How do I calculate the residual standard error in R?

Authored by

stats writer

April 20, 2024

The residual standard error in R is a measure of the variation of the data points around the fitted regression line. It is calculated by taking the square root of the sum of squared residuals, divided by the degrees of freedom. This value represents the average distance between the observed data points and the predicted values from the regression model. To calculate the residual standard error in R, you can use the "residuals()" function to obtain the residuals, and then use the "sqrt()" and "sum()" functions to compute the square root and sum of squared residuals. Finally, divide this value by the degrees of freedom, which can be found using the "df.residual()" function. This will give you the residual standard error value, which can be used as an indicator of the accuracy of the regression model.

Calculate Residual Standard Error in R

Whenever we fit a linear regression model in R, the model takes on the following form:

$Y = \beta 0 + \beta 1X + ... + \beta iX + ?$

where ? is an error term that is independent of X.

No matter how well X can be used to predict the values of Y, there will always be some random error in the model. One way to measure the dispersion of this random error is to use the residual standard error, which is a way to measure the standard deviation of the residuals ?.

The residual standard error of a regression model is calculated as:

Residual standard error = √SSresiduals / dfresiduals

where:

SSresiduals: The residual sum of squares.dfresiduals: The residual degrees of freedom, calculated as n - k - 1 where n = total observations and k = total model parameters.

There are three methods we can use to calculate the residual standard error of a regression model in R.

Method 1: Analyze the Model Summary

The first way to obtain the residual standard error is to simply fit a linear regression model and then use the summary() command to obtain the model results. Then, just look for "residual standard error" near the bottom of the output:

```
#load built-in mtcars dataset
data(mtcars)

#fit regression model
model <- lm(mpg~disp+hp, data=mtcars)

#view model summary
```

```
summary(model)

Call:
lm(formula = mpg ~ disp + hp, data = mtcars)

Residuals:
Min 1Q Median 3Q Max
-4.7945 -2.3036 -0.8246 1.8582 6.9363

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 30.735904 1.331566 23.083 < 2e-16 ***
disp -0.030346 0.007405 -4.098 0.000306 ***
hp -0.024840 0.013385 -1.856 0.073679 .
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.127 on 29 degrees of freedom
Multiple R-squared: 0.7482, Adjusted R-squared: 0.7309
F-statistic: 43.09 on 2 and 29 DF, p-value: 2.062e-09
```

We can see that the residual standard error is 3.127.

Method 2: Use a Simple Formula

Another way to obtain the residual standard error (RSE) is to fit a linear regression model and then use the following formula to calculate RSE:

sqrt(deviance(model)/df.residual(model))

Here is how to implement this formula in R:

```
#load built-in mtcars dataset
data(mtcars)

#fit regression model
model <- lm(mpg~disp+hp, data=mtcars)

#calculate residual standard error
sqrt(deviance(model)/df.residual(model))
```

3.126601

We can see that the residual standard error is 3.126601.

Method 3: Use a Step-By-Step Formula

Another way to obtain the residual standard error is to fit a linear regression model and then use a step-by-step approach to calculate each individual component

of the formula for RSE:

```
#load built-in mtcars dataset
data(mtcars)

#fit regression model
model <- lm(mpg~disp+hp, data=mtcars)

#calculate the number of model parameters - 1
k=length(model$coefficients)-1

#calculate sum of squared residuals
SSE=sum(model$residuals**2)

#calculate total observations in dataset
n=length(model$residuals)

#calculate residual standard error
sqrt(SSE/(n-(1+k)))

3.126601
```

We can see that the residual standard error is 3.126601.

How to Interpret the Residual Standard Error

As mentioned before, the residual standard error (RSE)

is a way to measure the standard deviation of the residuals in a regression model.

The lower the value for RSE, the more closely a model is able to fit the data (but be careful of overfitting). This can be a useful metric to use when comparing two or more models to determine which model best fits the data.

How to Interpret Residual Standard Error
How to Perform Multiple Linear Regression in R
How to Perform Cross Validation for Model Performance in R
How to Calculate Standard Deviation in R