χ^2 - TEST

While concluding experiments particularly related with various types of genetical studies, it has been found that the results of actual observations deviate from the results calculated on the basis of some hypothesis or expectation. If deviation is found to be very less, then it is assumed to be simply due to chance but when it is found to be more, it is considered to be due to some reasons. The situation attracts attention that upto what extent this deviation should be considered to be due to chance alone. The method adopted to compare the experimentally observed result with these expected on the basis of probabilities is known as χ^2 - test.

χ^2 test is a widely used and most important non-parametric or distribution free statistic. It is generally used when the data are to be expressed either in frequency or percentage or proportion. It is applied only for qualitative data (eg, colour, health, intelligence, cure, response of drug, etc.) lacking numerical values. It was developed by Fischer (1870) and its improvement in modern form was done was Karl Pearson (1900). χ (chi or Kye) is the 22nd letter of Greek alphabet.

Definitions: 1. χ^2 is the statistic by which obtained frequencies are compared with expected frequencies on the basis of a hypothesis.

2. χ^2 is a sum of ratio in which each ratio lies between a squared difference and an expected frequency (Guilford -1973).

3. χ^2 test is of overall deviation to square in the observed and expected frequencies divided

Formulation:

$$\chi^2 = \Sigma \frac{(O-E)^2}{E} \quad \text{or} \quad \Sigma \frac{(f_o - f_e)^2}{f_e}$$

Where, O or f_o = obtained or observed frequency.

E or f_e = hypothetical or expected frequency.

Prerequisite: 1. Sample must be random or binomial

2. Data should be qualitative

3. Observed frequency should not be less than 5.

Characteristics: 1. It is directly dependent on the discrepancies of observed and expected frequencies. If $O = E$, $\chi^2 = 0$.

2. It does not indicate about the direction of discrepancies but compare several proportions

3. It is a positive quotient and its value ranges upward from zero.

4. Yule and Kendall regarded a better minimum theoretical frequency to be 10.

5. Variables should not be more than 30.

Limitations: 1. When O or E are less than 5, χ^2 value is faulty. In this condition, Yates correction is required. Conversely, when O or E are >30, probability of normal distribution increases.

2. It gives the idea of corelation but not the cause.

3. It requires only absolute number.

4. The value of co-efficient of contingency is limited.

5. Observations should be independent.

6. Constraints on cell frequencies.

7. No. theoretical cell frequency should be less than 5.

Applications: 1. It tests equal probability distribution ($O = E$)

2. It tests independence hypothesis in a 2×2 contingency table

3. It tests goodness of fit (closeness of observed and expected frequencies).

4. It establishes a relation with ϕ (phi) coefficient

5. It tests the hypothesis of normal distribution

6. It finds significance of difference into 2 or more proportions

7. It is more useful than standard error method and is used when data fit into yes or no type of categories.

8. To establish association of attributes between two events in binomial or multinomial samples.

9. To test homogeneity of independent estimates of population variance and population correlation coefficient

10. To combine various probabilities to give a single test of significance.

Yates (1934) correction: When any of the cell frequency is less than 5, Yates correction is needed for continuity. In this correction, 0·5 is added to the smallest cell frequency and remaining frequencies are adjusted in such a way by which the grand total and marginal total remain unaffected. The correction may be applied at f_o level or $(f_o - f_e) - 0{\cdot}5$ level.

Critical values of χ^2-test: The critical values of χ^2-test may be determined by knowing the values of acceptance region, rejection region, upper ... and degree of freedom as shown in the graph.

rejection region upper and degree of freedom as shown in the graph

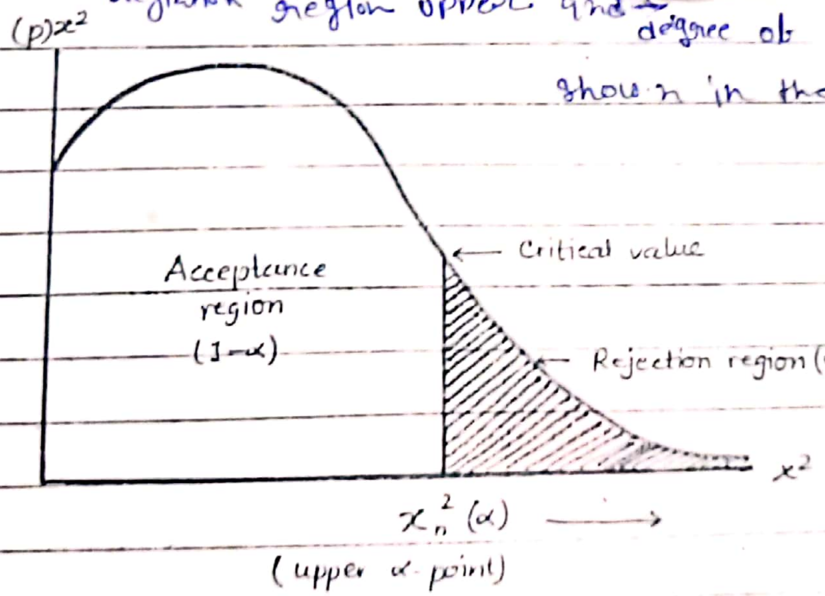

Fig: Graph showing significant values of χ^2 n degree of freedom.

The graph indicates that:

1. It is skewed to the right side so its skeweness is positive.
2. As the number of degree of freedom increases, χ^2 approaches a symmetric distribution
3. There is a family of χ^2 distribution (similarity with t test).