

Logistic regression predicts the probability of the outcome variable being true.

In this example, a logistic regression model would predict the probability that the patient is receiving poor care.

Or if we denote the PoorCare variable by y , the probability that $y = 1$.

For the remainder of this lecture, we will denote poor care by 1, and good care by zero.

So, since the outcome is either 0 or 1, this means that the probability that the outcome variable is 0 is just 1 minus the probability that the outcome variable is 1.

So by predicting the probability that $y = 1$, we also get the probability that $y = 0$.

Just like in linear regression, we have a set of independent variables, x_1 through x_k , where k is the total number of independent variables we have.

Then to predict the probability that $y = 1$, we use what's called the Logistic Response Function.

This seems like a complicated, nonlinear equation, but you can see the familiar linear regression equation in this Logistic Response Function.

The Logistic Response Function is used to produce a number between 0 and 1.

Let's understand this function a little better.

This plot shows the logistic response function for different values of the linear regression piece.

The logistic response function always takes values between 0 and 1, which makes sense, since it equals a probability.

A positive coefficient value for a variable increases the linear regression piece, which increases the probability that $y = 1$, or increases the probability of poor care.

On the other hand, a negative coefficient value for a variable decreases the linear regression piece, which in turn decreases the probability that $y = 1$, or increases the probability of good care.

The coefficients, or betas, are selected to predict a high probability for the actual poor care cases, and to predict a low probability for the actual good care cases.

Another useful way to think about the logistic response function is in terms of Odds, like in gambling.

The Odds are the probability of 1 divided by the probability of 0.

The Odds are greater than 1 if 1 is more likely, and less than 1 if 0 is more likely.

The Odds are equal to 1 if the outcomes are equally likely.

If you substitute the Logistic Response Function for the probabilities in the Odds equation on the previous slide, you can see that the Odds are equal to "e" raised to the power of the linear regression equation.

By taking the log of both sides, the $\log(\text{Odds})$, or what we call the Logit, looks exactly like the linear regression equation.

This helps us understand how the coefficients, or betas, affect our prediction of the probability.

A positive beta value increases the Logit, which in turn increases the Odds of 1.

A negative beta value decreases the Logit, which in turn, decreases the Odds of one.

In the next video, we'll build a logistic regression model in R, and predict the quality of care.