

How can I calculate the Residual Sum of Squares in R?

Authored by
stats writer

April 24, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I calculate the Residual Sum of Squares in R?*.

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=138685>

The Residual Sum of Squares (RSS) is a measure of the variation between the predicted values and the actual values in a regression model. In R, the RSS can be calculated by first fitting a regression model using the "lm" function and then using the "residuals" function to obtain the residuals of the model. These residuals can then be squared and summed to get the RSS value. This value can be used to assess the goodness of fit of the model and make comparisons between different models.

Calculate Residual Sum of Squares in R

A is the difference between an observed value and a predicted value in a regression model.

It is calculated as:

Residual = Observed value - Predicted value

One way to understand how well a regression model fits a dataset is to calculate the residual sum of squares, which is calculated as:

Residual sum of squares = $\sum(e_i)^2$

where:

Σ : A Greek symbol that means "sum" e_i : The i th residual

The lower the value, the better a model fits a dataset.

We can easily calculate the residual sum of squares for

a regression model in R by using one of the following two methods:

```
#build regression model
```

```
model <- lm(y ~ x1 + x2 + ..., data = df)
```

```
#calculate residual sum of squares (method 1)
```

```
deviance(model)
```

```
#calculate residual sum of squares (method 2)
```

```
sum(resid(model)^2)
```

Both methods will produce the exact same results.

The following example shows how to use these functions in practice.

Example: Calculating Residual Sum of Squares in R

For this example, we'll use the built-in mtcars dataset in R:

```
#view first six rows of mtcars dataset
```

```
head(mtcars)
```

```
mpg cyl disp hp drat wt qsec vs am gear carb
```

```
Mazda RX4 21.0 6 160 110 3.90 2.620 16.46 0 1 4 4
```

```
Mazda RX4 Wag 21.0 6 160 110 3.90 2.875 17.02 0 1 4 4
Datsun 710 22.8 4 108 93 3.85 2.320 18.61 1 1 4 1
Hornet 4 Drive 21.4 6 258 110 3.08 3.215 19.44 1 0 3 1
Hornet Sportabout 18.7 8 360 175 3.15 3.440 17.02 0 0 3
2
Valiant 18.1 6 225 105 2.76 3.460 20.22 1 0 3 1
```

The following code shows how to fit a multiple linear regression model for this dataset and calculate the residual sum of squares of the model:

```
#build multiple linear regression model
model <- lm(mpg ~ wt + hp, data = mtcars)

#calculate residual sum of squares (method 1)
deviance(model)

195.0478

#calculate residual sum of squares (method 2)
sum(resid(model)^2)

195.0478
```

If we have two competing models, we can calculate the

residual sum of squares for both to determine which one fits the data better:

#build two different models

```
model1 <- lm(mpg ~ wt + hp, data = mtcars)
```

```
model2 <- lm(mpg ~ wt + disp, data = mtcars)
```

#calculate residual sum of squares for both models

```
deviance(model1)
```

195.0478

```
deviance(model2)
```

246.6825

We can see that the residual sum of squares for model 1 is lower, which indicates that it fits the data better than model 2.

We can confirm this by calculating the of each model:

#build two different models

```
model1 <- lm(mpg ~ wt + hp, data = mtcars)
```

```
model2 <- lm(mpg ~ wt + disp, data = mtcars)
```

```
#calculate R-squared for both models  
summary(model1)$r.squared
```

0.8267855

```
summary(model2)$r.squared
```

0.7809306

The R-squared for model 1 turns out to be higher, which indicates that it's able to explain more of the variance in the response values compared to model 2.