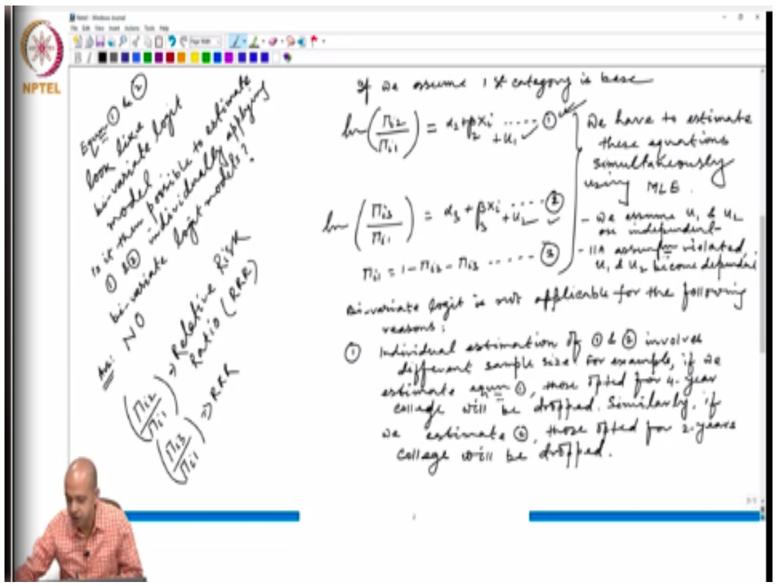
Applied Econometrics
Prof. Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology-Madras

Lecture-36
Multinomial Regression Model-Part III

(Refer Slide Time: 00:16)



Now We will take a dataset and then we will try to estimate an equation applying multinomial logit model. **(Video Starts: 00:24)** So, let us go back to our dataset. This is basically a dataset on the school choice problem taken from the USA context. A college choice problem, for a +2 graduate, whether to go for 2 year college or 4 year college or no college.

And that basically depends on whether you are a catholic or not? What is your average grade point, up to +2, what is your family income, what is your family size so on and so forth. So, we will take this data and then we will run the model in statistical software stata or the multinomial logit model. So, I am just copying the data from the excel file, so we have 1000 observations, so 1000 students are basically thinking about their college choice.

Whether don't go for a college at all, whether to go for a 2 year college, whether to go for a 4 years college and that depends on the chooser's specific or individual specific information. We will then put it in stata. so this is data, we will go to data editor and we will put it here. And then stata is asking whether the first row now is a variable name or not?

We have to say yes, it is the variable name, so this is our data set now I have pasted it. Now we can estimate the model. So, for estimating the model that means for estimating a multinomial logit model the command is very simple. In logit model what was the command we use? Logit, since it is multinomial we have to put m logit and then your dependent variable which is basically PAC choice.

And let us say that this college choice is dependent on your grades, your family income, family size and whether you are male or female, whether you are black or white? So, this is basically the command, command is very simple mlogit. Now if we put enter then look at the output. In the output this is the estimated result and multinomial logit also like the bivariate logit model, it involves the maximum likelihood method.

So, that means we are applying the likelihood function, the concept of likelihood function we are selecting the parameters to maximize this likelihood function. It means basically we are trying to maximize the probability of observing a given sample from the population. And the parameters that maximizes that probability, probably to observing that sample, likelihood of that sample those parameters are reported here.

So, if we do not specify anything here as we discussed, while discussing the theory since we cannot estimate all the 3 probabilities separately, 1 probability we must consider as the base. Here if we do not mention anything then stata will automatically consider the third category the 4 year college as base. Now suppose you want to estimate the model using no college as your base because it would be easier for you to make better sense when I compare 2 years college over no college and 4 years college over no college.

What is the relative probability of selecting 2 years college over no college? What is the relative probability of choosing 4 years college over no college? That makes better sense than this. So, what we will do now? We will put the same command multinomial logit, mlogit then we have to put what is your base? Base is let us say 1, now if you put base, now stata is considering first category that means no college is the base category and then stata is gives 2 panels separately.

What are these 2 panels basically? The first panel is corresponding to the equation wherein, what was an equation? Our equation was this,

$$Ln\left(\frac{\pi_{i2}}{\pi_{i1}}\right) = \alpha + \beta x_i$$

and if you want to convert this into an econometric model of course we have to include the error term, let us say this is u 1, this is u 2.

So, in stata the panel 2 is basically estimating the results from equation 1 and panel 2 is basically from equation 2. Now the question is how are we going to interpret these coefficients? Look at the grades, so average grades the coefficient is -0.31 and that is highly significant also. So, what is the interpretation of this? How will you interpret? To interpret the coefficient always we need to go back and look at what is the equation we are estimating.

So, if we go back and look at the equation it would be easier for us to interpret. So, let us say this is grade point, x_i indicate grade point and $\widehat{\beta_2}$ is basically the coefficient attached with the grade point. So, here the estimated value is negative that means what I am saying that grade point is negatively correlated with the individual's choice relative probability of selecting 2-year college over no college that is the interpretation.

What am I saying? That grade point is negatively correlated with the relative probability of selecting 2 years college over no college. But we cannot interpret them as for a unit change in grade probability of selecting 2 years college over no college reduces by this much amount that we cannot say. So, we cannot say for improvement in grade point by 1 unit probability of selecting 2 years college over no college reduces to 0.3107 unit because if we interpret the coefficient in this way that means we are directly considering these coefficients as marginal effect.

But I have already told you earlier that whenever you are estimating qualitative response model whether it is a logit, probit, tobit, binomial or multinomial logit, conditional logit or any type of qualitative response model, we should never ever take these coefficients directly as a measure of marginal effect, we should not commit that mistake. Because our equation is

$$Ln\left(\frac{\pi_{i2}}{\pi_{i1}}\right) = \alpha + \beta x_i$$

so if we differentiate.That means in terms of the probability, what was our probability?

$$\Pi_{i,j} = \frac{e^{\alpha_j+\beta_j x_i}}{\sum_{j=1}^{n} e^{\alpha_j+\beta_j x_i}}$$

So, obviously $\frac{d\Pi_{i,j}}{dx_i}$ is not β. So, that is why we need to calculate the marginal effect separately by applying specific command. So, that means the equation will guide you what specific interpretation we have to take from these values.

So, here I can only say whether a specific explanatory variable is positively or negatively correlated with the relative probability. So, that means logarithm of relative probability, that much only you can say. Since log transformation is monotonic transformation, we can consider them as relative probability itself. So, we can say that grade is negatively correlated with the relative probability of second year college over no college.

Family income is positively related with second year college with no college. As family income increases students are more likely to go for a 2-year college over no college. Family sizes also, as family size increases students are less likely to go for a 2-year college. Obviously as family size increases probably there is the family requires more amount of income for which probably the students are entering directly into the job market instead of going for the college.

And when you come to the third panel, this panel then how do you take the interpretation? Once again you go back that means third panel gives you the result for this equation 2,

$$ln\left(\frac{\Pi_{i,3}}{\Pi_{i,1}}\right) = \alpha_3 + \beta_3 x_i$$

So, that means I can say whether x_i is positively or negatively correlated with the relative probability of a 4-year college over no college. Like 2-year college even grade is negatively correlated with the relative probability of a 4-year college of over no college and then family income is again positively correlated with 4-year college than no college.

But if we compare the coefficient, the coefficient in panel 3 is 0.02 and 0.01, so that means comparing these 2 what we can say? That as income increases students are more likely for a 4-year college over no college than 2-year college. We can say as income increases students are more likely to go for a 4-year college than 2-year college, that much we can say. Family size is again negative.

So, this is how we need to interpret this coefficient, whether we have positive or negative correlation with a specific explanatory variable with the relative probability, we can never take these values as marginal effect. Then again look at here the LR chi square, so that means overall significance of the model is again given by LR chi square**.** how LR chi square is defined?

$$LR\ Stat = 2(L1 - L0)$$

that is how we define. Earlier we have already discussed in the context of simple logit model and the degrees of freedom here it is 10, why this is 10 when you have 5 explanatory variable? Because we have to say that since there are 2 panels basically, we have to consider 5 into 2, so degrees of freedom would be 10 for this multinomial logit model even though there are only 5 explanatory variables. Because we are considering equation 1 and 2 simultaneously that we have to keep in mind.

We are estimating equation 1 and 2 simultaneously, so that is why the degrees of freedom is 10 here and the probability of chi square is again 0000, so that means model is highly significant. And once again like the simple logit and probit model here also we get the pseudo R^2. So, which is basically 1 - l 0 by l 1 where l 0 is the log likelihood of the restricted model and l 1 is the log likelihood of unrestricted model.

What is the restricted model? This model is the restricted where we do not include any expanded variable which is given by iteration 0, so -1018 that is the restricted version of the log likelihood and unrestricted is basically 845.625. Same formula, what we have learned earlier in the context of multinomial, a simple logit or probit model the same is applicable here. Only difference is that in the context of multinomial the options for the dependent variables they take different values, they take more than 2 values that is what is given here, other things are same.

Now if we want to get the marginal effect and one more thing before we go to the marginal effect this $\left(\frac{\Pi_{i,2}}{\Pi_{i,1}}\right)$, $\left(\frac{\Pi_{i,3}}{\Pi_{i,1}}\right)$ they are called, relative risk ratio. Now suppose I want to get the marginal effect, so the marginal effect of let us say the first variable grade, for unit change in grade, what is the change in relative probability of a 2 year college over no college?

That is what I am interested in. So, the command would be margins then we have to give comma $\frac{dy}{dx}$ then in a bracket you have to put the variable name for which you are interested to estimate the marginal effect. And then after that what we need to put is probability, that means there are 2 panels. Are you interested in panel 2 or panel 3? If you are interested in panel 2 that means basically you are interested in for a unit change in grades, change in relative probability of a 2-year college over no college.

So, here you have to put pr out to, this is a command, margins dy dx, then you have to put grades out to. So, now what I am getting? Corresponding to grades I am getting 0, 2, 6, 4. So, that means for the unit change in grades you need change in grades probability of selecting 2-year college over no college increases by 0.02 unit, changes by 0.02 unit that is what we said. And if we want to estimate this as for the third one then here instead of 2 we have to put 3.

Now one thing you have to be very, very clear, see here in both panel 2 and panel 3 the sign of grade is negative however when you get the marginal effect here it is positive, here it is negative which is possible you should not get worried about the sign of the marginal effect. Sign of the marginal effect may or may not match with the sign of the coefficient because marginal effect is constructed in a completely different way than this.

Because here it is $\Pi_{i,2}$ that means $\Pi_{i,j}$ here you are estimating by the equation

$$\frac{e^{\alpha_j+\beta_j x_i}}{\sum_{j=1}^{n} e^{\alpha_j+\beta_j x_i}}$$

If you differentiate that $\Pi_{i,j}$ with respect to x_i, so that means differentiating that function when you are estimating

$$ln\left(\frac{\Pi_{i,2}}{\Pi_{i,1}}\right) = \alpha_2 + \beta_2 x_i$$

So, that means delta log of that delta pi ij delta x_i is basically this.

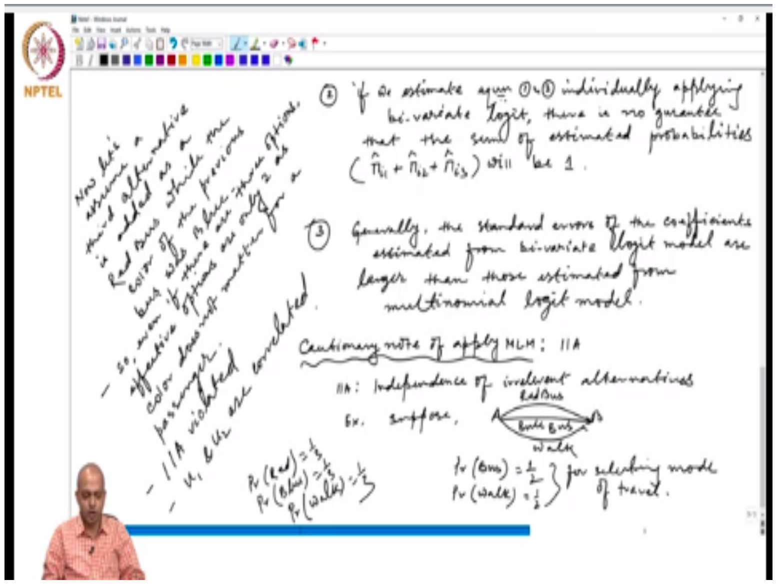
So, even though the sign of the coefficient is negative since the marginal effects are constructed by applying a different formula which is derived from this, they may or may not match in with the signs with the coefficient. So, you should not get worried about this, so that is why I say that as grade increases probability of a 2-year college over no college increases by 0.02 unit while as grade increases probability of a 4 year college over no college decreases by this much unit, that we have to keep in mind.

Marginal effects may or may not match with the sign of the coefficient that is very simple. So, this is how we can calculate the marginal effects. Likewise you can calculate the marginal effects for family income, marginal effects for family size, marginal effect for even the qualitative variable also. So, that means if I put this variable margins dy dx then in the bracket let us say I am putting female which is a dummy variable, qualitative variable female.

And then PR out to, what is the meaning? I cannot say that for a unique change in female. That means we have to see how females are defined. If female = 1 and male = 0 that is how you have defined your gender, that means we can say that females are 0.0364 unit more likely to participate in the level to go for a 2-year college compared to their male counterpart. So, that means when the relative probability of a 2-year college over no college that is higher for female by this much unit.

So, this is how we can actually interpret the coefficients in a multinomial logit model. Now lastly before we conclude we have to make a cautionary note. **(Video Ends: 26:10)**

(Refer Slide Time: 26:11)



What is the cautionary note? Cautionary note applying multinomial logit model, and this is called IIA. What is IIA? IIA stands for Independence of Irrelevance Alternatives. So, I will give an example to understand this. Suppose we are traveling from A to B and there are only 2 ways you can go by pass or you can go by walk. So, that means probability of selecting buses, so probability or bus equals to half, probability of walk equals to half for selecting mode of travel.

Now let us assume a third alternative is added as a red bus while the colour of the previous bus was blue. So, that means how many options now? This is a red bus; this is blue bus, so now you might be thinking the probability of red equals to one third, probability of blue equals to one third and probability of walk equals to one third. But the question is when a commuter is thinking about these 3 options, does a commuter really bother about whether the bus is red or blue?

Whether the colour of a bus matter significantly while is choosing the mode of transport from A to B? So, that means here even though in terms of options there are 3, this red bus and blue bus these are actually irrelevant alternatives, this is irrelevant, colour of the bus is irrelevant to a particular passenger. That is why apparently even though there are 3 options actually or effectively there are only 2 options bus and walk because colour does not matter to a commuter.

So, even if there are 3 options, effective options are only 2 as colour does not matter for a commuter or a passenger. If we consider this as also another alternative, so while the in reality a commuter of passenger does not bother about the colour of a bus. If you think that the red bus, blue bus and walk, red bus and blue bus actually they are 2 different alternatives then what will happen?

In terms of our equation, we assume this u 1 and u 2 they are independent but in the context we assume u 1 and u 2 are independent to each other. But in the context of this red bus, blue bus problem IIA ensures, that means independence of if IIA is IIA assumption violated then u 1 and u 2 actually become dependent to each other. So, our model should be independent of irrelevant assumption.

So, that means there should not be any irrelevant options or alternatives available while choosing the model before applying multinomial logit model we should carefully think of whether the options what we are thinking whether they are relevant or not. So, IIA assumption violated, u 1 and u 2 are correlated. So, we should never apply multinomial logit model in this type of situation, this is called the classic red bus, blue bus paradox, this is the example of classic red bus, blue bus paradox.

So, there might be several other cases, so we have to only think about the options in terms of their functionality, here colour does not matter. But that does not mean that any other context where yes, for example while selecting a shirt you might be very particular about red shirt or blue shirt or white shirt. So, probably the customers are more bothered or interested about colour while choosing the dress and not while selecting a particular mode of transport.

So, we have to be very, very careful whether the alternatives a real alternative to the respondents or not? Then only we can ensure the independence of irrelevant assumption alternatives. And that is required because while estimating the probability we assume the error terms are independent actually. If they are not independent, then we will have biased estimates. With this we are closing our discussion on multinomial logit model.