

What is Chi-Square Test of Independence?

Authored by
stats writer

December 26, 2025

RECOMMENDED CITATION

stats writer (2025). *What is Chi-Square Test of Independence?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=109094>

The Chi-Square Test of Independence is a fundamental non-parametric statistical tool employed to analyze relationships within data. Its primary function is to ascertain whether there is a statistically significant association between two categorical variables. In essence, this test helps researchers determine if the distribution of one variable is contingent upon the distribution of the other.

This powerful technique operates by comparing the actual counts observed in a sample (the observed frequencies) against the counts that would be expected if the two variables were truly independent (the expected frequencies). A significant discrepancy between these observed and expected values suggests that the variables are indeed related. Understanding this test is crucial for fields ranging from social sciences and market research to biology and epidemiology, where quantifying relationships between qualitative data types is essential.

The **Chi-Square Test of Independence** is specifically designed to assess whether two categorical variables display a significant statistical association. If an association exists, the variables are deemed dependent; otherwise, they are considered independent.

Outline of This Tutorial

This comprehensive tutorial will guide you through the process of conducting a Chi-Square Test of Independence. We will cover the core theoretical motivations, the mathematical formula, and a detailed practical example.

Explore the motivation and real-world scenarios requiring a Chi-Square Test of Independence.

Detail the fundamental formula used to calculate the Chi-Square test statistic.

Provide a step-by-step, practical example demonstrating how to execute the test and interpret its results.

Understanding the Need for the Chi-Square Test

The Chi-Square Test of Independence is indispensable when researchers are dealing exclusively with categorical variables--variables that take on fixed, defined values, often classifications or labels, rather than numerical measurements. This test allows us to move beyond simple descriptive statistics and formally evaluate whether observed patterns are likely due to chance or if they represent a genuine underlying relationship between the categories. If we find that the pattern of responses for one category level differs significantly based on the category level of the second variable, we conclude they are associated.

Consider these diverse scenarios where quantifying the association between categorical data is paramount for drawing valid conclusions:

Political Science Research: We aim to investigate if an individual's **gender** is associated with

their declared **political party preference**. We survey 500 eligible voters, classifying them by two nominal variables.

Market Research/Psychology: We wish to determine if a person's preferred **color** (e.g., Red, Blue, Green) is linked to their favorite **sport** (e.g., Soccer, Basketball, Tennis). We poll 100 individuals regarding their preferences for both categories.

Sociology/Demographics: We are testing whether **education level** (e.g., High School, Bachelor's, Graduate) and **marital status** (e.g., Single, Married, Divorced) are associated within a population. Data is collected from a simple random sample of 50 people.

In all these examples, the primary goal is hypothesis testing: determining if the observed relationship is strong enough to statistically reject the notion of independence. The Chi-Square test provides the statistical rigor necessary to make this determination, effectively quantifying the discrepancy between what we see in our data and what we would expect under conditions of zero relationship.

Foundations of the Test: Hypotheses and Assumptions

Before calculating the test statistic, it is crucial to formally define the statistical hypotheses that the Chi-Square Test of Independence is designed to evaluate. Like all hypothesis tests, it is anchored by a null hypothesis and an alternative hypothesis that must be mutually exclusive and exhaustive.

H0: (The Null Hypothesis) The two variables under investigation are statistically **independent**. This means that knowing the category of one variable provides no useful information about the likely category of the second variable.

H1: (The Alternative Hypothesis) The two variables are **not independent**. This implies they are statistically **associated**.

For the test results to be reliable, several assumptions must be met, primarily concerning sampling and expected frequencies. The most important assumption is that the data must be collected via a simple random sample from the population. Furthermore, the test is generally considered reliable only if the expected frequency (E) for at least 80% of the cells in the contingency table is greater than 5, and no cell has an expected frequency less than 1.

Calculating the Chi-Square Test Statistic

The core of the Chi-Square test involves calculating the test statistic, denoted as X^2 . This statistic measures the cumulative difference between the observed counts (O) from the survey data and the expected counts (E) calculated under the assumption that the null hypothesis (H0) is true. If the observed data closely matches the expected data (if the variables are truly independent), the X^2 value will be small. Conversely, a large X^2 value indicates a significant difference, suggesting the variables are associated.

The formula used to calculate the Chi-Square test statistic X^2 is:

$$X^2 = \sum(O - E)^2 / E$$

Where the components are defined as follows:

Σ : This is the summation operator, indicating that we must sum the results of the calculation across all cells (categories) within the contingency table.

O: Represents the **observed value**, which is the actual count recorded for a specific cell from the survey or experiment.

E: Represents the **expected value**, which is the count we would anticipate for that specific cell if the two variables were perfectly independent.

The term $(O - E)^2$ ensures that positive and negative differences contribute equally to the statistic, and dividing by E standardizes the difference relative to the size of the expected count, preventing cells with large counts from dominating the total X^2 value unnecessarily.

Degrees of Freedom and P-Value Interpretation

The calculated test statistic, X^2 , must be compared to a critical value derived from the Chi-Square distribution to determine the p-value. This comparison requires knowing the appropriate degrees of freedom (df) for the test. The degrees of freedom represent the number of values in the final calculation of a statistic that are free to vary.

For the Chi-Square Test of Independence, the degrees of freedom are calculated based on the dimensions of the contingency table:

$$df = (\#rows - 1) * (\#columns - 1)$$

Once X^2 and the degrees of freedom are known, we calculate the p-value. The p-value is the probability of observing a test statistic as extreme as, or more extreme than, the one calculated, assuming that the null hypothesis is true. If this p-value is less than your predetermined significance level (alpha, usually 0.05), you have sufficient statistical evidence to reject the null hypothesis, thereby concluding that the variables are associated.

Practical Application: A Detailed Example

To illustrate the methodology, let us examine a practical example focusing on the association between gender and political party preference. We assume a scenario where we are testing whether or not these two categorical variables are dependent. We utilized a simple random sample of 500 registered voters and collected data on their gender and stated political preference. The raw observed results of this survey are presented in the contingency table below:

	Republican	Democrat	Independent	Total
Male	120	90	40	250
Female	110	95	45	250
Total	230	185	85	500

We will now use a sequence of five defined steps to meticulously perform the Chi-Square Test of Independence and draw a statistically supported conclusion regarding the association between gender and political party preference in this sample.

Step 1: Defining the Hypotheses

The first formal step is to clearly establish the statistical framework by stating the null hypothesis and alternative hypotheses based on our research question:

H0: Gender and political party preference are statistically **independent**. (The distributions across parties are the same for males and females.)

H1: Gender and political party preference are **not independent** (they are associated). (The distribution across parties differs by gender.)

We set our significance level, alpha (α), typically at 0.05. If the resulting p-value is less than 0.05, we will reject H0.

Step 2: Calculating the Expected Values (E)

The next critical step involves calculating the expected frequency for every cell in the contingency table, assuming that the null hypothesis of independence is true. This value represents the count we would expect purely by chance, given the marginal totals (row and column sums).

The expected value (E) for any cell is calculated using the formula:

$$\text{Expected value} = (\text{Row Sum} \times \text{Column Sum}) / \text{Grand Total (Table Sum)}$$

For instance, to find the expected count for **Male Republicans**, we use the total number of Republicans (230), the total number of Males (250), and the total sample size (500): $(230 \times 250) / 500 = 115$. We repeat this calculation for all six cells to derive the full expected count table:

	Republican	Democrat	Independent	Total
Male	115	92.5	42.5	250
Female	115	92.5	42.5	250

Total	230	185	85	500
--------------	-----	-----	----	-----

We confirm that all expected values are greater than 5, thus satisfying the necessary assumptions for the test.

Step 3: Calculating the Chi-Square Components $(O-E)^2 / E$

Next we will calculate $(O - E)^2 / E$ for each cell in the table, where:

O: observed value

E: expected value

For example, **Male Republicans** would have a value of: $(120 - 115)^2 / 115 \approx \mathbf{0.2174}$.

Repeating this formula for each cell in the table yields the following contribution results:

	Republican	Democrat	Independent
Male	0.2174	0.0676	0.1471
Female	0.2174	0.0676	0.1471

Step 4: Calculating the Test Statistic (X^2) and P-Value

The total Chi-Square test statistic (X^2) is the sum of all the individual cell contributions calculated in Step 3.

$$X^2 = \sum(O - E)^2 / E = 0.2174 + 0.2174 + 0.0676 + 0.0676 + 0.1471 + 0.1471 = \mathbf{0.8642}$$

We determine the degrees of freedom (df) as: $(2 \text{ rows} - 1) \times (3 \text{ columns} - 1) = \mathbf{2 \text{ degrees of freedom}}$.

According to the Chi-Square Score to P Value Calculator, the p-value associated with $X^2 = 0.8642$ and $df = 2$ is **0.649198**.

Step 5: Drawing the Conclusion

The final step involves comparing the calculated p-value against our chosen significance level ($\alpha = 0.05$).

Since P-value (0.649198) is greater than the Significance Level (0.05), we **fail to reject the null hypothesis**.

This means we do not have sufficient evidence to conclude that there is a statistically significant

association between gender and political party preference in the population. The observed differences are likely due to random chance.

Note: You can also perform this entire test by simply using the [Chi-Square Test of Independence Calculator](#).

Resources for Further Implementation

The methodology for the Chi-Square Test of Independence remains consistent, but its implementation varies depending on the computational platform used. The following tutorials offer guidance on performing this test using various statistical programs and calculators:

[How to Perform a Chi-Square Test of Independence in Stata](#)

[How to Perform a Chi-Square Test of Independence in Excel](#)

[How to Perform a Chi-Square Test of Independence in SPSS](#)

[How to Perform a Chi-Square Test of Independence in Python](#)

[How to Perform a Chi-Square Test of Independence in R](#)

[Chi-Square Test of Independence on a TI-84 Calculator](#)

[Chi-Square Test of Independence Calculator](#)