

How can I calculate the coefficient of determination (R-squared) in R?

Authored by
stats writer

April 21, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I calculate the coefficient of determination (R-squared) in R?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=137680>

The coefficient of determination, also known as R-squared, is a statistical measure that indicates the proportion of the variance in the dependent variable that can be explained by the independent variable(s). In order to calculate the R-squared in R, one can use the "summary" function on a linear regression model object. This will provide the R-squared value as well as other important information about the model. Alternatively, the "rsq" function from the "psych" package can also be used to directly calculate the R-squared value. By understanding and utilizing these methods in R, one can accurately assess the strength of the relationship between variables in a linear regression model.

Find Coefficient of Determination (R-Squared) in R

The coefficient of determination (commonly denoted R^2) is the proportion of the variance in the response variable that can be explained by the explanatory variables in a regression model.

This tutorial provides an example of how to find and interpret R^2 in a regression model in R.

What is a Good R-squared Value?

Example: Find & Interpret R-Squared in R

Suppose we have the following dataset that contains data for the number of hours studied, prep exams taken, and exam score received for 15 students:

```
#create data frame
```

```
df <- data.frame(hours=c(1, 2, 2, 4, 2, 1, 5, 4, 2, 4, 4, 3, 6,  
5, 3),
```

```
prep_exams=c(1, 3, 3, 5, 2, 2, 1, 1, 0, 3, 4, 3, 2, 4, 4),  
score=c(76, 78, 85, 88, 72, 69, 94, 94, 88, 92, 90, 75, 96,  
90, 82))
```

```
#view first six rows of data frame  
head(df)
```

```
hours prep_exams score  
1 1 1 76  
2 2 3 78  
3 2 3 85  
4 4 5 88  
5 2 2 72  
6 1 2 69
```

The following code shows how to fit a multiple linear regression model to this dataset and view the model output in R:

```
#fit regression model  
model <- lm(score~hours+prep_exams, data=df)  
  
#view model summary  
summary(model)
```

Call:

```
lm(formula = score ~ hours + prep_exams, data = df)
```

Residuals:

```
Min 1Q Median 3Q Max
```

```
-7.9896 -2.5514 0.3079 3.3370 7.0352
```

Coefficients:

```
Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept) 71.8078 3.5222 20.387 1.12e-10 ***
```

```
hours 5.0247 0.8964 5.606 0.000115 ***
```

```
prep_exams -1.2975 0.9689 -1.339 0.205339
```

```
---
```

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 4.944 on 12 degrees of freedom

Multiple R-squared: 0.7237, Adjusted R-squared: 0.6776

F-statistic: 15.71 on 2 and 12 DF, p-value: 0.0004454

The R-squared of the model (shown near the very bottom of the output) turns out to be 0.7237.

This means that 72.37% of the variation in the exam scores can be explained by the number of hours

studied and the number of prep exams taken.

Note that you can also access this value by using the following syntax:

```
summary(model)$r.squared
```

0.7236545

How to Interpret the R-Squared Value

An R-squared value will always range between 0 and 1.

A value of 1 indicates that the explanatory variables can perfectly explain the variance in the response variable and a value of 0 indicates that the explanatory variables have no ability to explain the variance in the response variable.

In general, the larger the R-squared value of a regression model the better the explanatory variables are able to predict the value of the response variable.

Check out [this article](#) for details on how to determine whether or not a given R-squared value is considered "good" for a given regression model.

How to Calculate Adjusted R-Squared in R

ARABPSYCHOLOGY.COM