \frac{N}{2}\right]^2}{(a+c)(b+d)(a+b)(c+d)} = \frac{80\left[|4.5x50.5 - 19.5x5.5| - \frac{80}{2}\right]^2}{(10)(70)(24)(56)} = \frac{512000}{940800} = 0.5442$$

The tabulated value of χ^2 for 1 d.f. at 5% level of significance i.e., $\chi^2_{0.051}$ = 3.84. Since the calculated value of χ^2 is less than the tabulated value, we accept the null hypothesis and hence conclude that the vaccine is not effective in controlling susceptibility to tuberculosis.

Example 6: A milk product factory is bringing out a new product. In order to map out its advertising campaign, it wants to determine whether product appeals equally to all age groups. The following table gives the number of persons who liked or disliked the product in different age groups.

Table 19.7 No. of persons

Preference	Age group (in years)				Total
	< 20	20-30	30-40	> 40	
Liked	75	70	60	55	260
Disliked	25	30	40	45	140

Can it be reasonably concluded that the new product appeals equally to all age groups?

Solution:

N = 400

H_0: The new product appeals equally to all age groups.

H_1: The new product does not equally appeal to all the age groups.

Table 19.8

Preference	**Age group (in years)**				Total
	< 20	20-30	30-40	> 40	
Liked	a_1=75	a_2=70	a_3=60	a_4=55	n_1=260
Disliked	b_1=25	b_2=30	b_3=40	b_4=45	n_2=140
	m_1=100	m_2=100	m_3=100	m_4=100	N=400

$$X^2 = \frac{400^2}{260 \times 140}\left[\frac{75^2}{100} + \frac{70^2}{100} + \frac{60^2}{100} + \frac{55^2}{100} + \frac{260^2}{400}\right]$$

$= \quad 4.3956(2.5)=10.989$

The tabulated value of χ^2 for 3 d.f. at 5% level of significance i.e., $\chi^2_{0.053}$ = 7.815. Since the calculated value of χ^2 is greater than the tabulated value, we reject the null hypothesis and hence conclude that the new product did not equally appeal to all the age groups.