

How can I calculate AIC in SAS and can you provide an example?

Authored by
stats writer

June 24, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I calculate AIC in SAS and can you provide an example?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=149971>

The Akaike Information Criterion (AIC) is a statistical measure used to evaluate the quality of a statistical model. It is commonly used in the field of data analysis to select the best model among a set of competing models. In SAS, AIC can be calculated using the PROC GLMSELECT procedure, which allows for the comparison of multiple models based on their AIC values. An example of AIC calculation in SAS would involve selecting a dataset, specifying the model and variables of interest, and then using the AIC option in the PROC GLMSELECT statement to generate the AIC values for each model. This information can then be used to determine the model with the lowest AIC, indicating the best fit for the data.

Calculate AIC in SAS (With Example)

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.

The following example shows how to calculate the AIC for various regression models in SAS.

Example: How to Calculate AIC in SAS

Suppose we would like to fit three different to predict the exam score that students will receive in some class.

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

Predictor variables in Model 1: hours spent studying
Predictor variables in Model 2: practice exams taken
Predictor variables in Model 3: hours spent studying and practice exams taken

First, we'll use the following code to create a dataset that contains this information for 20 students:

```
/*create dataset*/  
data exam_data;  
input hours prep_exams score;  
datalines;
```

```
1 1 76  
2 3 78  
2 3 85  
4 5 88  
2 2 72  
1 2 69  
5 1 94  
4 1 94  
2 0 88  
4 3 92  
4 4 90  
3 3 75  
6 2 96  
5 4 90  
3 4 82  
4 4 85  
6 5 99  
2 1 83  
1 0 62  
2 1 76  
;  
run;
```

Next, we'll use proc reg to fit each of these regression

models and we'll use the statement `selection=adjrsq sse aic` to calculate the AIC values for each model:

```
/*fit multiple linear regression models and calculate AIC
for each model*/proc regdata=exam_data;
model score = hours prep_exams / selection=adjrsq sse
aic;
run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: score

Adjusted R-Square Selection Method

Number of Observations Read	20
Number of Observations Used	20

Number in Model	Adjusted R-Square	R-Square	AIC	SSE	Variables in Model
1	0.7121	0.7273	68.4537	501.90937	hours
2	0.7027	0.7340	69.9507	489.44312	hours prep_exams
1	0.0888	0.1368	91.4967	1588.53461	prep_exams

AIC with hours as predictor variable: 68.4537
AIC with hours and exams as predictor variables: 69.9507
AIC with exams as predictor variable: 91.4967

The model with the lowest AIC value is the one that only

contains hours as the predictor variable.

Thus, we would declare the following model to be the one that best fits the data:

$$\text{Score} = \beta_0 + \beta_1(\text{Hours Studied})$$

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 hours studied and final exam score.

The following tutorials explain how to perform other common tasks in SAS: