

How can I calculate the partial correlation in R?

Authored by
stats writer

April 17, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I calculate the partial correlation in R?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=136482>

Calculating the partial correlation in R involves determining the relationship between two variables while controlling for the effects of a third variable. This can be done using the "pcor.test" function in the "ppcor" package. The function takes in the three variables and returns the partial correlation coefficient and its associated p-value. It also allows for the visualization of the relationship using a scatterplot with a fitted line. This process can help to better understand the direct relationship between two variables by removing the influence of a third variable.

Calculate Partial Correlation in R

In statistics, we often use the to measure the linear relationship between two variables.

However, sometimes we're interested in understanding the relationship between two variables while controlling for a third variable.

For example, suppose we want to measure the association between the number of hours a student studies and the final exam score they receive, while controlling for the student's current grade in the class.

In this case, we could use a partial correlation to measure the relationship between hours studied and final exam score.

This tutorial explains how to calculate partial correlation in R.

Example: Partial Correlation in R

Suppose we have the following data frame that displays the current grade, total hours studied, and final exam score for 10 students:

#create data frame

```
df <- data.frame(currentGrade = c(82, 88, 75, 74, 93, 97, 83, 90, 90, 80),  
hours = c(4, 3, 6, 5, 4, 5, 8, 7, 4, 6),  
examScore = c(88, 85, 76, 70, 92, 94, 89, 85, 90, 93))
```

#view data frame

df

currentGrade hours examScore

1 82 4 88

2 88 3 85

3 75 6 76

4 74 5 70

5 93 4 92

6 97 5 94

7 83 8 89

8 90 7 85

9 90 4 90

10 80 6 93

To calculate the partial correlation between each pairwise combination of variables in the dataframe, we can use the `pcor()` function from the :

```
library(ppcor)
```

```
#calculate partial correlations
```

```
pcor(df)
```

```
$estimate
```

```
currentGrade hours examScore
```

```
currentGrade 1.0000000 -0.3112341 0.7355673
```

```
hours -0.3112341 1.0000000 0.1906258
```

```
examScore 0.7355673 0.1906258 1.0000000
```

```
$p.value
```

```
currentGrade hours examScore
```

```
currentGrade 0.00000000 0.4149353 0.02389896
```

```
hours 0.41493532 0.0000000 0.62322848
```

```
examScore 0.02389896 0.6232285 0.00000000
```

```
$statistic
```

```
currentGrade hours examScore
```

```
currentGrade 0.0000000 -0.8664833 2.8727185
hours -0.8664833 0.0000000 0.5137696
examScore 2.8727185 0.5137696 0.0000000
```

```
$n
```

```
10
```

```
$gp
```

```
1
```

```
$method
```

```
"pearson"
```

Here is how to interpret the output:

Partial correlation between hours studied and final exam score:

The partial correlation between hours studied and final exam score is .191, which is a small positive correlation. As hours studied increases, exam score tends to increase as well, assuming current grade is held constant.

The p-value for this partial correlation is .623, which is not statistically significant at $\alpha = 0.05$.

Partial correlation between current grade and final exam score:

The partial correlation between current grade and final exam score is .736, which is a strong positive correlation. As current grade increases, exam score tends to increase as well, assuming hours studied is held constant.

The p-value for this partial correlation is .024, which is statistically significant at $\alpha = 0.05$.

The partial correlation between current grade and hours studied and final exam score is -.311, which is a mild negative correlation. As current grade increases, final exam score tends to decrease, assuming final exam score is held constant.

The for this partial correlation is 0.415, which is not statistically significant at $\alpha = 0.05$.

The output also tells us that the method used to calculate the partial correlation was "pearson."

Within the `pcor()` function, we could also specify "kendall" or "pearson" as alternative methods to

calculate the correlations.

ARABPSYCHOLOGY.COM