

## What Is Logistic Regression?

In previous lessons, we introduced simple and multiple linear regression. In simple linear regression, the response is continuous and the predictor is continuous. In multiple linear regression, the response is continuous and the predictors can be continuous or categorical.

Logistic regression is used to model categorical response data. When multiple predictors are used, the predictors can be continuous or categorical.

Categorical responses are often binary or two-level nominal data. Common examples are outcomes, such as defective or non-defective, pass or fail, and true or false. Categorical responses might also be multinomial or ordinal.

Multinomial data has multiple unordered categories, such as the type of defect or the reason for failure. Ordinal data has ordered categories, such as severity ratings, degree of agreement scales, or sizes.

In this lesson, we focus on fitting logistic regression models for binary response variables.

As we will see in the following videos, logistic regression differs from least squares regression in a number of ways. A fundamental difference is that the logistic regression model predicts the probability of a particular response category rather than predicting a mean response value.

For example, we might estimate the probability that a part will be defective given the operating and environmental conditions, or we might want to estimate the probability that a customer places an order as a function of the quoted price and the estimated time to delivery.

Because the model predicts probabilities, the predicted value for a category must fall between zero and one, and the predicted probabilities for the different response categories must sum to one.

Another core difference is related to the concept of modeling error. In a linear model, we can compute a predicted value and calculate the error. The error is the difference between the actual response value and the predicted value. This error term is reflected in the linear regression model.

In logistic regression with a binary response, observations fall into one of two possible categories. The model either predicts the correct category or it doesn't. So, there is no error term in the logistic regression model.

However, logistic models estimate probabilities of membership in categories. Like all statistical models, there is some degree of uncertainty in our predictions.