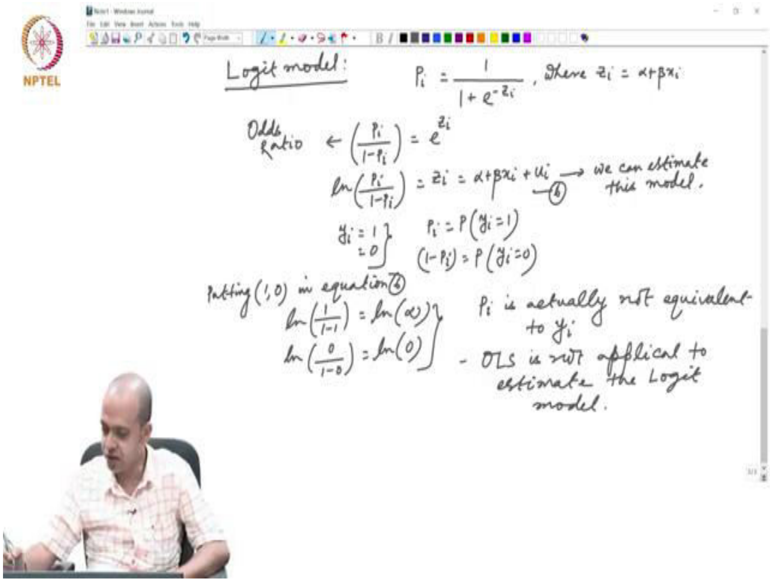
Environmental and Resource Economics
Professor Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology Madras
Lecture 41
Qualitative Response Models- Linear Probability Model, Logit and Probit Models Part - 2

(Refer Slide Time: 0:15)



P i since this is p i is probability of owning a car, or probability of owning a house, p i by 1 minus p i is called odds, in favor of happening the event, or odds in favor of owning a car. Since the numerator is p i we will say that, this is odds in favor of owning a car, if we calculate 1 minus p i by p i, that is also odds ratio, but then that will indicate odds against happening that event.

So, here p i by 1 minus p i is your dependent variable and we can you have to take log of this. Now, once if you think of estimating this model, see how will you estimate this model? You have information on y, a information on y and y can take two values, y i equals to either 1, or 0, this is y, this is y i, y i equals to 1 and which is basically indicates the individuals is owning a house and this is indicating not owning a house, that means what is p i and p i is probability y i equals to 1, and 1 minus p i is basically probability y i equals to 0.

Now, to estimate this model, this is let us say model, let us say this is model 6, to estimate this model first of all you need to have the information on the dependent variable. So, we do not have information on p i, rather we have information on y i. Now, apparently you may think of you can put y i value in this equation and you will construct the dependent variable.

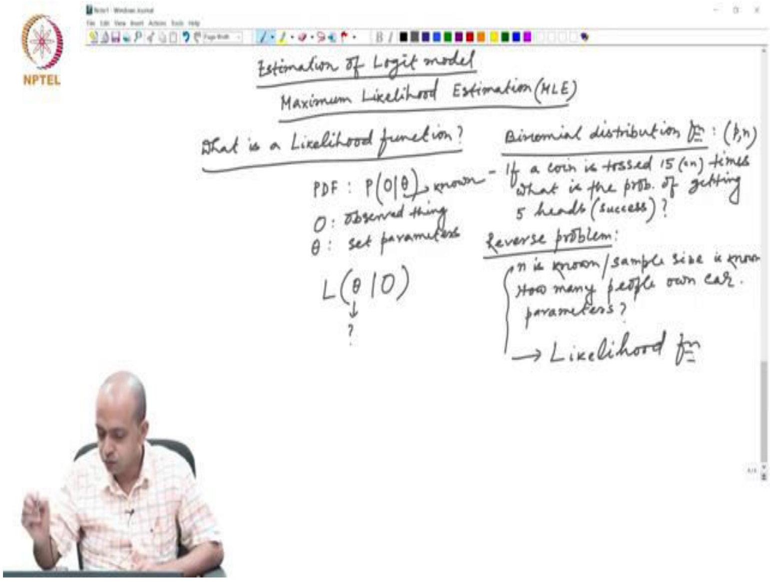
Now, if you put 1 and 0 here, what will happen? If you put 1 and 0, putting 1, 0, in equation 1, in equation 6, what you will get? You will get, so if you put 1, then that would become log of 1 by sorry log of 1 by 1 minus 1. So, that would become 1 by 0, so that means log of this would become your dependent variable. And if you put 0 here, then what will happen? This would become log of 0 by 1 minus 0 equals to log of 0.

So, this would become your dependent variable. Can we estimate this model? No, that means we cannot put 1 and 0, in this equation y, because see here, what we are thinking that p i is equivalent to y i, but that is not the case, p i indicates the probability, which will lie between 0 and 1. But here what you are, you what you observe is the realization, y i is the realization, some people they have owned the car, that is why you have put 1, some people they do not have car that is 0.

So, that is the realized thing, but what you are thinking of the probability which is unobserved probability of owning a car is not observed, rather what you observe is the ultimate realization, whether the person has taken or not, that is the decision, after taking the decision what you observe is a realized fact. But what we are thinking here, in terms of this model 6 is log of p i by 1 minus p i, which is not observable.

So, what we can say that p i is actually, not equivalent to y i. So, that is why in this model, we cannot estimate using the OLS method, because the dependent variable itself we do not have information, we cannot put 1 and 0 here, and we cannot estimate the model. So, that means OLS is not applicable to estimate the logit model, we need to go for a different route, we need to take a different route. And what is that route? Let us discuss now.

(Refer Slide Time: 6:07)



So, estimation of logit model, that we are going to discuss, estimation of logit model, how we are going to estimate, using maximum likelihood function or maximum likelihood estimates. So, let us say, I am writing maximum likelihood estimation, which in short known as MLE maximum likelihood estimation, that we are going to use.

But before we talk about maximum likelihood estimation, we need to know, what is a likelihood function? This is very important likelihood function can you think of, what is a likelihood function? The concept of likelihood function is closely related with probability distribution or probability density function.

So, if you know probability distribution function or probability density function, then you can easily understand the concept of probability, sorry likelihood function. Now, what is the probability distribution function? In any probability distribution function, if you think of let us say, binomial, Poisson or any other type of probability distribution function.

Suppose you are thinking of the simple probability distribution function, the binomial function. And the probability distribution function, how it is characterized? Let us say I am writing PDF, probability distribution or probability density function. PDF is characterized by a observed thing given the parameters, I will explain this, when I am specifying a particular probability density distribution function, let us say binomial, I can find out from there, if I cause, if I toss a coin, let us say 15 times, what is the probability of having, let us say 5 heads?

So, that means from PDF, let us say from binomial distribution function, let us say binomial distribution function, I am thinking of binomial distribution function which is given by 2 parameters n and p. So, I can answer this type of question, if a coin is tossed n times, that means let us say 15 times, what is the probability of getting 5 heads, let us say this getting a head, we are calling it this is success, that you can find out, if the probability distribution function is given.

Now, if I reverse this question, so if I reverse this question, here the parameters are known and you are trying to find out the probability of getting 5 heads out of 15 trials. So, you are tossing the coin 15 times, that means number of trials is 15, and you can find out what is the probability of getting 5 successes, or 5 heads in 15 trials. What is the reverse of this problem, can you think of, the reverse problem, reverse of this, that means I know in my sample, how many trials I have made, let us say 15, that means let us say I am asking 15 individuals.

So, my sample size n is 15, so the reverse problem is n is known, n is known that means sample size is known, I also know that, in that out of these 15 individuals, how many people they have car.

So, how many people own the car? Here I was trying to find out that, how many heads I will get, that means what is the number of success. Here if I define that owning a car is a success of that trial, that means I already know out of 15 individuals, how many of them are having a car. Then I what I do not know is the parameters, parameters are unknown.

So, that means in a sample size of 15 individuals, 5 individuals are having a car, that is my sample. And I have already observed that sample, here in terms of o, o is the observed thing, which in probability distribution function, you are trying to find out, you are trying to find out probability of getting 5 success, that is denoted by o. And what is known to you? Known to you is theta, which is the set of parameters.

So, here 0 in sorry, o, indicates observed thing, and theta basically indicates set of parameters. In binomial distribution function, this p the parameter is known for a single toss it is half, that is known to you. So, given theta, theta set of parameters in probability distribution function you are trying to find out probability of o, that means probability of 5 heads. But in a likelihood function, in a likelihood function I already know that, that is known to me what I try to find out is theta.

So, this is unknown, that means parameter are unknown, here this is known, that is why I say that likelihood function is just the reverse of probability density or distribution function. In likelihood function, I would like to get the parameter, because a set of parameter will only define the PDF.

So, that, I can observe that sample with maximum probability, that is why it is called maximum likelihood function, that means in short, what I want I already observed a sample, where out of 15 individuals, 5 individuals are owning a car, you give me the set of parameters, that will maximize the likelihood of observing that particular sample.

Because from a given population, I could have observed some other samples, some other samples of 15 individuals and in that sample, it would have happened that, out of 15, 2 individuals are having a car, or 14 individuals having a car, or 10 individuals having a car.

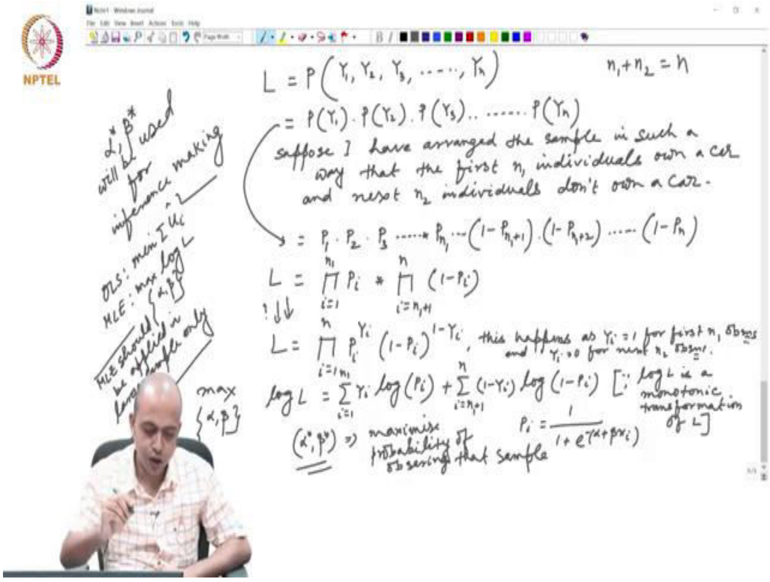
So, if you change the parameter, you will get different type of probability function, distribution function. So, I would like to get that set of parameter, that will maximize the likelihood of observing that particular sample. And that is the technique, what we apply here in maximum likelihood method, instead of minimizing the error term, that we used to do in ordinary least square method.

So, this approach is little different, maximum likelihood estimates is different from OLS in that regard, in OLS to get the parameters, we minimize some of the error terms, that means summation UI hat square. Here we are maximizing the likelihood. What is the likelihood? Likelihood of observing that particular sample, where out of 15 individuals, I already observed that 5 individuals they own a car, they own a car.

So, that is just the reverse of probability distribution function, in probability distribution function given theta I would like to estimate the probability of a observed thing. And what is the observe thing? That probability of 5 heads, that I would like to observe.

Here in likelihood function, I have already observed the o, but I would like to get theta set of parameters, that will maximize the probability of observing o, that means probability of observing a particular sample, in which 5 out of 15 individuals own a car, that is the likelihood function. So, this reverse problem will give you likelihood function.

(Refer Slide Time: 18:35)



Now, how will you apply this likelihood function in the context of logit model estimation. So, let us say that, the likelihood function is L denoted as probability of observing a sample. And as you know probability of observing a sample means, probability of observing Y 1, probability of observing Y 2, probability of observing Y 3, dot, dot, dot, probability of observing Y n.

So, your sample consists of n number of responses basically, Y takes the value 1 and 0, this is the response function. So, if you observe all your Y is that is basically the likelihood function, that is basically the likelihood function. And this so that means likelihood function is the joint probability of observing Y 1, Y 2, Y 3, dot, dot, dot, Y n.

So, that means this is nothing but probability of Y 1, then probability of Y 2, then probability multiplied by probability of Y 3, then probability of observing Y n. This is the likelihood function, which is nothing but the joint probability distribution function.

Now, suppose I have arranged the sample, suppose I have arranged the sample in such a way, that the first n 1 observation or n 1 individuals, own a car and next n 2 individuals do not own a car, that is how I have arranged. First that means where n 1 plus n 2, equals to, n 1 plus n 2 equals to n, which is the total sample size.

So, that means what I am saying, probability Y 1, Y 2, dot, dot, dot, n 1 up to n 1 it is actually, it is p i, because p i is the probability of observing the sample. So, from here what I can write equals to, equals to p 1, multiplied by p 2, multiplied by p 3, how long this will continue up to p of n 1, because first n 1 observations, first n 1 individuals will own a car, that is why p 1 p 2, p 3, dot, dot, p n 1.

And what would be the next term? Next term would be 1 minus p n 1 plus 1, because next person does not have a car, that is why 1 minus P, for first n 1, p 1, p 2, p 3, dot, dot, p n 1. Next observation, next individual does not own a car, that is why 1 minus p n 1 plus 1, then 1 minus p n 1 plus 2 dot, dot, dot, 1 minus p n, that is how I have arranged the data.

Now, if you write this expression in a concise format, then what you can write, this is nothing but product of i running from 1 to n 1, p i and then this is multiplied by i running from n 1 plus 1, to n, 1 minus p i. And this again you can write as 1 grand product p i to the power Y i into 1 minus p i to the power 1 minus Y i.

Now, this step is little interesting, how I could write this step, from this step to this step you need to understand how? See this happens, this happens because Y i equals to 1 for the first n 1 observations and if Y i equals to 1, then this would become p i and Y i equals to 0, for the next nth observation that is why it will become 1 minus p i, because this would become 0.

So, ultimately this will result in p i into 1 minus p i, this happens because as Y i equals to 1 for first n observations and Y i equals to 0, for next n 2 observations. Now, if you take log, then what will happen? If you take log, log of L equals to what you can write, this you can say that, log of Y i sorry, Y i log, p i, Y i log p i plus 1 minus Y i, log of 1 minus p i, i running from 1 to n 1, and here i running from n 1 plus 1 to n. Why I have done that? Because log L, why I did this, because log L is a monotonic transformation of L.

So, this is the thing log L equals to Y i, log p i and plus 1 minus Y i into log 1 minus p i. Now, you can substitute the value of p i. What is the value of p i? Where p i equals to, p i equals to 1 by 1 plus e to the power minus alpha plus beta sorry, this is your p i. So, you can substitute p i here and then you are trying to maximize this with respect to alpha and beta.

So, that means I am trying to maximize this with respect to alpha and beta. So, I want that particular set of alpha and beta, that will maximize this my likelihood function. And this likelihood function is basically probability of observing that particular sample, in which out of 15 individuals, 5 individuals own a car.

This is the mechanism, this is the mechanism of maximum likelihood estimates for estimating the logit model, instead of minimizing the errors sum summation ui hat square, we are maximizing log L with respect to alpha and beta. And you will get the optimum alpha star and beta star, that will maximize probability of observing that sample.

So, I will take that particular alpha hat and beta hat. So, that means alpha hat and beta hat, then will be used for inference making. So, in OLS, I was minimizing sum of ui hat square. So, OLS in case of OLS what you are doing, you are minimizing summation u hat square, but in MLE, what you are doing is maximizing log of L, with respect to alpha and beta, that is the difference between this and this.

And there are certain advantages of this MLE, the advantage is that, the this alpha star and beta star, what you get after maximizing log L, that is basically they asymptotically, they are asymptotically efficient and they approach normality, that is also asymptotically.

So, they are efficient asymptotically, that means what? That means they are efficient in large sample. And they approach normal, they follow the normal distribution, that also asymptotically that means, the MLE method should be applied, in large sample only.

So, whenever you are estimating logit model, you have to keep in mind, since the underlying estimation strategy is maximum likelihood and alpha hat and beta hat, they are efficient asymptotically, that means they show the efficiency property in large sample, you must have a large sample to estimate your logit model.

So, we should not apply the logit model in small sample like 3, 30, or 40, we should have minimum, minimum 200 to 250 observations, for logit model to estimate. So, with this, we are closing our discussion today, tomorrow we will discuss about another class of another particular model of binary response model, or qualitative response model. And then we will also estimate this will see how to estimate the model using a particular data set.