

How can I calculate cross correlation in R?

Authored by
stats writer

April 27, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I calculate cross correlation in R?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=140079>

Cross correlation is a statistical method used to measure the similarity between two variables over time. In R, the cross correlation function can be calculated by using the "ccf" command. This function takes in two time series data sets and produces a plot and numerical values showing the correlation between the two variables at various lags. It is a useful tool for analyzing the relationship between two variables and can provide valuable insights for time series analysis. To use this function effectively, it is important to have a basic understanding of time series data and the concept of correlation.

Calculate Cross Correlation in R

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:

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.

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.

Biology: Total ocean pollution is considered to be a leading indicator of the population of a certain species of turtle. For example, if pollution is higher during a given year then the total population of turtles is expected to be lower x years later.

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

Example: How to Calculate Cross Correlation in R

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

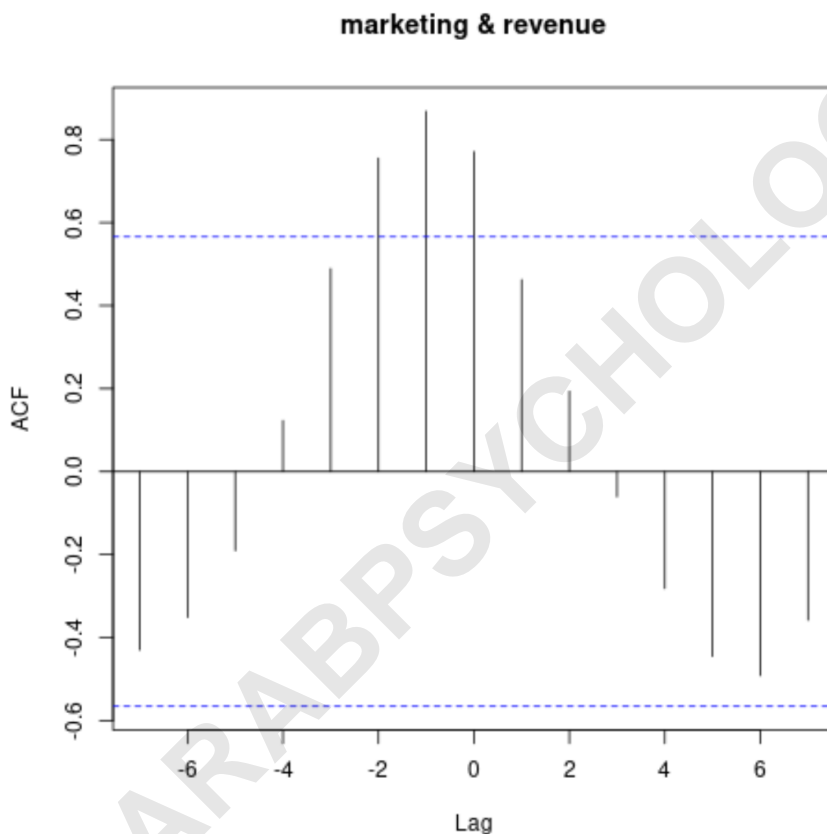
```
#define data
```

```
marketing <- c(3, 4, 5, 5, 7, 9, 13, 15, 12, 10, 8, 8)
```

```
revenue <- c(21, 19, 22, 24, 25, 29, 30, 34, 37, 40, 35, 30)
```

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

```
#calculate cross correlation  
ccf(marketing, revenue)
```



This plot displays the correlation between the two time series at various lags.

To display the actual correlation values, we can use the following syntax:

```
#display cross correlation valuesprint(ccf(marketing,  
revenue))
```

Autocorrelations of series 'X', by lag

```
-7 -6 -5 -4 -3 -2 -1 0 1 2 3  
-0.430 -0.351 -0.190 0.123 0.489 0.755 0.868 0.771 0.462  
0.194 -0.061  
4 5 6 7  
-0.282 -0.445 -0.492 -0.358
```

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 is quite positive within lags -2 to 2, which tells us that marketing spend during a given month is quite predictive of revenue one and two months later.

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.

ARABPSYCHOLOGY.COM