Chi – Square Test (χ^2)

Srh University

Abhishek Negi1, Aakash Vashist2, and Praful3

Msc. - Computer Science (Cohort - 2025)

Introduction:

A Chi-Square (χ^2) Test is a non-parametric statistical test used to evaluate whether observed categorical data differ significantly from expected frequencies under a specific hypothesis. It measures the discrepancy between observed counts (O) and expected counts (E) by summing the squared differences divided by the expected counts.

Or

A chi-squared test (also chi-square or χ^2 test) is a statistical hypothesis test used in the analysis of contingency tables when the sample sizes are large. In simpler terms, this test is primarily used to examine whether two categorical variables (*two dimensions of the contingency table*) are independent in influencing the test statistic (*values within the table*).

Two Main Types of Chi-Squared Test (χ^2):

The Chi-Square (χ^2) test is widely used for analyzing categorical data, and it can be applied in two main ways: the Chi-Square Goodness-of-Fit Test and the Chi-Square Test of Independence. Below is a detailed explanation of each type, including their purposes, calculations, and examples.

1. **Chi-Square Goodness-of-Fit Test**

1.1. Purpose

The **Chi-Square Goodness-of-Fit Test** is used to determine if the distribution of a single categorical variable follows a specific, expected distribution. This test compares the observed frequencies of categories with the frequencies that would be expected if the null hypothesis is true.

1.2. Hypotheses

- **Null Hypothesis (H_0):** The observed data follow the expected distribution.
- **Alternative Hypothesis (H_1):** The observed data do not follow the expected distribution.

1.3. Calculation:

$$\chi^2 = \sum \frac{(f_O - f_E)^2}{f_E}$$

f_O = observed frequencies

f_E = expected frequencies

Fig1. Formula to calculate chi-square

1.4. Decision Rule:

- If the χ^2 statistic is greater than the critical value from the χ^2 distribution table (based on the desired significance level, usually $\alpha = 0.05$), **reject the null hypothesis**.

- If the χ^2 statistic is less than the critical value, **fail to reject the null hypothesis**, meaning there is no significant difference between the observed and expected frequencies.

2. **Chi-Square Test of Independence:**

2.1. Purpose:

The **Chi-Square Test of Independence** is used to determine whether there is a significant association or relationship between two categorical variables. In other words, it tests whether the occurrence of one variable is independent of the occurrence of another variable.

2.2. Hypotheses:

- **Null Hypothesis (H_0):** The two categorical variables are independent (no association).
- **Alternative Hypothesis (H_1):** The two categorical variables are dependent (there is an association).

2.3. Calculation:

$$E(r, c) = \frac{n(r) \times c(r)}{n}$$

where:

$r = \text{row in question}$

$c = \text{column in question}$

$n = \text{corresponding total}$

Fig2. Formula to calculate Independence of variables

2.4. Decision Rule:

- If the χ^2 statistic is greater than the critical value from the χ^2 distribution table (based on the desired significance level, usually $\alpha = 0.05$), **reject the null hypothesis**.

- If the χ^2 statistic is less than the critical value, **fail to reject the null hypothesis**, meaning there is no significant difference between the observed and expected frequencies.

Example:

Data: Let us assume we are giving a multiple-choice exam, and the exam department assures us that all the options are equally distributed, i.e. they have an equal probability, that means probability of A = B = C = D = 25%. Now assuming the exam has 100 questions, and we have taken a sample, and we get some observed values. The table for this is given below:

Correct choice	Expected value	Actual results
A	25	20
B	25	20
C	25	25
D	25	35

Hypothesis:

H0 (null): Equal distribution of correct choices

HA (alternative): Not equal distribution of correct choices

Method:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

χ^2 = chi squared

O_i = observed value

E_i = expected value

Fig3. Formula to calculate chi-square

α= significance level

Let us assume α=0.05

Using the formula given above

$$\chi^2 = 6$$

Degree of freedom (df) = Total choices -1

So, in this case df = 3

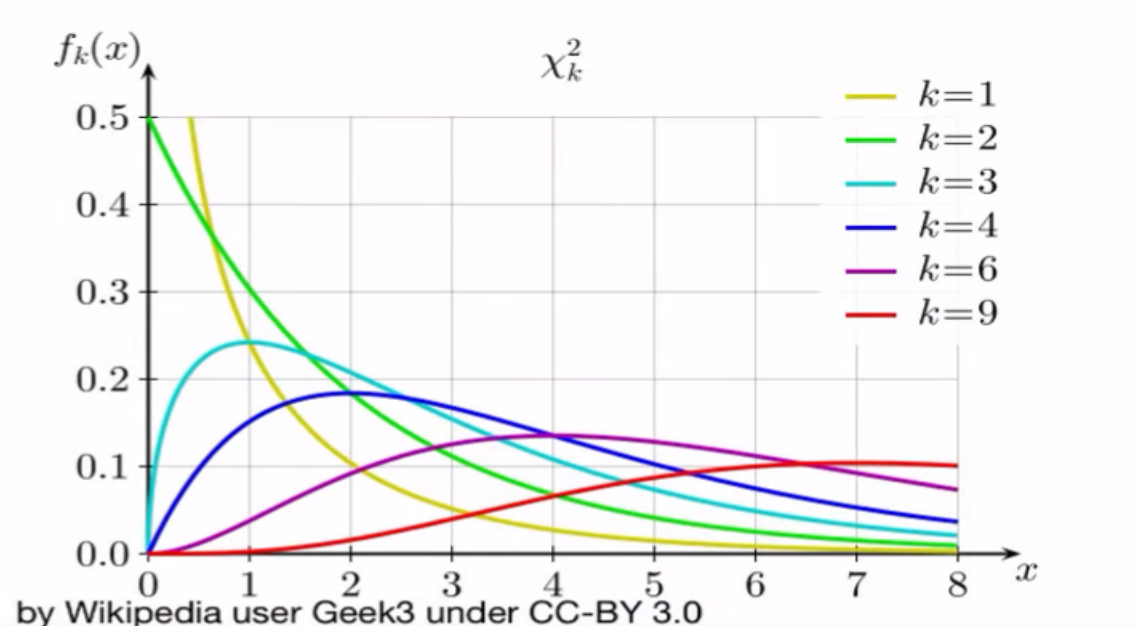

Fig4. Chi-Square distribution for different degrees of freedom

Or

	P										
DF	0.995	0.975	0.2	0.1	0.05	0.025	0.02	0.01	0.005	0.002	0.001
1	.0004	.00016	1.642	2.706	3.841	5.024	5.412	6.635	7.879	9.55	10.828
2	0.01	0.0506	3.219	4.605	5.991	7.378	7.824	9.21	10.597	12.429	13.816
3	0.0717	0.216	4.642	6.251	7.815	9.348	9.837	11.345	12.838	14.796	16.266
4	0.207	0.484	5.989	7.779	9.488	11.143	11.668	13.277	14.86	16.924	18.467
5	0.412	0.831	7.289	9.236	11.07	12.833	13.388	15.086	16.75	18.907	20.515

Fig5. Degree of Freedom vs p-value or chi-square table

Therefore, Probability of getting $\chi^2 = 6$

P ($\chi^2 = 6$) > 10%

This is also known as p value.

P value: The p value, or probability value, tells you how likely it is that your data could have occurred under the null hypothesis

Conclusion:

Since our p-value>0.1 or 10% > α=0.05, we cannot say for certainty that our initial hypothesis was incorrect.

Hence, we fail to reject the H0 (null) hypothesis.

References:

- **Wikipedia contributors. (n.d.). *Chi-squared test*. In *Wikipedia, The Free Encyclopedia*. Retrieved April 24, 2025, from https://en.wikipedia.org/wiki/Chi-squared_test**

- **Khan Academy. (n.d.). Chi-square statistics for hypothesis testing | AP Statistics. YouTube. https://www.youtube.com/watch?v=jABsbNBPXIk**

- **SAS Institute Inc. (n.d.). *Chi-square test*. JMP® Statistical Knowledge Portal. Retrieved April 24, 2025, from https://www.jmp.com/en/statistics-knowledge-portal/chi-square-test**

- **SAS Institute Inc. (n.d.). *Chi-square test*. JMP® Statistical Knowledge Portal. Retrieved April 24, 2025, from https://www.jmp.com/en/statistics-knowledge-portal/chi-square-test**

- **ChatGPT. (n.d.). *Conversation 6809f007-02dc-800f-8a53-7ab73a0c9207*. Retrieved April 24, 2025, from https://chatgpt.com/c/6809f007-02dc-800f-8a53-7ab73a0c9207**