Print

Logistic Regression with Interactions

In the previous videos, we fit logistic regression models with only main effects. But it is possible that there are important interaction effects. In other words, it is possible that the effect of one variable on the predicted probability depends on the values of other variables.

As we saw in multiple linear regression, we can add interaction terms to the model.

In this video, we see how to fit a model with interactions using a new example, Metal Coatings. In this scenario, we work for a company that produces metal parts.

The parts are coated with an enamel paint to prevent rust. Workers visually inspect the parts for uniformity of the coating, and defective parts fail the inspection. We're tasked with improving the coating process so that fewer parts fail. We've compiled historical data for a random sampling of 405 parts produced within the past 60 days.

The response variable of interest is the Outcome (defective or good), and we have data for seven input variables: Humidity in the factory, Paint Viscosity, Pump Pressure, Part Temp, Belt Speed, Nozzle Size, and the Paint Supplier. We use logistic regression to identify potential causes of defective parts.

We are interested in studying the input variables, but we believe that there may be important two-way interactions. We fit a model with all of the main effects and all two-way interaction terms, and slowly remove nonsignificant terms from the model.

Because we are using logistic regression here to identify potential causes, we use a p-value stopping rule for removal of 0.10 rather than 0.05 to retain more terms in the model.

Here's the reduced model. The most significant effects are two of the interaction terms. There is a highly significant interaction between Nozzle Size and Paint Supplier. This means that the effect of Nozzle Size on the probability that a batch is Defective depends on the Paint Supplier. There is also a highly significant interaction between Paint Viscosity and Belt Speed.

We can visualize these interactions in the Prediction Profiler. Let's focus on the interaction between Nozzle Size and Paint Supplier. When the Nozzle Size is large, the defective rate for the Paint Supplier Magic is high. But when the nozzle size is small, the defective rate for Magic is much lower. This implies that the best nozzle to use, to maintain uniform coating of the parts, depends on the paint supplier.

This example demonstrates how we can use logistic regression to identify potential causes of industrial problems.

Next, we see how to fit and interpret logistic regression models with interactions in JMP. We return to this example in a Think About It exercise at the end of this lesson.

Close