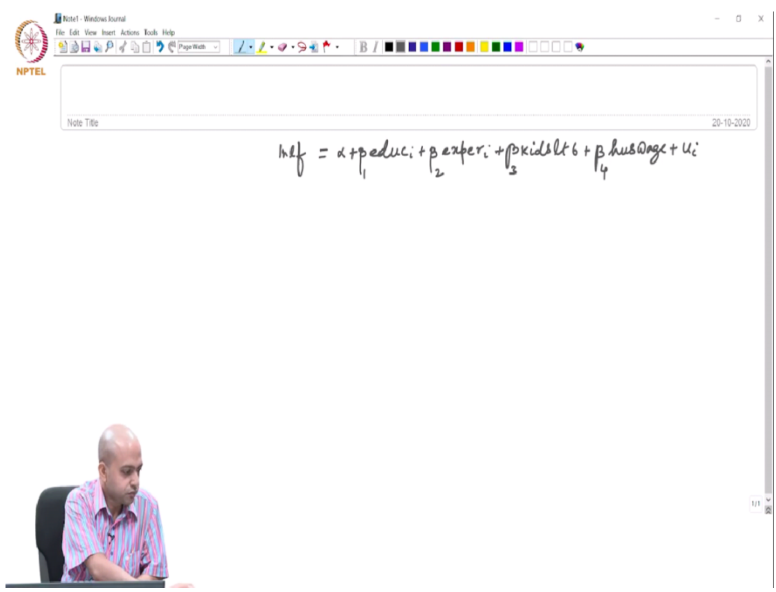
Introduction to Econometrics
Professor. Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology, Madras
Qualitative Response Models- Linear Probability Model,
Logit and Probit Models Part - 5

Welcome once again to our discussion of qualitative response model. And yesterday we were discussing about the estimation part. We are using the statistical software to estimate the Logit model and Linear Probability model that we are discussing. And what is the model we were estimating we are basically estimating the model or where the dependent variable is married woman's labor force participation.

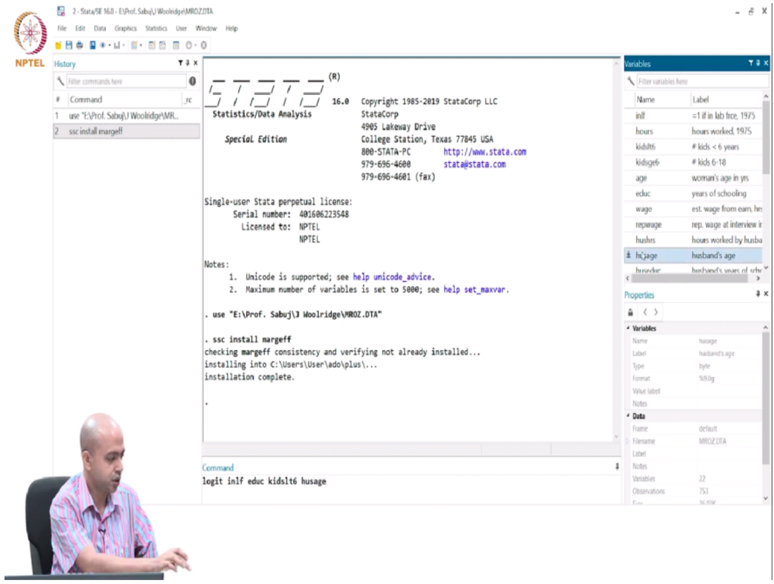
Whether a particular married woman will participate in the labor force or not that is basically a function of several factors, but we are considering only four factors. Number of kids below 6 years of age, then husband's wage, education of the married woman and then experience.

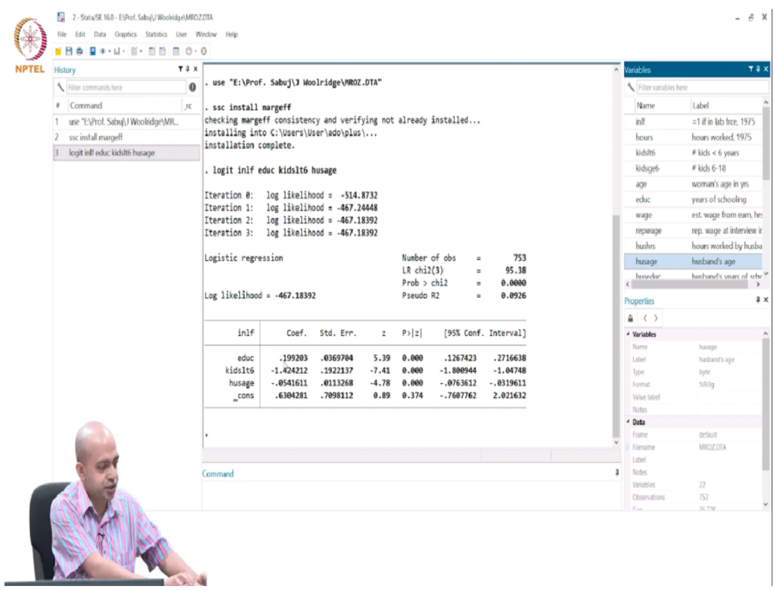
(Refer Slide Time: 1:20)

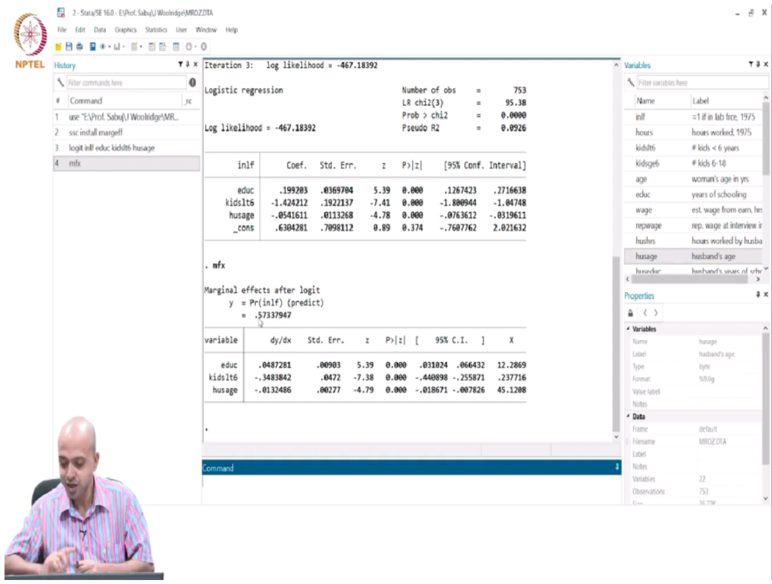


So, basically your dependent variable was in labor force Inlf, which is equal to alpha plus beta 1 education plus beta 2 experience plus beta 3 number of kids less than 6 years of age plus beta 4 which is husband's wage plus the error term, this was the model we were discussing. I will once again estimate the model.

(Refer Slide Time: 02:17)







So, let us see how we can estimate this. So, logit and then all your dependent variables- in labor force, Inlf then instead of typing the variables name here, you can directly click here also, Stata will take then you have to put education, if you double click on the variables name, then the variable will automatically appear here.

Education and then kidslt6 and then husband's wage and if you put enter then this is the model that you have estimated, which shows that this is the result. That means from this output we can say whether there is positive or negative relationship between a particular variable let us say education and the labor force participation which shows for example, education variable which is a 0.19 and it is highly significant also, which indicates that as education increases probability of labor force participation also increases.

But we cannot take these values as a marginal effect because our dependent variable is basically log of Pi by 1 minus pi. So, that means, you can say for a unit change in education log odds ratio changes by 0.19. And then we say that you have to specifically put a command to estimate the marginal effect. And what is the marginal effect command? The command is very simple mfx command, which is if you put enter then you will get the marginal effect.

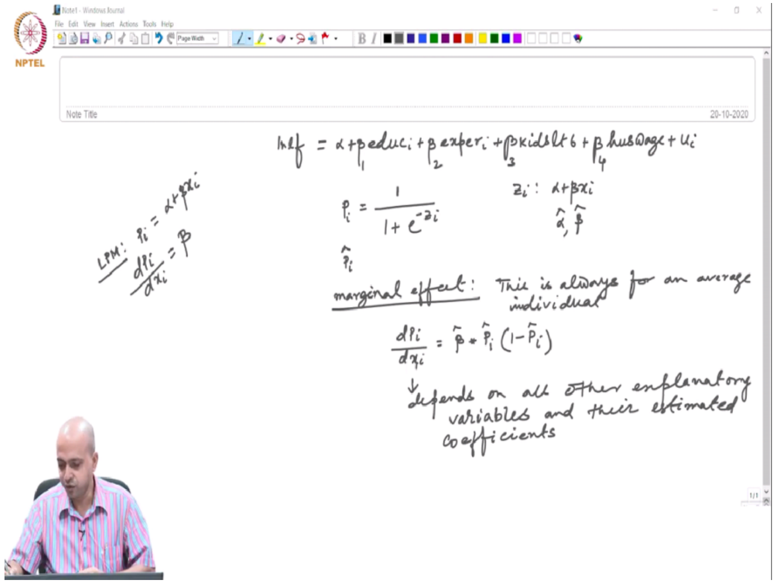
So now, you can say that for unit change in education, your labor force participation changes by 0.04 unit. Now, we need to discuss little more on this marginal effect. First of all, when I am saying that for a unit change in education, your probability of labor force participation of a married woman changes by 0.04 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.

So, 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 or Stata is showing, this is actually your yi hat that means, you can say this is your Pi hat equals 0.57.

Now, what is this probability? As you know that for any econometric model beta coefficient is always interpreted as a change, that means for a unit change in any explanatory variable or a percentage change in any explanatory variable, what is the change in your dependent variable. And for that change concept as I told that you must have a base point.

For the standard regression model, what is that base point that means, you estimate the model then you must calculate yi hat and then, when you say that when income changes from 1 unit, what is the base point from the point where, how will you calculate, that means, when your model is yi equals alpha plus beta xi that means, when income changes by 1 unit your consumption changes by this much unit. And what is the reference point for consumption? That is actually yi hat.

(Refer Slide Time: 07:08)



Similarly in this context, when you are saying that Pi equals to let us say for Logit model Pi equals to 1 by 1 plus e to the power minus zi, this is your Logit model where zi is basically your alpha plus beta xi. So, once you estimate the model then immediately you must calculate the Pi hat. So, the change in probability is from this particular base point which is Pi hat that you have to remember and Stata is using this Pi hat as 0.5733.

And also, you have to also remember that all this change, all this interpretation for the coefficient is for an average individual that means, the mfx command the marginal effect please keep in mind that this is always an average individual. Why this is so? Because your marginal effects shows dy/dxi equals to beta hat, I will say beta hat for beta hat into Pi into 1 minus Pi. And once you estimate then all these are hat.

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

That 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.

That means this individual is actually an imaginary concept because in your data set, 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, I do not know.

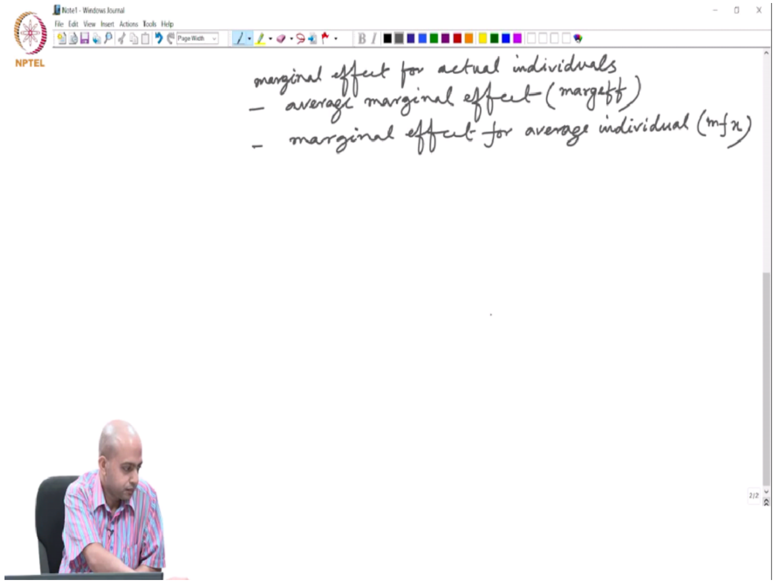
It may so happen that by chance there is an individual whose number of kids as at sample average, education and all other factors is at sample average, which is quite unlikely, you may or may not get. But the interpretation is always only for average. That is applicable for the standard econometric model as well as in this case of qualitative response model.

So, that means, since Pi depends on your xi and xi is basically all these independent variables, even though you are interested to calculate the marginal effect for let us say ith, x1i, let us say this is x1i, it all depends on all other explanatory variables, which is not the case in linear probability model. Because in linear probability model your Pi equals alpha plus beta xi, so if you differentiate in linear probability model in LPM, in LPM what was happening your Pi is basically alpha plus beta xi. So that means, if you take these it would be only beta.

So, what is beta? Beta is attached to only xi. But here, it is Pi and Pi is based on all other explanatory variables 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 Pi hat you have to calculate based on your beta1 hat x1i plus beta 2 hat x2i 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 a 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 effects. Please try to understand, what I am saying. In this context, when I am getting mfx command to get the marginal effect. These marginal effects data is for the average individual. That means the representative married woman whose education, experience, number of kids and husband wage all are fixed at the sample average.

(Refer Slide Time: 15:02)

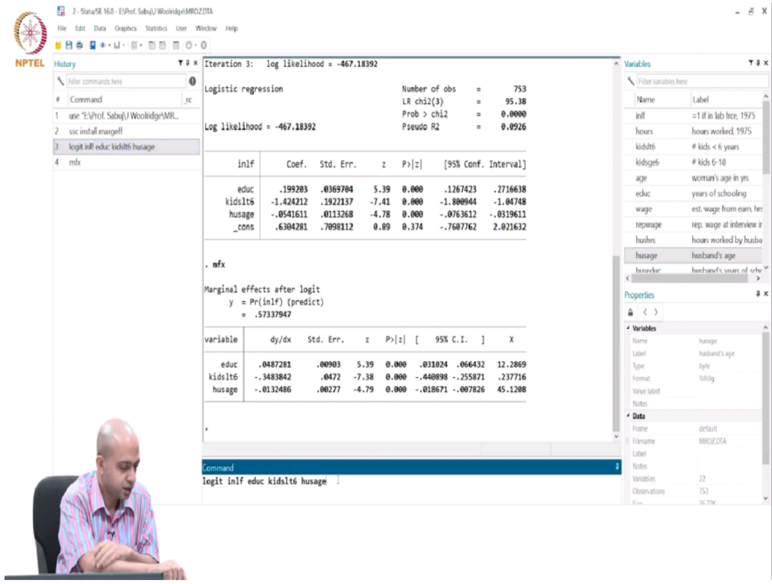


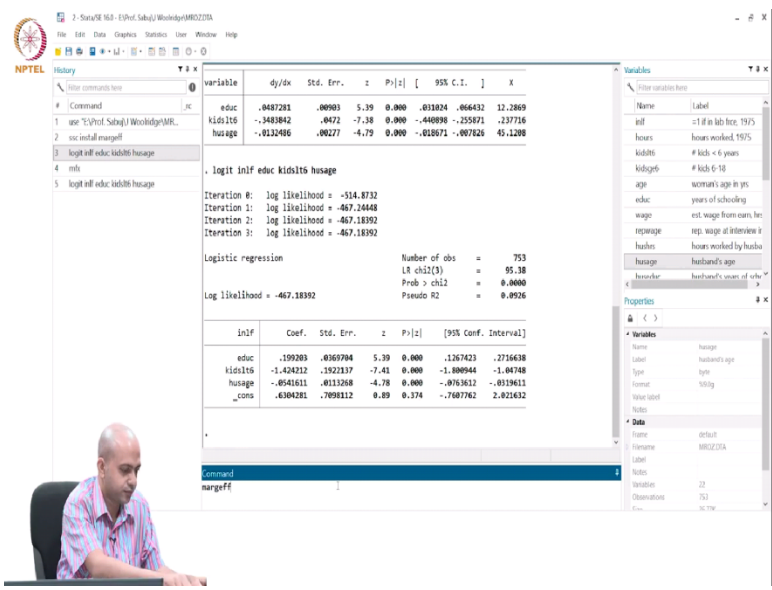
Instead of doing that, I want to get another marginal effect concept. So this is marginal effect concept for actual individual that means marginal effect. So, that means, if you have n number of individuals, you will calculate marginal effect for each and every individual. And then Stata can run another routine for which you can get the marginal effect for each and every individual calculated and then, get the average of that.

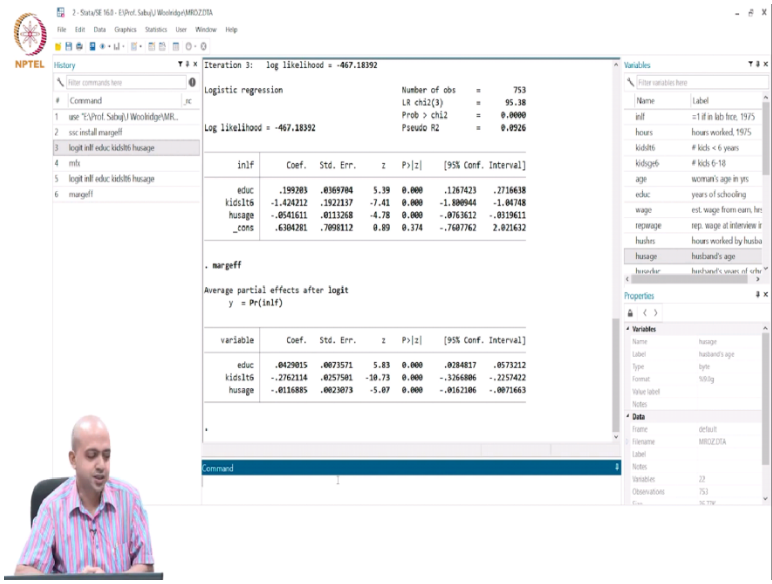
So, when education changes for one unit, what is the change in probability for the first individual? For the second individual? For the third individual? Dot, dot, dot, up to n number of individuals. And then, I can calculate the average of all those marginal effects that is also another marginal effect based on. So, this is called average marginal effect.

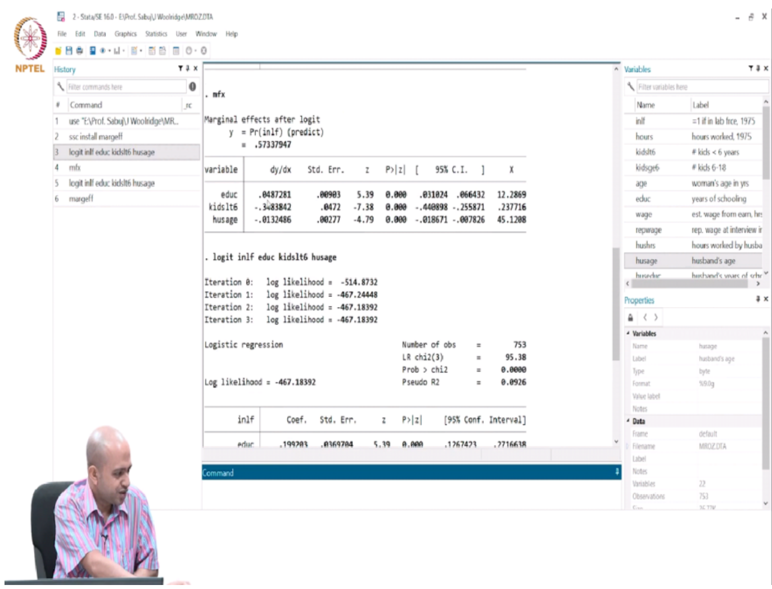
So, that means, there are two marginal effects- marginal effect for the average individual and this is called average marginal effect. So, average marginal effect and earlier what we have calculated that was marginal effect for average individual. So, this we have already calculated by mfx command. But this, average marginal effect calculation requires a separate command which is called margeff.

(Refer Slide Time: 17:49)









So after estimating the model, I will once again estimate the model, so that means, this is the model Logit and then I will put margeff. This is called average marginal effect. This is slightly lower, see the marginal effect earlier model, this is 0.0429 and earlier it was 0.0487. So, little lower. 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 what in which way these two marginal effects are actually different.

So, whenever you are estimating your model, this type of qualitative response model please remember to report the marginal effects and both types of marginal effects and you should also interpret. In standard model, since the beta hat coefficients are direct measure of marginal effect.

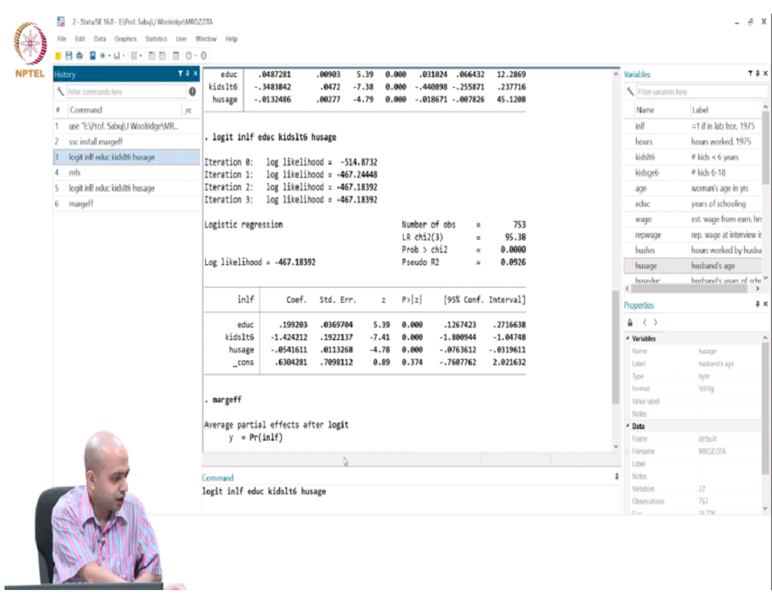
We do not need to mention anything else. But in Logit model, in qualitative response model, since beta coefficients are not a direct measure of marginal effect, we need to specifically mention or report the marginal effect otherwise your research would be incomplete.

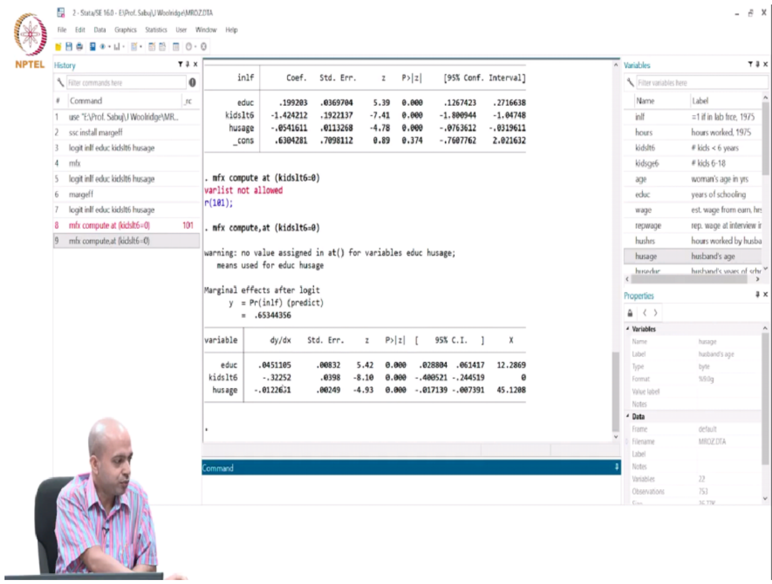
Many times after estimating this we forget to report the marginal effects. But then your research would be incomplete because at the end of the day, what we are interested in specifically for the marginal effect, for the unit change in any of this explanatory variable, on an average what is the change in the probability of labor force participation that is what we want and that is why you need to specifically report these 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 Pi hat. So, that means, I am saying when number of kids changes by one unit, that means from that sample average to one unit change, probability of change for that woman who is having that many number of kids.

But suppose my interest is a little different, I want to calculate the change in probability of labor force participation for the married woman who does not have any kids and then I want to see from 0 to 1 kid, what is the change in probability of labor force participation. So that means, that also you can calculate.

(Refer Slide Time: 21:15)





Suppose you estimate this Logit model and then, mfx compute, we have to put a comma and then at kidslt6 equals to 0. Now look at what is happening. So here, Stata has calculated the marginal change assuming kids equals to 0. So when the kids when the married woman does not have any kids at that situation, when education changes by one unit what is the change in probability?

When husband's wage changes by one unit what is the change in probability? And when kids changes by one unit, this is very important, now see, from 0 key to one kid, what is the change in probability? 0.32. what is the earlier one if you compare, look at this in the earlier model, 0.3483 and here it is 0.32 it is a little different. So, from this model, once again, you just see, this is your original model. This is the margeff command and this is key. So, slight change in probability when you specify x equals to 0. When x equals to 0, this is 0.3483.

I forgot to mention the experience let me put experience also. Logit in labor force, and then I will put experience also. This is the model and then you calculate mfx. Then this is 0.33. And now, if you put mfx compute at in the bracket kidslt6 equals to 0 you have to end the bracket and then enter. So, that means when you have 0 kid, then as you get the first kid, your probability of labor 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 keeps on increasing. So, it is quite difficult to manage more kids than when you have the first one. So, that means you can specify a specific level of value for the x and calculate the marginal effect by this command mfx compute at kids equals to 0. And similarly, you can also put education level equals to 0 or a specific value of education husband's wage so on and so forth, based on your research interest.

All right, these are all based on your marginal effect, but when you estimate any econometric model like previous cases, you need to know how good is your model in terms of goodness of it that means R square.