**Chi – Square Test (χ²)** Srh University

**Abhishek Negi1, Aakash Vashist2, and Praful3** 

Msc. - Computer Science (Cohort - 2025)

**Introduction:**

A Chi‑Square (χ²) 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 χ2 test) is a [statistical hypothesis test](https://en.wikipedia.org/wiki/Statistical_hypothesis_testing) used in the analysis of [contingency tables](https://en.wikipedia.org/wiki/Contingency_table) 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 (χ²):**

The Chi-Square (χ²) 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 (H₀):** The observed data follow the expected distribution.
* **Alternative Hypothesis (H₁):** The observed data do not follow the expected distribution.

**1.3. Calculation:**

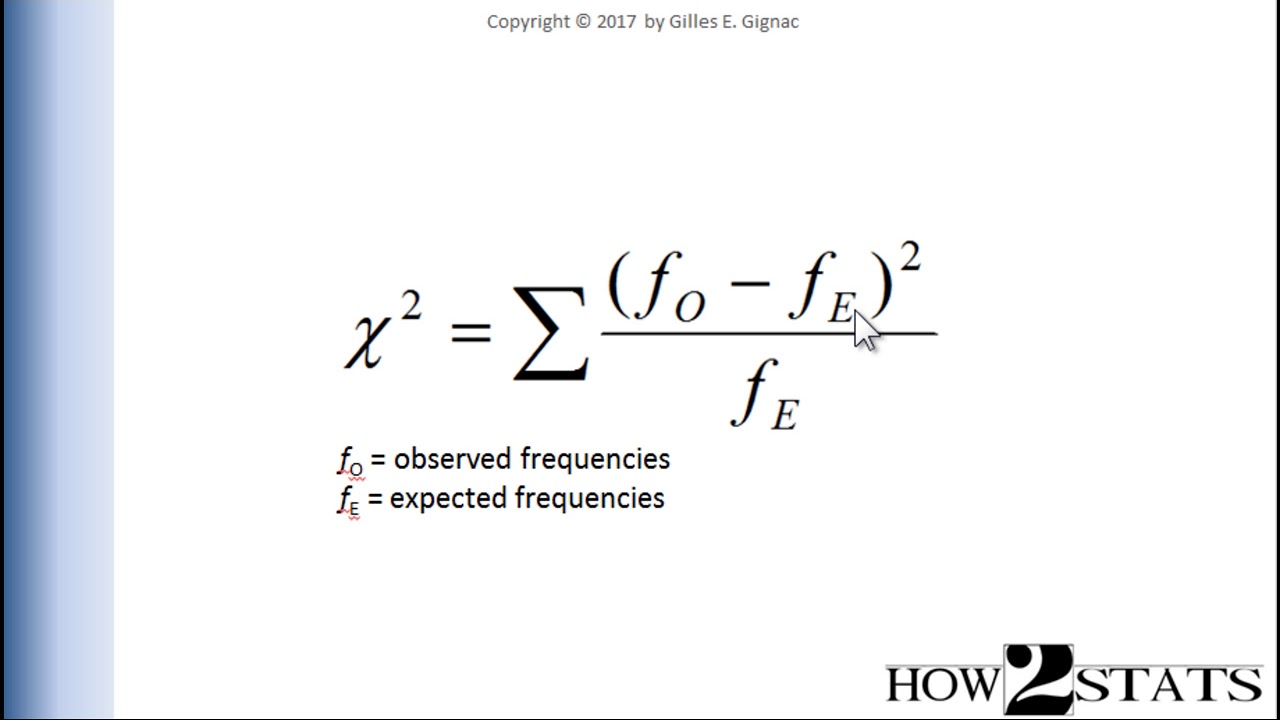


Fig1. Formula to calculate chi-square

**1.4. Decision Rule:**

* If the χ² statistic is greater than the critical value from the χ² distribution table (based on the desired significance level, usually α = 0.05), **reject the null hypothesis**.
* If the χ² 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.

1. **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 (H₀):** The two categorical variables are independent (no association).
* **Alternative Hypothesis (H₁):** The two categorical variables are dependent (there is an association).

**2.3. Calculation:**

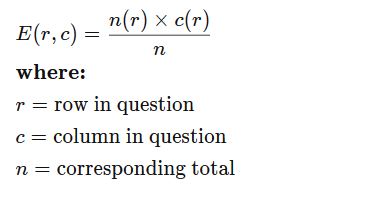


Fig2. Formula to calculate Independence of variables

**2.4. Decision Rule:**

* If the χ² statistic is greater than the critical value from the χ² distribution table (based on the desired significance level, usually α = 0.05), **reject the null hypothesis**.
* If the χ² 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:**

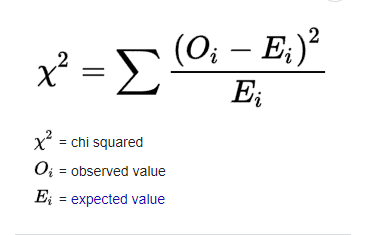


Fig3. Formula to calculate chi-square

α= significance level

Let us assume α=0.05

Using the formula given above

χ² = 6

Degree of freedom (df) = Total choices -1

So, in this case df = 3

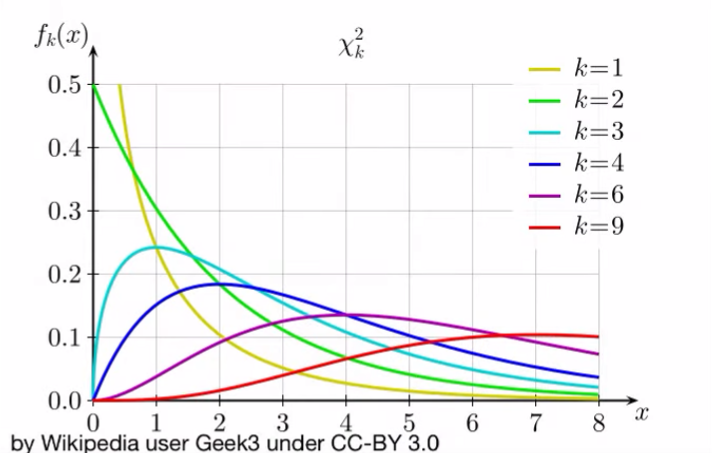


Fig4. Chi-Square distribution for different degrees of freedom

Or



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

Therefore, Probability of getting χ² = 6

P (χ² = 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**](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**](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**](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**](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**](https://chatgpt.com/c/6809f007-02dc-800f-8a53-7ab73a0c9207)