

How can a Jarque-Bera test be performed in Python?

Authored by
stats writer

April 17, 2024

RECOMMENDED CITATION

stats writer (2024). *How can a Jarque-Bera test be performed in Python?*.

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=136158>

A Jarque-Bera test is a statistical test used to determine if a dataset follows a normal distribution. In Python, this test can be performed using the "scipy.stats.jarque_bera" function. This function takes in a dataset as its input and returns two values - the Jarque-Bera statistic and the p-value. The Jarque-Bera statistic represents the deviation of the dataset from a normal distribution, while the p-value indicates the significance level of the test. A low p-value suggests that the dataset does not follow a normal distribution. This test can be useful in identifying any potential outliers or anomalies in the dataset.

Perform a Jarque-Bera Test in Python

The Jarque-Bera test is a goodness-of-fit test that determines whether or not sample data have skewness and kurtosis that matches a normal distribution.

The test statistic of the Jarque-Bera test is always a positive number and the further it is from zero, the more evidence that the sample data does not follow a normal distribution.

This tutorial explains how to conduct a Jarque-Bera test in Python.

How to Perform a Jarque-Bera test in Python

To conduct a Jarque-Bera test in Python we can use the from the Scipy library, which uses the following syntax:

```
jarque_bera(x)
```

where:

x: an array of observations

This function returns a test statistic and a corresponding p-value.

Example 1

Suppose we perform a Jarque-Bera test on a list of 5,000 values that follow a normal distribution:

```
import numpy as np  
import scipy.stats as stats  
  
#generate array of 5000 values that follow a standard  
normal distribution  
np.random.seed(0)  
data = np.random.normal(0, 1, 5000)  
  
#perform Jarque-Bera test  
stats.jarque_bera(data)  
  
(statistic=1.2287, pvalue=0.54098)
```

The test statistic is 1.2287 and the corresponding p-value is 0.54098. Since this p-value is not less than .05,

we fail to reject the null hypothesis. We don't have sufficient evidence to say that this data has skewness and kurtosis that is significantly different from a normal distribution.

This result shouldn't be surprising since the data that we generated is composed of 5000 random variables that follow a normal distribution.

Example 2

Now suppose we perform a Jarque-Bera test on a list of 5,000 values that follow a uniform distribution:

```
import numpy as np
import scipy.stats as stats

#generate array of 5000 values that follow a uniform
distribution
np.random.seed(0)
data = np.random.uniform(0, 1, 5000)

#perform Jarque-Bera test
stats.jarque_bera(data)

(statistic=300.1043, pvalue=0.0)
```

This result also shouldn't be surprising since the data that we generated is composed of 5000 random variables that follow a uniform distribution, which should have skewness and kurtosis that are much different than a normal distribution.

When to Use the Jarque-Bera Test

The Jarque-Bera Test is typically used for large datasets ($n > 2000$) in which other normality tests (like the Shapiro-Wilk test) are unreliable.

This is an appropriate test to use before you perform some analysis in which it's assumed that the dataset follows a normal distribution. A Jarque-Bera test can tell you whether or not this assumption is satisfied.