**Ordinal regression_Part 2: Multiple ordinal regression**

**Author: Dr Heini Väisänen**

**Transcript of:** https://youtu.be/aiQcFCFtT5k

**This video is part of an NCRM Online Resource**

Hi everyone and welcome to uh the second ordinal regression video um of the NCRM ordinal regression series. My name is Dr Heini Väisänen and I'm a lecturer at the University of Southampton.

So today we will talk about multiple ordinal regression models so that is an ordinal regression model where you have more than one exponential variable. We will look at the statistical significance and how to determine that in such a model and then we will look at interpretation of the results so we will look at cumulative watch ratios as well as predicted probabilities. There is a third video video of the series as well which I encourage you to look at because it talks about the importance of the proportional orders assumption, but first let's look at um interpretation of these models.

Like any other regression model that you've run in the past these ordinal models as well can include more than one experimental variable. In first video we only included one variable because we wanted to show a simple example of how to use these models but in real life you would normally have more than one variable in your model and that is usually because you want to control for some variables while you're investigating the associations of others.

The invitation is largely similar to other logistic models that you've seen but there are some differences because of the cumulative nature of the model. So the alter rations behave slightly differently from the alter rations that you've seen in binary logistic question and multinomial logistic regression.

Today we will use the same data as you will use in the computer workshop that is associated with these lectures it comes from the crime survey for England and Wales collected in 2013 and 2014 and we will look at how worried people are that their homes could be being broken into in England and Wales and we're interested in investigating whether this worry seems to vary by gender and education.

First let's look at some descriptive statistics of our data set. So our outcome variable is how worried people are about burglary and we have four categories that are ordered going from one not at all worried two not very worried three fairly worried and four very worried so the higher the category is the more worried the respondent is.

From the distribution within our sample of 2181 respondents you can see that almost a half were not very worried about 15% were not at all worried 27% were fairly worried and about 10% were very worried. And you can also see the association associated frequencies on this slide. Our two explanatory variables are here so we have slightly more women than men in our sample about 55% women and 45% men when it comes to education we have four categories ranging from no education at all which is about 26% of our sample to O-level/GCSC level which is about 19% of the sample A-level education among about 18 and a degree or a diploma among around 37% of the sample.

If we put both of these variables into an ordinal regression model, this is what we get. I am showing the results in cumulative auto ratios which is what you get from an original model so first we have a variable for gender we have chosen men as the reference category and then we have a ratio for women we have an associated p-value and a 95% confidence interval. Then we have the education variable we've chosen no education as the reference category and we compare those with all level, A level and decree your diploma to those with no education. Then we have three cut points or thresholds or insets depending on which name you would like to use for these which are not in the odds scale they are in the logo scale and these are the three different intercepts if you were to write down the equation of this model you would have three equations because we have four categories in our outcome, like you hopefully remember from the first video of the series, and if he was write down the equation the first one would have the lowest intercept minus 1.74 and then you would take the values from the model otherwise to write down the equation. And remember that the equation for all the equations are exactly the same other than the intercept changes for each of them which is different from multinomial regression where everything is different from all for all the equations. But before we go into that in more detail let's look at the statistical significance.

The statistical significance that most statistical software, these results are from Stata, but other or many of the software would do the same, are based on a wall test and that is similar to t-test in linear regression. And that is a very useful test especially if you want to know whether a continuous variable or a binary dummy variable is significant in the model. So here we have gender with this which is the binary categorical variable and you can see inside the red circle that it has an associated p-value of smaller than 0.001 which means that p-value is very small. So we can say that gender is significantly associated with being worried about burglary in our model. When it comes to the second variable in our model education the world test doesn't as clearly tell us what's going on we have three different p values for the three different categories of the four category dummy variable first one is very large 0.9 then we have a kind of one one that is fairly small but still not quite within the 0.05 threshold for the a level group and then we have a small about p value for the decree or diploma group which you can all see inside the blue circle on the slide.

And these p-values like you might remember from the other videos in for instance around binary logistics regression tell you whether that category is different from the reference category. So here we know that those with all-level education are not different from those with no education statistically significantly, those with A level might be but it is still a relatively large p-value whereas those with degree or diploma seem to be statistically significantly different from those with no education.

If you want to know whether education as a whole is significant in the model we can use the likelihood ratio test. You might remember that what likelihood ratio test does it tests two nested models so models which have some variables in common but then one model is larger than the other so it has more exponential variables than the other. In here the nested models that we want to test are the one where we only have gender as an expansive variables the one that you can see here on the slide and the one that you just saw in the previous slide where we also have education so the models are nested because both of them have gender but then the second one of the models is bigger because because it also has education and then it means that the likelihood ratio test tells us whether adding education improves the fit of the model significantly. And if you're using Stata you might use the LR test command as you will in the computer workshop and results might look something like this and you can see the associated p value inside the red circle and the p value is quite small is smaller than 0.01 so at one percent level of significance we can say that education is significant and should be included in the model.

Okay so now that we've established that both gender and education should be included in the model we can move on to interpretation which is the most interesting part of any statistical analysis.

So as as with any other logistic model we could interpret the results using the logic scale but that's usually not done because it's not very intuitive so I'm going to skip straight through to cumulative odds ratios.

So the odds ratios that you see in in your results when you run these models in your statistical software might look like something like this. Let's start from gender so men are the reference category and then we have a cumulative ratio seal 1.35 for women and that means that women have a 35% higher odds of being in higher rather than lower categories of the outcome. That means in other words that they are more likely to be worried about their homes being broken into when we control for education. And when I say women have higher loads of being higher rather than lower categories of the outcome it means that if we compared the likelihood of being in category one compared to two three or four then women would be more likely to be in two three or four rather than one or if we compared the likelihood of being in either in category one or two to three or four then women would be more likely to be in three or four and if you compare to likelihood of being one two or three to the likelihood of being in four then women would be more likely to be in category four. And it's always how much more likely is always 35% percent higher odds because of the proportional auto assumption that we make in an ordinal regression model and that basically just means that we assume that all of these different cut points that we have for our outcome variable the odds are the same and this is an assumption that we can test and I will show you how in my next video.

When it comes to education we have more odds ratios to deal with than this one because it is a dummy variable with four categories but luckily the odd ratios show quite a clear pattern of association. So it looks like more educated groups are less likely to be in the higher rather than the

lower categories of the outcome that means that they are less likely to be worried about their homes being broken into when we are controlling for gender.

You might want to pick an example or two from these odds ratios if you're writing down a report for instance or explaining your results to someone else and in that case you could say something like the odds are 28% lower among those with a degree than those without education of being in the higher rather than lower categories of the outcome.

However, it often makes sense to interpret these results using predicted probabilities and that is because it gets odds rations are sometimes quite difficult to grasp anyway and when we are talking about cumulative odds ratios is even harder so to make your results clearer you might want to calculate some predicted probabilities and use those to show what's going on in your model. And if you wanted to do this by hand you would have to look at the cumulative logic values of your model as you can see in the table here and then you would need to use the equation that you saw in the first video that is inside the red square here and the thing with ordinal regression is that most statistical software changed the signs of the explanatory variables or the estimates associated with your explanatory variables so unlike other equations that you might use to calculate fitted or predicted values we have the intercept A K and then we have minus whatever the peta is for each of these variables that you have included in your model. So you have to remember to change the sign of the values that you see in your table. And the reason this is done is to make interpretation more intuitive so unless we change this sign or unless statistical software change this sign when they calculate the results it would mean that higher odds increase the likelihood of being in the lower categories of outcome which is less intuitive than higher higher odds being associated with the likelihood of being in the higher categories of the outcome.

So if we wanted to calculate the first cumulative probability so the cumulative probability of being in category 1 we would take the lowest intercept or the lowest cut point which in our equation happens to be -1.74 then we've decided that we want to calculate probabilities for women who have a level education so we need to take the coefficient values from our table for these two categories. So the coefficient for women is 0.30 and since we have this negative sign in front of the the betas in our equation we have to change the sign so instead of writing down plus 1.3 we write down minus 1.3 as you can see in the slide. The same goes for a level education the coefficient in the table is negative 0.22 but when we transform that to the equation we change the sign so it becomes plus 0.22 we exponentiate that divided by 1 plus the exponentiated equation again and if we solve this equation we get 0.138 so that means that the predicted probability according to a model of a woman who has A level education of being in the first category so not at all worried is about 14%.

We could do the same calculation for the other cumulative categories of the outcome and the only thing that would change in this equation that you saw in the first slide is the intercept so the minus 4.3 plus 0.22 would change the same we would just instead of having negative 1.74 we would have 0.52 for cp2 equation and 2.18 for cp3 equation. And if we solve these equations this is what we get so the cumulative probability cp2 so the cumulative probability of being in either category one or two is about 61 percent for women with A level education the cumulative probability three so the

probability of being either in category one, two or three is about eighty nine percent for women with A level education and the cumulative probability four is one because that is the highest category we have four categories in our outcome and the probability that you're in one of these four categories is one because everyone is in at least in exactly one category of this outcome so we don't need to calculate that we always know that the cumulative probability for the highest category is one.

This is maybe not the most useful thing to report what we actually want to know is the probability that someone is in the exact category 1 or exact category 2 etc. And we can do that now that we have calculated the cumulative probabilities. Like you might remember from the first video the cumulative probability 1 is the same as the probability of being in the category 1 because there's nothing below it so that is 14% the cumulative probability 2 we can calculate by taking cp1 minus p1 or cp1 it's the same thing and we get 40 0.468 so the probability for women with a level education of being a little bit worried is about 47%.

The probability of being in third category are fairly worried is cp3 minus cp2 which is about 28% and the probability of being in the highest category zone not at all worried is cp4 minus cp3 which is about 11%.

And here is just an example what you can do with statistical software. So most of the time you wouldn't actually calculate these different things by hand it's useful to do that when you're learning about this method so that you understand where the probabilities are coming from but when you're actually running these models in your day-to-day life then you usually just use statistical software. And that means that in a second you can get different combinations of probabilities. Here I've calculated the probability of being in each exact category of the outcome one two three or four by gender and education and I am showing you the results of the two more two extreme categories so being not at all worried and being very worried on the left hand side you have not at all right and on the right hand side you see the very worried results. The green lines are for men the orange lines are for women and when we go from left to right on the x-axis education increases and we can see from these probabilities the same thing that we saw from the from the odds ratios. First of all women are more worried about crime than men and that we know because the line for women is below men in the not at all worried and above men in the very worried craft so women are more likely to be very worried than men are and they are less likely to be not at all worried. From the odds ratios we also saw that when education increases then people are less likely to be worried about crime and we can see the same pattern here in the predicted probabilities as education increases their probability of being not at all worried increases as you can see from the lines that are increasing in the first graph and the second graf which is the very worried category when education increases the likelihood of being in that very worried category decreases.

Okay that's all from me in in this video. Thank you so much.