

# How can cross correlation be calculated in Python?

Authored by  
**stats writer**

April 27, 2024

## RECOMMENDED CITATION

stats writer (2024). *How can cross correlation be calculated in Python?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=140086>

Cross correlation is a mathematical technique used to measure the similarity between two signals or datasets. In Python, cross correlation can be calculated using the "correlate" function from the "numpy" library. This function takes in two input arrays and returns a cross-correlation array, which represents the correlation between the two inputs at different time lags. The result can be further visualized using the "matplotlib" library. By utilizing these tools, cross correlation can be easily calculated and analyzed in Python for various applications such as signal processing, data analysis, and pattern recognition.

## Calculate Cross Correlation in Python

**Cross correlation is a way to measure the degree of similarity between a time series and a lagged version of another time series.**

**This type of correlation is useful to calculate because it can tell us if the values of one time series are predictive of the future values of another time series. In other words, it can tell us if one time series is a leading indicator for another time series.**

**This type of correlation is used in many different fields, including:**

**Business: Marketing spend is often considered to be a leading indicator for future revenue of businesses. For example, if a business spends an abnormally high amount of money on marketing during one quarter, then total revenue is expected to be high x quarters later.**

**Economics:** The consumer confidence index (CCI) is considered to be a leading indicator for the gross domestic product (GDP) of a country. For example, if CCI is high during a given month, the GDP is likely to be higher  $x$  months later.

The following example shows how to calculate the cross correlation between two time series in Python.

Example: How to Calculate Cross Correlation in Python

Suppose we have the following time series in Python that show the total marketing spend (in thousands) for a certain company along with the the total revenue (in thousands) during 12 consecutive months:

```
import numpy as np

#define data
marketing = np.array()
revenue = np.array()
```

We can calculate the cross correlation for every lag between the two time series by using the `ccf()` function from the `statsmodels` package as follows:

```
import statsmodels.api as sm

#calculate cross correlation
sm.tsa.stattools.ccf(marketing, revenue,
adjusted=False)

array()
```

Here's how to interpret this output:

The cross correlation at lag 0 is 0.771. The cross correlation at lag 1 is 0.462. The cross correlation at lag 2 is 0.194. The cross correlation at lag 3 is -0.061.

And so on.

Notice that the correlation between the two time series becomes less and less positive as the number of lags increases. This tells us that marketing spend during a given month is quite predictive of revenue one or two months later, but not predictive of revenue beyond more than two months.

This intuitively makes sense - we would expect that high marketing spend during a given month is predictive of increased revenue during the next two

**months, but not necessarily predictive of revenue several months into the future.**

ARABPSYCHOLOGY.COM