Applied Econometrics

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

Lecture-34
Multinomial Regression Model-Part I

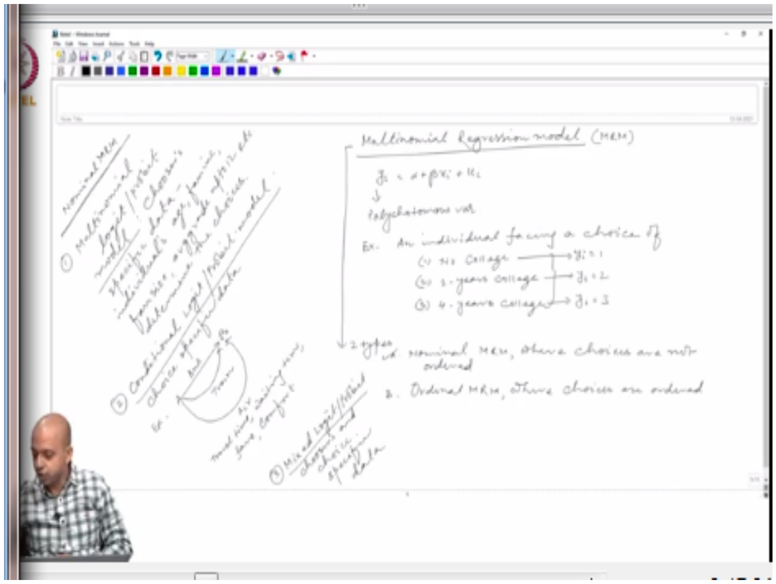
In our last lecture we were discussing about qualitative response model. And in our few previous sessions we are discussing about the logit and profit models. And in those models the dependent variable what we assumed as binary that means the dependent variable takes only 2 values 0 or 1 because we were assuming dependent variable to be dichotomous in nature.

For example, we were trying to understand what are the factors that determine whether the individual will have their own house or not? So, that means there are only 2 choices in front of the individuals either to own a house or not to own a house. Also, when individuals are selecting about transport choice then there are 2 choices, either to use public transport or to use a private one, whether to go for an MBA or not?

So, in all these cases the dependent variable is dichotomous in nature that means there are 2 options available. But in reality, many a times the individuals they face a problem where they have multiple choice, that means the dependent variable becomes polychotomous in nature. And in those cases, we need to develop a different type of model not the simple logit and profit model to understand those situations better.

So, if the econometric model that deals with the cases wherein the dependent variable takes more than 2 values or individuals, they face more than 2 options are called multinomial regression model. So, today we are going to discuss about multinomial.

(Refer Slide Time: 02:38)



So, this is multinomial regression model. So, our model

$$y_i = \alpha + \beta x_i + \mu_i$$

where y_i is basically polychotomous variable. For example, let us say that a student immediately after getting his or her +2 degree is deciding about whether to go for a college of 2 years, whether to go for 4 years or whether not to go for college degree at all. So, here for example an individual facing a choice of, first of all no college, this is choice number 1, no college, then choice number 2 is 2 years college and number 3 let us say this is 4 years college.

So, this y_i, let us say y_i = 1 when there is no college, then y_i = 2 for 2 years college and y_i = 3 for 4 years college. So, in this case the dependent variable becomes a polychotomous variable, it takes 3 values 1, 2, 3 depending on which particular choice the individual makes. Now these multinomial regression models are of 2 types, what are those 2 types? Number 1, a is nominal this is let us say MRM multinomial regression model, then MRM again categorized as nominal MRM, where choices are not ordered.

That means here we are not attaching any value judgment over these choices; we are not saying that 2 years college is better than no college or 4 years college is better than 2 years college. So, we are not saying that 2 is better than 1 and 3 is better than 2, this type of value judgment we are not attaching here. If we do not attach any value judgment, if we do not order the choices then that is called nominal multinomial regression model.

But if we say no, this 2 years college, at least 2 years college is better than no college at all and 4 years college is better than 2 years college, if we order these choices in this way then it would be called ordinal MRM, where choices are ordered. Here we are going to first talk about this case nominal multinomial regression model. And again, this nominal multinomial regression model let us say that this is nominal MRM multinomial regression model that is again of different types.

First type of nominal multinomial regression model is called multinomial let us say multinomial logit or probit model. Now in this case what is this multinomial logit or probit model? Here basically when we say that the individuals are facing these 3 choices, no college, 2 years college and 4 years college. Let us assume that we want to understand how the chooser's specific data can actually determine these choices. What are chooser's specific data?

That means I am not attaching any attribute for these choices rather I want to understand how the individual specific data, individual's age, individual's family income, individual's family size, individual's average grade up to 12^{th} standard, all this data can determine these choice. if that is the situation that is called multinomial logit or probit model, so this is basically chooser's specific data.

That means individual's age, family income, family size, let us say average grade up to 12^{th} etc. determine the choices. So, this is called multinomial logit or probit model. Now second type of nominal multinomial regression model is called conditional logit or probit model. What is this conditional logit or probit model? Here this is called choice specific data.

For example, let us say you are travelling from place A to B and there are 3 alternative ways you can travel by bus, you can travel by train, and you can also travel by flight, by air. So, I am trying to understand the factors that determine whether an individual will take bus, train or air, mode of transport while travelling from A to B. And I have choice specific data that means what are the choice specific data?

Travel time, let us say travel time, waiting time, then fare, then comfort, these are all not the individual specific data rather these are choice specific data. If I take bus, then I have a specific travel time required to reach B from A. I have certain amount of waiting time at the bus stop to get the bus and to board it. I have also specific amount of fare that I have to incur while travelling from A to B.

So, you can understand while the travel time in bus is much more, there would be less waiting time in the bus stop. But if you take a flight from travelling from A to B, travel time is very less but there is a huge waiting time in the airport. That means before taking the flight you have to reach the airport at least 2 hours before that and then you have so many security processes so on and so forth.

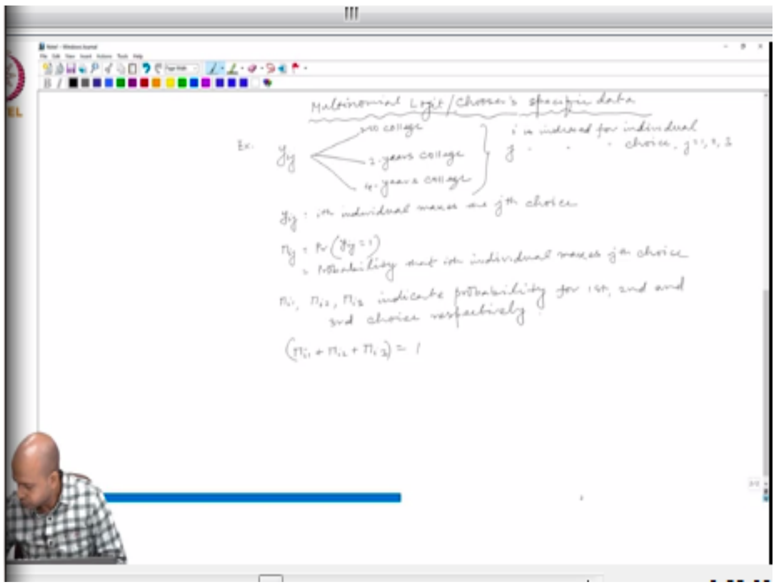
Actually it takes around 3 and half to 4 hours before to you actually board the flight and fare is also very high compared to bus. But the level of comfort is very high compared to bus or even train also. So, these are called choice specific data that will determine whether the individual will take bus, train or flight while travelling from A to B. And if we combine these 2 then we will have a mixed logit or probit model that means choosers and choice specific data.

If you have chooser's and choice specific data to determine, that means in this particular case I will consider travel time, waiting time, fare, comfort as well as individual's education, individual's average income whether the individual is male or female, individual's age so on and so forth, to determine whether the individual will take bus train or air mode of transport while travelling from A to B.

So, you must understand and differentiate all these cases, so that you can actually apply appropriate econometric model, in this situation wherein you dependent variable is polychotomous in nature. So, we should not always apply the multinomial logit or probit model which is the most popular models in the literature. Many a times the students, the applied researchers they fail to understand the difference between multinomial, conditional, and mixed kind of logit and probit model.

So, irrespective of whether it is a choice specific or chooser's specific data they simply apply the multinomial logit model. So, as a researcher we need to understand what our data structure is? how we are going to explain the variation independent variable? and based on the situation we will apply these types of models. So, this is basically the difference between multinomial conditional and mixed logit. So, for today's discussion we will be taking first the chooser's specific data or multinomial logit model.

(Refer Slide Time: 17:08)



For today's discussion this is multinomial logit or chooser's specific data. So, the example what we are taking as we discussed earlier that this is a choice of college, so that means we are saying that y_i actually is polychotomous in nature, no college, 2 years collage or 4 years of college. So, the choice is instead of y_i let us say we denote this dependent variable as y_{ij}, here i is indexed for individual and j is indexed for choice.

So, j = 1, 2 and 3, we have 3 choices here, no college, 2 years college, 4 years college. y_{ij} indicates ith individual makes the jth choice. And P_i ij equals to the probability that y_{ij} = 1.

$$\prod_{i,j} = Pr\ (Y_{i,j} = 1)$$

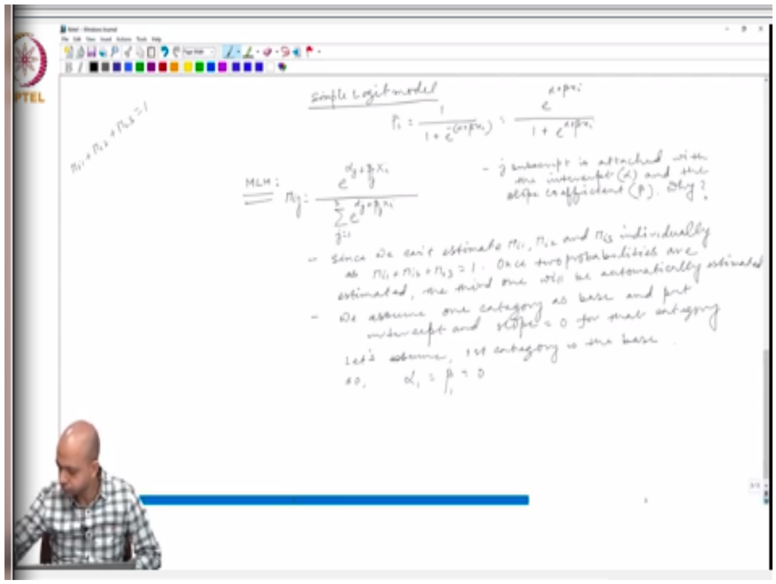
So, P_i ij that means basically indicates probability that ith individual makes jth choice. P_{ij} = probability that y_{ij} = 1. So, that means we will have P_{i1} $P_{i2}$and P_{i3} indicate probability for first, second and third choice respectively.

And we know that since this choices are not mutually exclusive or you can say that

$$(\Pi_{i,1} + \Pi_{i,2} + \Pi_{i,3}) = 1$$

sum of these 3 probabilities equals to 1, So, what we will do? In the logit model what you observe?

(Refer Slide Time: 22:09)



In case of simple logit model, so this is simple logit model, what we assume that

$$P_i = \frac{1}{1 + e^{-(\alpha + \beta x_i)}}$$

Now if you multiply both numerator and denominator by $e^{(\alpha + \beta x_i)}$

$$= \frac{e^{\alpha + \beta x_i}}{1 + e^{\alpha + \beta x_i}}$$

That is how we model simple logit model that means when the dependent variable was dichotomous in nature.

Now if we extend this simple multinomial logit and make is a more generalized case when the individuals are facing more than 2 choices. Then the multinomial logit model or MLM, multinomial logit model this would become

$$= \frac{e^{\alpha_j + \beta_j x_i}}{\sum_{j=1}^{3} e^{\alpha_j + \beta_j x_i}}$$

you have 3 choices. This is the generalized logit model which we call multinomial logit model.

Now one's striking feature of this multinomial logit model if you look at closely, we have attached j subscript for the Intercept α and the slope coefficient β, why this is so? That is something we need to understand. So, j subscript is attached with the intercept α and the slope coefficient β, why this is so? Because in this model if you look at in simple logit model there is no j over here, here it is only $\alpha + \beta x_i$ what does it mean?

That means while facing this type of choices whether to go for no college, 2 years college or 4 years college the same individual i is attaching differential weight to the explanatory variable. For example, let us say I have only one explanatory variable here income, so that means the coefficient, the ith individual, the weight attached to that particular coefficient income by the ith individual who is let us say not going for college at all is different from the other individual who is actually going for a 4 years college.

I will repeat again, here differential this β which if we assume that β's are basically the weight attached to the explanatory variable then the individuals are attaching differential weight to different explanatory variables. So, that while no college becomes more probable for ith individual, a 2 years college becomes more probable for another individual. Why this is so? Because ultimately each and every individual will derive some amount of utility by choosing no college, 2 years college and 4 years college.

So, a particular individual will choose no college, if choosing no college gives more utility than 2 years or 4 year college while for the other individual choosing 2 years college will give more utility compared to no college or 4 years college. How are these individuals deriving different utility? Because given the role of these β's they are attaching differential weight to the explanatory variable.

So, when a particular individual is deciding no college, 2 years college or 4 years college, let us say I have the only one expanded variable income. While considering 2 years college I will give a different weight to the income, while considering no college at all I will put a different weightage, while choosing 4 years college I will give another type of weight to the explanatory variable income.

So, my intercept and my slope are actually varying from one choice to another, that is why one choice becomes more probable compared to others. That is the reason here the individuals they are attaching differential weight for different choices to make and that is why choice specific weightage is attached which particular choice I will make and that depends on x_i.

That means here if we assume x is a vector then I have more than 1 explanatory variable in the model. So, that is the reason we attach a differential impact, differential weight to the explanatory variable. But earlier in this case the impact is same, the same weight is attached for all the individuals, for all the individuals they are choosing same α and same β. So, that means for if we extend this multinomial logit model for P_i that means I can write this is let us say P_{ij} instead of P_i I am writing P_{ij}= e to the power α j + β j x_i divided by this.

So, since

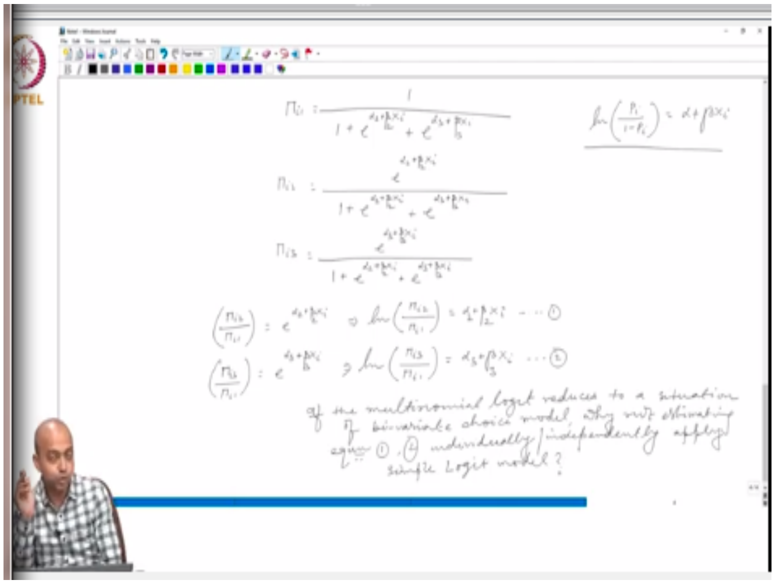
$$(\Pi_{i,1} + \Pi_{i,2} + \Pi_{i,3}) = 1$$

we cannot actually estimate all these 3 probabilities separately because if we estimate P_{i1} and P_{i2}, P_{i3} will automatically be estimated by 1 minus this. That is the reason we will select one category as our base category, for example let us say no college is a base category. Since we cannot estimate P_{i1}, P_{i2} and P_{i3} individually as

$$(\Pi_{i,1} + \Pi_{i,2} + \Pi_{i,3}) = 1$$

Once 2 probabilities are estimated the third one will be automatically estimated. So, what we assume? We assume 1 category as base and put intercept and slope = 0 for that category. So, let us assume that first category is the base, let us assume that α_1= β_1 = 0, let us assume that.

(Refer Slide Time: 34:56)



So,

$$\Pi_{i,1} = \frac{1}{1 + e^{\alpha_2+\beta_2 x_i} + e^{\alpha_3+\beta_3 x_i}}$$

$$\Pi_{i,2} = \frac{e^{\alpha_2+\beta_2 x_i}}{1 + e^{\alpha_2+\beta_2 x_i} + e^{\alpha_3+\beta_3 x_i}}$$

$$\Pi_{i,3} = \frac{e^{\alpha_3+\beta_3 x_i}}{1 + e^{\alpha_2+\beta_2 x_i} + e^{\alpha_3+\beta_3 x_i}}$$

And then what you will do?

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

and this implies if we take log of both sides log of

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

Similarly

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

equals to that implies

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

let us say this is equation 1 and this is equation 2.

Now if we look at equation 1 and 2 then it appears exactly like the simple logit model wherein, we were estimating.

$$ln\left(\frac{P_i}{1 - P_i}\right) = \alpha + \beta x_i$$

So, this is the simple logit model what we are discussing, where $\left(\frac{P_i}{1 - P_i}\right)$ is basically the odds ratio, odds in favour of happening that event. So, that means this is basically the relative probability of success $\left(\frac{P_i}{1 - P_i}\right)$.

If we assume P_i is the probability of success then $\left(\frac{P_i}{1 - P_i}\right)$ is a relative probability of success and that we have estimated using the simple bivariate model. Now here in the context of multinomial logit also ultimately the kind of model what we derive is basically the bivariate. Because $\frac{\Pi_{i,2}}{\Pi_{i,1}}$ is the relative probability of second choice over the first one, that means in this example we can think of this is basically we are estimating a relative probability of a 2 years college over no college.

And equation 2 tells us relative probability of a 4 years college over no college and that depends on the explanatory variable x_i. So, if we derive this type of bivariate model, so that means if the multinomial logit reduces to a situation of bivariate choice model why not estimating equation 1 and 2 individually and independently in a applying simple logit model? That is the question.