How can the polychoric correlation be calculated in R?

Authored by

stats writer

May 12, 2024

The polychoric correlation is a statistical measure used to assess the relationship between two ordinal variables. In R, this can be calculated using the "polycor" package, which provides functions for estimating the polychoric correlation coefficient. To calculate the polychoric correlation in R, the user needs to first install and load the "polycor" package, then use the "polychor" function and specify the two ordinal variables of interest. The output of this function provides the estimated polychoric correlation coefficient, along with other relevant information such as p-value and confidence interval. This method allows for the assessment of the strength and direction of the relationship between two ordinal variables, making it a useful tool in various research and data analysis scenarios.

Calculate Polychoric Correlation in R

Polychoric correlation is used to calculate the correlation between ordinal variables.

Recall that are variables whose possible values are categorical and have a natural order.

Some examples of variables measured on an ordinal scale include:

Satisfaction: Very unsatisfied, unsatisfied, neutral, satisfied, very satisfiedIncome level: Low income, medium income, high incomeWorkplace status: Entry Analyst, Analyst I, Analyst II, Lead AnalystDegree of pain: Small amount, medium amount, high amount

The value for polychoric correlation ranges from -1 to 1 where:

-1 indicates a perfect negative correlation0 indicates no correlation1 indicates a perfect positive correlation

We can use the polychor(x, y) function from the polycor package to calculate the polychoric correlation between two ordinal variables in R.

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

Example 1: Calculate Polychoric Correlation for Movie Ratings

Suppose want to know whether or not two different movie ratings agencies have a high correlation between their movie ratings.

We ask each agency to rate 20 different movies on a scale of 1 to 3 where:

1 indicates "bad"2 indicates "mediocre"3 indicates "good"

We can use the following code in R to calculate the polychoric correlation between the ratings of the two agencies:

library(polycor)

```
#define movie ratings for each agency
agency1 <- c(1, 1, 2, 2, 3, 2, 2, 3, 2, 3, 3, 2, 1, 2, 2, 1, 1, 1, 2, 2)
agency2 <- c(1, 1, 2, 1, 3, 3, 3, 2, 2, 3, 3, 3, 2, 2, 2, 1, 2, 1, 3, 3)

#calculate polychoric correlation between ratings
polychor(agency1, agency2)

0.7828328
```

The polychoric correlation turns out to be 0.78.

This value is quite high, which indicates that there is a strong positive association between the ratings from each agency.

Example 2: Calculate Polychoric Correlation for Restaurant Ratings

We randomly survey 20 customers who ate at both restaurants and ask them to rate their overall satisfaction a scale of 1 to 5 where:

1 indicates "very unsatisfied"2 indicates "unsatisfied"3 indicates "neutral"4 indicates "satisfied"5 indicates "very satisfied"

We can use the following code in R to calculate the polychoric correlation between the ratings of the two restaurants:

```
library(polycor)

#define ratings for each restaurant
restaurant1 <- c(1, 1, 2, 2, 2, 3, 3, 3, 2, 2, 3, 4, 4, 5, 5, 4, 3, 4, 5, 5)
restaurant2 <- c(4, 3, 3, 4, 3, 3, 4, 5, 4, 4, 4, 5, 5, 4, 2, 1, 1, 2, 1, 4)

#calculate polychoric correlation between ratings
polychor(restaurant1, restaurant2)

-0.1322774
```

The polychoric correlation turns out to be -0.13.

This value is close to zero, which indicates that there is very little (if any) association between the ratings of the restaurants.

The following tutorials explain how to calculate other common correlation coefficients in R: