

How can I create a ROC curve in SAS for my data analysis?

Authored by
stats writer

June 29, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I create a ROC curve in SAS for my data analysis?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=160117>

Creating a Receiver Operating Characteristic (ROC) curve in SAS for data analysis involves plotting the true positive rate against the false positive rate for a given set of data. This curve helps in evaluating the performance of a predictive model and determining the optimal threshold for classification. To create a ROC curve in SAS, one must first import the data, then use the PROC LOGISTIC procedure to fit a logistic regression model. Next, the PROC ROCCURVE procedure is used to generate the ROC curve and calculate the area under the curve (AUC). This information can then be used to assess the accuracy of the model and make informed decisions. Overall, creating a ROC curve in SAS is a useful tool for analyzing data and evaluating the effectiveness of predictive models.

Create a ROC Curve in SAS

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

To assess how well a logistic regression model fits a dataset, we can look at the following two metrics:

Sensitivity: The probability that the model predicts a positive outcome for an observation when indeed the outcome is positive. This is also called the "true positive rate."**Specificity:** The probability that the model predicts a negative outcome for an observation when indeed the outcome is negative. This is also called the "true negative rate."

One way to visualize these two metrics is by creating a ROC curve, which stands for "receiver operating characteristic" curve. This is a plot that displays the

sensitivity and specificity of a logistic regression model.

The following step-by-step example shows how to create and interpret a ROC curve in SAS.

Step 1: Create the Dataset

First, we'll create a dataset that contains information on the following 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;
```

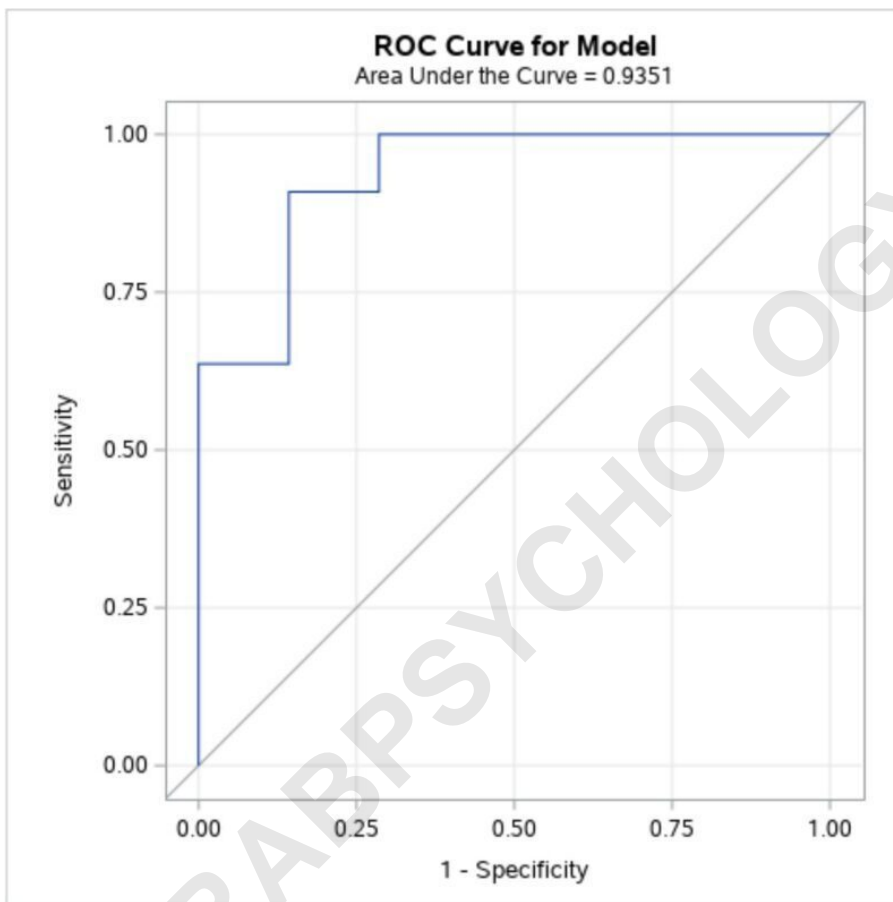
Step 2: Fit the Logistic Regression Model & Create ROC Curve

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.

We will specify `descending` so SAS knows to predict the probability that the response variable will take on a value of 1.

We will also use `plots(only)=roc` to create the ROC curve for the model:

```
/*fit logistic regression model & create ROC curve*/  
proc logistic data=my_data descending  
plots(only)=roc; model acceptance = gpa act;  
run;
```



Step 3: Interpret the ROC Curve

The more that the ROC curve hugs the top left corner of the plot, the better the model does at predicting the value of the response values in the dataset.

To quantify how well the logistic regression model fits the data, we can calculate the AUC - area under the curve - which tells us how much of the plot is located under the curve.

The closer AUC is to 1, the better the model. A model with an AUC equal to 0.5 is no better than a model that makes random classifications.

Under the title of the plot above we can see that the AUC for this model is 0.9351.

Since this value is close to one, it confirms that the model does a good job of predicting the value of the response values.

We can also use the AUC value to compare the fit of different logistic regression models.

For example, suppose we fit two different logistic regression models and calculate the AUC values for each:

AUC of model 1: 0.9351 AUC of model 2: 0.8140

Since the AUC of model 1 is greater, this tells us that

model 1 does a better job of fitting the data than model 2.

Additional Resources

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

ARABPSYCHOLOGY.COM