The Chi – Squared Test
(X^2 – test)

By
Dr. Asaad Al-Yassen
Assist. Professor of dermato-epidemiology
& occupational health
Department of Community Medicine
University of Basrah

The distribution of a categorical variable in a sample often needs to be compared with the distribution of a categorical variable in another sample.

- It is a test for qualitative data.
- Based on counts or frequencies.
- Chi – squared test measures the difference between **actual frequencies** and **expected frequencies** (as expected under the null hypothesis)

$$X^2 = \text{sum of} \quad \frac{(\text{Observed frequency} - \text{Expected frequency})^2}{\text{Expected frequency}}$$

$$X^2 = \Sigma \frac{(O - E)^2}{E}$$

Procedure:

1.State the null hypothesis (Ho):
There is no relation ship between the two variables.
2.Arrange the data in a table.
3.Calculate the expected frequencies:

$$\text{Expected frequency (E)} = \frac{\text{Row total X Column total}}{\text{Grand total}}$$

4.Calculate X^2 value:

$$X^2 - \text{test} = \Sigma \frac{(O - E)^2}{E}$$

5.Determine degree of freedom:

$$df = (\text{Rows} - 1)(\text{Columns} - 1)$$

6.Compare the **calculated X^2 value** with the **tabulated critical value**.

7.Conclusion:

At 95% level

If the **calculated X^2 value < tabulated critical value**

P > 0.05

So **accept** the null hypothesis

If the **calculated X^2 value > tabulated critical value**

P < 0.05

So **reject** the null hypothesis

Example: The following data were obtained from a study on the association between smoking and lung cancer in men:

Smoking status	No. of persons who developed lung cancer	No. of persons who did not develop lung cancer	Total
Smokers	30	120	150
Non – smokers	10	100	110
Total	40	220	260

Perform a complete X^2 - test on the data in the table above to show whether an association does exist between smoking and lung cancer.

1. **Null hypothesis:** There is no relation ship or association between smoking and lung cancer, and if there is association is due to chance or sampling error.
2. **Arrange the table.**
3. **Calculate the expected frequency for each cell.**

$$\textbf{Expected frequency (E)} = \frac{\textbf{Row total X Column total}}{\textbf{Grand total}}$$

$$E(30) = \frac{150 \times 40}{260} = 23.08$$

$$E(120) = \frac{150 \times 220}{260} = 126.92$$

$$E(10) = \frac{110 \times 40}{260} = 16.92$$

$$E(100) = \frac{110 \times 220}{260} = 93.08$$

4. **Calculate X^2 value:**

$$\mathbf{X^2 - test} = \Sigma \frac{\mathbf{(O - E)^2}}{\mathbf{E}}$$

$$= \frac{(30 - 23.08)^2}{23.08} + \frac{(120 - 126.92)^2}{126.92} + \frac{(10 - 16.92)^2}{16.92} + \frac{(100 - 93.08)^2}{93.08}$$

$$= 2.08 + 0.37 + 2.83 + 0.51$$

= 5.79

5. Calculate degree of freedom:

df = (Rows – 1) (Columns – 1)

= (2 – 1) (2 – 1)

= 1

6. Tabulated critical X^2 value:

Df	0.05	0.01
1	3.84	6.63

At 95% level

5.79 > 3.84

P < 0.05

So **reject** the null hypothesis

There is **significant relationship** between smoking and development of lung cancer.

At 99% level

5.79 < 6.63

P > 0.01

So **accept** the null hypothesis

No highly significant relationship between smoking and development of lung cancer.

0.05 > P > 0.01

X^2 – test (facts and limitation)

- It shows whether a relationship exists between two variables of interest.
- It does not show the nature of the relationship.
- The expected frequency in each cell should not be less than 5.
- The calculation of X^2 must always be based on absolute numbers not on percentage or proportions.
- If all the observed cell frequencies coincide with the expected frequencies $X^2 = 0$. the greater the differences between the observed and the expected frequencies the larger the value of X^2.
- X^2 – test does not measure the strength of association between two factors.
- It does not show causality.