STATISTICS

The Chi-Square (χ^2) Test

www.DiligentEdu.in
email: eduinfo.inbox@gmail.com

Pearson's **chi-square test** is a statistical test for categorical data. *Chi-square* is often written as χ^2 and is *pronounced* "***kai-square***". It is also called *chi-squared*.

The *chi-square* tests are among the most common **non-parametric tests**.

▪ **USES**

▪ It is used to determine whether your data are *significantly different* from what you expected.

▪ Chi Square test has a large number of applications where parametric tests cannot be applied.

- It can be used in businesses
- For various analysis in Offices
- In Scientific experiments
- In Economic Studies

▪ This is a **non-parametric test** which is being extensively used for the following reasons:

1. This test is a Distribution free method, which does not rely on assumptions that the data are drawn from a given parametric family of probability distributions.

2. This is easier to compute and simple enough to understand as compared to parametric test.

3. This test can be used in the situations where parametric tests are not appropriate or measurements prohibit the use of parametric tests.

Quick Review

▪ **Chi-square test (χ^2) is defined as:**

$$\chi^2 = \sum \frac{(O-E)^2}{E} \qquad \chi^2 = \sum \frac{(\text{observed}-\text{expected})^2}{\text{expected}}$$

Where **O** refers to the observed frequencies and **E** refers to the expected frequencies.

▪ **Draw a conclusion**

If **It.s.I < c.v.**, do not reject the *null hypothesis*.

If **It.s.I** $\geq$ **c.v.**, *reject the null hypothesis*.

Terms - Test statistic (**t.s.**); Critical value (**c.v.**) of the test.

▪ **TYPES**

(1) A Test of Independence (or a test of any sort of association / relation)

This test (chi-square - χ^2 test) is helpful in **detecting the association** between two or more attributes (detecting the association means - whether the two or more attributes are somehow related *or* not).

Suppose we have **N observations** classified according to **two attributes**. By applying this test (chi-square - χ^2 test) on the given **observations (data)** we try to **find out** whether the attributes have some association (whether they are related *or* not) OR they are independent (no relation). This association may be *positive*, *negative* or *absence of association*.

For example: we can find out

1. whether there is any association between **regularity in class** and **division of passing** of the students,
2. Similarly, we can find out whether **quinine** is effective in controlling **fever** or not.

Process of *chi-square test*

1. After **computing** the value of chi square, we **compare** the calculated value with its corresponding **critical value** for the given *degree of freedom* at a certain level of significance.
2. If calculate value of χ^2 is ***less*** **than critical or table value**, *null hypothesis* is said to be accepted and it is concluded that **two attributes have no association** that means they are **independent**.
3. On the other hand, if the calculated value of χ^2 is ***greater*** **than the critical or table value**, it means that the results of the experiment *do not support the hypothesis* and **hypothesis is rejected**, and it is concluded that the **attributes are associated**.

The Five Steps of Hypothesis Testing

Recall the question asked earlier: How large should ***chi-square*** be so that we can conclude that a statistically significant relationship exists between gender and year of promotion or, in other words, so that we can reject the null hypothesis and accept the alternate hypothesis? All statistical tests follow the same five steps of hypothesis testing:

1. State the null hypothesis (in Greek letters).

2. Choose a statistical test.

3. Calculate the test statistic (**t.s.**) and evaluate test assumptions.

4. Look up the critical value (**c.v.**) of the test.

5. Draw a conclusion:

If **|t.s.| < c.v.**, do not reject the null hypothesis.

If **|t.s.|** $\geq$ **c.v**., reject the null hypothesis.

Critical value - The critical value is the minimum value that a test statistic must be in order to rule out chance as the cause of a relationship. Technically, the critical value is the value above which the test statistic is sufficiently large to reject the null hypothesis at a user-specified level of significance.

Example 1: From the data given in the following table, find out whether there is any relationship between gender and the preference of colour.

Colour	Male	Female	Total
Red	25	45	70
Blue	45	25	70
Green	50	10	60
Total	120	80	200

(Given: For ν =2, $\chi^2_{0.05}$ = 5.991)

Solution:

Let us take the following hypothesis:

Null Hypothesis H_0 : There is no relationship between gender and preference of colour.

Alternative Hypothesis Ha : There is relationship between gender and preference of colour.

Now calculate the *chi square* value for the observed frequencies.

Colour	Gender	O	E	O-E	$(O-E)^2$	$(O-E)^2/E$
Red	M	25	42	-17	289	6.88
	F	45	28	17	289	10.32
Blue	M	45	42	3	9	0.21
	F	25	28	-3	9	0.32
Green	M	50	36	14	196	5.44
	F	10	24	-14	196	8.16
Total						χ^2 = 31.33

O – Observed values; E – Expected values

The *degrees of freedom* are *(r-1) (c-1)* = (3-1) (2-1) = 2.

c – number of columns
r -number of rows

Result / Meaning

The *critical value* (c.v.) of χ^2 for **2 degrees of freedom** at **5% level of significance** is 5.991.

Since the calculated χ^2 =31.33 exceeds the *critical value* of χ^2, the **null hypothesis is rejected**.

Hence, the **conclusion** is that there is a definite relationship between gender and preference of colour.

(2) Goodness of fit test

It is the most important utility of the Chi Square test. This method is mainly used for testing of goodness of fit. It attempts to set up whether an observed frequency distribution differs from an estimated frequency distribution. When an ideal frequency curve whether normal or some other type is fitted to the data, we are interested in finding out how well this curve fits with the observed facts.

***Example 2*:**

In an anti-malaria campaign in a certain area, ***quinine*** was administered to **812** persons out of a total population of **3248**. The number of fever cases is shown below:

Treatment	Fever (A)	No fever (a)	Total
Quinine (B)	140(AB)	30 (aB)	170 (B)
No Quinine (b)	60(Ab)	20 (ab)	80 (b)
Total	200(A)	50 (a)	250 (N)

Discuss the usefulness of quinine in checking malaria.

(Given: For ν =1, $\chi^2_{0.05}$ = 3.84)

Solution: Let us take the following hypotheses:

Null Hypothesis H_0: Quinine is not effective in checking malaria.

Alternative Hypothesis H_a: Quinine is effective in checking malaria.

Applying χ^2 test:

Expected frequency of AB = $\frac{(A)X(B)}{N} = \frac{(200)X(170)}{250} = 136$

The table of expected frequencies shall be:

Treatment	Fever	No fever	Total
Quinine	136	34	170
No Quinine	64	16	80
Total	200	50	250

Computation of Chi Square value:

O	E	O-E	$(O-E)^2$	$(O-E)^2/E$
140	136	4	16	0.11764706
60	64	-4	16	0.25
30	34	-4	16	0.47058824
20	16	4	16	1

$$\chi^2 = \sum \frac{(\text{observed}-\text{expected})^2}{\text{expected}}$$

$$\chi^2 = \Sigma \frac{(O-E)^2}{E} = 1.839$$

$\chi^2 = \Sigma \frac{(0-E)^2}{E}$ = 1.839

Degree of freedom ν = (r-1) (c-1) = (2-1) (2-1) = 1

Table Value: For ν =1, $\chi^2_{0.05}$ = 3.84

The calculated value of χ^2 i.e. 1.839 is less than the table value i.e. 3.84, the null hypothesis is accepted.

Result / Meaning

Hence quinine is not useful in checking malaria.

The Chi-Square Test - Comparing Observed to Expected Values

In **Biology** problems - *Genotype frequencies* are calculated from *allele frequencies* using the *Hardy–Weinberg equation*. These genotype frequencies represent what should be *expected* in the *sampled* population if that population is in Hardy–Weinberg equilibrium.

Then, the *observed* genotype frequencies found within that population were calculated.

Question: It can now be asked - do the *observed genotype frequencies* differ significantly from the *expected genotype frequencies*?

This question can be addressed using a statistical test called ***chi-square*** χ^2.

Chi-square is used to test whether or not some observed distributional outcome fits an expected pattern. Since it is unlikely that the observed genotype frequencies will be exactly as predicted by the Hardy–Weinberg equation, it is important to look at the nature of the differences between the observed and expected values and to make a judgment as to the "***goodness of fit***" between them.

In the chi-square test, the expected value is subtracted from the observed value in each category, and this value is then squared. Each squared value is then weighted by dividing it by the expected value for that category. The sum of these squared and weighted values, called chi-square (denoted as χ^2), is represented by the following equation:

$$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

▪ Null Hypothesis (H_0) and Alternative Hypothesis (H_a)

In the chi-square test, two hypotheses are tested.

▪ **Null Hypothesis (H_0)** - The null hypothesis (**H_0**) states that there is ***no difference between the two observed and expected values***; they are statistically the same and any difference that may be detected is due to chance.

▪ **Alternative Hypothesis (H_a)** - The alternative hypothesis **(H_a)** states that the two sets of data, the observed and expected values, ***are different***; the difference is statistically significant and must be due to some reason other than chance.

In populations, a *normal distribution* of values around a consensus frequency should be represented by a *bell-shaped curve*. The degree of distribution represented by a curve is associated with a value termed the **degrees of freedom (df)**.

The smaller the **df** value (*the smaller the sample size*), the larger the dispersion in the distribution, and a bell-shaped curve will be more difficult to distinguish.

The larger the **df** value (*the larger the sample size*), the closer the distribution will come to a normal, bell-shaped curve.

For most problems, the **degrees of freedom** is equal to ***one less than the number of categories*** in the distribution. For such problems, the number of categories is equal to the number of different alleles (or the number of different genotypes, depending upon which is the object of the test). Statistically significant differences may be observed in chi-square if the sample size is small (less than 100 individuals sampled from a population) or if there are drastic deviations from Hardy–Weinberg equilibrium.

How chi-square test is performed - level of significance

A chi-square test is performed at a certain ***level of significance***, usually **5%** (α = 0.05; p = .95). At a 5% significance level, we are saying that there is less than a 5% chance that the null hypothesis will be rejected even though it is true. Conversely, there is a **95%** chance that the null hypothesis is correct, that there is no difference between the observed and expected values.

A *chi-square distribution table* will provide the expected chi-square value for a given probability and **df** value. For example, suppose a chi-square value of 2.3 is obtained for some data set having 8 degrees of freedom and we wish to test at a 5% significance level. From the chi-square distribution table, we find a value of 15.51 for a probability value of 0.95 and 8 degrees of freedom. Therefore, there is a probability of 0.95 (95%) that the chi-square will be less than 15.51. Since the chi-square value of 2.3 is less than 15.51, the *null hypothesis is not rejected*; the data do not provide sufficient evidence to conclude that the observed data differ from the expected.

At 8 degrees of freedom, for example, there is a probability of.95 (95%) that the chi-square value will be less than 15.51. If the chi-square value you obtain is less than 15.51, then there is insufficient evidence to conclude that the observed values differ from the expected values. Below 15.51, any differences between the observed and expected values are merely due to chance.

The chi-square test can be used to compare observed genotype frequencies with those expected from **Hardy–Weinberg** analysis and to compare local allele frequencies (observed) with those maintained in a larger or national database (expected values).

Example 3 – Suppose, observed and expected genotype frequencies were calculated for a fictitious population. These are shown in the following table. Using the *chi-square test* at a 5% level of significance (95% probability), determine if the observed genotype frequencies are significantly different from the expected genotype frequencies.

Genotype	Observed frequency	Expected frequency
AA	0.03	0.053
AB	0.17	0.136
AC	0.23	0.218
BB	0.06	0.087
BC	0.30	0.280
CC	0.21	0.226

Solution 3

A table is prepared so that the steps involved in obtaining the chi-square value can be followed. So that we are not dealing with small decimal numbers, the frequencies given in the table will each be multiplied by 100 to give a percent value.

Genotype	Observed frequency(O)	Expected frequency(E)	Difference ($O-E$)	Square of difference $(O-E)^2$	$(O-E)^2/E$
AA	3	5.3	−2.3	5.3	1.00
AB	17	13.6	3.4	11.6	0.85
AC	23	21.8	1.2	1.4	0.06
BB	6	8.7	−2.7	7.3	0.84
BC	30	28	2.0	4	0.14
CC	21	22.6	−1.6	2.6	0.12
Total					χ^2=3.01

Therefore, a *chi-square value* χ^2 of 3.01 is obtained. The degrees of freedom is equal to the number of categories (6, because there are six genotypes) minus 1:

$$df = 6 - 1 = 5$$

At 5 degrees of freedom and 0.95 probability, a ***chi-square value*** of 11.07 is obtained from the *chi-square distribution table*. Therefore, there is a probability of 95% that χ^2 will be less than 11.07 if the null hypothesis is true, if there is not a statistically significant difference between the observed and expected genotype frequencies.

Result / Meaning

Since our derived chi-square value χ^2 of 3.01 is less than 11.07, we do not reject the null hypothesis; there is insufficient evidence to conclude that the observed genotype frequencies differ from the expected genotype frequencies. This suggests that our population may be in Hardy–Weinberg equilibrium.

++++

Important Points to Remember

▪ **Statistics** means, the science of collecting, analyzing, presenting, and interpreting data.

▪ **Data** are the facts and figures that are collected, analyzed, and summarized for presentation and interpretation.

▪ **Quantitative / Qualitative Data** - Data may be classified as either **quantitative** or **qualitative**.

Quantitative data measure either how much or how many of something, and Qualitative data provide labels, or names, for categories of like items.

Quantitative data - how much? OR how many?
Qualitative data provide - labels, or names, for categories of like items.

For example,

- **Age** – 28 years; **Annual income** - $30,000 are **quantitative** variables;

- **Gender** (male / female); **Marital status** (single, married, divorced, and widowed) are **qualitative** variables.

▪ **Types of studies** - observational studies, experimental studies.

▪ **Observational studies** - **Sample survey** methods are used to collect data from observational studies, and

▪ **Experimental studies** - experimental design methods are used to collect data from experimental studies.

The area of *descriptive statistics* is concerned primarily with methods of presenting and interpreting data using graphs, tables, and numerical summaries. Whenever statisticians use data from a **sample**—i.e., a subset of the population—to make statements about a population, they are performing statistical ***inference***.

Estimation and **hypothesis testing** are procedures used to make statistical inferences. Fields such as health care, biology, chemistry, physics, education, engineering, business, and economics make extensive use of statistical inference.

inference means- a conclusion or opinion that is formed because of known facts or evidence.

▪ **Methods of statistical inference**

- some of these methods are used primarily for **single-variable** studies,
- others, such as **regression** and **correlation** analysis, are used to make inferences about relationships among **two or more variables**.

▪ **Continuous Data Type and Categorical Data Type**

▪ **Continuous Data Type** – Continuous data types are ones that are infinite numerical value between any two values. For example, salary, time.

▪ **Categorical Data Type** – Categorical data types are ones that contain a finite set of distinct categories or groups. For example, gender, marital status.

Source / References

1. https://www.britannica.com/science/statistics
2. https://www.mygreatlearning.com/blog/chi-square-test-explained/
3. https://www.sciencedirect.com/topics/medicine-and-dentistry/chi-square-test
4. https://www.scribbr.com/statistics/chi-square-tests
5. Chi Square - Test by Dr. Anoop Kumar Singh, Dept. of Applied Economics, University of Lucknow