How can I use R to calculate the point-biserial correlation?

Authored by
stats writer

April 17, 2024

ARABPSYCHOLOGY.COM

RECOMMENDED CITATION

stats writer (2024). *How can I use R to calculate the point-biserial correlation?*. PSYCHOLOGICAL SCALES. Retrieved from https://scales.arabpsychology.com/?p=136487

R is a statistical programming language that is commonly used for data analysis and manipulation. One of the many functions of R is the ability to calculate the point-biserial correlation coefficient, which measures the strength and direction of the relationship between a continuous and a binary variable. To use R for this purpose, one needs to first import the necessary data and then use the "cor.test" function with the arguments specifying the continuous and binary variables. This will generate the correlation coefficient, along with its significance level and confidence interval. R provides a user-friendly and efficient way to calculate the point-biserial correlation, making it a useful tool for researchers and statisticians.

Calculate Point-Biserial Correlation in R

Point-biserial correlation is used to measure the relationship between a binary variable, x, and a continuous variable, y.

Similar to the , the point-biserial correlation coefficient takes on a value between -1 and 1 where:

-1 indicates a perfectly negative correlation between two variables0 indicates no correlation between two variables1 indicates a perfectly positive correlation between two variables

This tutorial explains how to calculate the point-biserial correlation between two variables in R.

Example: Point-Biserial Correlation in R

Suppose we have a binary variable, x, and a continuous

ARABPSYCHOLOGY.COM

variable, y:

x <- c(0, 1, 1, 0, 0, 0, 1, 0, 1, 1, 0)

y <- c(12, 14, 17, 17, 11, 22, 23, 11, 19, 8, 12)

We can use the built-in R function cor.test() to calculate the point-biserial correlation between the two variables:

#calculate point-biserial correlation
cor.test(x, y)

Pearson's product-moment correlation

data: x and y
t = 0.67064, df = 9, p-value = 0.5193

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:
-0.4391885 0.7233704

sample estimates:
cor
0.2181635

ARABPSYCHOLOGY.COM

From the output we can observe the following:

The point-biserial correlation coefficient is 0.218The corresponding p-value is 0.5193

Since the correlation coefficient is positive, this indicates that when the variable x takes on the value "1" that the variable y tends to take on higher values compared to when the variable x takes on the value "0."

However, since the p-value of this correlation is not less than .05, this correlation is not statistically significant.

Note that the output also provides a 95% confidence interval for the true correlation coefficient, which turns out to be:

95% C.I. = (-0.439, 0.723)

Since this confidence interval contains zero, this is further evidence that the correlation coefficient is not statistically significant.

Note: You can find the complete documentation for the cor.test() function .

ARABPSYCHOLOGY.COM

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

ARABPSYCHOLOGY.COM