4.08 Multiple regression: Interpreting results

Until now we looked at the more technical aspects of different regression models, tests and assumptions, but we haven't put everything together yet. In this video we'll go through all the steps of performing a regression analysis, from checking assumptions to interpreting results. We'll also compare the different regression models we've discussed so far.

We'll use the same examples that we've used before, so you can focus on the steps and interpretation. We'll start with the example where we predicted popularity of cat videos–measured as number of page views–with the predictors cat age and hairiness – both rated on a scale from zero to ten.

The steps to performing multiple regression are as follows: First we inspect the data and check assumptions. If they're met, we perform the analysis and interpret the overall test. If this is significant, we interpret the tests for the individual regression coefficients. And finally, if we've specified several, nested regression models we compare them to see which model provides the best fit. By the way, we're assuming explicit expectations and an analysis plan were specified beforehand. Also, we're assuming the basic data preparation, like checking for data entry errors, has already been done.

Suppose I've collected data on 152 online cat videos that were also submitted to the international cat video festival last year. I want to test whether cat age and hairiness can be used to predict video popularity, and if so, whether the model improves by adding the indicator 'costume' (whether the cat is dressed up). I have the feeling that cat age will decrease popularity and hairiness will increase it, but let's use two-sided tests just to be safe.

First, we look at the data by eyeballing the **scatterplot matrix**. This gives the scatterplots for each combination of the response variable and predictors. And we can also look at the **correlation table**. This gives us a general idea of the simple relations between all the variables.
Of course we're hoping to find high correlations between the response variable and each predictor. The correlation between popularity and age looks promising, the correlation between popularity and hairiness doesn't. Generally, it helps if the correlations among the predictors are low, because it means that there is less overlap between predictors and more room for unique contribution to the explained variance of the response variable. The relation between age and hairiness is quite high. Still the data look promising: there are no nonlinear patterns or weird outliers and the correlation between popularity and age looks good.

To make sure we'll look at the assumptions and technical requirements in more detail. The first and easiest thing to check is whether we have **enough observations**. We have two predictors so we should have at least twenty observations. We have 152, so this requirement is met. Remember that ten cases per predictor is a rule of thumb and therefore somewhat arbitrary; more is better!

Next we check for **regression outliers** by looking at the standardized residuals. To obtain the residuals we have to perform the regression analysis, but only to calculate the residuals; we don't look at results of the analysis yet! In our example there are no extreme cases to consider. If we check for absolute values more extreme than two, we see seven cases, which is to be expected with 152 observations.
If there are standardized residuals with an absolute value larger than three, we inspect these cases and remove invalid data. But what if they're valid? Should they be removed just to make the model fit better? That depends on a lot of issues that are beyond the scope of this course. The safest course of action is to do the analysis with and without the problematic outliers and report both outcomes.

Next we check for **linearity** and **homoscedasticity** by checking the residuals plotted against each predictor. The plots for both predictors look ok, no nonlinearity or non-constant error. If we had time-series data we would check for **independence of errors** by plotting the residuals against time. In our example we can't check this assumption. We check the last assumption of **normality**, by looking at a histogram of the residuals. The histogram looks ok, and even if it hadn't, we would have been ok because our sample is relatively large with only two predictors.

Since all requirements and assumptions seem to be met, we can interpret the results of our regression analysis. First we look at the overall F-test and multiple R-squared. As you can see our overall test is significant, the p-value is smaller than 0.05. Multiple R-squared is quite small however, so even though the model explains a significant portion of the variation in popularity, this portion is small. There are probably better predictors out there than age and hairiness.

Before we consider another predictor, let's look at the current ones. If the F-test had not been significant - if the model did not explain a significant part of the variation, then it wouldn't make sense to interpret the individual tests, but here the overall test *was* significant, so let's see which of the predictors contribute. As we expected from the correlations, only age contributes significantly; age is negatively related to popularity, controlling for hairiness. Hairiness is *not* significantly related to popularity, while controlling for age.

What would happen if we included the indicator costume? Normally you would recheck the assumptions if you enter a new variable unexpectedly - or if extension of the model was planned, you check the assumptions based on the most elaborate model. For now we'll skip these checks.
If we run the analysis again including costume, we see that the model is still significant - obviously. Relatively speaking R-squared has grown quite a bit, but in absolute terms it's still small. The unique contribution of the indicator costume is significant. The cat wearing a costume is positively related to popularity, while controlling for hairiness and age. The contribution of the other predictors is about the same as before.

We can even perform an F-test to see whether the increase in R-squared differs significantly from zero. We don't have time to go into the details of the test and how it's calculated, but you can see the increase is significant here; adding costume helps.
It's important to note that you can only compared *nested* models in this way. Nested means that the predictors in the smaller model are *all* present in the larger model. For example, we can't directly compare a model containing age and hairiness with a model containing age and costume, we can only compare these with the large three-predictor model.
Besides testing pre-specified models based on solid hypotheses, it's also possible to use statistical algorithms to iteratively add or delete predictors to come up with the best fitting model. These methods are beyond the scope of this course though.