

# How can I calculate the correlation between two columns in Pandas?

Authored by  
**stats writer**

April 28, 2024

## RECOMMENDED CITATION

stats writer (2024). *How can I calculate the correlation between two columns in Pandas?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=140598>

Calculating the correlation between two columns in Pandas involves using the built-in "corr" function, which takes in two columns of data and produces a correlation coefficient. This coefficient represents the strength and direction of the linear relationship between the two columns. The resulting value ranges from -1 to 1, with 0 indicating no correlation, -1 indicating a perfect negative correlation, and 1 indicating a perfect positive correlation. This function is useful for identifying any potential relationships between variables in a dataset and can aid in data analysis and decision making.

## Calculate Correlation Between Two Columns in Pandas

You can use the following syntax to calculate the correlation between two columns in a pandas DataFrame:

```
df.corr(df)
```

The following examples show how to use this syntax in practice.

### Example 1: Calculate Correlation Between Two Columns

The following code shows how to calculate the correlation between columns in a pandas DataFrame:

```
import pandas as pd

#create DataFrame
df = pd.DataFrame({'points': ,
'assists': ,
```

```
'rebounds': })
```

```
#view first five rows of DataFrame
```

```
df.head()
```

```
points assists rebounds
```

```
0 25 5 11
```

```
1 12 7 8
```

```
2 15 7 10
```

```
3 14 9 6
```

```
4 19 12 6
```

```
#calculate correlation between points and assists
```

```
df.corr(df)
```

```
-0.359384
```

The correlation coefficient is **-0.359**. Since this correlation is negative, it tells us that points and assists are negatively correlated.

In other words, as values in the points column increase, the values in the assists column tend to decrease.

**Example 2: Calculate Significance of Correlation**

To determine whether or not a correlation coefficient is statistically significant, you can use the `pearsonr(x, y)` function from the library.

The following code shows how to use this function in practice:

```
import pandas as pd
from scipy.stats import pearsonr

#create DataFrame
df = pd.DataFrame({'points': ,
'assists': ,
'rebounds': })

#calculate p-value of correlation coefficient between
points and assists
pearsonr(df, df)

(-0.359384, 0.38192)
```

The first value in the output displays the correlation coefficient (-0.359384) and the second value displays the p-value (0.38192) associated with this correlation coefficient.

**Since the is not less than  $\alpha = 0.05$ , we would conclude that the correlation between points and assists is not statistically significant.**

ARABPSYCHOLOGY.COM