Chi Square Test

-Mukesh Kumar

- The Chi-Square Test is a statistical method used to determine if there is a significant association between categorical variables.

ChiSquare Test Types

- **Chi-Square Test of Independence**: Tests whether two categorical variables are independent.
- **Chi-Square Test of Goodness of Fit**: Tests whether a sample distribution matches an expected distribution.

When to use ChiSquare Test

- When we have two categorical variable and we want to check whether there is a relationship

Process to perfrom ChiSquare

- 1. **Formulate Hypotheses**
- 2. **Create a Contingency Table**
- 3. **Calculate Expected Frequencies**

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

Where:

- R_i = total for row i
- C_j = total for column j
- N = grand total of observations

4. **Compute the Chi-Square Statistic**

Now, calculate the Chi-Square statistic using:

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

5. **Determine Degrees of Freedom**

6. **Compare the Chi-Square Statistic to the Critical Value**

- If ChiSquare > Critical value – reject H0
- If ChiSquare < Critical Value – fail to reject H0

Example

- A teacher wants to know if there is a relationship between students' preferred study methods and their exam performance. The study methods are categorized as "Group Study," "Self-Study," and "Online Resources," and the exam performance is categorized as "Pass" or "Fail." The data collected from a sample of 120 students is as follows:

Study Method	Pass	Fail	Total
Group Study	20	10	30
Self-Study	50	20	70
Online Resources	10	10	20
Total	80	40	120

Formulate the hypothesis

- H0 = no relationship
- Ha= there is relation

- Outcome:
- - if P>alpha: fail to reject Null i.e . No relation
- - if p< alpha : reject null i.e. there is relation

Solution

- Contingency table is given :

Study Method	Pass	Fail	Total
Group Study	20	10	30
Self-Study	50	20	70
Online Resources	10	10	20
Total	80	40	120

Calculate the expected frequencies

Let's calculate the expected frequencies:

- **Group Study, Pass:**

$$E_{11} = \frac{(30 \times 80)}{120} = \frac{2400}{120} = 20$$

- **Group Study, Fail:**

$$E_{12} = \frac{(30 \times 40)}{120} = \frac{1200}{120} = 10$$

- **Self-Study, Pass:**

$$E_{21} = \frac{(70 \times 80)}{120} = \frac{5600}{120} = 46.67$$

- **Self-Study, Fail:**

$$E_{22} = \frac{(70 \times 40)}{120} = \frac{2800}{120} = 23.33$$

- Online Resources, Pass:

$$E_{31} = \frac{(20 \times 80)}{120} = \frac{1600}{120} = 13.33$$

- Online Resources, Fail:

$$E_{32} = \frac{(20 \times 40)}{120} = \frac{800}{120} = 6.67$$

Expected Frequency table

The expected frequency table is:

Study Method	Pass	Fail
Group Study	20	10
Self-Study	46.67	23.33
Online Resources	13.33	6.67

Observed Vs Expected values

Study Method	Pass	Fail	Total
Group Study	20	10	30
Self-Study	50	20	70
Online Resources	10	10	20
Total	80	40	120

The expected frequency table is:

Study Method	Pass	Fail
Group Study	20	10
Self-Study	46.67	23.33
Online Resources	13.33	6.67

4. Compute the Chi-Square Statistic

Now, calculate the Chi-Square statistic using:

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Let's compute it:

- **Group Study, Pass:** $\frac{(20-20)^2}{20} = 0$
- **Group Study, Fail:** $\frac{(10-10)^2}{10} = 0$
- **Self-Study, Pass:** $\frac{(50-46.67)^2}{46.67} = \frac{(3.33)^2}{46.67} \approx 0.238$
- **Self-Study, Fail:** $\frac{(20-23.33)^2}{23.33} = \frac{(-3.33)^2}{23.33} \approx 0.475$
- **Online Resources, Pass:** $\frac{(10-13.33)^2}{13.33} = \frac{(-3.33)^2}{13.33} \approx 0.832$
- **Online Resources, Fail:** $\frac{(10-6.67)^2}{6.67} = \frac{(3.33)^2}{6.67} \approx 1.664$

Summing these values gives:

$$\chi^2 \approx 0 + 0 + 0.238 + 0.475 + 0.832 + 1.664 = 3.209$$

Determine Degree of Freedom

- $df=(r-1)\times(c-1)=(3-1)\times(2-1)=2\times1=2$

Compare the Chi-Square Statistic to the Critical Value

- Since $\chi 2=3.209$ is less than the critical value of 5.991, we **fail to reject** the null hypothesis.

- H0 = no relationship
- Ha= there is relation