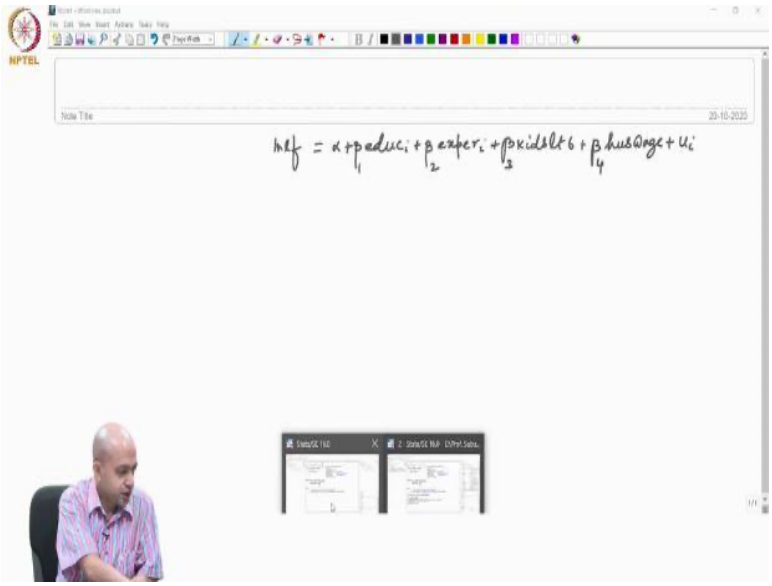
Applied Econometrics
Prof. Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology-Madras

Lecture-32
Qualitative Response Model Part-V

In our discussion of qualitative response model , we were discussing about the estimation part**.** We were using the statistical software stata to estimate the logit and linear probability model. We are basically estimating the model where the dependent variable is married women's labour force participation.

Whether, a particular married woman will participate in the labour force or not is a function of several factors. But we are considering only four factors, number of kids below 6 years of age, husband's wage, education of the married woman and then experience if any of the married woman.

(Refer Slide Time: 01:21)



So, basically our model is,

$$Lnlf_i = \alpha + \beta_1 educ_i + \beta_2 exper_i + \beta_3 kidslt6_i + \beta_4 huswage_i + \mu_i$$

and, we estimated the above model. **(Video Starts: 02:19)**

So, let us see, how we can estimate this model. The command for logit model is logit and then your dependent variable in labour force i.e inlf . instead of typing the variables name in command bar you can directly click on variable also, stata will take them. If you have to put education, if you double click on the variable's name then the variable will automatically appear in command bar. Education and then kidslt6 and then husband's wage and put enter.

Then, you will get results for your estimation. From this output what we can say whether there is positive or negative relationship between a particular variable. Let us say education and the labour force participation, which shows for, education is 0.19 and it is highly significant also which indicates that as education increases, probability of labour force participation also increases.

But, we cannot take these values as a marginal effect. That means, we cannot say that for one unit change in education what is the change in probability, because our dependent variable is basically $ln\left(\frac{P_i}{1-P_I}\right)$. So, that means you can say that for a unit change in education log odds ratio changes by 0.19**.** And then we have to specifically put a command to estimate the marginal effect.

What is the marginal effect command? As we discussed, the command for marginal effect is mfx. If you enter mfx, then you will get the marginal effect**.** Now we can say that for unit change in education your labour force participation changes by 0.04 unit. here, we need to discuss little more on this marginal effect. First of all, when we are saying a unit change in education, probability of labour force participation of a married woman changes by 04 unit.

That means whenever you want to measure a change, the change has to be made based on some reference point. If you do not know the reference, then the change concept does not make any sense**.** What is the reference point? That means, you need to calculate a base probability, based on which all these changes are actually calculated. And the base probability is calculated using the y value.

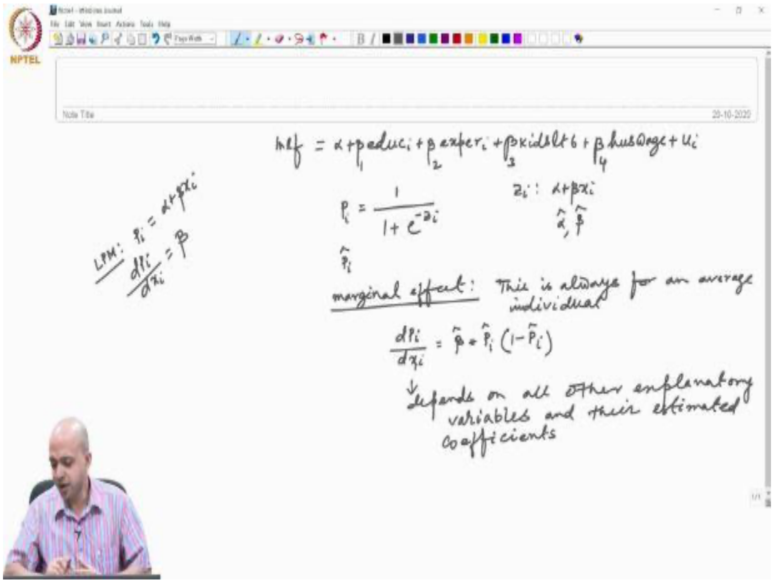
This is actually your $\hat{Y}_i$ that means, you can say this is your $\hat{P}_i$ equals to 0.57. Now, what is this probability? As you know that for any econometric models, whenever you are going to interpret the beta (β) coefficient is always interpreted as a change. That means, for a unit change in any explanatory variable or a percentage change in any exploratory variable what is the change in your dependent variable**.** And for that change concept as I said, you must have a base point.

For a standard regression model what is that base point that means, you estimate the model then you must calculate $\hat{Y}_i$. When you say that, when income changes from 1 unit, what is the base point for that? From the point, how will you calculate? when your model is

$$y_i = \alpha + \beta X_i$$

when income changes by 1 unit your consumption changes by (β) unit and what is the reference point for consumption that is actually $\hat{Y}_i$**. (Video Ends: 07:08)**

(Refer Slide Time: 07:09)



Similarly, in this context when you are saying that P_i equals to let us say for logit model

$$P_i = \frac{1}{1 + e^{-z_i}}$$

This is your logic model, where,

$$Z_i = \alpha + \beta x_i$$

So, once you estimate the model then immediately you must calculate the $\widehat{P_i}$ So, the change in probability is from this particular base point, which is $\widehat{P_i}$; that you have to remember. And stata is showing this $\widehat{P_i}$ as 0.5733.

We have to remember that all this change, all this interpretation for the coefficient is for an average individual. The marginal effect mfx command provide is always for an average individual. Why this is so? because, your marginal effect shows

$$\frac{dp_i}{dx_i} = \beta * P_i\,(1 - P_i)$$

and once you estimate these then all these are hat.

Now, why I am saying that marginal effect concept is only for the average individual because, how will you calculate $\widehat{P_i}$? To get $\widehat{P_i}$ you must specify some x_i because, from this model you will get $\hat{\alpha}$ and $\hat{\beta}$. Once you know your $\hat{\alpha}$ and $\hat{\beta}$ you must plug in some value for x to get your $\widehat{P_i}$. And what value of x you will put, the average value of x. That means, this estimated probability $\widehat{P_i}$ is calculated as at sample mean.

It means, when you have these variables education, experience, kids and husband's wage; this marginal concept is actually applicable to that particular married woman, whose education level is at sample average, whose experience is at sample average, whose number of kids are at sample average and whose husband's wage is at sample average. That is why that particular married woman is called an average individual or a representative individual.

This average individual this is actually an imaginary concept, because, in your data set no one actually, no one is having that level of education. Every individual is having her own level of education, own level of experience, kids and husband's wage, but interestingly my marginal effect concept is related to only that average individual that may or may not exist which I do not know.

It may so happen that by chance there is an individual, whose number of kids as at sample average education and all other factors is at sample average, which is quite unlikely you may or may not get, but the interpretation is always only for average. That is applicable for the standard econometric model as well as in this case of qualitative response model. So, that means since P_i depends on your x_i and x_i is basically all these independent variables.

Even though, you are interested to calculate the marginal effect for let us say ith x_{1i}. Let us say this x_{1i} it all depends on all other explanatory variables, which is not the case in linear probability model. Because, in linear probability model what was happening your $P_i = \alpha + \beta x_i$. So, if you differentiate in linear probability model in LPM what was happening your

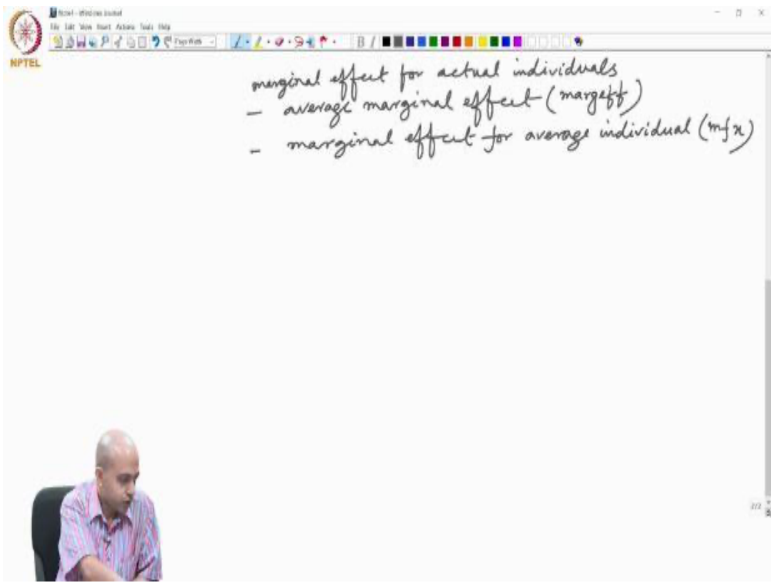
$$P_i = \alpha + \beta x_i$$
$$\frac{dP_i}{dx_i} = \beta$$

what is beta? Beta is attached to only x_i but here it is P_i and P_i is based on all other explanatory variable. That is why while linear probability model marginal effect is calculated based on only that particular variable for which you are interested to calculate marginal effect.

In the context of logit model, it all depends on other factors as well. Because $\widehat{P_i}$ you have to calculate based on your $\widehat{\beta_1}x_{1i} + \widehat{\beta_2}x_{2i}$…. like that. That is one thing you have to all other explanatory variables and their estimated coefficient. Now, suppose I do not want that type of marginal effect that means, I do not want the marginal effect for the average or representative individual who is an imaginary person.

Rather, I want to calculate the marginal effect for each and every individual and then I can take the average of that marginal effect; please try to understand what I am saying. In this context when I am getting result of mfx command to get the marginal effect, this marginal effects stata is calculating for the average individual that means, that a representative married woman, whose education, experience, number of kids and husband wage; all are fixed at the sample average. Instead of doing that, I want to get another marginal effect concept.

(Refer Slide Time: 15:02)



This is marginal effect concept 2; marginal effect for actual individual. That means marginal effect. If you have n number of individuals you will calculate marginal effect for each and every individual And then you can run another routine in stata for which, you can get the marginal effect for each, and every individual calculated and then get the average of that.

Thus, when education changes for 1 unit what is the change in probability for the first individual, for the second individual, for the third individual up to n number of individuals. And then I can calculate the average of those all that marginal effect. That is also another marginal effect based on average of each individual so, this is called average marginal effect.

It means there are two marginal effects. Marginal effect for each average individual which is called average marginal effect and earlier what we have calculated that was marginal effect for average individual**.** So, this we have already calculated by mfx command but, this average marginal effect calculation requires a separate command, which is called margeff. **(Video Starts: 17:50)**

After estimating the model if you put margeff command let us say I will once again estimate then I will put margeff. This is called average marginal effect. This is slightly lower; see the marginal effect of this model is 0.0429 and earlier it was 0.0487. So, average marginal effect is little lower than the marginal effect for the average individual, but you should know both these two types of marginal effect and you should also know in what sense this marginal effects are actually different.

Whenever you are estimating qualitative response model; please remember to report both types of marginal effects and you should also interpret it. In standard model, since the $\widehat{\beta_1}$ coefficients they are direct measure of marginal effect we do not need to mention anything else, but in logit model or in qualitative response model since, beta coefficients are not a direct measure of marginal effect.

We need to specifically mention or report the marginal effect. Otherwise, your research should be incomplete. Many a times after estimating this, we forget to report the marginal effects, because at the end of the day what we are interested in specifically for the marginal effect. For a unit change in any of this explanatory variable on an average what is the change in the probability of labour force participation.

That is what we want and that is why you need to specifically report this type of marginal effects. Now, once you are done with the marginal effects then the other important things about the marginal concept. See here, when I am calculating marginal effect then what you need to do you need to specify the x values at the sample average, because you need to calculate $\widehat{P_i}$

It means, when number of kids changes by 1 unit, from that sample average to 1 unit change. So, that means I am calculating the probability of change for that woman, who is having that many number of kids. But suppose my interest is little different I want to calculate the change in probability of labour force participation for the married woman, who does not have any kids and then I want to see from 0 kid to 1 kid what is the change in probability of labour force participation.

It also you can calculate. Suppose, you estimate this logit model and then you can put this command mfx compute at kidslt6 = 0, sorry, my command I made a mistake here mfx compute we have to put a comma and then add kidslt6 = 0. Now, look at what is happening. So, here stata has calculated the marginal change assuming kids = 0. So, when the married woman does not have any kids at that situation when education changes by 1 unit what is the change in probability.

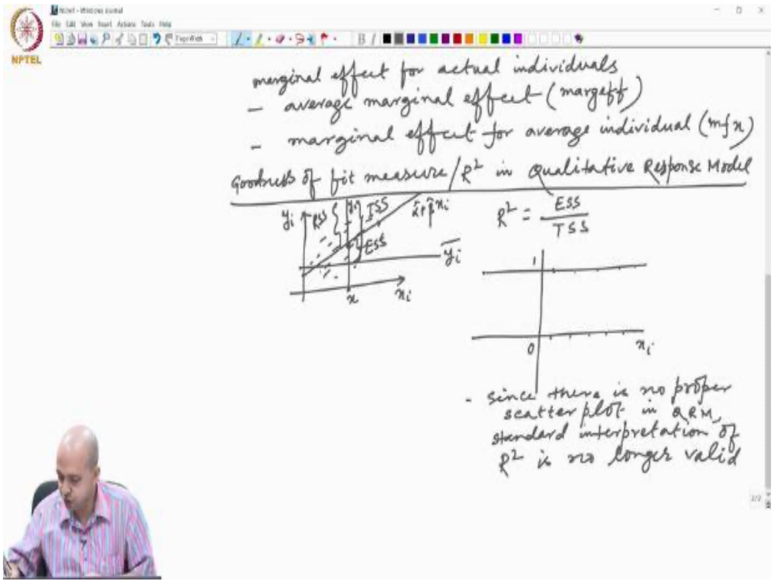
When husband's wages by 1 in what the change in probability is? And when kids changes by 1 unit this is very important. Now, see from 0 kid to 1 kid what is the change in probability 0.32 and what is the earlier one if you compare. Look at this any earlier model 0.3483 and here it is 0.32 it is a little different 0.3483 and this is 0.32. So, from this model, once again you just see this is your original model.

This is the margin command, and this is kids. So, there is slight changes in probability when you specify x = 0 this is 0.3483 and I forgot to mention the experience. Let me put experience also logit in labour force and then I will put experience also experience this is the model and then you calculate mfx then this is 0.33. And now, if you put mfx compute what is the command I put at in the bracket kidslt6 = 0, mfx compute you have to end the bracket.

When you have 0 kid, then as you get the first kid probability of labour force participation decreases by 0.30. But, when you have the average number of kids 0.23 it is 0.33. So, that means lower reduction in the probability when you get the first kid and then it keep on increases. So, which makes sense also. So, it is quite difficult to manage more kids than when you have the first one. So, that means you can calculate you can specify a specific labour of value for the x and calculate the marginal effect by this command mfx compute at kids = 0.

And similarly, you can also put education labour = 0 or a specific value of education husband wage so on and so forth based on your research interest. These are all based your marginal effect, but when you estimate any econometric model like previous cases you need to know how good your model in terms of goodness of fit is. That means R^2. **(Video Ends: 26:29)**

(Refer Slide Time: 26:30)



So, let us discuss something about goodness of fit measure or R^2 in qualitative response model**.** Now, you have to keep in mind that the standard interpretation of R^2, the R^2 which we derived earlier in case of classical linear regression model is no longer applicable in qualitative response model. So, standard R^2 makes little sense in the context of qualitative response model.

Why this is so? Because, in the case of linear regression model what is the R^2 measure; let us say this is your x_i, this is your y_i and let us say you have this type of scatter plot**.** And what is the regression line first of all? The regression line $\hat{\alpha} + \hat{\beta}x_i$ is basically a representation of your data. That means, this is some kind of average line, which represents this data point.

And what is your R^2? Your R^2 is for any given value of x_i what is your predicted value, what is your observed value, let us say this is your observed value and let us say this is my $\overline{y_i}$**.** So, this is my total variation TSS and out of which I have explained this much. This is $\widehat{y_i}$, this is ESS and this is RSS and your R^2 you have defined as ESS by TSS.

$$R^2 = \frac{ESS}{TSS}$$

So, out of your total variation in y, which is measured by $\sum(y_i - \overline{y_i})^2$ divided by total variation and how much you have explained $\sum(\widehat{y_i} - \overline{y_i})^2$and then your R^2 is ESS by TSS. So, when you have a scatter plot like this, so that means given x_i your y_i can take any value and you have this proper scatter plot and you have a regression line, which basically represents that scatter plot.

And what is goodness of feet, how good my regression line is in representing this scatter plot, which is nothing but the raw data. But, suppose we are in a context, where this scatter plot itself does not exist. Then what you will do, that means this regression line itself make little sense we cannot get this type of a regression line which will call the average line, average representation of this scatter plot.

Then how will you take this RSS, ESS by TSS it does not make any sense and that is what is happening in the context of qualitative response model. How? Let us see here. This is in this axis let us say I am measuring x_i and this is y_i and this is 0, this is 1. So, that means for any value of x you are either here or here that means, your probability is here or here either the event will happen or not happen depending on your x value let us say that is income.

So, that means there is no scatter plot and if that is the case if the scatter plot does not exist in the context of qualitative response model standard R^2 interpretation also is no longer valid. So, what I am saying since there is no proper scatter plot in qualitative response model, I will write QRM standard interpretation of R^2 is no longer valid, because I cannot get a line like this and unless I get that type of regression line, I cannot apply the ESS by TSS formula to get the R^2. So, what we need to know, we need to derive alternative measures of R^2.