How can I calculate Spearman Rank Correlation in Python?

Authored by

stats writer

April 23, 2024

Spearman Rank Correlation is a statistical method used to measure the strength and direction of the relationship between two variables. In Python, the scipy library provides a function called "spearmanr" which can be used to calculate the Spearman Rank Correlation coefficient. This function takes in two arrays of data and returns the correlation coefficient as well as a p-value to determine the significance of the correlation. The correlation coefficient ranges from -1 to 1, where a value of 1 indicates a perfect positive correlation, 0 indicates no correlation, and -1 indicates a perfect negative correlation. By using the "spearmanr" function in Python, one can easily and accurately calculate the Spearman Rank Correlation and determine the strength and direction of the relationship between two variables.

Calculate Spearman Rank Correlation in Python

In statistics, correlation refers to the strength and direction of a relationship between two variables. The value of a correlation coefficient can range from -1 to 1, with the following interpretations:

-1: a perfect negative relationship between two variables0: no relationship between two variables1: a perfect positive relationship between two variables

One special type of correlation is called Spearman Rank Correlation, which is used to measure the correlation between two ranked variables. (e.g. rank of a student's math exam score vs. rank of their science exam score in a class).

This tutorial explains how to calculate the Spearman rank correlation between two variables in Python

Example: Spearman Rank Correlation in Python

Suppose we have the following pandas DataFrame that contains the math exam score and science exam score of 10 students in a particular class:

```
import pandas as pd

#create DataFrame
df = pd.DataFrame({'student': ,
'math': ,
'science': })
```

To calculate the Spearman Rank correlation between the math and science scores, we can use the spearmanr() function from scipy.stats:

```
from scipy.statsimport spearmanr

#calculate Spearman Rank correlation and corresponding p-value
rho, p = spearmanr(df, df)

#print Spearman rank correlation and p-value
print(rho)
```

-0.4181818181818815

print(p)

0.2291128409828189 2

From the output we can see that the Spearman rank correlation is -0.41818 and the corresponding p-value is 0.22911.

This indicates that there is a negative correlation between the science and math exam scores.

However, since the p-value of the correlation is not less than 0.05, the correlation is not statistically significant.

Note that we could also use the following syntax to just extract the correlation coefficient or the p-value:

#extract Spearman Rank correlation coefficient
spearmanr(df, df)

-0.4181818181818815

#extract p-value of Spearman Rank correlation coefficient

spearmanr(df, df)

0.2291128409828189 2

How to Calculate Spearman Rank Correlation in R

How to Calculate Spearman Rank Correlation in Excel

How to Calculate Spearman Rank Correlation in Stata