The Chi – Squared Test
(X^2 – test)

Dr. Asaad Q. Al-Yassen
Assist. Professor of Dermato-Epidemiology
& Occupational Health
Department of Family &Community Medicine
College of Medicine
University of Basrah

Learning Objective:
1. What are parametric and non - parametric test?
2. What is a Chi Square test?
3. What are the different situations where chi square test can be used?
4. How to use the Chi square test?

There are many common situations where the underlying distribution is **Non – normal or unknown,** where it is not possible for the researcher to make rigid assumptions about the shape of the population from which the samples have been drawn for a study.

Statisticians have developed a class of alternative techniques known as **Non – parametric or distribution – free methods**.

A Non – parametric method may be defined as a test in which **no hypothesis is made about specific values of population parameters**.

Parametric tests implies that the scores should achieve either an interval or ratio scale of measurement.
i.e. The test in which, the population constants like mean, standard deviation, standard error, correlation coefficient, etc. and the data tend to follow one assumed or established distribution such as normal, binomial, poisson etc.

Non – parametric test statistics are used when the data are measured in nominal or categorical scales.
i.e. The test in which no constant of a population is used. Data not follow any specific distribution and no assumption is made in these tests. **e.g.** to classify good, better and best we just allocate arbitrary numbers or marks to each category.

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

The chi – square test is an important test among the several tests of significance developed by statisticians.

The Greek letter X^2 was first used to describe this situation by Karl Pearson in early 1900's.

The chi - square test is a non parametric test not based on any assumption or distribution of any variable.

This statistical test follows a specific distribution known as chi square distribution.

Chi square distribution

1. Positively skewed distribution
2. X^2 values will never be negative; minimum is zero
3. X^2 of close to 0 indicate that the variables are independent of one another.

The chi - square test

- 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)

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

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

Applications of a chi-square test

This test can be used in:

1. **Goodness of fit of distributions**
   This test enables us to see how well does the assumed theoretical distribution fit to the observed data.
   e.g. A researcher has chosen 25 participants (10 of whom are males and 15 are females) and wishes to know if there are significantly more female than male participants. Assuming that the participants were chosen from a population where the number of females and males is equal.

2. **Test of independence of attributes**

   This test enables us to explain whether or not two attributes are associated.
   e.g. we may be interested in knowing whether a new medicine is effective in controlling fever or not.

3. **Test of homogensity**

   This test can be also used to test whether the occurrence of events follows uniformity or not.
   e.g. the admission of patients in a governmental hospital in all days of the week is uniform or not can be tested with the help of chi square test.

Procedure:

1.State the null hypothesis (Ho):

There is no relationship between the two variables.

2.Arrange the data in a table.

3.Calculate the expected frequencies:

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

4.Calculate X^2 value:

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

5.Determine degree of freedom:

df = (Rows – 1) (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 relationship 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.**

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

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

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

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

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

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

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

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

$$+ \frac{(100 - 93.08)^2}{93.08}$$

$= 2.08 + 0.37 + 2.83 + 0.51$

$= 5.79$

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 **a significant relationship** between smoking and the development of lung cancer.

At 99.7% level

$$5.79 < 6.63$$
$$P > 0.01$$

So **accept** the null hypothesis

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

$$\mathbf{0.05 > P > 0.01}$$

Special case analyzing 2 X 2 table

In general form a 2 x 2 table is:

	Column 1	Column 2	Total
Row 1	A	B	R1
Raw 2	C	D	R2
Total	C1	C2	n

In this case, the Chi square statistic has the following simplified form,

$$X^2 = \frac{n\,(AD - BC)^2}{R1R2C1C2}$$

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.
- The X^2 – test does not measure the strength of association between two factors.
- It does not show causality.