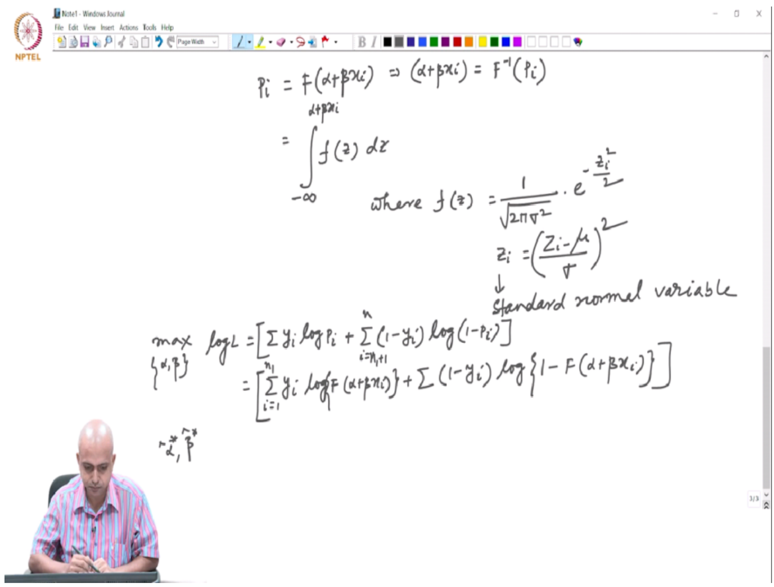
Introduction to Econometrics
Professor. Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology, Madras
Qualitative Response Models- Linear Probability Model, Logit and Probit Models Part - 4

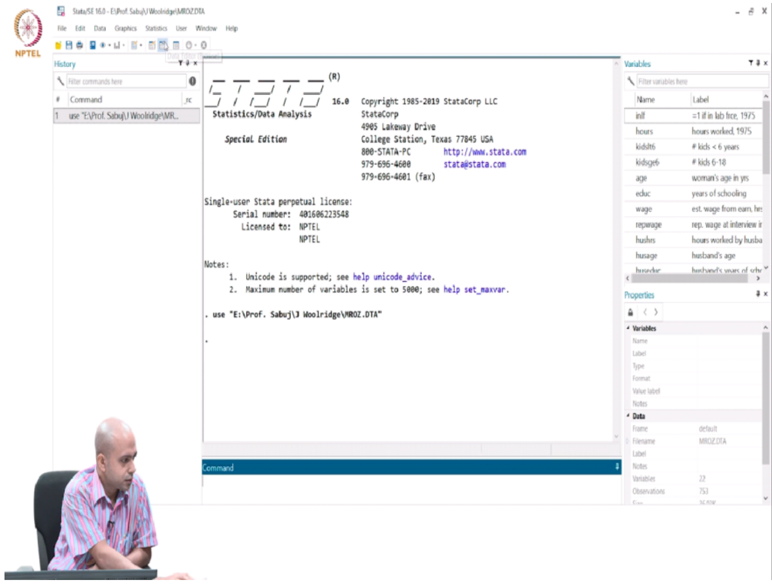
(Refer Slide Time: 00:14)

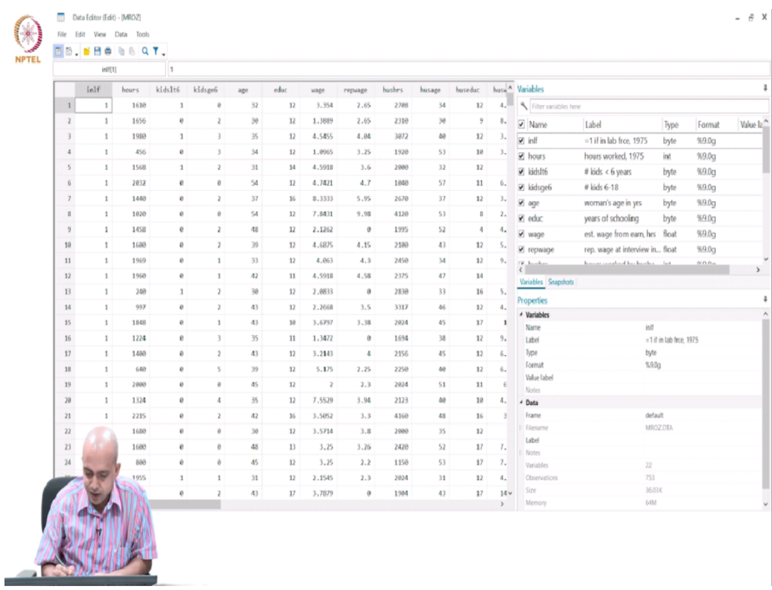


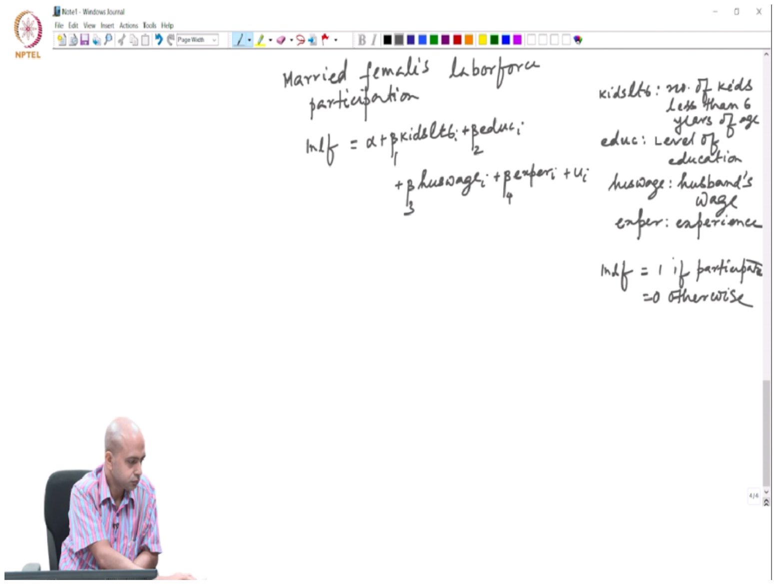
So, that means this is an alternative derivation of the Probit model. And if you follow then you can derive the Logit model also in this way because up to this when Pi equals to F(alpha plus beta xi) that is same and depending on which particular cumulative density function you will get it will define whether it is a Logic model or Probit model or Linear Probability model. So, F(alpha plus beta xi) equals to alpha plus beta xi in the context of linear probability model. F(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 mean 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 fz equals to 1 by root over 2 pi e to the power minus z square by 2. And zi, standard normal variable, it has 0 mean, and sigma squared 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 Logit model and we have Probit model. We will see how to estimate.

(Refer Slide Time: 02:03)







First of all let us look at the data set. This is a data set on married woman's labor force participation. This is the female labor force participation or I would say that married females labor force participation. And our dependent variable is 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.

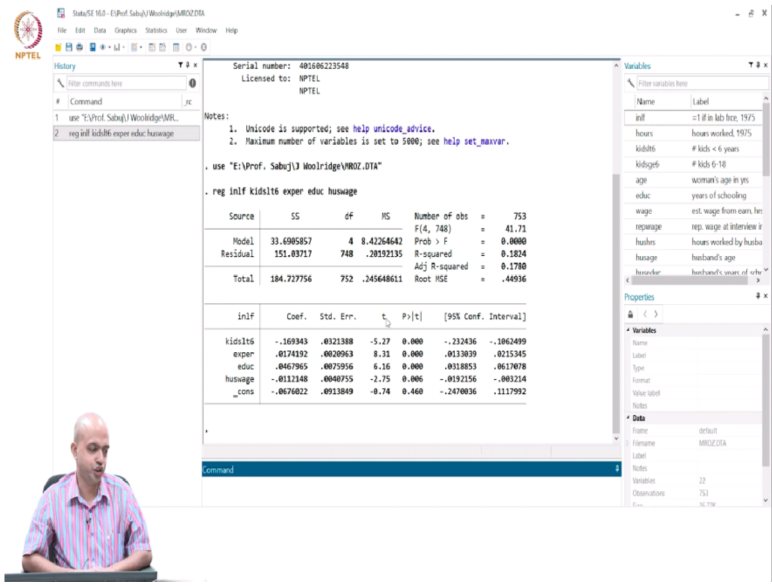
So, the number of kids below 6 years of age is my first explanatory variable, which is Kidslt6. Then we are to beta 2. Educ is the level of education of that married woman. Then plus beta3 whether your husband is working and what his husband's salary? So, beta 3 huswage, if your husband is earning a 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 were 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 4 explanatory variables. So, this is huswage is basically husband's wage and exper is basically your experience.

These are the 4 variables you have included in your model. And here 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 and 0 otherwise. So, this is the model that we are going to estimate. So first, we will estimate the linear probability model or LPM.

(Refer Slide Time: 07:09)



So, in LPM, what you basically do you try to estimate the model using standard OLS method. So, reg inlf kidslt6 exper educ huswage. So, this is a purely OLS kind of model that you are trying to estimate using in this data set. But the results show that kidslt6 is negatively related with the probability. So when the number of kids increases by 1 your probability of labor force participation decreases by 0.16. That is how you can interpret.

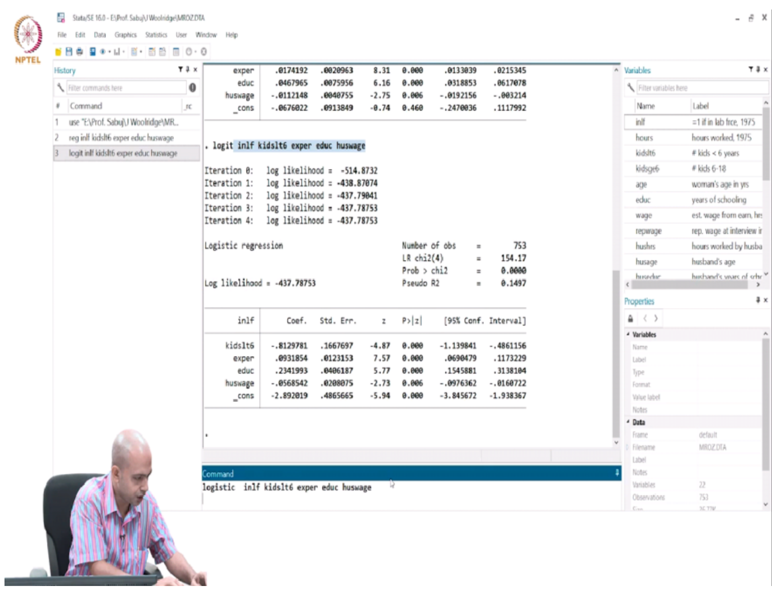
Similarly, as experience goes up by 1 year, probability of labor force participation increases by 0.01 unit. Then when your education level goes up by 1 year then your probability increases by 0.04 unit. And when your husband's wage increases by 1 unit, then again, your probability of labor force participation decreases. So, that means all these explanatory variables are giving signs according to what you expected.

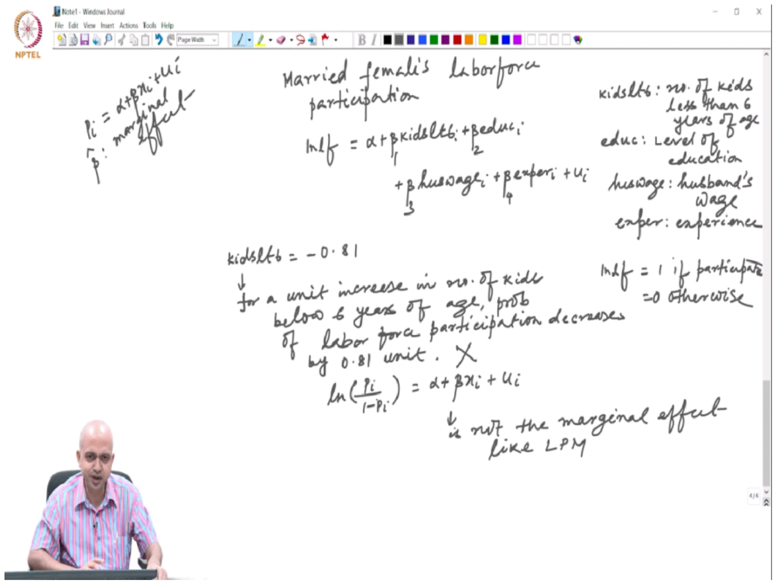
But where is the problem? Problem if you recall, in our linear probability model, what we said that the ui basically not following a normal distribution because Pi equals to alpha plus beta xi plus ui that means what do we model yi equals to alpha plus beta xi plus ui and depending on what value yi takes, 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 the t statistic, what we are getting 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 these kind of reliabile 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 and 1 as we have discussed earlier. So, that is why our next model the same regression function, we are going to estimate now using the Logit model and how will you estimate the Logit model?

(Refer Slide Time: 11:03)





The command is logit and then again what is your dependent variable? in labor force and then kidslt6, then experience then education, huswage, these are the four then you put in 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 a likelihood function that is why you see in the output they have given the log likelihood, the maximum value of the log likelihood.

But once you estimate the model how will you interpret the coefficient? For example, coefficient of kidslt6 which is minus 0.81 can we say that as number of kids increases by one unit, probability of labor force participation goes down by minus 0.01 unit, so that means, I am trying to interpret this coefficient as so now, what I will do, I will now try to estimate the coefficient of kidslt6. Kidslt6, 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 the unit increase in number of kids below 6 years of age, probability of labor force participation decreases by 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 a 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 a 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 actually the marginal effect 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.

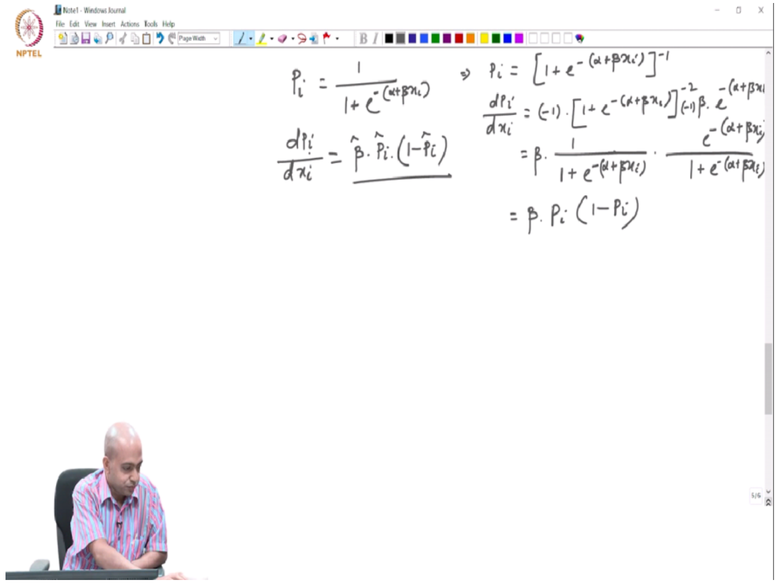
But here, your model is log of pi by 1 minus pi equals to alpha plus beta xi plus ui that is why this 0.81, what is the interpretation? As number of kids increases by one unit log odds ratio goes down by 0.81 unit, it is not the direct marginal effect. But, if you interpret the coefficient in this way,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

So instead, if you run this way that means, if you run this model, let us say, logistic and then after logistic what you do, logistic in labor force. See, 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. So, that means, it is saying that as number of kids increases all odds ratio goes up by this and if you take log then it will give a negative sign. So, you will get all 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 odds ratio and in the logistic command you will get the odds ratio that mean relationship with the independent variable and odds ratio.

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. How will you calculate?

(Refer Slide Time: 19:19)



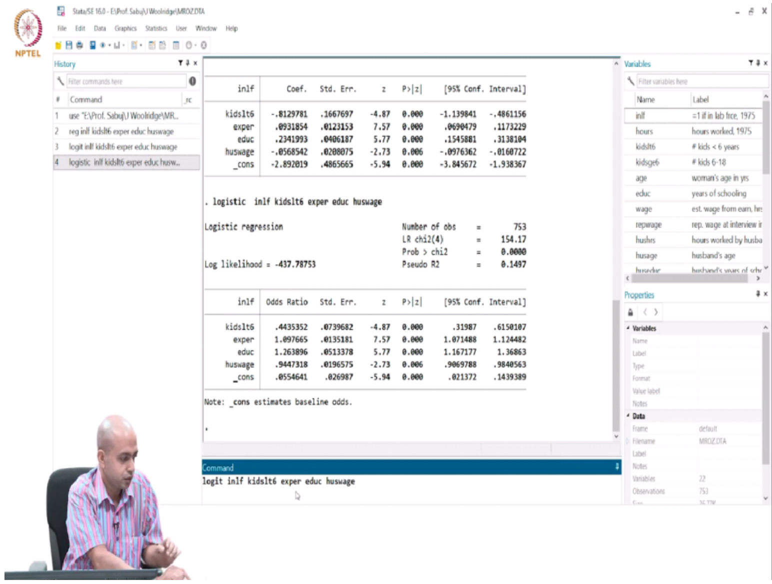
In the context of Logit pi equals to 1 by 1 plus e to the power minus alpha plus beta xi. This is the model. So, what do you want? You want dpi xi that is what you want. Change in xi how much change it is causing to bi that is nothing what dpi xi. And if you differentiate this then what you get is basically beta into pi into one minus pi. If you differentiate, you can check you will get beta into pi into 1 minus pi.

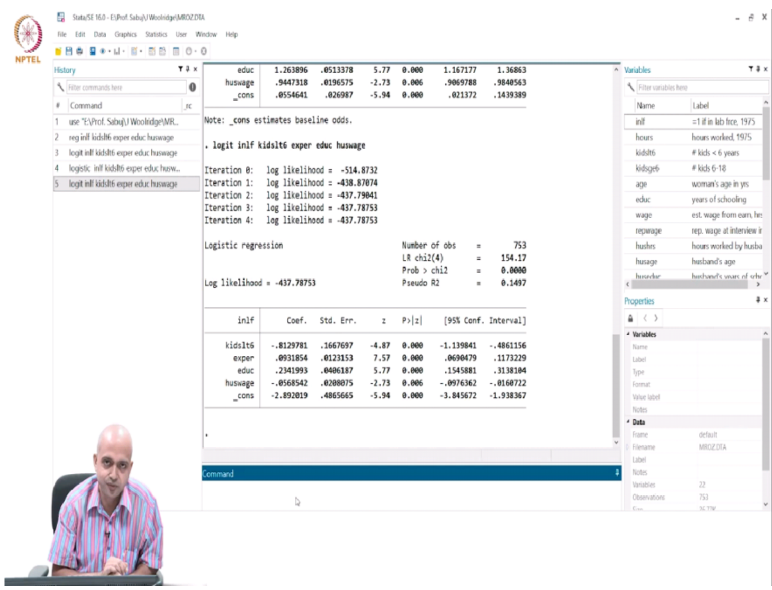
And I am not showing the differentiation you can you can get it easily this implies that pi equals to basically 1 plus e to the power minus alpha plus beta xi entire thing to the power minus 1. Then if you do dpi/dxi then you will see that this is minus 1 into 1 plus e to the power minus alpha plus beta xi minus 2 into beta into e to the power minus alpha plus beta xi. 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 what you will get? You will get e to the power, this is 1 by 1 plus e to the power minus alpha plus beta xi to 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 once you estimate you will get your beta hat and that you need to multiply with the Pi and 1 minus Pi, this will also become Pi hat and 1 minus Pi hat. So, if you get this way then only you will get dpi xi.

(Refer Slide Time: 23:04)

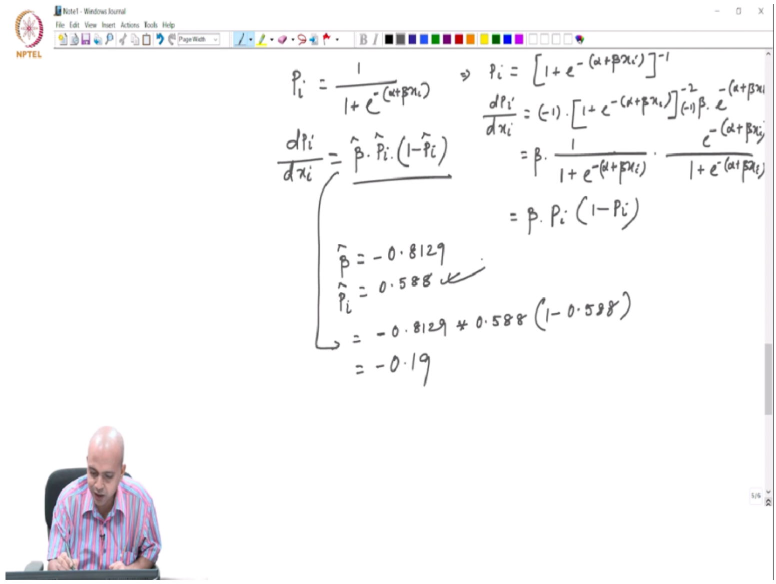




So now, you estimate this model Logit and then in labor force and then all your explanatory variable kidslt6 then exper then education then huswage and after that you need to give a specific command to get the marginal effect.

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. 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 or unit change of any of this explanatory variable that you cannot get from this? So for that what you need to do? You put a specific command which is called mfx.

(Refer Slide Time: 25:10)



After putting the mfx you are getting the probability. Now you see, here they are saying is dydx. dydx means actually dpdx in our model. Now 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 have. I will give you one 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.

So, minus 0.8129 and then what is your pi hat? I will write beta hat equals to minus 0.8129. What is your pi hat? Pi hat is the probability of labor force participation predicted that means that is actually your yi hat that is point 0.588.

So, now, if you use this beta hat value and Pi hat value here, you will get dpi/dxi. So that means, 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 your dpi/dxi 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 0.19. So, this is your marginal effort.

Likewise, you can use this 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. This is how we have to estimate the marginal effects. 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 explained 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 dpdxi 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.