

How to Perform a Chi-Square Test of Independence in SPSS: A Step-by-Step Guide

Authored by
stats writer

March 15, 2026

RECOMMENDED CITATION

stats writer (2026). *How to Perform a Chi-Square Test of Independence in SPSS: A Step-by-Step Guide*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=135853>

Understanding the Fundamentals of the Chi-Square Test of Independence

The **Chi-Square Test of Independence** is a cornerstone of **bivariate analysis**, specifically designed to examine the relationship between two **categorical variables**. Unlike tests that compare means, such as the t-test or ANOVA, this statistical procedure focuses on **frequencies** and counts within a **contingency table**. By evaluating whether the observed distribution of data differs significantly from what would be expected if no relationship existed, researchers can draw conclusions about the **independence** of the variables in question.

When conducting research in fields like sociology, marketing, or healthcare, it is often necessary to determine if membership in one category, such as **gender**, influences the likelihood of belonging to another category, such as **political party** preference. The **Chi-Square Test of Independence** provides a mathematical framework to test these associations. It operates under the **null hypothesis** that the two variables are completely independent, meaning that knowing the value of one variable does not help predict the value of the other.

To perform this test effectively within **SPSS** (Statistical Package for the Social Sciences), a researcher must ensure that their data meets several underlying assumptions. Primarily, the data must be collected through a **simple random sample** to ensure representativeness. Additionally, the variables must be **nominal or ordinal** in nature, and the **expected frequencies** in each cell of the **contingency table** should generally be five or greater to maintain the validity of the test results.

The power of **SPSS** lies in its ability to automate the complex calculations involved in determining **expected values** and the final **test statistic**. By following a structured workflow--including data entry, case weighting, and the execution of the **crosstabs** procedure--users can generate detailed reports that include the **Pearson Chi-Square** value and the corresponding **p-value**, which are essential for making data-driven decisions.

Establishing the Research Scenario and Hypothesis

To illustrate the application of this test, consider a study aimed at determining whether **gender** is significantly associated with **political party** preference. This is a classic research question where both variables are categorical. In this scenario, we might collect data from a **simple random sample** of 500 voters. The objective is to see if men and women distribute themselves differently across political affiliations like Republican, Democrat, and Independent, or if their preferences are essentially the same.

Before proceeding with the analysis in **SPSS**, it is vital to define the **null hypothesis** (H_0) and the **alternative hypothesis** (H_a). In our case, the **null hypothesis** posits that gender and political party preference are independent; in other words, there is no significant association between the

two. The **alternative hypothesis** suggests that there is a significant association, meaning that party preference does indeed depend, to some extent, on gender.

The following table summarizes the raw data collected from the 500 survey participants. This **contingency table** provides the **observed frequencies** that will serve as the foundation for our **Chi-Square Test of Independence**:

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

By examining these raw numbers, we can see slight differences in the counts. For example, 120 males identified as Republican compared to 110 females. However, raw numbers alone are insufficient to determine if these differences are due to random **sampling error** or a genuine underlying pattern. This is precisely why we employ the **Chi-Square Test** to provide **statistical significance** to our observations.

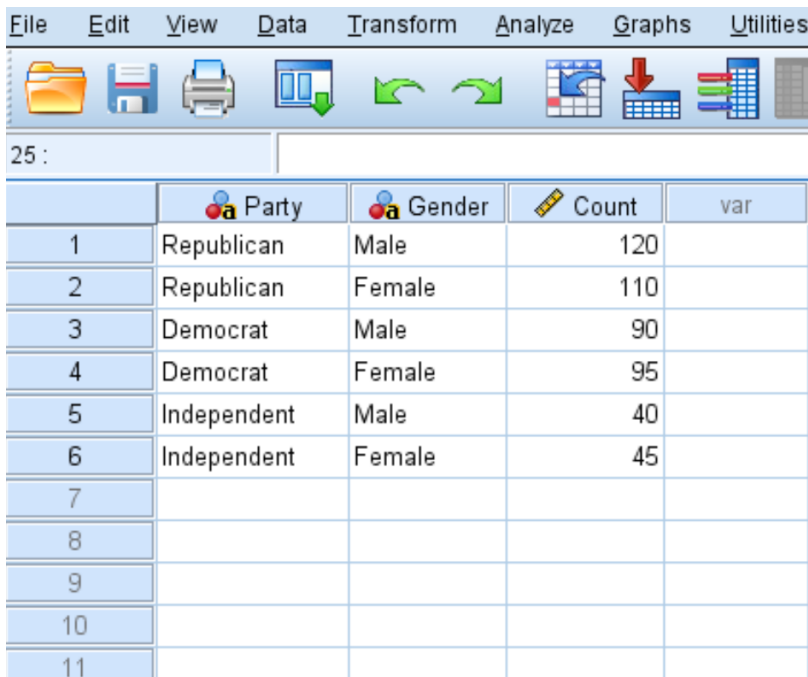
Step 1: Data Entry Procedures within the SPSS Interface

The first practical step in performing the **Chi-Square Test of Independence** is to enter the data into **SPSS** correctly. There are two primary ways to enter data for this test: raw individual responses or weighted frequency counts. For efficiency, many researchers use the weighted method, where each row represents a unique combination of the categories and a third variable indicates the count of occurrences for that specific combination.

In our **SPSS Data View**, we should create three distinct columns: one for Gender, one for Party, and one for Count. For the categorical variables, it is standard practice to use numeric **coding**. For instance, we might code Male as 1 and Female as 2. Similarly, Republican could be 1, Democrat 2, and Independent 3. This numeric approach allows the software to process the categorical data more effectively.

Once the variables are defined in the **Variable View** tab, you can input the values based on our table. For example, the first row would show a "1" for Gender (Male), a "1" for Party (Republican), and "120" for the Count. This structured format ensures that every possible interaction between the variables is accounted for in the **dataset**.

Refer to the image below to visualize the correct layout of the data within the **SPSS** environment. Ensuring the accuracy of this step is paramount, as any errors in **data entry** will propagate through the analysis and lead to incorrect conclusions regarding the **statistical significance** of the results.

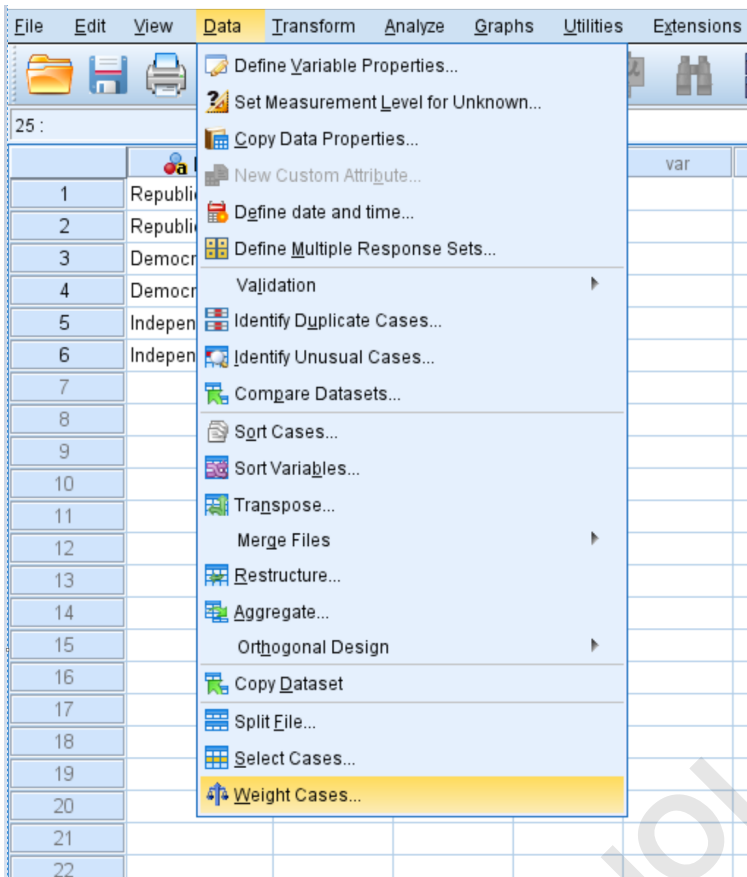


	Party	Gender	Count	var
1	Republican	Male	120	
2	Republican	Female	110	
3	Democrat	Male	90	
4	Democrat	Female	95	
5	Independent	Male	40	
6	Independent	Female	45	
7				
8				
9				
10				
11				

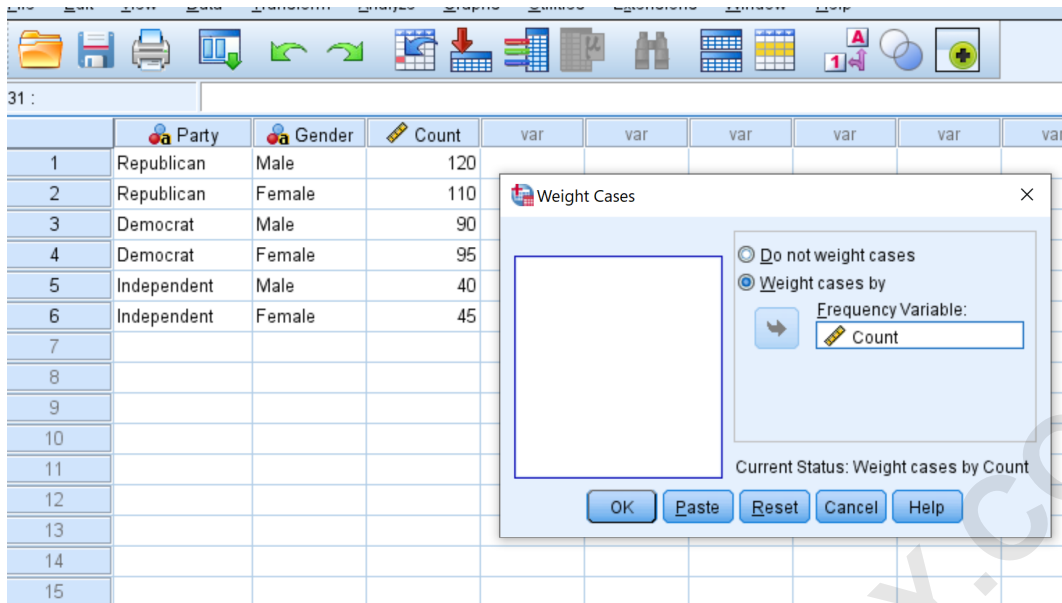
Step 2: The Critical Requirement of Weighting Cases

After the data has been entered, a crucial but often overlooked step in **SPSS** is **weighting cases**. By default, **SPSS** treats every row in the **Data View** as a single observation or a single person. Since we have entered aggregated counts (e.g., one row representing 120 individuals), we must instruct the software to treat that "Count" variable as the actual number of observations for each category combination.

To begin this process, navigate to the **Data** menu at the top of the interface and select the **Weight Cases** option. This will open a new dialog box where you can specify how the weighting should be applied. Failure to perform this step will result in **SPSS** calculating the test based on only 6 observations (the number of rows in our spreadsheet) instead of the actual 500 survey participants, which would yield completely invalid results.



In the **Weight Cases** dialog box, select the radio button labeled **Weight cases by**. Then, select your "Count" variable from the list on the left and move it into the **Frequency Variable** box. Once you click **OK**, you will notice "Weight On" appearing in the status bar at the bottom right of the **SPSS** window. This confirms that the software is now ready to perform the **Chi-Square Test of Independence** using the full volume of your sample data.

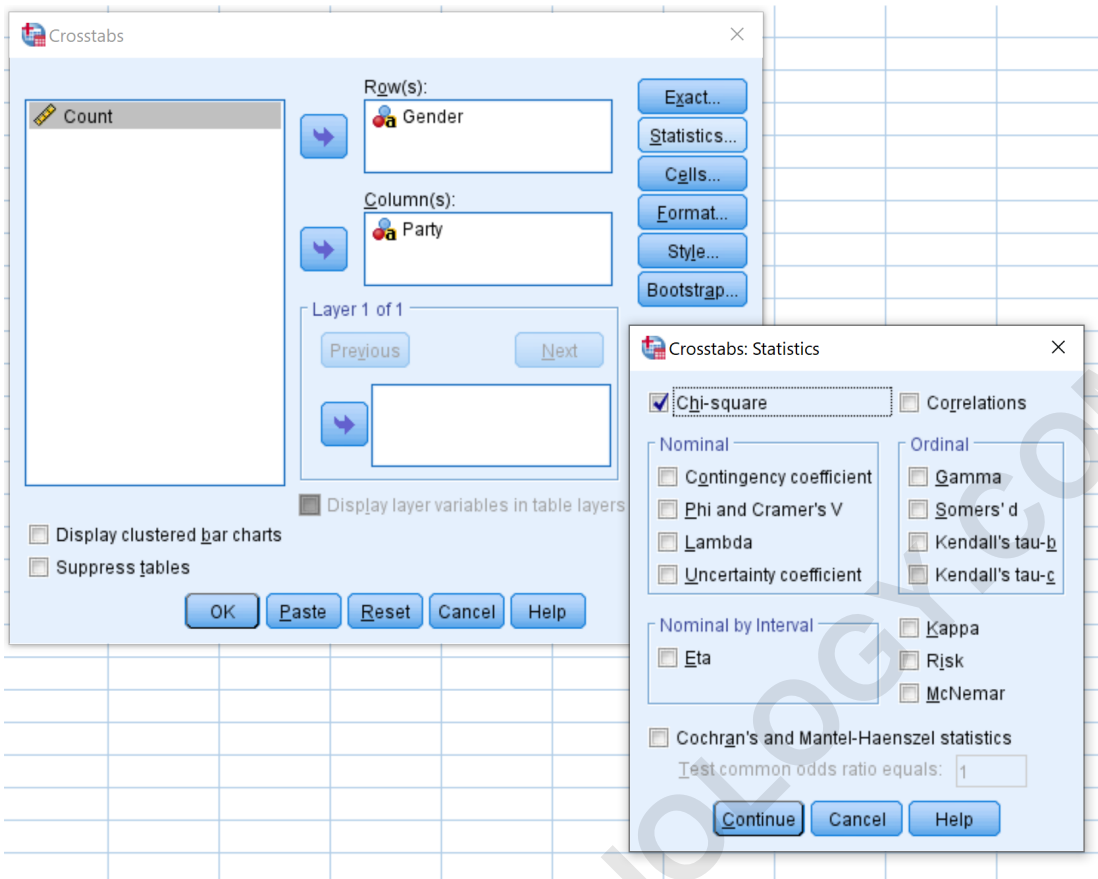


Step 3: Navigating the Crosstabs Procedure in SPSS

With the data properly weighted, we can now proceed to the core analysis. The **Chi-Square Test of Independence** is found within the **Crosstabs** procedure. To access this, click on the **Analyze** tab, hover over **Descriptive Statistics**, and then select **Crosstabs**. This menu is used for generating **contingency tables** and performing various tests of association for **categorical variables**.

Inside the **Crosstabs** dialog window, you will need to assign your variables to the **Rows** and **Columns** boxes. Traditionally, the **independent variable** (in this case, Gender) is placed in the Rows, and the **dependent variable** (Political Party) is placed in the Columns. However, for a **Chi-Square Test**, the mathematical result remains the same regardless of which variable is in the row or column.

The next critical sub-step is to click the **Statistics** button located on the right side of the dialog box. This will open a smaller window containing various statistical tests. You must check the box next to **Chi-square** to ensure that **SPSS** calculates the **Pearson Chi-Square** statistic. After checking the box, click **Continue** to return to the main **Crosstabs** window and then click **OK** to run the analysis.



By executing these commands, **SPSS** will process the **frequency** data and produce a comprehensive output in the **SPSS Viewer**. This output contains the necessary information to either reject or fail to reject our **null hypothesis** regarding the independence of gender and political affiliation.

Step 4: Analyzing the Statistical Output and Frequencies

Once the analysis is complete, **SPSS** will display several tables in the output window. The first table, **Case Processing Summary**, is essential for **data cleaning**. It shows the total number of cases processed and identifies any **missing data**. In our example, it indicates that 500 cases were included with 0 missing cases, confirming that our **dataset** is complete and the analysis is based on the full sample.

The second table produced is the **Gender * Party Crosstabulation**. This table is a reflection of the **contingency table** we initially discussed. It provides the **observed frequencies** for every combination of gender and party. Reviewing this table is a good way to double-check that the data was entered and weighted correctly. It offers a clear visual breakdown of how the 500 voters are distributed across the categories.

The most vital part of the output is the **Chi-Square Tests** table. This table provides the actual **Pearson Chi-Square** value, the **degrees of freedom** (df), and the **asymptotic significance** (2-sided p-value). For our specific example, the test statistic is reported as **.864**, with 2 **degrees of freedom**, and a **p-value** of **.649**.

→ 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 ^a	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.

In addition to these values, **SPSS** often includes a footnote below the table indicating the percentage of cells that have **expected counts** less than 5. For the **Chi-Square Test** to be reliable, this percentage should ideally be zero. If this assumption is violated, the researcher might need to use **Fisher's Exact Test** instead, especially in smaller samples.

Final Interpretation of the P-Value and Hypothesis Testing

The final step in the **Chi-Square Test of Independence** is interpreting the **p-value** to make a statistical decision. In the context of **hypothesis testing**, we compare our **p-value** to a pre-determined **significance level** (alpha), which is commonly set at 0.05. The **p-value** represents the probability of observing a test statistic as extreme as the one calculated, assuming the **null hypothesis** is true.

In our analysis, the **p-value** is **.649**. Since this value is significantly greater than the threshold of 0.05, we **fail to reject the null hypothesis**. This indicates that the observed differences in political party preference between males and females are likely due to chance rather than a systematic relationship. In formal terms, there is no **statistically significant** association between gender and political party preference in this sample.

Concluding a study with a non-significant result is just as important as finding a significant one. It tells the researcher that, based on the evidence, the two variables operate independently of one another. Therefore, knowing an individual's gender does not provide a statistically reliable basis for predicting their political affiliation within this specific population. This information can help guide future research or influence how organizations target their outreach efforts.

By mastering the **Chi-Square Test of Independence** in **SPSS**, you gain a powerful tool for analyzing **categorical data**. Whether you are validating survey results or exploring social trends, following these structured steps ensures that your **statistical analysis** is rigorous, accurate, and reproducible. Always remember to check your **expected frequencies** and ensure your **data entry** matches the requirements of the software for the best results.