How do you perform a Chi-Square Test of Independence in SAS?

Authored by

stats writer

July 1, 2024

The Chi-Square Test of Independence is a statistical method used to determine the relationship between two categorical variables. In SAS, this test can be performed by using the PROC FREQ procedure. First, the data must be organized in a contingency table format with the two variables of interest. Then, the PROC FREQ statement is used to specify the variables and the CHISQ option is added to the TABLES statement to request the Chi-Square test. The results of the test, including the Chi-Square statistic and the p-value, will be displayed in the output. This test is useful for determining if there is a significant association between two variables, and can provide valuable insights for further analysis.

Perform a Chi-Square Test of Independence in SAS

A Chi-Square Test of Independence is used to determine whether or not there is a significant association between two .

The following example shows how to perform a Chi-Square Test of Independence in SAS.

Example: Chi-Square Test of Independence in SAS

Suppose we want to know whether or not gender is associated with political party preference. We take a of 500 voters and survey them on their political party preference.

The following table shows the results of the survey:

	Republican	Democrat	Independent	Total
Male	120	90	40	250
Female	110	95	45	250

Total	230	185	85	500

Use the following steps to perform a Chi-Square Test of Independence in SAS to determine if gender is associated with political party preference.

Step 1: Create the data.

First, we will create a dataset in SAS to hold the survey responses:

```
/*create dataset*/
data my_data;
input Gender $ Party $ Count;
datalines;
Male Rep 120
Male Dem 90
Male Ind 40
Female Rep 110
Female Dem 95
Female Ind 45
;
run;

/*print dataset*/
```

```
proc printdata=my_data;
```

Obs	Gender	Party	Count
1	Male	Rep	120
2	Male	Dem	90
3	Male	Ind	40
4	Female	Rep	110
5	Female	Dem	95
6	Female	Ind	45

Step 2: Perform the Chi-Square Test of Independence.

Next, we can use the following code to perform the Chi-Square Test of Independence:

```
/*perform Chi-Square Test of Independence*/
proc freqdata=my_data;
tables Gender*Party / chisq;
weight Count;
run;
```

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Gender by Party			
	Party			
Gender	Dem	Ind	Rep	Total
Female	95 19.00 38.00 51.35	45 9.00 18.00 52.94	110 22.00 44.00 47.83	250 50.00
Male	90 18.00 36.00 48.65	40 8.00 16.00 47.06	120 24.00 48.00 52.17	250 50.00
Total	185 37.00	85 17.00	230 46.00	500 100.00

Statistics for Table of Gender by Party

Statistic	DF	Value	Prob
Chi-Square	2	0.8640	0.6492
Likelihood Ratio Chi-Square	2	0.8644	0.6491
Mantel-Haenszel Chi-Square	1	0.5464	0.4598
Phi Coefficient		0.0416	
Contingency Coefficient		0.0415	
Cramer's V		0.0416	

Sample Size = 500

There are two values of interest in the output:

Chi-Square Test Statistic: 0.8640Corresponding p-value: 0.6492H0: The two variables are independent.HA: The two variables are *not* independent.

Since the (0.6492) of the test is not less than 0.05, we fail to reject the null hypothesis.

This means we do not have sufficient evidence to say that there is an association between gender and political party preference.

In other words, gender and political party preference are independent.

Additional Resources

The following tutorials provide additional information about the Chi-Square test of independence: