How can the BIC (Bayesian Information Criterion) be calculated in R?

Authored by

stats writer

May 6, 2024

The Bayesian Information Criterion (BIC) is a statistical measure used in model selection to evaluate the goodness of fit of a model. It takes into account both the complexity of the model and the goodness of fit to determine the best model among a set of competing models. In order to calculate the BIC in R, the user must first specify the model using the appropriate functions and then use the BIC function from the package "stats". The BIC function calculates the BIC value based on the likelihood of the model, the number of parameters, and the number of observations. This value can then be compared to other models to determine the most appropriate model for the given data. The BIC is a useful tool for selecting the most parsimonious and accurate model in various fields such as economics, psychology, and biology.

Calculate BIC in R

The Bayesian Information Criterion, often abbreviated BIC, is a metric that is used to compare the goodness of fit of different regression models.

In practice, we fit several regression models to the same dataset and choose the model with the lowest BIC value as the model that best fits the data.

We use the following formula to calculate BIC:

BIC: (RSS+log(n)dσ?2) / n

where:

d: The number of predictorsn: Total observationsσ?: Estimate of the variance of the error associate with each response measurement in a regression

modelRSS: Residual sum of squares of the regression modelTSS: Total sum of squares of the regression model

The following step-by-step example shows how to calculate BIC values for regression models in R.

Step 1: View the Data

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

#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

Step 2: Fit Several Models

Next, we'll fit several different regression models using this dataset:

```
#fit three different regression models
model1 <- lm(mpg ~ disp + hp, data = mtcars)model2 <- lm(mpg ~ disp + qsec, data = mtcars)model3 <- lm(mpg ~ disp + wt, data = mtcars)
```

Step 3: Choose Model with Lowest BIC

To calculate the BIC value for each model, we can use the BIC() function from the flexmix package:

```
library(flexmix)

#calculate BIC of model1
BIC(model1)
```

174.4815

```
#calculate BIC of model2
BIC(model2)
```

177.7048

```
#calculate BIC of model3
BIC(model3)
```

170.0307

We can see the BIC values for each model:

BIC of model 1: 174.4815BIC of model 2: 177.7048BIC of model 3: 170.0307

Since model 3 has the lowest BIC value, we will choose it as the model that best fits the dataset.

The following tutorials explain how to fit common regression models in R: