

How do I perform a Chi-Square Goodness of Fit Test in Python?

Authored by
stats writer

April 16, 2024

RECOMMENDED CITATION

stats writer (2024). *How do I perform a Chi-Square Goodness of Fit Test in Python?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=136151>

The Chi-Square Goodness of Fit Test in Python is a statistical method used to determine if there is a significant difference between observed data and expected data. It is commonly used to analyze categorical data and determine if the observed frequencies match the expected frequencies. To perform this test in Python, one must first import the necessary libraries and data into the program. Then, using the appropriate function or method, the test can be conducted and the results can be interpreted to determine the significance of the data. This test is useful for analyzing data in various fields such as marketing, social sciences, and biology. It is a powerful tool for understanding the relationship between categorical variables and can provide valuable insights for decision making.

Perform a Chi-Square Goodness of Fit Test in Python

A is used to determine whether or not a categorical variable follows a hypothesized distribution.

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

Example: Chi-Square Goodness of Fit Test in Python

A shop owner claims that an equal number of customers come into his shop each weekday. To test this hypothesis, a researcher records the number of customers that come into the shop in a given week and finds the following:

**Monday: 50 customers Tuesday: 60 customers
Wednesday: 40 customers Thursday: 47 customers
Friday: 53 customers**

Use the following steps to perform a Chi-Square goodness of fit test in Python to determine if the data is consistent with the shop owner's claim.

Step 1: Create the data.

First, we will create two arrays to hold our observed and expected number of customers for each day:

expected =

observed =

Step 2: Perform the Chi-Square Goodness of Fit Test.

Next, we can perform the Chi-Square Goodness of Fit Test using the `chi2` function from the SciPy library, which uses the following syntax:

```
chi2(f_obs, f_exp)
```

where:

f_obs: An array of observed counts.
f_exp: An array of expected counts. By default, each category is assumed to be equally likely.

The following code shows how to use this function in our specific example:

```
import scipy.stats as stats

#perform Chi-Square Goodness of Fit Test
stats.chisquare(f_obs=observed, f_exp=expected)

(statistic=4.36, pvalue=0.35947)
```

The Chi-Square test statistic is found to be 4.36 and the corresponding p-value is 0.35947.

Recall that a Chi-Square Goodness of Fit Test uses the following null and alternative hypotheses:

H0: (null hypothesis) A variable follows a hypothesized distribution.
H1: (alternative hypothesis) A variable does not follow a hypothesized distribution.

Since the p-value (.35947) is not less than 0.05, we fail to reject the null hypothesis. This means we do not have sufficient evidence to say that the true distribution of customers is different from the distribution that the shop owner claimed.