

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

Authored by
stats writer

April 16, 2024

RECOMMENDED CITATION

stats writer (2024). *How do you perform a Chi-Square Test of Independence in Python?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=136155>

The Chi-Square Test of Independence is a statistical method used to determine the relationship between two categorical variables. In Python, this test can be performed by using the `chi2_contingency` function from the `scipy.stats` library. First, the data must be organized in a contingency table format with the rows representing one variable and the columns representing the other variable. The function then calculates the chi-square statistic and the corresponding p-value, which can be used to determine if there is a significant relationship between the variables. If the p-value is less than the chosen significance level, it can be concluded that there is a significant association between the variables. The results of the test can also be visualized using a contingency table or a bar graph. Overall, performing a Chi-Square Test of Independence in Python involves organizing the data, using the `chi2_contingency` function, and interpreting the results to determine the relationship between the variables.

Perform a Chi-Square Test of Independence in Python

A is used to determine whether or not there is a significant association between two categorical variables.

This tutorial explains how to perform a Chi-Square Test of Independence in Python.

Example: Chi-Square Test of Independence in Python

Suppose we want to know whether or not gender is associated with political party preference. We take a simple random sample 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 Python to determine if gender is associated with political party preference.

Step 1: Create the data.

First, we will create a table to hold our data:

```
data = ,  
]
```

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

Next, we can perform the Chi-Square Test of Independence using the `chi2_contingency` from the SciPy library, which uses the following syntax:

```
chi2_contingency(observed)
```

where:

observed: A contingency table of observed values.

The following code shows how to use this function in

our specific example:

```
import scipy.stats as stats
```

```
#perform the Chi-Square Test of Independence  
stats.chi2_contingency(data)
```

```
(0.864,  
0.649,  
2,  
array(  
))
```

The way to interpret the output is as follows:

Chi-Square Test Statistic: 0.864p-value: 0.649Degrees of freedom: 2 (calculated as #rows-1 * #columns-1)Array: The last array displays the expected values for each cell in the contingency table.

Recall that the uses the following null and alternative hypotheses:

H0: (null hypothesis) The two variables are independent.H1: (alternative hypothesis) The two variables are *not* independent.

Since the p-value (.649) 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.

ARABPSYCHOLOGY.COM