Binary logistic regression: Interactions (video 3 of 3)

Hi everyone, my name is Heini Vaisanen, I work at the University of Southampton and today I will talk to you about interactions in binary logistic regression.

So the outline for today's lecture is we will talk about interactions and how to interpret them using predictive probabilities and then we will talk about how to interpret interactions using odds ratios. Our example today is a relatively simple example of a regression model, a logistical question model where we want to study that whether the probability of having access to a car, varies by gender. Our response variable or our and dependent variable is whether someone normally has access to a car, yes or no, and we have three explanatory variables, gender, whether someone is a man or a woman, whether they are alone, married, or cohabiting, so their martial status and finally their age as an continuous frame.

Here are our model results both in low good scale and odds ratios. As you can see from where the arrow is pointing, we have added an interaction between gender and and marital status. So the only variable in this model that is not involved in the interaction effect is age and therefore the interpretation of age is exactly the same as you would have in any logistic regression model without interactions and so we will not talk about that. Instead we will talk about how might we go about saying something about the relationship between the probability of having access to a car, gender and marital status. We will first look at this by calculating predicted probabilities, and for that purpose it is quite useful to figure out what our model equation in the logit scale is. So basically the logarithm of the odds and that we're interested in, or the logic of the probability, on the left hand side equals our intercept and then all the coefficients and the associated estimate values for each of the coefficient. So we have coefficients for gender, age, marital status, and then the interaction effects for women who are married and women who are cohabiting. So when it comes to our reference categories this means that for gender, the reference category is man, and for marital status the reference category is people who live alone.

So the interpretation of marital status and gender is a bit tricky because there's the interaction effect, and about 90% of the time, I'd say the easiest way to interpret an interaction in a logistic regression model is to calculate the probabilities for a range of different values that we might be interested in and then cut them in a table and use that to interpret the results.

So here we're interested in the interaction between gender and marital status. So we will vary these two variables and calculate different probabilities for different levels of gender and marital status, but we will keep age constant because we're not really, we're just controlling for age, but we're not really interested in the effect of age, so we're just going to use the average age for every probability that we count. The average age in this data happens to be 46.21 years so that's what we're plugging for age every time. So if you want to calculate the probability of normally having access to a car for a man who is living alone and who is of average age in this data, this is how our equation looks like. So as you might remember from the previous lecture we calculate predicted probabilities by exponentiating the equation of interest, dividing that by 1 plus the same thing. So the equation of interest in this case means that we take into account the intercept, which we always take into account, and then we take into account the coefficient for age. The reason that we don't take anything else into account, we can see here, we actually don't have anything in our model equation for men, or for people who live alone, because those were the reference categories for our tiny variables. The coefficient for age is minus 0.035, so that means that we multiply that by the average

age which is 46.22, and then that is our equation. When we solve this we get 0.672, so that means that the probability of normally having access to a car for a man who lives alone and is of average age is around 67%.

If you wanted to calculate the probability for a married man of average age to have an access to a car, we would have to take into account more things. So when we plug in numbers for our equation we would take, again, into account the intercept and then the effects for age, which are the things that don't vary actually when we are calculating these probabilities in this example, and then in addition we would take into account the effect of being married. So because we are talking about men here, we only need to worry about the coefficient associated with being married, so the 1.608, because men are not involved in the interaction effect that we have in the last two rows here, because they are the reference category for gender. If we had been interested in calculating the predicted probability for a married woman, we would have had to take into account the coefficient for married, and then the coefficient for the interaction between being a woman and being married, so the 0.476. If we solve this equation for married men we get 0.911 which means that the probability of having access to a car for a married man of average age is about 91%.

If you use the same logic to calculate the probabilities for every different possible combination of marital status and gender that we had in this data, this is the following the table that we would get for people who were average, of average age in this data. The ones that I've highlighted with red are the ones that we calculated and the rest are probabilities that you could try and calculate yourself by hand and see if you get the same result. When it comes to the interpretation of the interaction effect, now it's much easier to do than it would have been by just looking at the odds ratios or the local values in our regression table. We can see that the probability that a single man or a man who lives alone of having access to a car is around 67%. For a woman who lives alone this probability is around 52%, so much lower than for a man who lives alone. However, if we look at married men and married women the probability of having access to a car is very similar, so it's 91% for married men and 90% for married women. The same goes for cohabiting men and women, so the probability of having access to a car for a cohabiting man is about 77% and about 78% for women, and that means that the interaction here, is really comes from the single category. So there is a quite a big difference between the two genders in the first category of those who live alone, but as soon as we have people with partners these gender differences disappear and that's our interaction effect.

Now we could have also ignored predictive probabilities and used the odds ratio to interpret these interactions, however, this is usually a bit complicated so I would, I normally just use predictive probabilities, but in case you want to know how this works without ratios I will show you. So if we want to know what are the odds that a man has access to a car, we are lucky in the sense that we don't have to worry about the interaction effect because men was the reference category here, so we can ignore their interaction for now. So if we want to know what the odds are that a married man has access to a car, we just look at the odds ratio row for married, and we can see that their probability of having access to a car are about five times higher than the probability, sorry the odds of a single man having access to a car, so these are odds not probabilities. When it comes to cohabiting men we can again, just look at the odds ratio that we see in the cohabiting row, and we can say that the odds for a cohabiting man to have access to a car are about 1.63 times higher than the odds of a single man when we're controlling for age. The reason that these odds ratios only correspond to men now is that we have this interaction effect that we need to take into account if we are talking about women.

So what are the odds that a woman has access to a car? If you're talking about single women, again, we can ignore, we can keep ignoring the interaction effect because single, or living alone is the reference category for marital status, so there is no interaction effect associated with the combination of being woman and living alone. We can just look at the odds ratio table and we can say that the odds for a single woman to have access to a car, are about half of the odds of a single man when we control for age. When we are interested in women who are married or cohabiting then things get a bit more complicated. So let's look at married women. So I'm going to claim that the odds for a married woman of having access to a car are about 4.3 times higher than the odds of a single man when we control for age, and you might be wondering how I came up with this because there is no 4.33 anywhere in the table. Well when we are looking at the odds for married women, we first need to take into account the odds for women, but remember that this now only corresponds to single women because we have the introduction effect in the model as well. So that's why we need to also take into account the effect of being married and the interaction effect of being both woman and married, and then we multiply all of these things together, so we multiply 0.54 times 4.99 which is the coefficient for married, times 1.61 so that's the coefficient for the interaction effect, and when we calculate this we get 4.33 and that's how I came up with that odds ratio. We can do the same thing for cohabiting women, but then we take into account the odds ratios for cohabiting people and the interaction for women times cohabiting. If we do that the odds ratio that we get from the calculation is 1.76, you could try do this in your wrong time by hand and make sure that you get the right result. And if we then put all of these odds ratios in a table this is how it looks like, and essentially the interpretation is very similar to what we saw with the probabilities, so the odds for a married man and a married woman of having access to a car are very similar, as are the odds for a cohabiting man and a cohabiting woman to having access to a car, again very similar, but if we compare women who live alone and men who live alone, its much more likely that a man has access to a car than a woman and that is our interaction effect in this model.