

# Chi-Square Test of Independence in SPSS: A Step-by-Step Guide

Authored by  
**Mohammed looti**

November 8, 2025

## RECOMMENDED CITATION

Mohammed looti (2025). *Chi-Square Test of Independence in SPSS: A Step-by-Step Guide*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=12878>

The [Chi-Square Test of Independence](#) is a fundamental non-parametric statistical technique utilized to determine whether a statistically significant association exists between two [categorical variables](#). This test relies on comparing the observed frequencies in a contingency table with the frequencies that would be theoretically expected if the two variables were truly independent within the population. If the discrepancy between the observed and expected values is sufficiently large, we reject the assumption of independence, thereby concluding that an association or relationship exists.

Understanding when and how to apply this test is crucial for anyone engaging in statistical analysis, particularly within the social sciences, market research, or healthcare. This comprehensive tutorial provides a step-by-step guide on how to accurately perform the [Chi-Square Test of Independence](#) using the widely adopted statistical software package, [SPSS](#) (Statistical Package for the Social Sciences). We will walk through data preparation, execution of the test, and the subsequent interpretation of the resulting statistical output tables to draw robust conclusions.

## Understanding the Chi-Square Test of Independence

The primary goal of the [Chi-Square Test of Independence](#) is to test the independence of the classification criteria. When dealing with data summarized into frequency counts across two dimensions, this test provides the mechanism to evaluate the strength of evidence against the [null hypothesis](#). The [null hypothesis](#) ( $H_0$ ) always states that the two variables are independent (i.e., there is no association between them), while the alternative hypothesis ( $H_a$ ) posits that the two variables are dependent (i.e., there is a statistically significant association).

For the test to be valid, the data must satisfy certain assumptions. Crucially, the observations must be independent, and the variables must be [categorical variables](#) (nominal or ordinal). Additionally, while the test is quite robust, it is generally recommended that no more than 20% of the expected cell counts should be less than five, and no expected cell count should be less than one. Failure to meet this requirement may necessitate combining categories or utilizing Fisher's Exact Test, although [SPSS](#) typically provides warnings or automatic corrections when these assumptions are violated.

By following the steps outlined below, analysts can confidently determine whether differences observed in their sample data are likely due to a true relationship between the variables in the population or merely due to random sampling variation. This rigorous process ensures that statistical conclusions drawn are both accurate and reliable for reporting and decision-making purposes.

## Case Study: Gender and Political Preference

To illustrate the procedure, consider a scenario where a researcher wishes to investigate whether

a person's gender is associated with their political party preference. This is a classic application for the [Chi-Square Test of Independence](#), as both gender (Male/Female) and political preference (Republican, Democrat, Independent) are [categorical variables](#).

A simple random sample of 500 eligible voters was surveyed, and their responses were tabulated based on these two variables. The resulting frequency counts summarize the observed data, establishing the foundation for our analysis. The goal is to determine if the distribution of political preferences differs significantly between the male and female respondents.

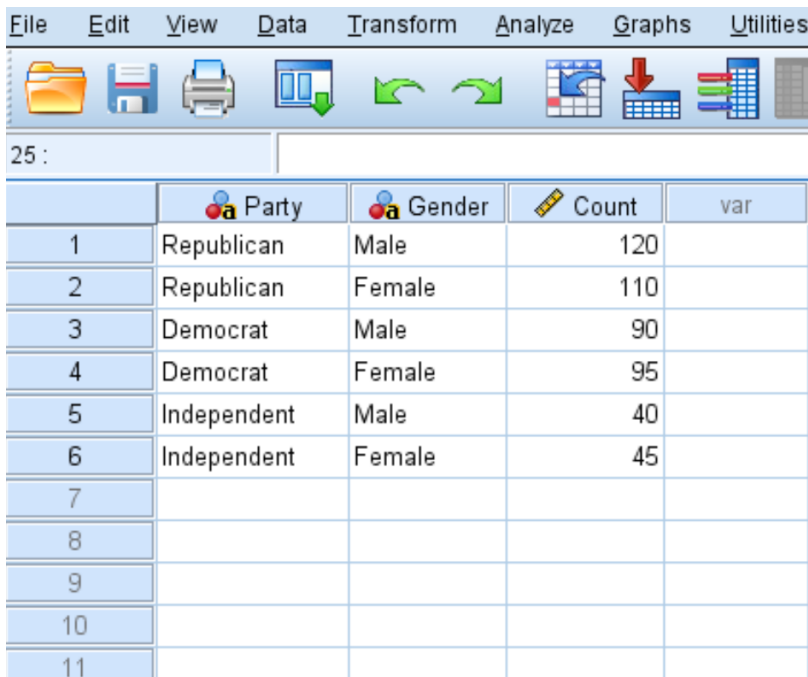
The summary of the survey results is presented in the contingency table below. Note that this table represents the raw, summarized data that we will use as input for the [SPSS](#) analysis. The test aims to quantify the degree to which these observed counts deviate from what we would expect if gender and party preference were entirely independent of each other.

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

## Preparing Data and Weighting Cases in SPSS

When inputting summarized frequency data, such as the table above, into [SPSS](#), the data must be structured carefully so that the program correctly interprets the counts. Unlike raw data where each row represents a single participant, summarized data requires three variables: the first categorical variable (Gender), the second categorical variable (Party), and a third variable representing the frequency or count of observations (Count). This format is essential for the software to correctly calculate the expected frequencies and the resulting Chi-Square statistic.

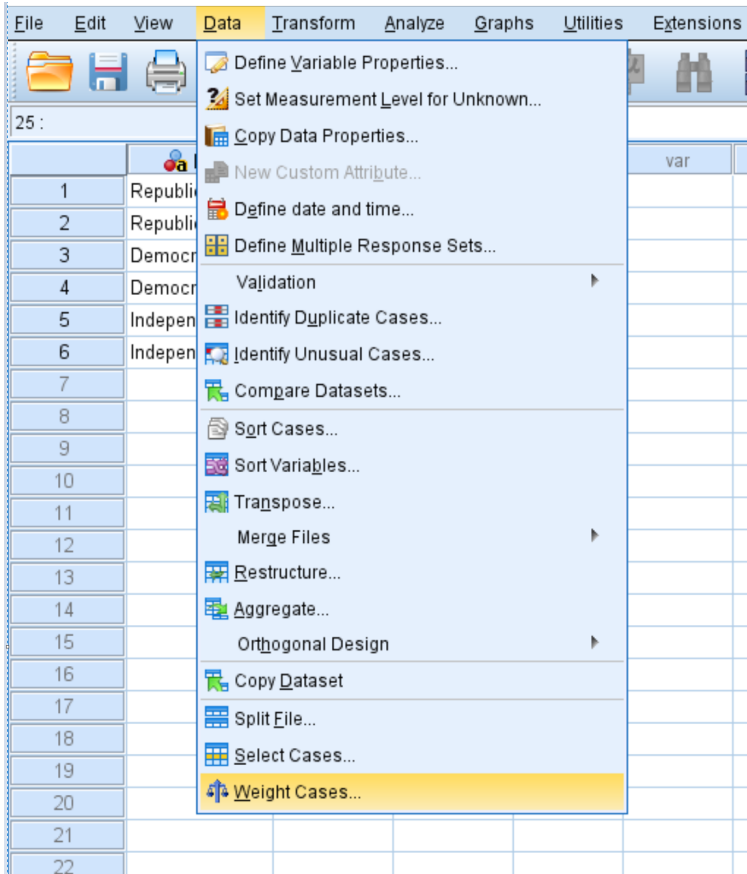
The first critical step in [SPSS](#) is to enter the data in this specific columnar format, ensuring that every unique combination of Gender and Party preference has a corresponding frequency count. This data entry setup effectively tells [SPSS](#) how many participants fall into each cell of the contingency table, enabling the subsequent calculations.



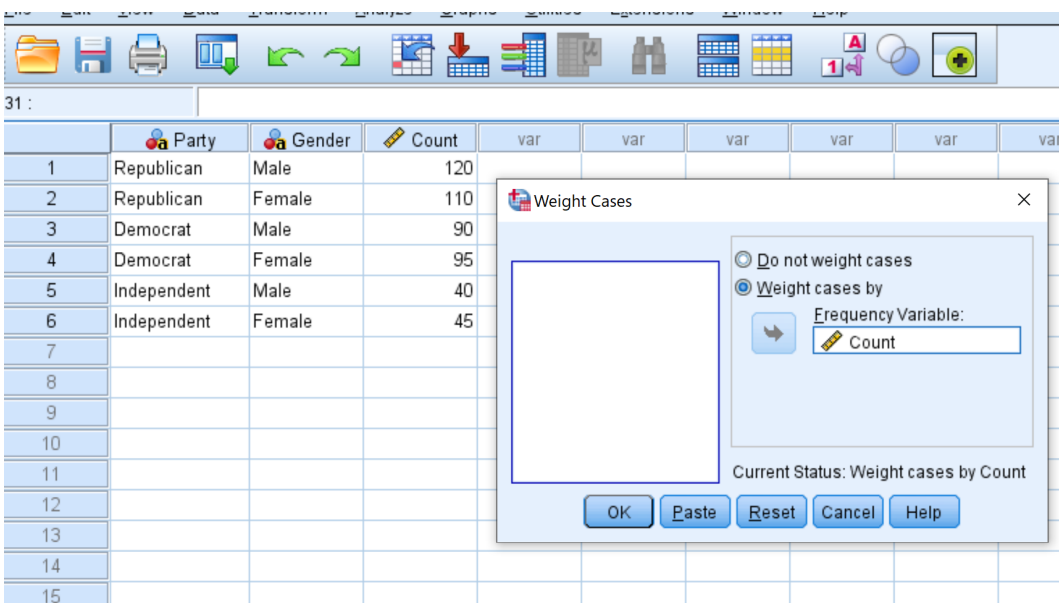
	Party	Gender	Count	var
1	Republican	Male	120	
2	Republican	Female	110	
3	Democrat	Male	90	
4	Democrat	Female	95	
5	Independent	Male	40	
6	Independent	Female	45	
7				
8				
9				
10				
11				

Once the data is entered, we must instruct [SPSS](#) to use the 'Count' variable to weight the cases. This process of weighting ensures that the statistical procedures treat the frequency values in the 'Count' column as actual case observations, rather than treating each entered row (representing a cell combination) as a single observation. If this step is omitted, the analysis will be based on only six cases (the six rows of data entered) instead of the total sample size of 500.

To implement the weighting, navigate to the **Data** tab in the [SPSS](#) menu bar, and then select **Weight Cases**. In the ensuing dialog box, choose the option to **Weight cases by** and drag the variable **Count** into the designated frequency variable box. After confirming this selection by clicking **OK**, the software will correctly weight the data, preparing it for the Chi-Square calculation.



The final step in the weighting process confirms that **Count** is correctly assigned as the frequency variable list, ensuring that all 500 observations are accounted for in the statistical test.

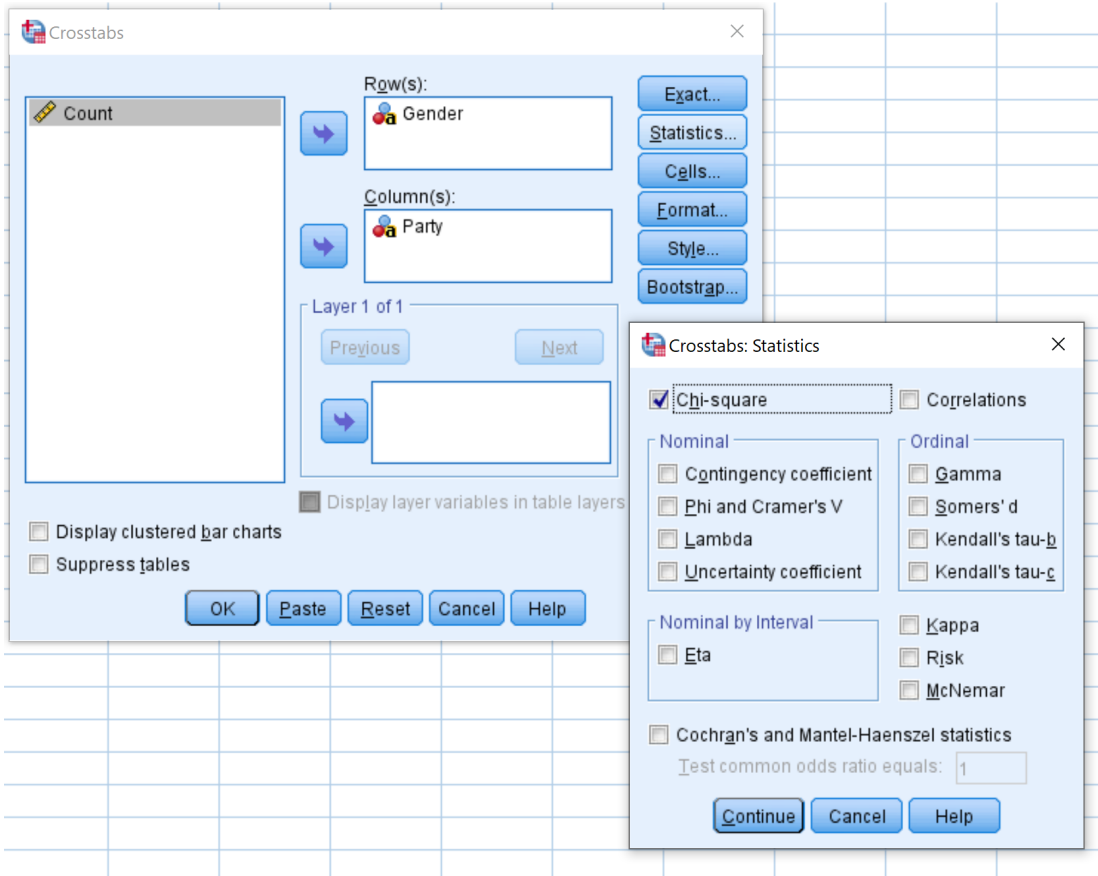


## Executing the Crosstabs Procedure

With the data correctly weighted, the next phase involves running the actual test using the appropriate menu commands in [SPSS](#). The [Chi-Square Test of Independence](#) is housed within the Crosstabs procedure, which is designed specifically for analyzing the relationship between two or more [categorical variables](#) by generating a contingency table and relevant statistics.

To begin the procedure, click the **Analyze** tab in the main menu, hover over **Descriptive Statistics**, and then select **Crosstabs**. This opens the main Crosstabs dialog box, which allows us to define which variable should represent the rows and which should represent the columns in our final output table. Conventionally, the independent variable (Gender) is often placed in the Rows box, and the dependent variable (Party Preference) is placed in the Columns box, although the statistical result remains symmetric regardless of placement.

In the Crosstabs dialog box, drag the variable **Gender** into the **Rows** box and the variable **Party** into the **Columns** box. Crucially, before executing the analysis, you must navigate to the **Statistics** option within the dialog box. Here, ensure that the checkbox next to **Chi-square** is selected. This step explicitly requests [SPSS](#) to calculate the Pearson Chi-Square statistic and its associated [P-value](#), along with the degrees of freedom, which are essential for interpreting the test results. After selecting the required statistic, click **Continue** and then **OK** to run the analysis and generate the output.



## Interpreting the Statistical Output

Upon clicking **OK**, [SPSS](#) will immediately generate the output viewer, presenting several tables that summarize the analysis. These tables provide comprehensive information regarding case processing, the crosstabulation of observed counts, and the critical results of the [Chi-Square Test of Independence](#).

The first table, the **Case Processing Summary**, confirms that all cases (N=500) were included in the analysis, indicating zero missing values, which is ideal. The second table is the **Crosstabulation** table, which simply reproduces the observed counts for each category combination, allowing the researcher to visually inspect the distribution of respondents across gender and political party preference. This table is vital for understanding the raw frequencies that formed the basis of the test.

The third and most important table is the **Chi-Square Tests** output, which contains the statistical test results necessary for hypothesis testing. This table lists various forms of the Chi-Square statistic, but we primarily focus on the **Pearson Chi-Square** value and its associated [P-value](#) (labeled as Asymp. Sig. (2-sided)).

## → Crosstabs

### Case Processing Summary

	Valid		Cases Missing		Total	
	N	Percent	N	Percent	N	Percent
Gender * Party	500	100.0%	0	0.0%	500	100.0%

### Gender \* Party Crosstabulation

Count

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

### Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	.864 <sup>a</sup>	2	.649
Likelihood Ratio	.864	2	.649
N of Valid Cases	500		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 42.50.

The results from the table show that the Pearson Chi-Square test statistic is approximately **.864**, with 2 degrees of freedom, and the corresponding two-sided **P-value** (Asymptotic Significance) is **.649**. The degrees of freedom (df) are calculated as  $(R-1) \times (C-1)$ , where R is the number of rows and C is the number of columns (in this case,  $(2-1) \times (3-1) = 2$ ).

## Drawing Conclusions from the P-value

The final step in the analysis is to compare the calculated **P-value** to the predetermined significance level (alpha, typically set at 0.05). This comparison dictates the decision regarding the **null hypothesis**. As stated earlier, the **null hypothesis** ( $H_0$ ) for this test is that gender and political party preference are independent; that is, there is no association between these two variables.

The decision rule is straightforward: If the **P-value** is less than the significance level ( $\alpha = 0.05$ ), we reject the **null hypothesis**, concluding that a significant association exists. Conversely, if the **P-value** is greater than or equal to 0.05, we fail to reject the **null hypothesis**, concluding that there is insufficient evidence to claim an association.

In this specific example, the calculated [P-value](#) is **0.649**. Since  $0.649$  is substantially greater than the conventional significance level of  $0.05$ , we must fail to reject the [null hypothesis](#). Therefore, based on the statistical evidence gathered from this sample, we conclude that there is not enough sufficient evidence to support the claim that a statistically significant association exists between gender and political party preference among the surveyed voters. The distributions of political preferences observed across male and female respondents are statistically similar, suggesting independence between the two variables.