

How to Perform a Chi-Square Test of Independence in R (With Examples)

Authored by
stats writer

December 19, 2025

RECOMMENDED CITATION

stats writer (2025). *How to Perform a Chi-Square Test of Independence in R (With Examples)*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=108002>

The ability to analyze relationships between different characteristics is fundamental to statistical research. Among the various tools available, the Chi-Square Test of Independence stands out as the standard method for examining associations between two categorical variables. This powerful inferential test helps researchers determine whether the distribution of one variable is dependent upon the distribution of the other.

The primary goal of this test is to assess the null hypothesis, which states that the two variables are statistically independent. If the test reveals a significant result (typically indicated by a small p-value), we reject the null hypothesis and conclude that an association exists. This tutorial provides an in-depth guide on how to implement and interpret this crucial statistical procedure specifically within the R statistical environment, utilizing the native `chisq.test()` function.

We will demonstrate the complete workflow using a practical, real-world inspired scenario: investigating whether a voter's **gender** is associated with their **political party preference**. By the end of this guide, you will be equipped to structure your data, execute the necessary R commands, and confidently interpret the resulting statistical output.

1. Introduction to the Chi-Square Test of Independence

The Chi-Square Test of Independence is specifically designed for analyzing data presented in a contingency table, which cross-classifies observations based on two nominal or ordinal variables. It calculates the discrepancy between the observed frequencies in the table and the frequencies that would be expected if the variables were truly independent. The resulting test statistic, known as the Chi-square statistic (χ^2), quantifies the magnitude of this difference.

Understanding the underlying hypotheses is essential before performing the test. We operate under a defined set of competing statements regarding the relationship between our two variables, A and B . These hypotheses guide our statistical decision-making process.

H0 (Null Hypothesis): The two variables (e.g., Gender and Political Preference) are statistically independent. Knowing the category of one variable provides no information about the category of the other.

H1 (Alternative Hypothesis): The two variables are **not** independent; there is a significant association or relationship between them.

The core objective is to calculate the test statistic and use the corresponding p-value to determine whether the observed data provides strong enough evidence to reject H_0 . If the probability of observing our data (or more extreme data) under the assumption that H_0 is true is very low (typically $p < 0.05$), we conclude that the relationship is significant.

2. Prerequisites and Assumptions for Valid Testing

While the Chi-Square Test is robust, its validity depends on satisfying several critical statistical assumptions. Ignoring these assumptions can lead to unreliable results and flawed conclusions. The primary assumptions relate to the data structure and the expected counts within the contingency table.

First, the data must consist of a **simple random sample** from the population of interest. Each observation must be independent of all others. Furthermore, the variables being analyzed must be categorical variables, meaning they classify individuals into distinct groups (e.g., Male/Female, Republican/Democrat/Independent). The test is not appropriate for continuous or interval data unless it has been appropriately categorized.

Secondly, and most importantly, the assumption of **expected frequency counts** must be met. This assumption states that, for the test statistic to follow the Chi-Square distribution reliably, the expected count in each cell of the contingency table should be at least 5. If more than 20% of the cells have an expected count less than 5, or if any cell has an expected count less than 1, the standard Chi-Square Test of Independence may be inappropriate, and alternatives like Fisher's Exact Test should be considered.

3. Understanding the `chisq.test()` function in R

The R environment provides a straightforward way to execute this analysis using the built-in function `chisq.test()`. This function is highly versatile and capable of accepting various forms of input, most commonly a two-way contingency table, matrix, or array.

When you supply the contingency table to `chisq.test()`, the function automatically performs several key calculations. It calculates the necessary expected frequencies based on the assumption of independence (H_0), computes the χ^2 statistic, determines the appropriate degrees of freedom (df), and finally derives the associated p-value.

The basic syntax is simply `chisq.test(x)`, where `x` is the contingency table. Crucially, the function returns a list object containing all these critical statistical outputs, allowing for detailed examination of the results beyond just the p-value, including access to the expected counts via `$expected` and the observed statistic via `$statistic`.

4. Setting Up the Example: Gender and Political Preference

To provide a clear, practical demonstration of the Chi-Square Test of Independence in R, we will analyze a hypothetical dataset. Our research question is: **Is there a significant association between a voter's gender and their preferred political party?**

Suppose a simple random sample of 500 eligible voters was surveyed, and their responses were cross-tabulated based on two categorical variables: Gender (Male, Female) and Political Affiliation (Republican, Democrat, Independent). The resulting observed frequencies are summarized in the contingency table below. Note that this table represents the raw data that we must input into R for analysis.

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

The goal is now to utilize R to formally test the Null Hypothesis (H_0): Gender and Political Preference are independent. We will proceed through the necessary steps to structure this data appropriately within the R environment and run the core statistical test.

5. Step 1: Data Preparation and Matrix Creation

The first crucial step in performing the Chi-Square Test of Independence in R is transforming the raw counts from the contingency table into a suitable object format, typically a matrix or a table object. This ensures that the `chisq.test()` function correctly interprets the data structure as a cross-classification.

We use the `matrix()` function to input the cell counts. It is standard practice to enter the frequencies either row by row or column by column, defining the number of columns (`ncol`) and whether the data should be filled by row (`byrow=TRUE`). For our example, we input the male counts (120, 90, 40) followed by the female counts (110, 95, 45).

After creating the numerical matrix, we apply meaningful labels using the `colnames()` and `rownames()` functions, corresponding to the political parties and gender categories, respectively. Finally, converting the matrix to a `table` object using `as.table()` is good practice, though not strictly required for `chisq.test()`, as it ensures proper display and handling of categorical data.

The following code block demonstrates this preparation process, resulting in the final two-way table object named `data`:

```
# Create the contingency matrix, inputting counts row by row
data <- matrix(c(120, 90, 40, 110, 95, 45), ncol=3, byrow=TRUE)
```

```
# Assign column names (Political Parties)
colnames(data) <- c("Rep", "Dem", "Ind")
```

```
# Assign row names (Gender)
rownames(data) <- c("Male", "Female")

# Convert to a formal table object
data <- as.table(data)

# View the resulting table structure
data

Rep Dem Ind
Male 120 90 40
Female 110 95 45
```

6. Step 2: Executing the Chi-Square Test in R

Once the data has been correctly formatted as a contingency table named `data`, executing the statistical test is remarkably simple, thanks to the integrated nature of the R environment. We simply pass the table object to the `chisq.test()` function.

This single command triggers the internal calculation of the expected counts, the sum of squared differences, the degrees of freedom, and finally, the exact p-value based on the Chi-Square distribution.

The output provides the essential components needed for hypothesis testing. We receive the test type ("Pearson's Chi-squared test"), the dataset used, the calculated Chi-square statistic (X^2), the degrees of freedom (df), and the ultimate measure of significance, the p-value.

```
# Perform Chi-Square Test of Independence on the 'data' table
chisq.test(data)
```

```
Pearson's Chi-squared test
```

```
data: data
```

```
X-squared = 0.86404, df = 2, p-value = 0.6492
```

7. Interpreting the Results and Drawing Conclusions

The statistical output provides three crucial metrics that determine our conclusion regarding the relationship between gender and political preference. These metrics must be interpreted relative to the defined null hypothesis (H_0 : The variables are independent).

Chi-Square Test Statistic (χ^2): 0.86404. This value quantifies the total difference between the observed data and the data expected under independence. Larger values indicate greater evidence against the null hypothesis.

Degrees of Freedom (df): 2. The degrees of freedom are determined by the dimensions of the table: $(R-1) \times (C-1)$, where R is the number of rows (2 genders) and C is the number of columns (3 parties). Thus, $(2-1) \times (3-1) = 1 \times 2 = 2$.

P-Value: 0.6492. This is the probability of observing a Chi-square statistic of 0.86404 (or something more extreme) if the null hypothesis of independence were actually true.

The final step involves comparing the calculated p-value to a predetermined level of significance (α), typically set at 0.05. The decision rule is simple: If $p \leq \alpha$, we reject H_0 ; if $p > \alpha$, we fail to reject H_0 .

In our analysis, the calculated p-value is 0.6492. Since $0.6492 > 0.05$, we **fail to reject the null hypothesis**. Statistically, this means that the observed differences in political preferences between male and female respondents are likely due to random sampling variability, not a true association in the underlying population. We conclude that there is insufficient evidence to claim a significant association between gender and political party preference based on this sample data.

8. Further Exploration and Related Resources

While the standard Chi-Square Test of Independence provides a binary conclusion regarding independence, further analysis can be conducted to explore the specific nature of the relationship, should one be found (i.e., if H_0 was rejected). Techniques such as residual analysis or post-hoc comparisons can help pinpoint which specific cells in the contingency table contribute most significantly to the overall Chi-square statistic.

For researchers working extensively with categorical variables, R offers numerous related functions and packages. For instance, if the sample size is very small, leading to low expected counts, the `chisq.test()` function automatically defaults to using a correction (Yates' continuity correction for 2x2 tables) or requires the use of `fisher.test()` for exact probability calculation.

Mastering the application of the Chi-Square Test of Independence in R is a foundational skill for data analysis, providing a reliable method for assessing fundamental dependencies in data structures derived from surveys and observational studies.

[An Introduction to the Chi-Square Test of Independence](#)

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

[How to Calculate the P-Value of a Chi-Square Statistic in R](#)

[How to Find the Chi-Square Critical Value in R](#)