Print

Simple Logistic Regression Example

In a previous lesson, we introduced the Impurity example. Recall that, in this scenario, a polymer is being produced. A catalyst is required for chemical reactions to occur to produce the polymer. The catalyst contains a chemical that can create an impurity. The upper specification limit for the impurity for a batch of polymer is 7.0%.

In the multiple linear regression lesson, we fit a model for Impurity, and used regression to identify the important predictors and interactions.

What if we only had a measure of whether a batch passed or failed? That is, what if the only data we collected regarding the quality of each batch of polymer was a pass or fail measure?

We can use logistic regression to model the probability that a batch will fail as a function of the available predictors. For illustration, we create a new variable, Outcome.

If the Impurity is greater than 7.0%, the Outcome is Fail. If the Impurity is less than or equal to 7.0%, the Outcome is Pass. We've saved the file with this new variable, Outcome, as Impurity Logistic.jmp.

In this video, we'll use simple logistic regression to model the probability that a batch will fail as a function of one predictor. Later, we'll see how to include multiple predictors in the same model using multiple logistic regression.

Before we proceed, remember that continuous response measures have more information content than categorical measures, and it's never a good idea to throw away raw measurements.

Statistical methods using continuous measures generally require much less data, and provide richer graphical and statistical results than analogous methods using categorical measures.

Let's return to our example. We will use logistic regression to understand the relationship between Outcome and one predictor, Temp.

The logistic curve has an upward slope, indicating that as Temp increases, the probability of Fail increases. The data points are plotted at the value of Temp (on the X axis) and appear either above or below the curve based on the outcome.

Batches that failed to meet the Impurity specification are plotted below the curve and are randomly scattered relative to the Y axis. This is easier to see when we change the markers and marker colors.

All of the points below the curve are for batches that failed to meet the impurity specification, and all of the points above the curve passed.

Note that behind the scenes, JMP recodes the response, Outcome, to a binary 0/1 response. This scatterplot is created to allow us to more easily see the distribution of the observations as a function of Temp.

The intercept and slope for the logistic model are reported in the parameter estimates table. The slope coefficient is positive, confirming what we see in the logistic curve.

In the next video, we take a closer look at logistic regression analysis results and see how to use the logistic model to calculate predicted probabilities.

Statistical Thinking for Industrial Problem Solving

Close