CHI-SQUARE DISTRIBUTION

11.1. Chi-square distribution

Let $X_1, X_2, \ldots, X_n$ be n independent normal variates each is distributed normally with mean zero and S.D. unity then $X^2_1+X^2_2+\ldots+X^2_n=\Sigma X^2_i$ is distributed as Chi-square (χ^2) with n degrees of freedom (d.f.) where n is large. The probability that χ^2 lies in the interval $d\chi_2$ is given by

$$f(\chi^2) = \frac{1}{2^{n/2}\sqrt{n/2}} e^{-\chi^2/2} (\chi^2)^{n/2-1} d(\chi^2), \quad 0<\chi^2<\infty \qquad \ldots (11.1)$$

The expression (11.1) is called the chi-square distribution with n.d.f. This was first given by Helmert in 1875, later it was independently given by Karl Pearson in 1900 along with chi-square test of goodness of fit. The chi-square probability curves for d.f., n. = 1, 2, ..., 6 are given in Fig. 11.1

The areas under the chi-square curve for different fixed value χ^2 on the X-axis are tabulated.

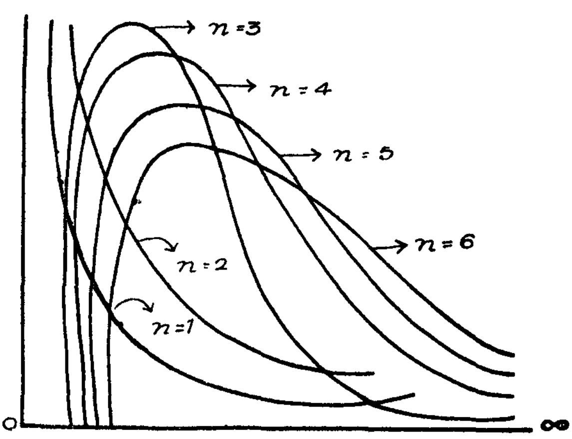

Fig. 11.1. Chi-square distribution.

11.2. Properties

1. For $n>2$, the χ^2—distribution has mode at $(n-2)$. For $n=2$, the distribution is J-shaped with maximum ordinate at zero, while for $0<n<2$ the distribution is J-shaped and has infinite ordinate at the origin.

2. Chi-square distribution is having additive property.

If $\chi^2_1, \chi^2_2, \ldots, \chi^2_k$ are k independent chi-square variates with $n_1, n_2, \ldots, n_k$ d.f. respectively, then $\chi^2_1+\chi^2_2+\ldots+\chi^2_k$ is a χ^2 — variate with $n_1+n_2+\ldots+n_k$ d.f.

3. The different central moments are $\mu_2=2n$, $\mu_3=8n$, $\mu_4=48n+12n^2$.

4. The coefficients of skewness and kurtosis are given as $\gamma_1=(8/n)^{\frac{1}{2}}$, $\gamma_2=\beta_2-3-12/n$

5. As n tends to infinity, the chi-square distribution tends to normality. The distribution function is an incomplete Gamma-function. Fisher showed that when n is large $\sqrt{2\chi^2}$ is approximately normally distributed with mean $\sqrt{2n-1}$ and variance unity. Wilson and Hilferty showed that $(\chi^2/n)^{1/3}$ is approximately normally distributed with mean, $1-2/9n$ and variance, $2/9n$. The latter one is more accurate approximation.

7. The ratio of two chi-square variates χ^2_1 and χ^2_2 with n_1 and n_2 d.f. respectively is a β-variate, i.e., $\beta(n_1/2, n_2/2)$.

11.3. Chi-Square Test of Goodness of Fit

Chi-square test is used to know whether the given sampling distribution is in agreement with the known theoretical distribution or whether the given objects are segregating in a theoretical ratio or whether the two attributes are independent in a contingency table, etc.

11.3.1. Measurement data: The data obtained by actual measurement is called measurement data. For example, height, weight, age, income, area, etc.

11.3.2. Enumeration data: The data obtained by enumeration or counting is called enumeration data. For example, number of blue flowers, number of intelligent boys, number of curled leaves, etc.

11.3.3. χ^2-test is used for enumeration data which generally relate to discrete variable whereas t-test and standard normal deviate tests are used for measurement data which generally relate to continuous variable.

The expression for χ^2-test for goodness of fit is

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} = \Sigma \frac{O_i^2}{\Sigma_i} - N \qquad \text{... (11.2)}$$

where O_i = observed frequency of i-th cell, E_i = expected frequency of i-th cell, and k, the number of cells.

Here the term 'cell' is used for class interval in the case of frequency districution and compartment (or category) in the case of enumeration data.

CONCLUSION : If χ^2 (calculated) $\leq \chi^2$ (tabulated) then the hypothesis that the observed frequencies are in agreement with the expected frequencies is accepted. Otherwise, the hypothesis is rejected.

The expression (11.2) has been obtained from the expression $\sum_{ij} \alpha^{ij} \phi_i \phi_j$ where f_i is the likelihood function of the multinominal probability function, α^{ij} is the inverse of the covariance matrix.

11.4 2 × 2 Contingency Table

When the individuals (objects) are classified into two categories with respect to each of the two attributes then the table showing frequencies distributed over 2 × 2 classes is called 2 × 2 contingency table.

EXAMPLE : Suppose the individuals are classified according to two attributes colour (B) and intelligence (A). The distribution of frequencies over cells is shown in Table 11.1

TABLE 11.1

B \ A	A_1	A_2	
B_1	a	b	R_1
B_2	c	d	R_2
	C_1	C_2	N

where a, b, c and d are frequencies of the different cells. R_1, R_2, C_1 and C_2 are the respective marginal totals, and N is the grand total.

Null Hypothesis: The two attributes are independent.

$$\chi^2 = \Sigma \frac{(O_i - E_i)^2}{E_i}$$

If the colour is not dependent on intelligence, then

$$\frac{a}{C_1} = \frac{b}{C_2} = \frac{R_1}{N} \quad \text{or}$$

$$\frac{c}{C_1} = \frac{d}{C_2} = \frac{R_2}{N}$$

Similarly if the intelligence is nothing to do with colour, then $a/R_1 = c/R_2 = C_1/N$
and $b/R_1 = d/R_2 = C_2/N$

The expected frequencies are obtained as follows

$$E(a) = \frac{C_1 R_1}{N}$$

$$E(b) = \frac{C_2R_1}{N} \qquad \text{or } E(b) = R_1 - C_1R_1/N$$

$$E(c) = \frac{C_1R_2}{N} \qquad \text{or } E(c) = C_1 - C_1R_1/N$$

$$E(d) = \frac{C_2R_2}{N} \qquad \text{or } E(d) = C_2 - E(b)$$

where, E(a), E(b), etc., are the expected frequencies of the cells containing observed frequencies 'a', 'b', etc. Since the marginal totals are fixed we need to compute only one expected frequency in 2×2 contingency table and the rest are obtained by subtraction as shown in R.H.S. above.

The degrees of freedom for testing the χ^2 value is derived as total number of items—No. of restrictions from rows—No. of restrictions from columns—No. of restrictions from the grand total, i.e., $4-(2-1)-(2-1)-1=1$.

Substituting the expected frequencies in (11.3), we have

$$\chi^2 = \frac{\left(a - \frac{C_1R_1}{N}\right)^2}{\frac{C_1R_1}{N}} + \frac{\left(b - \frac{C_2R_1}{N}\right)^2}{\frac{C_2R_1}{N}} +$$

$$\chi^2 = \frac{(ad - bc)^2 . N}{R_1 R_2 C_1 C_2} \qquad \text{... (11.3)}$$

CONCLUSION : If χ^2 (calculated) < χ^2 (tabulated) with (2 – 1) × (2 – 1) d.f. at chosen level of significance, the null hypothesis is accepted, i.e., the two attributes are independent. Otherwise, the null hypothesis is rejected.

EXAMPLE : 100 individuals of a particular race were tested with an intelligence test and classified into two classes. Another group of 120 individuals belong to another race were administered the same intelligence test and classified into the same two classes. The following are the observed frequencies of the two races :

	Intelligence		
Race	Intelligent	Non-intelligent	
Race I	42	58	100
Race II	55	65	120
	97	123	220

Test whether the intelligence is anything to do with the race.

Null Hypothesis : Intelligence and Race are two independent attributes. Using (11.3), we have

$$\chi^2 = \frac{(42 \times 65 - 58 \times 55)^2 \times 220}{100 \times 220 \times 97 \times 123} = 0.33$$

CONCLUSION : χ^2 (calculated) < χ^2 (tabulated), (3.481) with (2 – 1) (2 – 1) d.f. at 5 per cent level of significance. Therefore, there is evidence to conclude that race and intelligence may be independent.

11.4.1 Yates correction for continuity : In the 2 × 2 contingency table, if the expected cell frequencies are large, the discrete distribution of the probabilities of the cell frequencies approximate to normal distribution and hence χ^2-statistics is distributed as χ^2 distribution with 1 d.f. On the other hand, if the expected frequencies are small (less than 5) the distribution of χ^2 cannot be used.

The corrected value of χ^2 can also be obtained directly using the expression

$$\chi^2 = \frac{(\mid ad-bc \mid - N/2)^2}{R_1R_2\,C_1C_2} \times N \qquad \ldots (11.6)$$

11.4.2. V.M. Dandekar's method: If χ^2_0 is the χ^2 value obtained from original observed frequencies, χ^2_{-1} is the one obtained by increasing the smallest frequency by unity keeping the marginal totals fixed and χ^2_{+1} is the value of χ^2 obtained by decreasing the smallest frequency by one keeping the marginal totals constant, then,

$$\chi^2 = \chi^2_0 - \frac{(\chi^2_0 - \chi^2_{-1})}{\chi^2_{+1} - \chi^2_{-1}} (\chi^2_{+1} - \chi^2_0) \qquad \ldots (11.7)$$

The value of χ^2 is to be obtained from χ^2 and is to be compared with tabulated value at 10 per cent level of significance of normal distribution.

EXAMPLE: The following were the data of 40 individuals classified according to smoking and residential background. Test whether the smoking habit is inherent with the residential background.

	Smokers	Non-smokers	
Rural	15	3	18
Urban	15	7	22
	30	10	40

Yates method: Here the expected frequency of cell containing observed frequency 3 is less than 5. Hence Yates correction for continuity is applied since $15 \times 3 < 15 \times 7$, subtracting $\frac{1}{2}$ from 15 and 7 and adding $\frac{1}{2}$ to 15 and 3, we have

	Smokers	Non-smokers	
Rural	14.5	3.5	18
Urban	15.5	6.5	22
	30.0	10.0	40

Null Hypothesis: Smoking and residential background are independent.

$$\chi^2 = \frac{(14.5\times6.5-15.5\times3.5)^2}{30\times10\times18\times22}\times 40=0.5387$$

$$\chi = 0.7340$$

CONCLUSION: The calculated value of χ is less than the tabulated value 1.645 at 10 per cent level of significance of normal table. Therefore, there is evidence to say that the smoking habit is independent of the residential background.

V.M. Dandekar's method:

$$\chi^2_0 = \frac{(105-45)^2\times40}{30\times10\times18\times22} = 1.2121$$

$$\chi^2_{-1} = \frac{(14\times6-16\times4)^2\times40}{30\times10\times18\times22} = 0.1347$$

$$\chi^2_{+1} = \frac{(16\times8-14\times2)^2\times40}{30\times10\times18\times22} = 3.3670$$

$$\chi^2 = \chi^2_0 - \frac{\chi^2_0-\chi^2_{-1}}{\chi^2_{+1}-\chi^2_{-1}}(\chi^2_{+1}-\chi^2_0)$$

$$\chi^2 = 1.2121 - \frac{(1.2121-0.1347)}{3.3670-0.1347}(3.3670-1.2121)$$

$$= 0.4938$$

$$\chi = 0.7027$$

CONCLUSION: The χ (calculated) $<\chi$ (tabulated), (1.645) at 10 per cent level of significance of normal table. Therefore, null hypothesis is accepted. The two attributes smoking and residential background may be considered independent.

11.5. r × s Contingency Table

When the number of individuals (or objects) are classified according to 'r' categories $A_1, A_2, \ldots, A_r$ with respect to attribute A and 's' categories $B_1, B_2, \ldots, B_s$ with respect to attribute B, then the arrangement of the frequencies in r × s cells is called as r × s contingency table.

TABLE 11.2

A \ B	B_1	B_2	..	B_j	..	B_s	
A_1	O_{11}	O_{12}	..	O_{1j}	..	O_{1s}	R_1
A_2	O_{21}	O_{22}	..	O_{2j}	..	O_{2s}	R_2
..	..	..		..		..	..
A_i	O_{i1}	O_{i2}	..	O_{ij}	..	O_{is}	R_i
..	..	..		..		..	..
A_r	O_{r1}	O_{r2}	..	O_{rj}	..	O_{rs}	R_r
	C_1	C_2	..	C_j	..	C_s	N

Null Hypothesis: The two attributes A and B are independent

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} (O_{ij} - E_{ij})^2 / E_{ij} \qquad \dots \quad (11.8)$$

The expected frequency of each cell can be obtained in the same way as in 2×2 contingency table. For example,

$$E_{ij} = E(O_{ij}) = \frac{R_i \times C_j}{N}$$

where $E(O_{ij})$ is the expected frequency of cell containing observed frequency O_{ij}. The expected frequencies and observed frequencies will be substituted in the expression (11.8) for arriving at the calculated value of χ^2.

CONCLUSION: If χ^2 (calculated) $\geq$ χ^2 (tabulated) with $(r-1)(s-1)$ d.f. at chosen level of significance, the null hypothesis is rejected.

11.6. χ^2-test for Genetic Problems

Genetic theory states that different genes segregate independently and in a particular phenotypic ratios depending upon the cross. A dominant gene segregates from its recessive allele in the ratio 3 : 1 in F_2 generation and in the ratio 1 : 1 in a back cross. If there are two genes responsible for a particular character then there are four type of phenotype combinations in F_2 generation and back cross. They are $(3:1)\times(3:1)=$ 9 : 3 : 3 : 1 and $1:1\times1:1=1:1:1:1$ for F_2 generations and back cross respectively.

EXAMPLE: A cross between the two varieties of Sorghum one giving high yield and the other for high amount of fodder was made. The number of plants in F_2 generation were observed as 79, 160, and 85. Test whether this sample data is in agreement with the Mendelian ratio 1 : 2 : 1 or not.

Null Hypothesis: The sample ratio is in agreement with 1 : 2 : 1 ratio.

Observed Freq. (O_i)	*Expected Freq.* (E_i)	(O_i-E_i)	$\frac{(O_i-E_i)^2}{(E_i)}$
79	$324\times1/4=81$	-2	0.0494
160	$324\times2/4=162$	-2	0.0247
85	$324\times1/4=81$	4	0.1975
324	324		0.2716

CONCLUSION: χ^2 (calculated) $<\chi^2$ (tabulated), (5.991) with (3−1) d.f. at 5 per cent level of significance. Therefore, the null hypothesis is accepted, i.e., the plants are segregating according to Mendelian ratio, 1 : 2 : 1 in F_2 generation.

11.6.1. If there are two or more families, first we compute the χ^2 individually with 1 d.f. for testing the deviation from the theoretical ratio and pool the different χ^2 values for obtaining the total heterogeneity. This χ^2 will be partitioned into two Chi-squares, one due to deviation from theoretical ratio for the totals of the observed frequencies and the other χ^2 due to heterogeneity among families. The different observed frequencies are shown in Table 11.3.

If the families segregate into $e_1 : e_2$ theoretical ratio, then

$$\chi^2_1 = \frac{[(O_{11})e_2-(O_{12})e_1]^2}{R_1\, e_1\, e_2}$$

$$\chi^2_2 = \frac{[O_{21})e_2-(O_{22})e_1]^2}{R_2\, e_1\, e_2}$$

$$\vdots \qquad\qquad \vdots$$

$$\chi^2_k = \frac{[(O_{k1})e_2-(O_{k2})e_1]^2}{R_k\, e_1\, e_2}$$

$$\chi^2_t=\chi^2_1+\chi^2_2+\ldots+\chi^2_k$$

TABLE 11.3

Family	Observed frequencies		
1	O_{11}	O_{12}	R_1
2	O_{21}	O_{22}	R_2
⋮	⋮	⋮	⋮
k	O_{k1}	O_{k2}	R_k
	C_1	C_2	N

The χ^2_t would be partitioned into two components, one due to deviation from $e_1 : e_2$ ratio and another due to heterogeneity between the families as $\chi^2_t=\chi^2_d+\chi^2_h$, where

$$\chi^2_d = \frac{(C_1e_2-C_2e_1)^2}{N\, e_1e_2} \text{ and}$$

$$\chi^2_h=\chi^2_t-\chi^2_d$$

The calculated values of χ^2_d and χ^2_h can be compared with tabulated values with 1 and (k−1) d.f. respectively.

EXAMPLE: In the rabi crop the segregation of the plants (phenotype) in F_2 generation for different parents with regard to one pair of genes (colour) is given as follows :

Parents	*Purple*	*Green*	*Total*
I	320	100	420
II	280	85	365
III	315	11C	425
IV	270	84	354
V	295	105	400
Total	1480	484	1964

Test whether the F_2 generation is segregating in the 3 : 1 ratio as a whole and test whether there exists homogeneity among different pairs of parents.

TABLE 11.4

Parent	χ^2 *(calculated)*	*d.f.*	χ^2 *(tabulated)*	*Remarks*
I	0.3175	1	3.841	Not significant
II	0.5708	1	,,	-do-
III	0.1765	1	,,	-do-
IV	0.3051	1	,,	-do-
V	0.3333	1	,,	-do-
Total	1.7032	5	11.070	

TABLE 11.5

	χ^2 *(calculated)*	*d.f.*	χ^2 *(tabulated)*	*Remarks*
Total	1.7032	5	11.070	Not significant
Deviation from 3 : 1 ratio	0.1331	1	3.841	-do-
Hetero-geneity	1.5701	4	9.488	-do-

CONCLUSION: It is observed from the Tables 11.4 and 11.5 that χ^2 values are not significant within each of the pairs of parents indicating thereby that phenotype frequencies are segregating in the ratio 3 : 1 and it is also seen that the F_2 generation for all the parents is segregating in the 3 : 1 ratio. The calculated value of χ^2 for the heterogeneity between the families is obtained by subtracting the χ^2 value of deviation from 3 : 1 ratio from the total value of χ^2. The calculated value of χ^2 for heterogeneity is also found not significant and hence it can be inferred that the heterogeneity between the families is not significant.

11.7. X^2-test for Linkage Problems

Linkage is defined as the tendency of genes belong to the same chromosome or linkage group to enter the gametes in the parental combinations.

The alternative to linkage is cross over. In this case the genes would tend to enter the gametes other than the parental combinations. These are called recombinations.

11.7.1. When the two pairs of genes (or characters) are under consideration whether the F_2 data (phenotype frequency) is segregating according to expected theoretical ratio or not can be tested with the help of chi-square test. If there is a significant departure from the theoretical ratio, the reason for this departure may be either due to the individual characters (pair of genes) may not be segregating into theoretical ratio or it may be due to significant linkage.

If χ^2_t is the departure of F_2 data from theoretical ratio (say) 9 : 3 : 3 : 1 for the two characters A and B with 3 d.f. and χ^2_A is the departure from the theoretical ratio 3 : 1 for the character A with 1 d.f. and χ^2_B is the departure from the theoretical ratio 3 : 1 for the character B with 1 d.f., we have

$\chi^2_l = \chi^2_t - (\chi^2_A + \chi^2_B)$ where χ^2_l is the value of χ^2 for linkage.

TABLE 11.6

	B	b	
A	AB	Ab	R_1
a	aB	ab	R_2
	C_1	C_2	N

The different chi-square values are presented in Table 11.7.

TABLE 11.7

Source	χ^2 *value*	*d.f.*	χ^3 *(tabulated)*
Character			
A	χ^2_a	1	3.841
B	χ^2_b	1	-do-
Linkage	χ^2_l	1	-do-
Total	χ^2_t	3	

EXAMPLE: In an experiment on sweet peas purple and long pollen grains were crossed with red and round pollen grains and progeny in F_2 generation were observed as follows. Purple and long pollen grains were 220, purple and round pollen 105, red and long pollen 90 and red and round pollen 17. Test departure of F_2 generation from the theoretical ratio 9 : 3 : 3 : 1 and also the linkage and the individual characters in the ratio 3 : 1.

Null Hypothesis: Observed frequencies are in the theoretical ratios.

TABLE 11.8

	Long	*Round*	
Purple	220	105	325
Red	90	17	107
	310	122	432

$$\chi^2_c \text{ (Colour)} = \frac{(325-3\times 107)^2}{3\times 432} = 0.012$$

$$\chi^2_s \text{ (Shape)} = \frac{(310-3\times 122)^2}{3\times 432} = 2.420$$

$$\chi^2_l \text{ (Linkage)} = \frac{(220-3\times 105-3\times 90+9\times 17)^2}{9\times 432} = 11.56$$

TABLE 11.9

Observed (O_i)	*Expected* (E_i)	$\frac{(O_i-E_i)^2}{(E_i)}$
220	243	2.177
105	81	7.111
90	81	1.000
17	27	3.704
432	432	13.992

χ^2 (Total) = 13.992

TABLE 11.10

Source	χ^2 *(calculated)*	*d.f.*	χ^2 *(tabulated)*	*Remarks*
Colour (Purple Vs Red)	0.012	1	3.841	Not significant
Shape (Long Vs Round)	2.420	1	3.841	-do-
Linkage	11.560	1	3.841	Significant
Total	13.992	3	7.815	Significant

CONCLUSION: Here the values of χ^2 for colour and shape are not significant, thereby indicating that the individual characters are segregating in the theoretical ratio 3 : 1. But the χ^2_t value is significant indicating that both the pairs of genes are not assorting independently. Therefore, the discrepancy may be due to linkage. The χ^2_l is found significant which confirms the earlier statement.

11.7.2. If there are three pairs of genes A, a, B, b, C, c segregating in a back cross (Aa Bb Cc×aa bb cc) and each gene is expected to show a 1 : 1 ratio. If there is no linkage the three pairs of genes segregate independently. There are in all eight classes ABC, ABc, AbC, aBC, Abc, aBc, abC, and abc. The expected frequencies will be same in all the classes i.e., 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 in a back cross. The total χ^2 will have 7 d.f. for the eight classes as a whole. The coefficients for the different pairs of comparisons can be obtained with the help of Fisher's evens Vs Odds rule.

TABLE 11.11

		ABC	*ABc*	*AbC*	*aBC*	*Abc*	*aBc*	*abC*	*abc*
Main effects	A	+	+	+	—	+	—	—	—
	B	+	+	—	+	—	+	—	—
	C	+	—	+	+	—	—	+	—
Interaction	AB	+	+	—	—	—	—	+	+
	BC	+	—	—	+	+	—	—	+
	AC	+	—	+	—	—	+	—	+
Second order Interaction	ABC	+	—	—	—	+	+	+	—

Since in the back cross, the ratio of segregation for the pairs of genes in F_2 generation is 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1, the divisor for each comparison is equal to N where N is the total frequency.

$$\chi^2_A = [(ABC)+(ABc)+(AbC)+Abc-(aBC)-(aBc)-(abC)-(abc)^2]/N$$

The total χ^2 with 7 d.f. will be partitioned into different χ^2s for each of the 7 comparisons with 1 d.f. each and these will be tested as follows.

TABLE 11.12

Item (Source)	χ^2 (*calculated*)	*d.f.*	χ^2 (*tabulated*)
A	χ^2_A	1	3.841
B	χ^2_B	1	,,
C	χ^2_C	1	,,
Interaction Linkage			
AB	χ^2_{AB}	1	,,
BC	χ^2_{BC}	1	,,
AC	χ^2_{AC}	1	,,
ABC	χ^2_{ABC}	1	,,
Total	χ^2_t	7	14.067

EXERCISES

1. A group of school children were classified according to intelligence level (I) and economic level (E) and the results were as under

	E_1	E_2	E_3	E_4
I_1	85	216	165	206
I_2	144	305	320	152
I_3	120	185	160	45

Test for the independence of the two factors at one per cent level.

2. The following table gives the number of literates and criminals in the three cities A, B and C. Compare the degree of association between criminality and illiteracy in each of the three cities.

		A	B	C
Total number (in ten thousands)	..	246	185	228
Literates (,,)	..	42	47	32
Literate criminals (,,)	..	4	2	3
Illiterate criminals (,,)	..	41	22	24

3. The probability for an animal to catch a particular infection is 0.20. In an experiment 60 animals were treated with a new vaccine, and 5 of them caught the infection. Applying the χ^2-test, arrive at a decision regarding the efficiency of the new vaccine at the 5 per cent level of significance.

4. In an experiment on chillies, the following results were obtained

Table

Shape	Pungent	Not-Pungent
Long	48	27
Short	32	73

Test whether there is any association between taste and shape of chillies at 5 per cent level of significance.

5. To prevent eye disease in children, an experimental diet was recommended. In an investigation the following results were obtained.

	Prevented the disease	Not-Prevented the disease
Experimental diet	10	6
Control diet	4	14

Test whether experimental diet has any effect in preventing eye disease in children.

6. Experimental plots were classified according to duration of the variety and infestation of particular pest and the following results were obtained in an investigation.

Duration	High infested	Medium infested	Low infested
Long	10	18	27
Medium	14	22	34
Short	38	20	12

Test whether duration of a variety has any effect in controlling the pest at 5 per cent level.

7. Tobacco leaves were classified according to shape and quality in a survey and the results are presented as follows.

		Shape	
Quality	Large	Medium	Narrow
Good	34	28	9
Satisfactory	30	16	11
Not-good	14	20	38

Test whether there is any association between "quality" and 'shape' of Tobacco leaves.

8. The following table gives the number of children vaccinated against polio disease. Test whether vaccination has any effect in controlling the disease at 1 per cent level of significance.

	Effected	**Not-effected**
Vaccinated	2	22
Not-vaccinated	18	10

9. In an experiment conducted on Tomatoes Red and round were crossed with Pink and elongated and the progeny in F_2 generation were observed as follows. Red and round tomatoes were 40, Red and elongated 84, Pink and round 69 and Pink and elongated 46. Test departure of F_2 generation from the theoretical ratio 9:3:3:1 and also the linkage and individual characters in the ratio 3:1.

10. The following are the number of seeds germinated in each pot when 10 seeds are sown in glass house experiment on sunflower crop.

	Pot									
	1	2	3	4	5	6	7	8	9	10
Seeds germinated	8	6	3	8	6	4	5	7	9	4

Test whether seeds germinated uniformly in all the pots at 5 per cent level of significance.

11. The following are the observed and expected frequencies obtained in fitting Binomial distribution.

Observed	6	10	19	28	13	9	5
Expected	4	12	20	27	14	7	6

Test whether expected frequencies are in close agreement with observed frequencies at 1 per cent level of significance.