

How can AIC be calculated in R, and can you provide examples?

Authored by
stats writer

April 30, 2024

RECOMMENDED CITATION

stats writer (2024). *How can AIC be calculated in R, and can you provide examples?*.
PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=141376>

AIC (Akaike Information Criterion) is a statistical measure used for model selection in order to determine the best fitting model for a given dataset. In R, AIC can be calculated using the "AIC" function, which takes in the model object as its argument. This function returns the AIC value for the specified model. Additionally, R also offers the "stepAIC" function which automates the process of selecting the best model by comparing AIC values for different models. This function can be used for both linear and generalized linear models. For example, to calculate the AIC for a linear regression model in R, the following code can be used:

```
model <- lm(dependent variable ~ independent variable, data = dataset)
AIC(model)
```

Similarly, the AIC value for a generalized linear model can be calculated using the "AIC" function after fitting the model using the appropriate function, such as "glm" for logistic regression. Overall, AIC is a useful tool in model selection and building and can be easily calculated in R using built-in functions.

Calculate AIC in R (Including Examples)

The Akaike information criterion (AIC) is a metric that is used to compare the fit of several regression models.

It is calculated as:

$$\text{AIC} = 2K - 2\ln(L)$$

where:

K: The number of model parameters. The default value of K is 2, so a model with just one predictor variable will have a K value of 2+1 = 3. **$\ln(L)$:** The log-likelihood of the model. Most statistical software can automatically calculate this value for you.

The AIC is designed to find the model that explains the most variation in the data, while penalizing for models that use an excessive number of parameters.

Once you've fit several regression models, you can compare the AIC value of each model. The lower the AIC, the better the model fit.

To calculate the AIC of several regression models in R, we can use the `aictab()` function from the `AICcmodavg` package.

The following example shows how to use this function to calculate and interpret the AIC for various regression models in R.

Example: Calculate & Interpret AIC in R

Suppose we would like to fit three different using variables from the `mtcars` dataset.

Here are the predictor variables we'll use in each model:

Predictor variables in Model 1: `disp`, `hp`, `wt`,
qsec
Predictor variables in Model 2: `disp`, `qsec`
Predictor variables in Model 3: `disp`, `wt`

The following code shows how to fit each of these regression models:

```
#fit three models
```

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

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

```
model3 <- lm(mpg ~ disp + wt, data = mtcars)
```

Next, we'll put the models into a list and use the `aictab()` function to calculate the AIC of each model:

```
library(AICcmodavg)
```

```
#define list of models
```

```
models <- list(model1, model2, model3)
```

```
#specify model names
```

```
mod.names <- c('disp.hp.wt.qsec', 'disp.qsec', 'disp.wt')
```

```
#calculate AIC of each model
```

```
aictab(cand.set = models, modnames = mod.names)
```

Model selection based on AICc:

```
K AICc Delta_AICc AICcWt Cum.Wt LL
```

disp.hp.wt.qsec 6 162.43 0.00 0.83 0.83 -73.53

disp.wt 4 165.65 3.22 0.17 1.00 -78.08

disp.qsec 4 173.32 10.89 0.00 1.00 -81.92

K: The number of parameters in the model.
AICc: The AIC value of the model. The lowercase 'c' indicates that the AIC has been calculated from the AIC corrected for small sample sizes.
Delta_AICc: The difference between the AIC of the best model compared to the current model being compared.
AICcWt: The proportion of the total predictive power that can be found in the model.
Cum.Wt: The cumulative sum of the AIC weights.
LL: The log-likelihood of the model. This tells us how likely the model is, given the data we used.

The model with the lowest AIC value is always listed first. From the output we can see that the following model has the lowest AIC value and is thus the best fitting model:

$$\text{mpg} = \beta_0 + \beta_1(\text{disp}) + \beta_2(\text{hp}) + \beta_3(\text{wt}) + \beta_4(\text{qsec})$$

Once we've identified this model as the best, we can proceed to fit the model and analyze the results including the R-squared value and the beta coefficients

to determine the exact relationship between the set of predictor variables and the .

ARABPSYCHOLOGY.COM