



MENU

🕒 APRIL 6, 2020 👤 BY ZACH

How to Calculate RMSE in R

The **root mean square error (RMSE)** is a metric that tells us how far apart our predicted values are from our observed values in a regression analysis, on average. It is calculated as:

$$\text{RMSE} = \sqrt{[\Sigma(P_i - O_i)^2 / n]}$$

where:

- Σ is a fancy symbol that means “sum”
- P_i is the predicted value for the i^{th} observation in the dataset
- O_i is the observed value for the i^{th} observation in the dataset
- n is the sample size

This tutorial explains two methods you can use to calculate RMSE in R.

Method 1: Write Your Own Function

Suppose we have a dataset with one column that contains the actual data values and one column that contains the predicted data values:

```
#create dataset
```

```
data <- data.frame(actual=c(34, 37, 44, 47, 48, 48, 46, 43, 32, 27, 26, 24),  
                    predicted=c(37, 40, 46, 44, 46, 50, 45, 44, 34, 30, 22, 23))
```

```
#view dataset
```

```
data
```

	actual	predicted
1	34	37
2	37	40
3	44	46
4	47	44
5	48	46
6	48	50
7	46	45
8	43	44
9	32	34
10	27	30
11	26	22
12	24	23

To compute the RMSE, we can use the following function:

```
#calculate RMSE
```

```
sqrt(mean((data$actual - data$predicted)^2))
```

```
[1] 2.43242
```

The root mean square error is **2.43242**.

Method 2: Use a Package

We could also calculate RMSE for the same dataset using the **rmse()** function from the **Metrics** package, which uses the following syntax:

```
rmse(actual, predicted)
```

where:

- **actual:** actual values
- **prediced:** predicted values

Here is the syntax we would use in our example:

```
#load Metrics package
library(Metrics)

calculate RMSE
rmse(data$actual, data$predicted)

[1] 2.43242
```

The root mean square error is **2.43242**, which matches what we calculated earlier using our own function.

How to Interpret RMSE

RMSE is a useful way to see how well a regression model is able to fit a dataset.

The larger the RMSE, the larger the difference between the predicted and observed values, which means the worse a regression model fits the data. Conversely, the smaller the RMSE, the better a model is able to fit the data.

It can be particularly useful to compare the RMSE of two different models with each other to see which model fits the data better.

Additional Resources