The Chi-Square

Definition

Chi- square test is the test of significance. It was first o
used by Karl Pearson in the year 1900. Chi-square tes
a useful measure of comparing experimentally obtaine
result with those expected theoretically and based on t
hypothesis. Following is the formula.

It is a mathematical expression, representing the ratio

$$\chi^2 = \sum \frac{(Observed - Expected)^2}{Expected}$$

Chi-Square

* It is a mathematical expression, representing the ratio between experimentally obtained result (O) and the theoretically expected result (E) based on certain hypothesis. It uses data in the form of frequencies (i.e., the number of occurrence of an event).
* Chi-square test is calculated by dividing the square of the overall deviation in the observed and expected frequencies by the expected frequency.

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

Chi-Square

* If there is no difference between actual and observed frequencies, the value of chi-square is zero.
* If there is a difference between observed and expected frequencies, then the value of chi-square would be more than zero.
* But the difference in the observed frequencies may also be due to sampling fluctuations and it should be ignored in drawing inference.

Degrees of Freedom

In test, while comparing the calculated value of with the table value, we have to calculated the degree of freedom. The degree of freedom is calculated from the no. of classes. Therefore, the no. of degrees of freedom in a test is equal to the no. of classes minus one.

Degrees of Freedom

If there are two classes, three classes, and four classes, the degree of freedom would be 2-1, 3-1, and 4-1, respectively. In a contingency table, the degree of freedom is calculated in a different manner: d.f. = (r-1) (c-1) where r = number of row in a table, c = number of column in a table. Thus in a 2×2 contingency table, the degree of freedom is (2-1) (2-1) = 1. Similarly, in a 3×3 contingency table, the number of degree of freedom is (3-1) (3-1) = 4. Likewise in 3×4 contingency table the degree of freedom is (3-1) (4-1) = 6, and so on.

Contingency table

The term contingency table was first used by Karl Pearson. *A contingency table is a type of table in a matrix format that displays the (multivariate) frequency distribution of the variables. They are heavily used in survey research, business intelligence, engineering and scientific research. They provide a basic picture of the interrelation between two variables and can help find interactions between them.

Contingency Table…..

The value of depends on the no. of classes or in other words on the number of degrees of freedom (d. f.) and the critical level of probability. *2×2 table when there are only two sample, each divided into classes and a 2×2 contingency table is prepared. It is also known ad four fold or four cell table.

Contingency table…..

2 rows × 2 columns

	Column 1	Column 2	ROW Total
ROW 1	+	+	RT 1
ROW 2	+	+	RT 2
Column Total	CT 1	CT 2	

Degree of freedom = (r-1) (c-1)

= (2-1) (2-1)

= 1×1

= 1

Characteristics of Chi-Square Test

The chi-square distribution has some important characteristics:

1. This test is based on frequencies, whereas, in theoretical distribution the test is based on mean and standard deviation.

2. The other distribution can be used for testing the significance of the difference between a single expected value and observed proportion. However this test can be used for testing difference between the entire set of the expected and the observed frequencies.

3. A new chi-square distribution is formed for every increase in is the number of degree of freedom

Assumptions For validity of Chi-square test

There are a few assumptions for the validity of chi-square test.

1. All the observation must be independent. No individual item should be included twice or a number of items in the sample.

2. The total number of observation should be large. The chi-square test should not be used if n>50.

3. All the events must be mutually exclusive.

4. For comparison purposes, the data must be in original units.

Application of Chi-square test

The chi-square test is applicable to varied problems in agriculture, biology and medical science

1. To test the goodness of fit.
2. To test the independence of attributes.
3. To test the homogeneity of independent estimates of the population variance.
4. To test the detection of linkage.

Determination of Chi-square Test

Following steps are required to calculate the value of c
square.

1. Identify the problem
2. Make a contingency table and note the observed
   frequency (O) is each classes of one event, row wise i.e.
   horizontally and then the numbers in each group of the
   other event, column wise i.e. vertically.
3. Set up the Null hypothesis (Ho); According to Null
   hypothesis, no association exists between attributes. This
   need s setting up of alternative hypothesis (HA). It assumes that an associa
exists between the attributes.
4. Calculate the expected frequencies (E).
5. Find the difference between observed and Expected
   frequency in each cell (O-E).