

Demo: Fitting a Simple Logistic Regression Model

In this video, we use the Impurity Logistic data to fit a logistic regression model for Outcome and Catalyst Conc using Fit Y by X.

We select Outcome as Y, Response and Catalyst Conc as X, Factor. The analysis key, in the bottom left corner, tells us that JMP will conduct a logistic regression analysis.

We also see that the target level, which is the outcome we are interested in modeling, is Fail. By default, JMP models the probability of the outcome that comes first alphanumerically. If desired, we can change the target category using this field. In this example, we'll model the probability of Fail.

Notice that we can also change the value that is modeled in all analyses by setting the Value Ordering column property. To do this, we right-click on the column name in the data table, select Column Properties, and then select Value Ordering. Then, we can select the values and use the Move Up or Move Down buttons to reorder the column.

Fail is already first in the list, so we'll simply cancel, return to the analysis dialog, and click OK to run the analysis.

The default graph is the logistic curve. From the curve, we can see that as the value of Catalyst Conc increases, the probability of Fail also increases. From the steepness of the curve, we can see that the relationship is fairly strong.

The data points are plotted at the value of Catalyst Conc (on the X axis), and appear either above or below the curve based on Outcome. Batches that failed to meet the Impurity specification are plotted below the curve and are randomly scattered relative to the Y axis.

The coefficients in the logistic model, the logit, are provided in the parameter estimates table. The positive coefficient for Catalyst Conc tells us that as the value of Catalyst Conc increases, the probability of Fail increases.

The p-values in the Parameter Estimates table and in the Whole Model Test confirm that the logistic model with Catalyst Conc is significant.

We can use this model to predict the outcome for different values of Catalyst Conc. When we select Save Probability Formula from the top red triangle, JMP adds four new columns to the data table.

Lin[Fail] is the logit, that is, the logistic model. When we look at the formula, we see the intercept and slope coefficients that are reported in the Parameter Estimates table.

The next two columns, Prob[Fail] and Prob[Pass], are predicted probabilities that are calculated from the logit. In the first row, we see that Prob[Fail], based on a Catalyst Conc of 1.5, is 0.11.

The fourth column, Most Likely Outcome, classifies Outcome as Fail or Pass, based on the outcome with the highest predicted probability.

In the first row, Prob[Pass] is the highest, so Outcome is classified as Pass.

Returning to the analysis window, we can look at the misclassification rate in the Fit Details report to see how well our model performs.

For the logistic model with only Catalyst Conc, the overall misclassification rate is 15%. That is, 15% of the observations were incorrectly classified by the model.

This simple logistic regression model is wrong 15% of the time. Can we do better? In the next video, we build a more complex model to see whether we can improve the overall misclassification rate.

Statistical Thinking for Industrial Problem Solving

Copyright © 2020 SAS Institute Inc., Cary, NC, USA. All rights reserved.

Close