AJIT KUMAR SINHA
Prof. & Head; Deptt of Zoology
H.D. Jain College, Ara

Chi - Square Test

In biological experiments, sometimes we get qualitative data like colour, health, intelligence etc. in which observations are classified in a particular category, class or group. For these qualitative data a non parametric test, called "Chi - Square Test" is generally used.

Chi - Square Test was first developed by Prof. A.R. Fisher in 1870. Later on, in the year 1900, Karl Pearson improved Fisher's Chi - Square Test.

Definition —
"Chi - Square Test" is the test of significance of overall deviation square in the Observed and Expected frequencies divided by Expected frequencies."

Formula for determining Chi - Square (χ^2)

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

or

$$\chi^2 = \Sigma \left\{ \frac{(f_o - f_e)^2}{f_e} \right\}$$

Where, O or f_o → Observed frequency in a class

E or f_e → Expected frequency in a class

Σ → Summation

Chi - Square test is used for two specific purposes —

(i) To test the hypothesis of no association between

two or more groups, population or criteria.
(i) to check independence between two variables)
(ii) to test how likely the observed distribution of data fits with the distribution that is expected (i.e. to test the goodness of fit)

Pre-requisites of Chi Square test —

There are three basic pre-requisites of χ^2 test such as :-

(a) Sample must be random
(b) data should be qualitative
(c) Preferably observed frequency should not be less than five.

Degrees of freedom (d.f.)

The degree of freedom is calculated from the number of classes. Therefore, the number of degrees of freedom in a χ^2 test is equal to the number of classes minus one. If there are two classes, three classes and four classes; the degree of freedom would be 2-1, 3-1, 4-1, respectively.

Limitations of Chi-square test —

1. The chi-square test does not give us much information about the strength of the relationship between the variables investigated.
2. It is sensitive to sample size. This may make a weak relationship statistically significant if the sample is large enough.
3. It is also sensitive to small expected frequencies; It can be used only when not more than 20% of cells have an expected frequency of less than 5.
4. It cannot be used when samples are related or matched.

1. In a grassland the earthworm population was sampled from ten randomly located plots of $1m^2$ area. The following table gives the number of earthworms obtained. Examine the distribution pattern of earthworms.

Area	1	2	3	4	5	6	7	8	9	10
No. of earthworms/m^2	25	32	17	23	15	39	27	19	22	26

Ho: It is presumed that the earthworm population in ground is normally distributed.

$$\Sigma X = 245$$

$$\bar{X} = \frac{\Sigma X}{N} = \frac{245}{10} = 24.5$$

Thus expected number of earthworms in each quadrates is the mean number of earthworms i.e. 24.5

O	E	(O−E)	$(O-E)^2$	$\frac{(O-E)^2}{E}$
25	24.5	0.5	0.25	0.01
32	24.5	7.5	56.25	2.29
17	24.5	−7.5	56.25	2.29
23	24.5	−1.5	2.25	0.09
15	24.5	−9.5	90.25	3.6
39	24.5	14.5	210.25	8.58
27	24.5	2.5	6.25	0.25
19	24.5	−5.5	30.25	1.23
22	24.5	−2.5	6.25	0.25
26	24.5	1.5	2.25	0.09

$$\Sigma\left[\frac{(O-E)^2}{E}\right] = 18.68$$

$$\boxed{\chi^2 = \Sigma\left[\frac{(O-E)^2}{E}\right]} = 18.68$$

Here d.f. = 10 − 1 = 9

Q.2. In a cross between Tall (TT) and dwarf (tt) 1574 tall and 554 dwarf plants were obtained. Suggest if a Mendelian ratio of 3:1 is suitable or not.

Total number = 1574 tall + 554 dwarf = 2128

Ho: It is presumed that the tall and dwarf plants are segregating in a 3:1 ratio. Therefore, expected 3:1 will be —

$$2128 \times \frac{3}{4} : 2128 \times \frac{1}{4} = 1596 : 532$$

Thus Observed ratio is 1574 : 554 and Expected ratio is 1596 : 532

Putting the values in the formula —

$$\chi^2 = \Sigma\left[\frac{(f_o - f_e)^2}{f_e}\right]$$

$$= \left[\frac{(1574 - 1596)^2}{1596} + \frac{(554 - 532)^2}{532}\right]$$

$$= \frac{(-22)^2}{1596} + \frac{(22)^2}{532} = \frac{484}{1596} + \frac{484}{532}$$

$$= 0.303 + 0.909 = 1.212$$

Here, d.f. = 2-1 = 1

Contingency Table — Association attributes are shown by contingency table.

In case of contingency table, degree of freedom is (row − 1) × (column − 1)

If there are 4 rows and 9 columns,

d.f = (4-1) × (9-1) = 24.

33. From the following data find out whether there is any relationship between sex [Male/Female] and preference of colours.

Colour	Males	Females	Total
Pink	10	40	50
Black	70	30	100
Yellow	30	20	50
Total	110	90	200

H_0 : colour preference by Male = colour preference by females

H_a : colour preference by Male ≠ colour preference by females.

Given data — calculation of Expected Value —

Colour	Male	Expected	Female	Expected	Total
Pink	10	27.5	40	22.5	50
Black	70	55	30	45	100
Yellow	30	27.5	20	22.5	50
Total	110		90		200

$$\text{Expected Value} = \frac{\text{Row Total} \times \text{Column Total}}{\text{Total}}$$

Calculation of χ^2

O	E	(O-E)	$(O-E)^2$	$\frac{(O-E)^2}{E}$
10	27.5	−17.5	306.25	11.14
70	55	+15	225	4.09
30	27.5	+2.5	6.25	0.23
40	22.5	+17.5	306.25	13.61
30	45	−15	225	5.00
20	22.5	−2.5	6.25	0.28

$$\Sigma\left[\frac{(O-E)^2}{E}\right] = 34.35$$

$$\chi^2 = \sum\left[\frac{(O-E)^2}{E}\right]$$

$$= 34.382$$

Here, d.f. $= (r-1) \times (c-1)$

$= (3-1) \times (2-1)$

$= 2 \times 1$

$= 2$