

How can I calculate the point-biserial correlation in Python?

Authored by
stats writer

April 17, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I calculate the point-biserial correlation in Python?*.

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=136328>

The point-biserial correlation is a statistical measure that determines the relationship between a continuous variable and a dichotomous variable. It is commonly used to examine the correlation between a numerical variable and a binary variable. In order to calculate the point-biserial correlation in Python, one can use the "pointbiserialr" function from the "scipy.stats" library. This function takes in the two variables and returns the correlation coefficient along with the corresponding p-value. By using this function, one can easily determine the strength and direction of the relationship between a continuous variable and a dichotomous variable in Python.

Calculate Point-Biserial Correlation in Python

Point-biserial correlation is used to measure the relationship between a binary variable, x , and a continuous variable, y .

Similar to the r , the point-biserial correlation coefficient takes on a value between -1 and 1 where:

**-1 indicates a perfectly negative correlation between two variables
0 indicates no correlation between two variables
1 indicates a perfectly positive correlation between two variables**

This tutorial explains how to calculate the point-biserial correlation between two variables in Python.

Example: Point-Biserial Correlation in Python

Suppose we have a binary variable, x , and a continuous variable, y :

x =

y =

We can use the function from the `scipy.stats` library to calculate the point-biserial correlation between the two variables.

Note that this function returns a correlation coefficient along with a corresponding p-value:

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

```
#calculate point-biserial correlation  
stats.pointbiserialr(x, y)
```

```
PointbiserialrResult(correlation=0.21816,  
pvalue=0.51928)
```

The point-biserial correlation coefficient is 0.21816 and the corresponding p-value is 0.51928.

Since the correlation coefficient is positive, this indicates that when the variable x takes on the value "1" that the variable y tends to take on higher values compared to when the variable x takes on the value "0."

Since the p-value of this correlation is not less than .05, this correlation is not statistically significant.

You can find the exact details of how this correlation is calculated in the `scipy.stats`.

ARABPSYCHOLOGY.COM