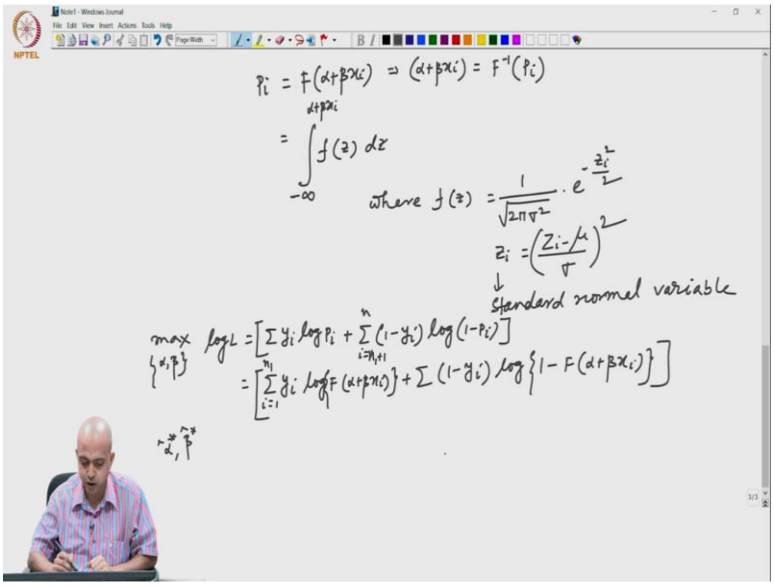
Environmental and Resource Economics
Professor Sabuj Kumar Mandal
Indian Institute of Technology, Madras
Lecture 43
Qualitative Response Models - Linear Probability Model, Logit and Probit Models
Part - 4

(Refer Slide Time: 00:15)

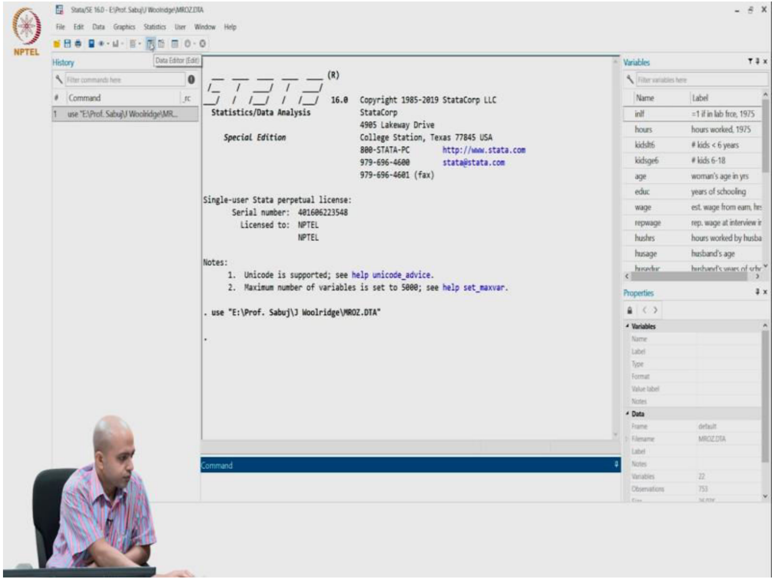


So that means this is an alternative derivation of the probit model and if you follow then you can derive the logit model also in this way because up to this when Pi equals to F of alpha plus beta xi that is same and depending on which particular cumulative density function you will get it will define whether it is a logit model or probit model or linear probability model.

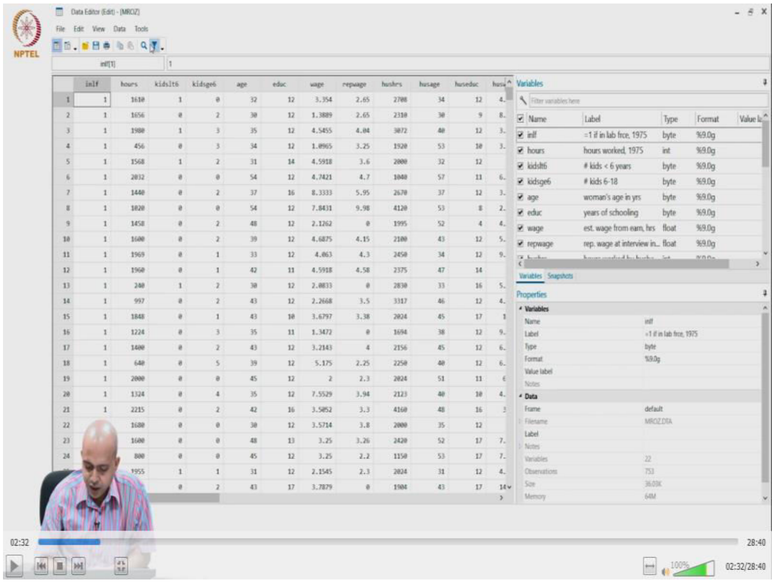
So, F of alpha plus beta xi equals to alpha plus beta xi in the context of linear model, linear probability model, F of alpha plus beta xi equals to 1 by 1 plus e to the power minus alpha plus beta xi in the context of logit. That means it assumes cumulative density function of a logistic distribution and here it is the cumulative density function of a normal distribution function, normal distribution function.

Where f z equals to 1 by root over 2 Pi e to the power minus z square by 2 and zi is actually standard normal variable defined as, and you know the standard normal variable it has 0 mean and 1 sigma square, sigma square equals to 1. So now what we will do, we will take one data set and then we will try to estimate the model using linear probability model then we have a logistic logit model and we have a probit model. We will see how to estimate.

(Refer Slide Time: 02:02)

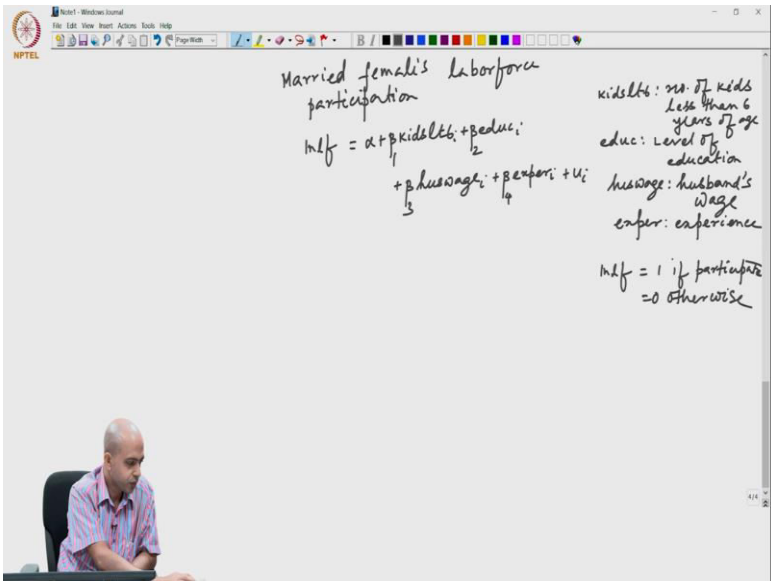


(Refer Slide Time: 02:21)



So, this is basically, this is a first of all let us, let us look at the data set. This is, this is a data set on married women's labor force participation. Married women's labor force participation. That means we are trying to understand, we are trying to understand this is the female labor force participation or if I would say that married females labor force participation.

(Refer Slide Time: 02:47)



This is married female labor force participation. And what is our model here? Our dependent variable is ln, sorry in inlf, that means in labor force, which is actually a function of several variables. Let us assume that whether the married woman has any kids below 6 years of age. Because if you have kids below 6 years of age, it is difficult for you to participate in the labor force.

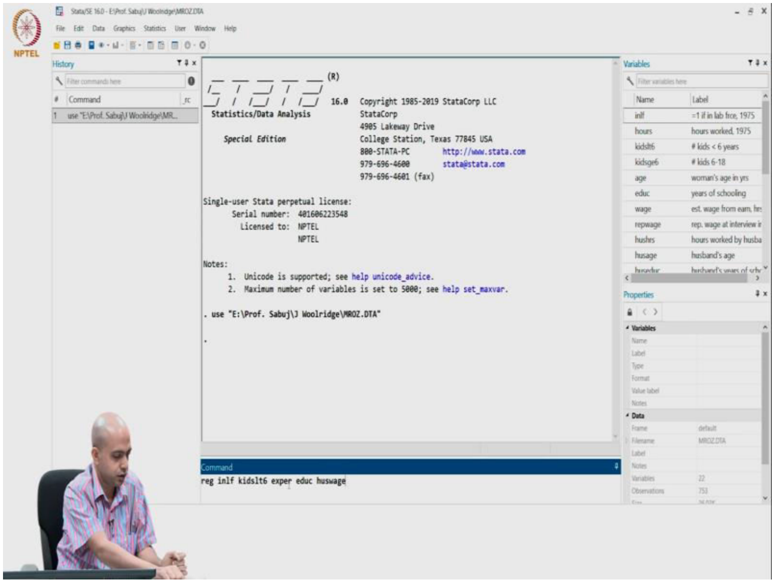
So let us say the number of kids below 6 years of age that is my first explanatory variable, which is kids. How I am defined in this model? Kids lt6, kids lt6, kids lt6, this is how I have defined kids lt6. Number of kids, kids less than, less than 6 years of age, 6 years of age that is your first variable. Then beta 2, beta 2, what is your education level, what is your education level?

So educ is the level of education of that married woman, level of education, level of education. Then plus beta 3, whether your husband is working? And what is husband's salary? So, beta 3 huswage. If your husband is earning a higher salary, then you are basically less likely to participate in the labor force. And then let us say that beta 4 indicates, whether you have any previous work experience or not.

If you are working earlier, then it is likely that again after marriage also you will continue to work and higher probability of labor force participation. There are so many other variables, but for the time being I have included only four variables, four explanatory variable. One is, so this is huswage, is basically husband’s wage, whether you, what is the your husband salary or wage.

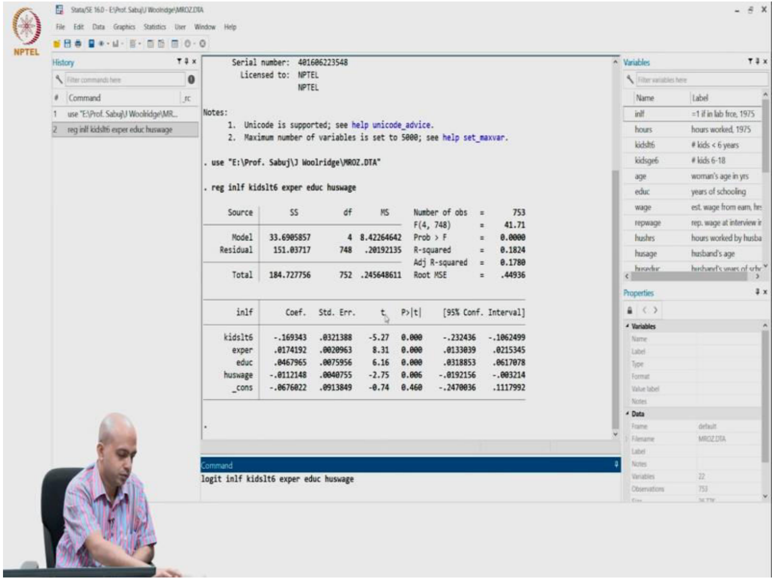
And exper is basically your experience. These are the four variables you have included in your model. And here in labor force in labor force this is the independent variables. And in labor force that is your dependent variable in labor force equals to 1 if participated, 0 otherwise. If you do not participate, then that is 0, that is your dependent variable. So dependent variable is a binary 1. So, this is the model that we are going to estimate. This is the model. So first we will estimate the linear probability model or LPM.

(Refer Slide Time: 07:10)



So, in LPM what you basically do? You try to estimate the model using standard OLS method. So reg in labor force then your dependent variable is kids lt6 and then you have experience, exper, then education, and then haswage, which is husband's wage.

(Refer Slide Time: 07:39)



So, this is a purely OLS kind of model that you are trying to estimate, using in this data set. But now what is the result showing? That kids lt6 that is negatively related with the probability. So, for unit change in number of kids, when number of kids increases by 1, your probability of labor force participation decreases by 0.16, that is how you can interpret.

Similarly, as experience goes by goes up by 1 year, probability of labor force participation increases by 0.01 unit. Then when your education level goes up by one year, then your probability increases by 0.04 unit. And when your husband's wage increases by 1 unit then again, your probability of labor force participation decreases.

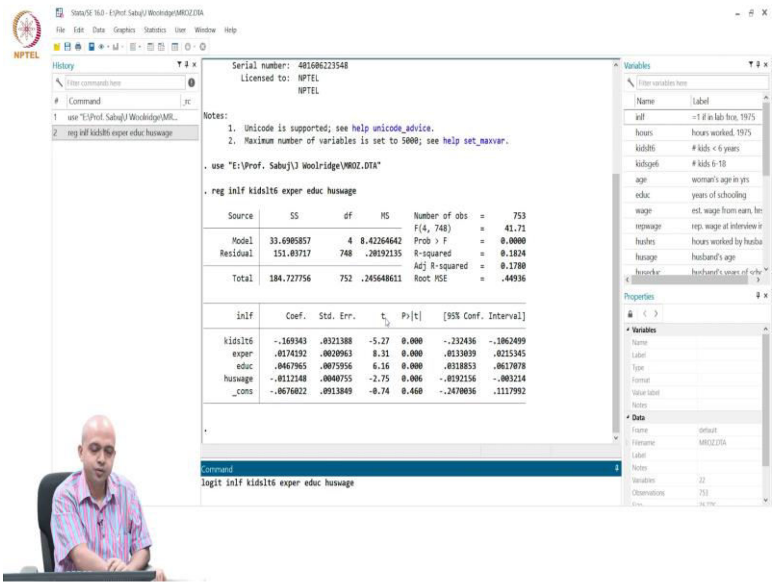
So that means, all these explanatory variables are giving signs according to what you expected. That is absolutely no problem. But where is the problem? Problem is if you recall in our linear probability model. What we said that the ui, ui basically not following a normal distribution. Because Pi equals to alpha plus beta xi plus ui, that means what do we model yi equals to alpha plus beta xi plus ui and depending on what value yi takes.

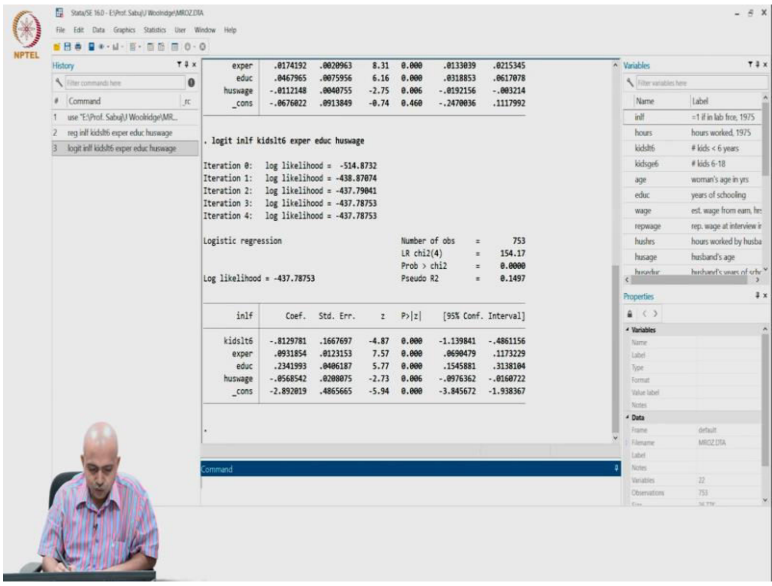
If yi equals to 1 then ui equals to 1 minus alpha minus beta xi. If yi equals to 0, then ui equals to minus alpha minus beta xi. So ui basically follows a discrete distribution. So that means this t statistic what we are getting in this, in this output that is actually a problematic t statistics. Because given ui follows a discrete distribution, actually we cannot rely on this t statistic. Because you have problem in your hypothesis testing.

All this t is assumed that ui actually follows normal distribution. Based on the normality of ui only, we construct the t statistic beta hat, by standard error or beta hat. But when ui itself does not follow a normal distribution, rather it follows a discrete distribution, in the context of LPM, how can you get this kind of reliable t statistics? So that is why we cannot rely on this t statistic for hypothesis testing.

And also, it has its own problem like Pi may not lie between 0 1 1, as we have discussed earlier. So that is why our next model, the same regression function we are going to estimate, now using, using the logit model. And how will you estimate the logit model? The command is l o g i t, logit. And then again, what is your dependent variable?

(Refer Slide Time: 11:04)

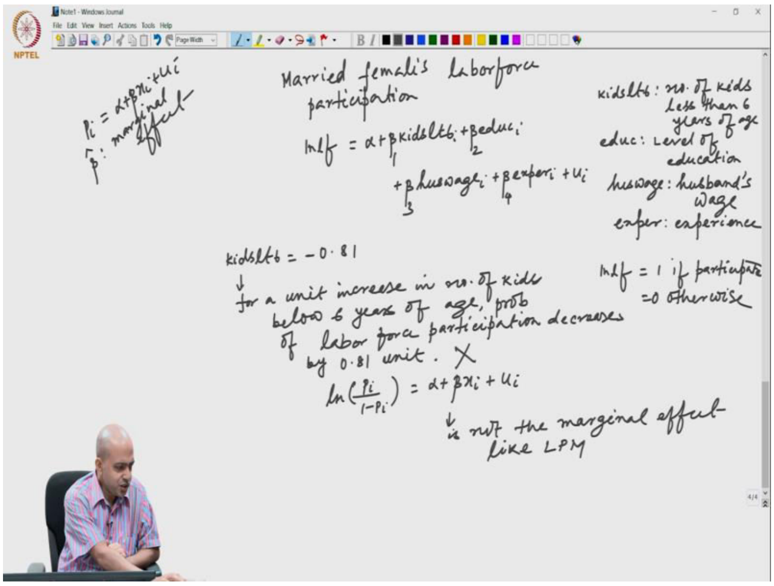




In labor force and then kids lt6, then experience, then education, huswage. These are the four. Then you put enter, then you put enter. And this is the output. And as we said, that in logit model and probit model, OLS is not applicable rather we are trying to estimate, by maximizing the likelihood function. That is why you see in the output that; they have given the log likelihood value.

The maximum value of the log likelihood. But once you estimate the model how will you interpret the coefficient? For example, coefficient of kids lt6 which is minus 0.81. Can we say that as, as number of kids increases by one unit, probability of labor force participation goes down by minus 0.1 in every unit.

(Refer Slide Time: 12:49)



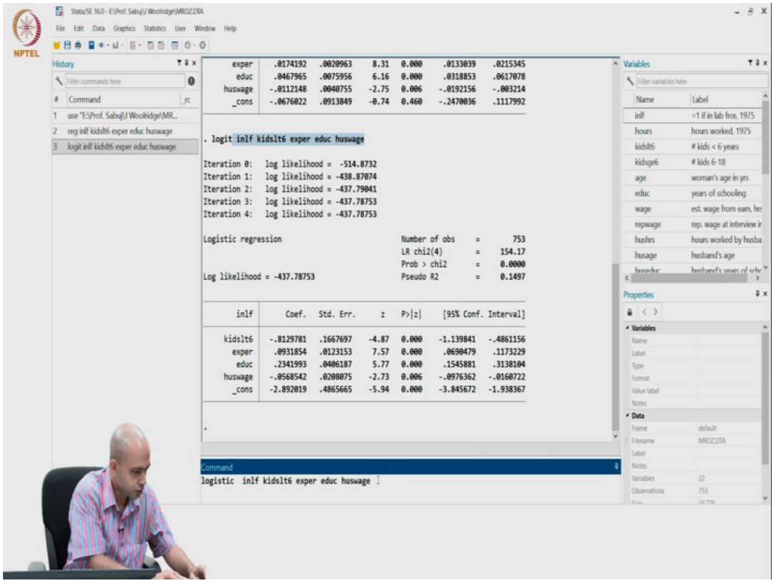
So that means I am trying to, I am trying to interpret this coefficient as, so now what I will do, I will now try to estimate the coefficient of kids lt6, kids lt6. Coefficient is, what is the coefficient? If you look at minus 0.81. And how we are going to interpret this? Let us say that I am saying for what is the interpretation, for a unit, increase in kids, in number of kids, below 6 years of age, below 6 years of age, probability of labor force participation, participation, labor force participation decreases by minus sorry 0.81 unit.

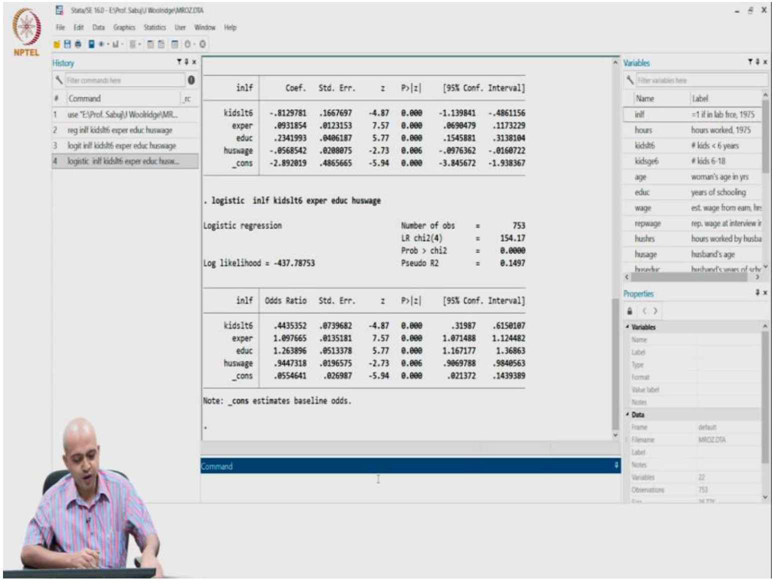
So that should be the interpretation. So that should be the interpretation of this, but if you interpret the coefficient in this way, your interpretation is totally wrong. You have to be very, very careful about interpreting the coefficient, when you estimate a logit model or probit model. Why this is so? Because you look at your model what you are estimating the model in logit, is log of Pi by 1 minus Pi equals to alpha plus beta xi plus ui. This is the model you estimated.

Now, if this is the case what is your dependent variable? Your dependent variable is actually log of Pi by 1 minus Pi. That means for a unit change in x there is a change in not Pi, rather log of Pi by 1 minus Pi. So that means for a unit change in number of kids, log odds ratio goes down by 0.81 unit. That is the interpretation. So, you cannot take beta as it is not, it is not, is not actually the marginal effect. Effect like, like LPM. So in LPM your model was Pi equals to alpha plus beta xi plus ui. In this model what is beta?

Beta hat actually directly the marginal effect, marginal effect. But here your model is log of Pi by 1 minus Pi equals to alpha plus beta xi plus ui. That is why this 0.81, what is the interpretation, as number of kids increases by 1 unit log odds ratio goes down by 0.81 unit. It is not the direct marginal effect. But if you interpret the coefficient in this way, it is little problematic to understand. As number of kids increases then you are saying that log odds ratio goes down by this much unit. It is little problematic in understanding.

(Refer Slide Time: 17:08)



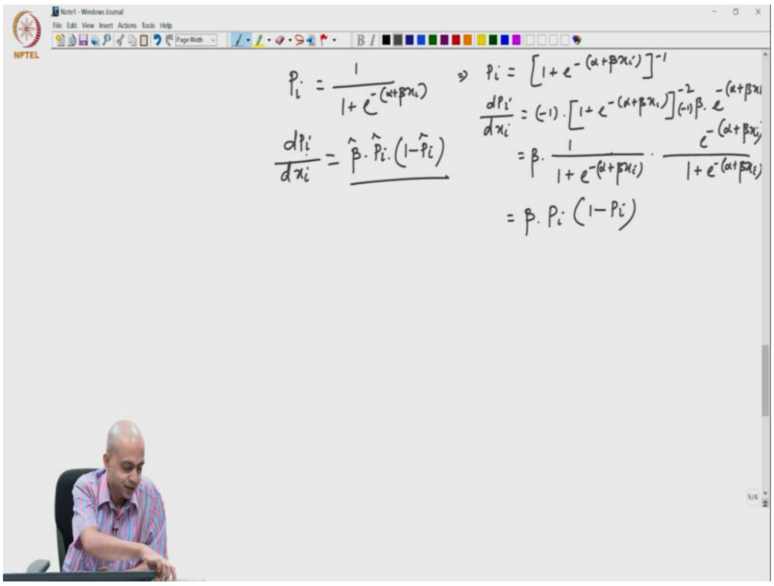


So instead, if you run this way, that means if you run this model, let us say l o g i s t i c logistic and then after logistic, what you do, sorry, logistic in labor force C. Now what you have estimated? Your dependent variable. When you put this logistic command is now odds ratio. And the coefficient is now 0.44, 0.44. So that means it is saying that as number of kids increases, odds ratio goes up by this and if you take log, then it will give a negative sign.

So, you will get odds ratio. A relationship between number of kids and odds ratio. But that is also not something which is easy to understand. That means one thing is very clear from this model that after estimating this, we need to separately calculate the marginal effect. It is not directly given. From the logit command, you will get log like, log odds ratio and in the logistic command you will get the odds ratio that means relationship with the independent variable and the odds ratio.

Earlier model when you put logit command, relationship with independent variable and log odds ratio that is what you are getting. But what I want is actually, a direct relationship between the explanatory variable with the probability. That is something what I want, that is something easy to interpret and easy to apply. But if you want to get that then, you need to specifically calculate. You need to specifically calculate.

(Refer Slide Time: 19:22)



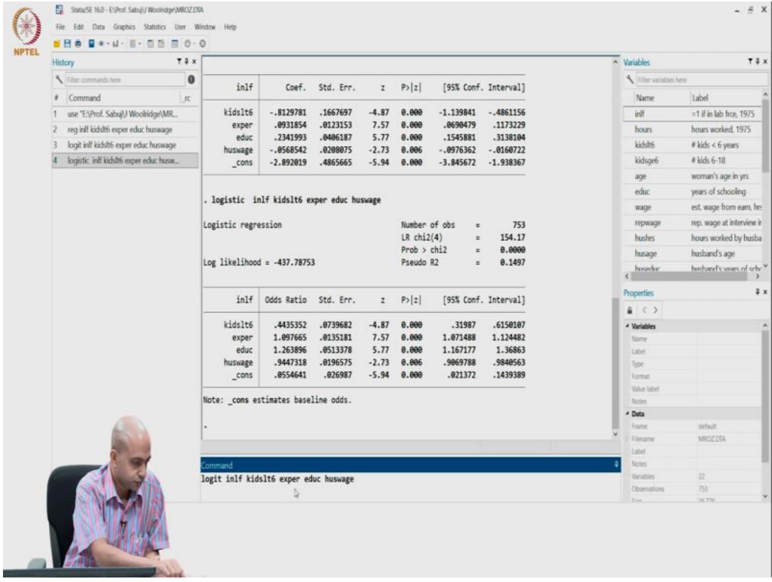
How will you calculate? In the context of logit Pi equals to 1 by 1 plus e to the power minus alpha plus beta xi. This is the model. So, what you want, you want d Pi d xi, that is what you want. Change in xi. How much change it is causing to y, Pi? That is nothing but d Pi d xi. And if you differentiate this, then what you get is basically beta into Pi into 1 minus Pi. If you differentiate, you can check you will get beta into Pi into 1 minus Pi, beta into Pi into 1 minus Pi.

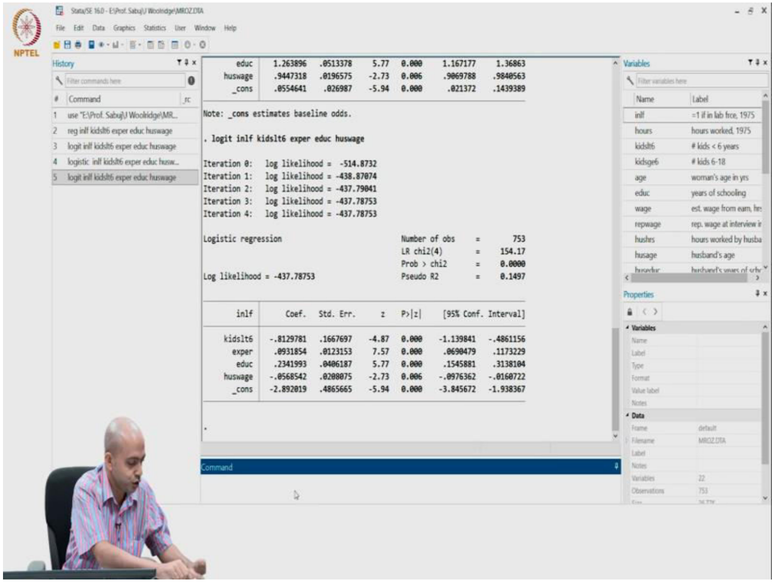
And I am not showing the differentiation, you can get it easily. This implies that Pi equals to basically 1 plus e to the power minus alpha plus beta xi, entire thing to the power minus 1. Then if you do d Pi d xi then you will see that this is minus 1 into 1 plus e to the power minus alpha plus beta xi minus 2 into beta into e to the power minus alpha plus beta xi minus alpha plus beta xi.

So that means from here, sorry, this is also minus beta. So that means you will get beta into 1 by 1 plus e to the power minus alpha plus beta xi into, into what you will get, you will get e to the power of this is 1 by 1 plus e to the power minus alpha plus beta xi into e to the power minus alpha plus beta xi. So, equals to beta, this equals to Pi. And this equals to 1 minus Pi. That you can get.

So, from here what I can understand that means to get the marginal effect, once you estimate you will get your beta hat and that you need to multiply with the Pi and 1 minus Pi. These will also become Pi hat and 1 minus Pi hat. So, if you get this way then only you will get d Pi d xi.

(Refer Slide Time: 23:03)

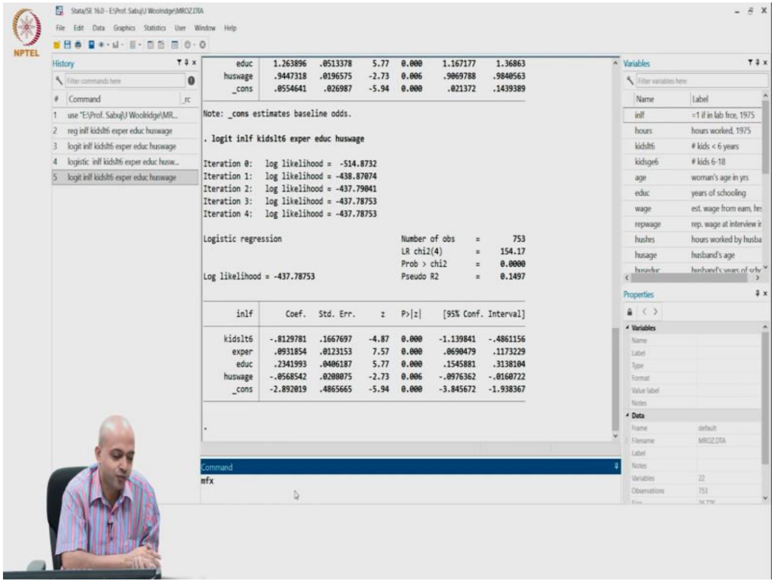


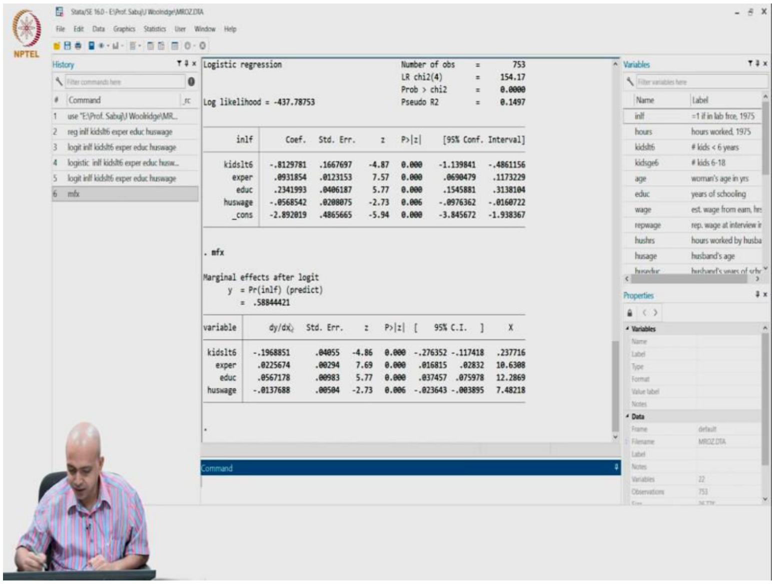


So now if you estimate this model, you estimate this model logit and then in labor force and then all your explanatory variable, kids lt6, then e x p e r, then education, then huswage, you estimate this model. And after that you need to give a specific command to get the marginal effect. So, from this result, at most you can say that there is some kind of negative relationship between number of kids and probability of labor force participation.

There is positive relationship between experience and probability of labor force participation. Like that you can say only whether there is a positive or negative relationship between a particular explanatory variable and probability of labor force participation. But how much does the probability change for a unit change of any of this explanatory variable, that you cannot get from this.

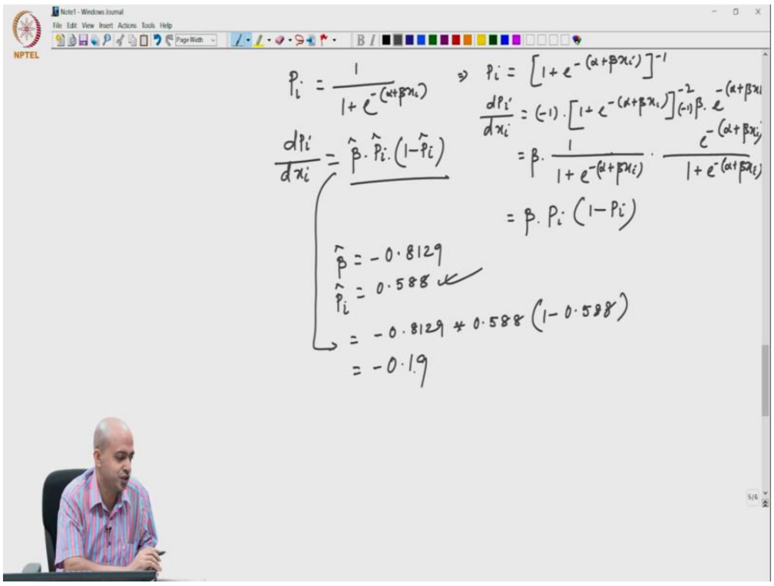
(Refer Slide Time: 24:20)

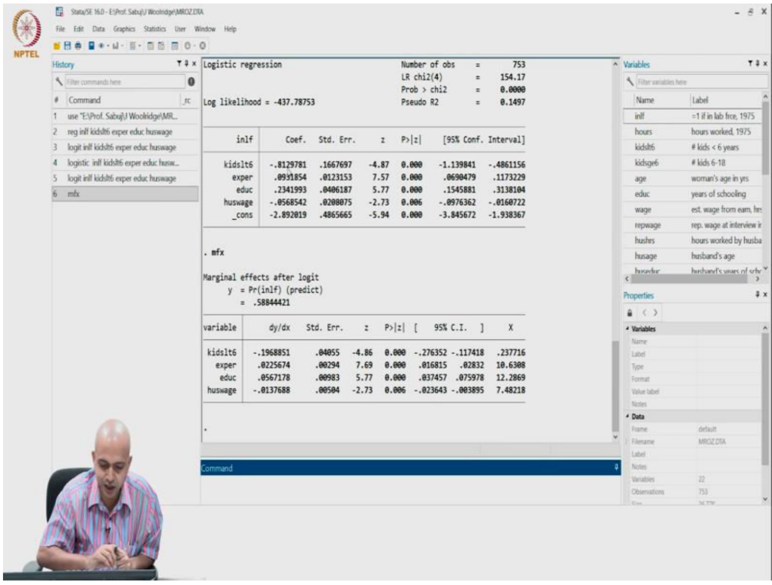




So, for that what you need to do? What you need to do? You put a specific command, which is called mfx, mfx. Now after putting the mfx, you are getting the probability. Now you see, here they are saying this is dy dx, dy dx means actually dP dx in our model.

(Refer Slide Time: 24:55)





Now as I said, as I said, the formula says, you need beta hat, you need Pi hat and then you need to multiply beta hat with Pi hat with 1 minus Pi hat. I will give you an example. So here, what is your beta hat for a particular variable? Beta hat is actually here. Look at minus 0.81, that is your beta hat, minus 0.8129, minus, so minus 0.81, 81, 8129, sorry 8129, 8129.

And then what is your Pi hat? So, this is actually your, I will write beta hat equals to beta hat equals to minus 0.8129. What is your Pi hat? What is your Pi hat? Pi hat is actually stata is showing, look at this, Pi had this probability of labor force participation predicted. That means that is actually your yi hat that is 0.588, 0.588, 0.588.

So now if you use this beta hat value and Pi hat value here, you will get d Pi d ix so, that means now what you do? You use this formula and this would become 0.8129 and then you multiply that with 0.588 and then 1 minus 0.588. You will get, if you do that then you will get your d Pi d xi, which is nothing but you can, you can calculate this at home and you can see that if you do so then your value would be minus 0.19.

So, this would become minus, minus 0.19. So, this is your marginal effect. Likewise, you can use this bit Pi hat value equals to this. And you can use all other beta hat value and you will arrive at this dy dx value as data is reporting, as data is reporting. This is how you have to estimate the marginal effect.

So that means one thing is very clear. While in linear model, the marginal effect is directly given, and that depends only on that particular experiment, coefficient of that particular explanatory variable, here even though you are actually estimating the marginal effect for a particular explanatory variable 'kids', since that involves beta hat into Pi into 1 minus Pi.

And what is this Pi hat? Estimated probability. Since Pi depends on all other factors dP d xi that means change in probability for a particular explanatory variable, depends on the estimated coefficient of all other explanatory variable. That is something different from the linear probability model. We will discuss other features of the logit model and probit model in our next class tomorrow. Thank you.