How can logistic regression be performed in SAS?

Authored by

stats writer

July 1, 2024

Logistic regression is a statistical method used to model the relationship between a categorical dependent variable and one or more independent variables. In SAS (Statistical Analysis System), logistic regression can be performed by using the PROC LOGISTIC procedure. This procedure allows users to specify the dependent and independent variables, as well as any necessary data transformations and model specifications. SAS also provides various options for model diagnostics and validation, making it a comprehensive tool for conducting logistic regression analysis. Overall, SAS offers a straightforward and efficient approach to performing logistic regression for data analysis and predictive modeling.

Perform Logistic Regression in SAS

Logistic regression is a method we can use to fit a regression model when the response variable is binary.

Logistic regression uses a method known as *maximum likelihood estimation* to find an equation of the following form:

$$\log = \beta 0 + \beta 1X1 + \beta 2X2 + ... + \beta pXp$$

where:

Xj: The jth predictor variableβj: The coefficient estimate for the jth predictor variable

The formula on the right side of the equation predicts the log odds of the response variable taking on a value of 1.

The following step-by-step example shows how to fit a logistic regression model in SAS.

Step 1: Create the Dataset

First, we'll create a dataset that contains information on the following three variables for 18 students:

Acceptance into a certain college (1 = yes, 0 = no)GPA (scale of 1 to 4)ACT score (scale of 1 to 36)

```
/*create dataset*/
data my_data;
input acceptance gpa act;
datalines;
1 3 30
0 1 21
0 2 26
0 1 24
1 3 29
1 3 34
0 3 31
1 2 29
0 1 21
1 2 21
```

```
0 1 15
1 3 32
1 4 31
1 4 29
0 1 24
1 4 29
1 3 21
1 4 34
;
run;

/*view dataset*/
proc printdata=my_data;
```

Obs	acceptance	gpa	act
1	1	3	30
2	0	1	21
3	0	2	26
4	0	1	24
5	1	3	29
6	1	3	34
7	0	3	31
8	1	2	29
9	0	1	21
10	1	2	21
11	0	1	15
12	1	3	32
13	1	4	31
14	1	4	29
15	0	1	24
16	1	4	29
17	1	3	21
18	1	4	34

Step 2: Fit the Logistic Regression Model

Next, we'll use proc logistic to fit the logistic regression model, using "acceptance" as the response variable and "gpa" and "act" as the predictor variables.

Note: We must specify descending so SAS knows to predict the probability that the response variable will take on a value of 1. By default, SAS predicts the probability that the response variable will take on a value of 0.

```
/*fit logistic regression model*/
proc logisticdata=my_data descending;
model acceptance = gpa act;
run;
```

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	26.057	16.595
SC	26.947	19.266
-2 Log L	24.057	10.595

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	13.4620	2	0.0012
Score	10.5311	2	0.0052
Wald	5.2807	2	0.0713

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-3.2839	4.2665	0.5924	0.4415
gpa	1	2.9665	1.6250	3.3324	0.0679
act	1	-0.1145	0.2369	0.2336	0.6289

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
gpa	19.423	0.804	469.398
act	0.892	0.561	1.419

From this table we can see the AIC value of the model, which turns out to be 16.595. The lower the AIC value,

the better a model is able to fit the data.

However, there is no threshold for what is considered a . Rather, we use AIC to compare the fit of several models fit to the same dataset. The model with the lowest AIC value is generally considered the best.

The next table of interest is titled Testing Global Null Hypothesis: BETA=0.

From this table we can see the Likelihood Ratio Chi-square value of 13.4620 with a corresponding p-value of 0.0012.

Since this p-value is less than .05, this tells us that the logistic regression model as a whole is statistically significant.

Next, we can analyze the coefficient estimates in the table titled Analysis of Maximum Likelihood Estimates.

From this table we can see the coefficients for gpa and act, which indicate the average change in log odds of getting accepted into the university for a one unit increase in each variable.

For example:

A one-unit increase in GPA value is associated with an average increase of 2.9665 in the log odds of getting accepted into the university.A one-unit increase in ACT score is associated with an average *decrease* of 0.1145 in the log odds of getting accepted into the university.

The corresponding p-values in the output also give us an idea of how effective each predictor variable is at predicting the probability of getting accepted:

P-value of GPA: 0.0679P-value of ACT: 0.6289

This tells us that GPA seems to be a statistically significant predictor of university acceptance while ACT score seems to not be statistically significant.

Additional Resources

The following tutorials explain how to fit other regression models in SAS: