Let us examine how to interpret the model we developed.

One of the things we should look after is that there might be what is called multicollinearity.

Multicollinearity occurs when the various independent variables are correlated, and this might confuse the coefficients-- the betas-- in the model.

So tests to address that involve checking the correlations of independent variables.

If they are excessively high, this would mean that there might be multicollinearity, and you have to potentially revisit the model, as well as whether the signs of the coefficients make sense.

Is the coefficient beta positive or negative?

If it agrees with intuition, then multicollinearity has not been a problem, but if intuition suggests a different sign, this might be a sign of multicollinearity.

The next important element is significance.

So how do we interpret the results, and how do we understand whether we have a good model or not?

For that purpose, let's take a look at what is called Area Under the Curve, or AUC for short.

So the Area Under the Curve shows an absolute measure of quality of prediction-- in this particular case, 77.5%, which means that, given that the perfect score is 100%, so this is like a B, whereas, as we'll see later, a 50% score, which is pure guessing, is a 50% rate of success.

So the area under the curve gives an absolute measure of quality, and it's less affected by various benchmarks.

So it illustrates how accurate the model is on a more absolute sense.

So what is a good AUC?

The area on the right shows the maximum possible of a perfect prediction, whereas the area on this curve now-- it is 0.5, and it's pure guessing.

Other outcome measures that are important for us to discuss is the so-called confusion matrix.

So the matrix here is formulas for the various terms we use.

The actual class is 0-- means, in our example, good quality of care, and actual class = 1 means poor quality of

care, whereas the predicted class = 0 means that will predict good quality, and the predicted class = 1 mean that we predict poor quality.

So we define true negatives, short by TN.

False positives, short by FP.

False negatives, FN, and true positives by TP.

So if N is the number of observations, the overall accuracy is basically the number of true negatives and true positives divided by N.

It's basically the terms in the diagonal of this two by two matrix divided by the total observations.

The overall error rate is the terms off-diagonal-- the false positives, plus the false negatives, divided by the total number of observations.

That's the overall measure of an error rate.

An important component is the so-called sensitivity, and sensitivity is TP, the true positives, whenever we predict poor quality, and indeed it is poor quality, divided by TP, these true positives, plus FN, which is the total number of cases of poor quality.

So this is the total number of times that we predict poor quality, and it is, indeed, poor quality, versus the total number of times the actual quality is, in fact, poor.

False negative rate is FN, the number of false negatives, divided by the number of true positives, plus the number of false negatives.

And specificity is TN, true negatives, the number of times we predict the quality is good, and, in fact, the quality is good, divided by this number, TN, plus false positives.

So specificity is the number of times we predict the quality is good, and it is indeed good, versus the total times we have good quality, and the false positive error rate is 1 minus the specificity.

So in this particular example that we have discussed, quality of care, just like in linear regression, we want to make predictions on a test set to compute out-of-sample metrics.

We develop the logistic regression model using data, but would like to make predictions out-of-sample.

So in our test, we utilized 32 cases, and the R command that makes the statements about the quality of a

prediction out-of-sample is illustrated here in the slide.

So in that way, we make predictions about probabilities, of course, simply because logistic regression makes predictions about probabilities, and then we transform them to a binary outcome-- the quality is good, or the quality is poor-- using a threshold.

In this particular example, we used a threshold value of 0.3, and in doing so, we obtain the following confusion matrix.

So there were, as I mentioned, there are 32 cases, out of which 24 of them are actually good care, and eight of them are actually poor care.

We observe that the overall accuracy of the model is 19 plus 6, is 25, over 32.

The false positive rate is, in this case, 5 over 24, 19 plus 5, whereas the true positive rate is 6 out of 8-- 6 plus 2.

Notice, if you compare this model with making always-- let's say one alternative is to say we predict good care all the time.

In that situation, we will be correct 19 plus 5, 24 times, versus 25 times, in our case.

But notice that predicting always good care does not capture the dynamics of what is happening, versus the logistic regression model that is far more intelligent in capturing these effects.