Faculty Science

Everest Shiwach

Department: Botany

B Sc III -Paper I (Plant Resource Utilization, Palynology, Plant Pathology and Biostatistics)

Unit- IV Topic- Chi square test (x^2)

Chi square test was developed by Professor A.R. Fischer in 1870. Karl Pearson improved Fisher's chi square test in its modern form in 1900. It is a useful measure of comparing experimentally obtained results with those expected theoretically and are based on hypothesis.

Chi square test is applied to those problems in which we study observed frequency is significantly different from expected frequency. In other word, it is used to find the degree of discrepancy between observed frequencies and expected frequencies. Whether this discrepancy so obtained between observed frequency and theoretically frequency is due to error of sampling or due to chance. It is defined as

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

where O = observed frequency E = expected frequency

It is not a parameter because its value is not derived from observations in a population. It is non parametric test.

Degree of freedom- If data is given in the form of a series of variables in a row or column.

The degree of freedom = n-1 n = number of items in series
when the frequencies are put in cells in the contingency table.
Degree of freedom will be the product of (number of rows less 1 and the number of columns less one). Degree of freedom = (R-1) (C-1)

Uses of Chi square:

1. Test of goodness of fit- Chi square test is used to test goodness of fit. Goodness of fit reveals the closeness between observed frequency and expected frequency.

It helps to answer whether something (physical or chemical factors) did or did not have effect. If observed and expected frequencies are in agreement with each other than chi square should be zero, null hypothesis. But it does not happen in biology. There is always some degree of deviation.

Pre-requisites of chi square test- There are three basic pre-requisites.
1. Sample must be random.
2. Data should be qualitative.
3. Observed frequency should not be less than 5.

Example- In an experiment among 1600 beans the number of four types were 882, 313, 287 and 118. Theory predicts proportion of beans in four groups A, B, C and D should be 9 : 3 : 3 : 1. Does the experiment results support the theory?
H_0- The theory fits well into experiment.

Type of Beans	Observed Frequencies	Expected Frequencies
A	882	$\frac{1600\times9}{16} = 900$
B	313	$\frac{1600\times3}{16} = 300$
C	287	$\frac{1600\times3}{16} = 300$
D	118	$\frac{1600\times1}{16} = 100$
Total	1600	

$$\chi^2 = \sum\left\{\frac{(882-900)^2}{900} + \frac{(313-300)^2}{300} + \frac{(287-300)^2}{300} + \frac{(118-100)^2}{100}\right\}$$

= 0.3600+0.5633+0.5633+3.240

= 4.7

n-1=3 , χ^2 for 3 degree of freedom is 7.815 from table. Since calculated value of χ^2 is less than the tabulated χ^2 value we accept our null hypothesis that there is no difference between theoretical and experimental result.

2. Test of independence of attributes- It is used to test the independence of two attributes of a population for example We want to know that a new medicine is effective in controlling fever or not.
H_0= medicine is not effective in controlling fever
H_1= effective in controlling fever

Now we calculate expected frequency and then Chi square if the value of calculated Chi square is less than the table value at a certain level of significance for given degree of freedom. We accept null hypothesis that new medicine is not effective in controlling the fever.

3. Test of difference of more than two proportion

Suppose assembly elections are announced in three major states. A major political party wishes to test if the proportions of its supporters in the three states are the same or not.The party conduct the survey of 1000 people in each states and finds that there are π_1, π_2 and π_3 supporters in the sample in the surveyed states. Here we wish to test

Null hypothesis $H_0 = \pi_1 = \pi_2 = \pi_3$, all proportion are equal against the alternate hypothesis $H_1 = \pi_1 \neq \pi_2 \neq \pi_3$, all proportion are not equal.

We only test if all the proportions are the same or not. In case we conclude that all the proportions are not the same, we cannot find which one is different.

Now we calculate expected frequency and then Chi square if the value of calculated Chi square is greater than the table value at a certain level of significance for given degree of freedom. We reject null hypothesis and conclude that all proportions of supporters of the political party in three states are not the same.