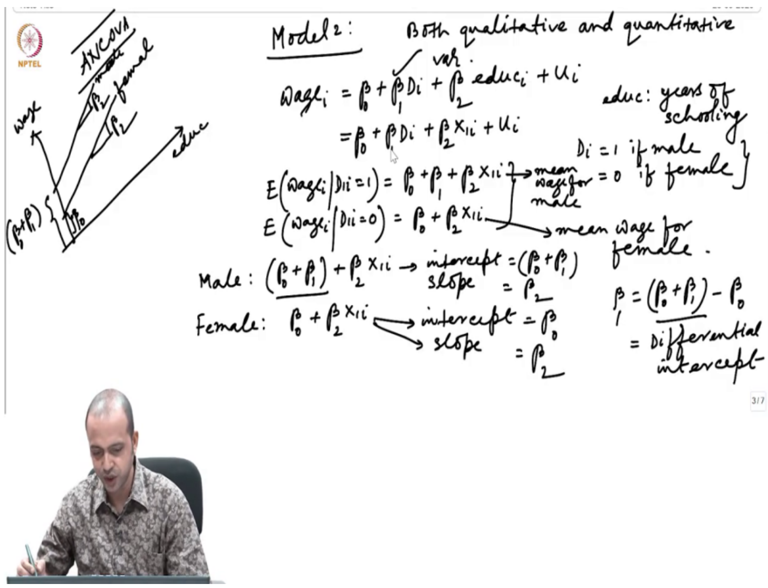
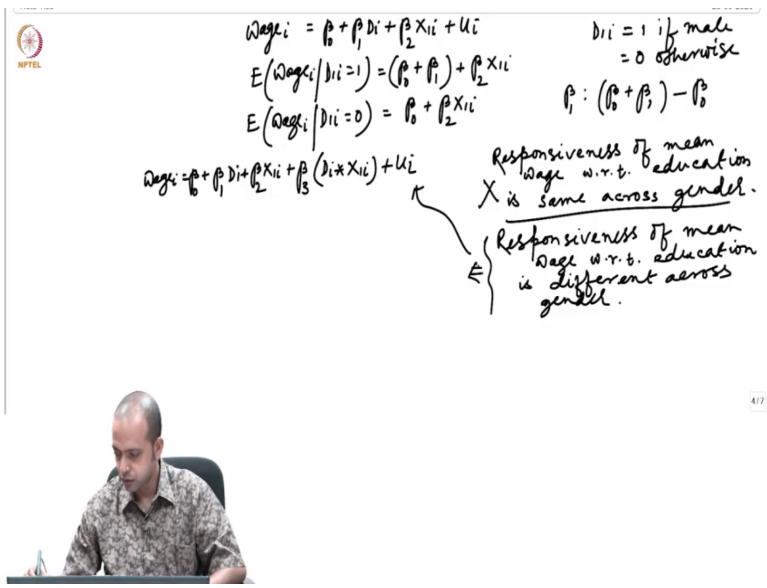
Introduction to Economics
Professor Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology Madras
Dummy Variable Analysis and Application of Difference- in Difference for Impact Evaluation Part-2

(Refer Slide Time: 00:16)



But in this model, when I have included one covariate, what I am assuming, let me tell you something more about this model.

(Refer Slide Time: 00:27)



So, your wage function is wage equals to beta 0 plus beta1Di plus beta2 x1i plus Ui. This is your model, and then you have derived the interpretation in this way, expectation of wage i given D1i equals to 1. So, D1i equals to 1 if male and 0 otherwise. so equals to beta 0 plus beta1 plus beta2 xi and expectation of wage i given D1i equals to 0 becomes beta 0 plus beta2 x1 i.

And we say that beta 1 equals to beta 1. How do you derive the interpretation? Beta 0 plus beta 1 minus beta 0. So, that means the differential intercept. Now, if I ask you, for a unit change in education, on an average, what would be the change in males, which that is beta 2. Similarly, if I asked for the unit change in education on an average, what would be the change in females, that is beta 2.

So, in both the cases the incremental change in education is producing same impact on wage. So, that means, if male person goes for one extra year of schooling and if the female laborer goes for one extra years of schooling, then both of them will generate same increment in their wage. Let us say that 100 rupees or 150 rupees or 1000 rupees. So, that means, what I am assuming here that, in this model, that responsiveness of mean wage with respect to education is same across, gender.

So, this is the assumption that we are making in this model, we have to very clearly keep this in mind both male and female are getting same increment when they are going for extra years of education, but suppose, I am assuming a different story. When male goes for one extra education, let us say the increment is 100.

And when female goes for one extra year of education, let us say that the increment is 75. That means, mean wage responds differently with respect to education across gender. Then how will you modify this equation? So, that means when the dummy variable model, we have one qualitative variable and one quantitative variable, but there is no interaction between the qualitative and the quantitative variable, this is the assumption we make implicitly.

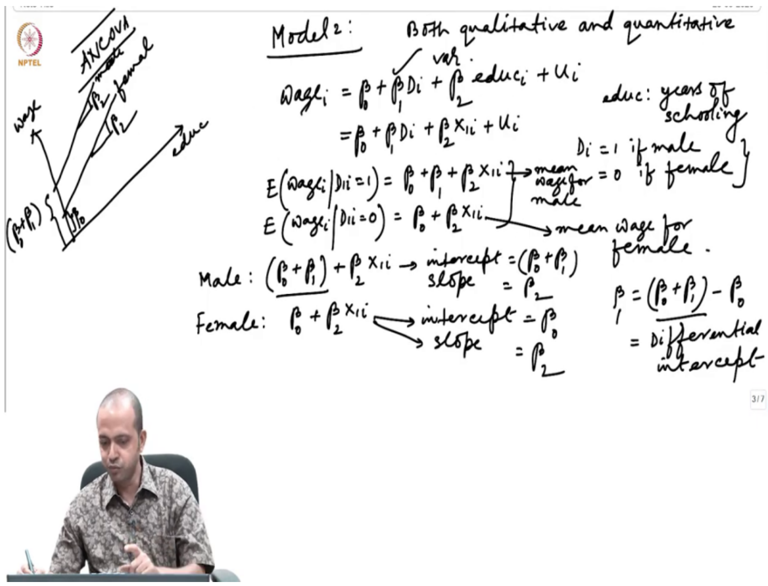
That means the incremental change in your wage, due to one extra years of education is actually same across gender, male and female will get same increment if they go for one extra years of education. Now, suppose I am relaxing this assumption. And I am saying that responsiveness of mean wage with respect to education is different across gender.

Let us say that this assumption I am relaxing, and I am saying that responsiveness of mean wage with respect to education is different across gender. This is the assumption I am making. And what should be the corresponding economic model then with this assumption? Simply you have to interact with this dummy variable with that qualitative information.

So, your model will look like wage equals to beta0 plus beta1 D1i or Di plus beta2 x1i plus beta 3 Di and x1i should interact plus Ui. So, this is going to be your model when you relax this assumption and see that mean wage responds quite differently with respect to education for male and female categories. How this is so?

We have already defined, derived the interpretation of beta 1 and beta 2. Now, additionally, we have to derive the interpretation of beta 3, how will you do that? So, I will do the interpretation in the next page.

(Refer Slide Time: 08:09)



So, your model is now wage equals to beta 0 plus beta1 Di plus beta2 x1i plus beta3 Di interacting with x1i plus Ui and Di equals to 1, if male and 0 if female, then you derive the interpretation wage. So, that means what I will do, I will do expectation of wage given D1 Di equals to 1, this would become beta0 plus beta1 plus beta2 x1i plus beta 3 x1i plus sorry, this Ui is not there since I have taken expectation.

And this again you can write as beta 0 plus beta 1 plus beta 2 plus beta 3 x1i. So, this is mean wage for male category. Now, from this equation you can clearly understand one thing that the intercept of the mean wage function for the male category is beta 0 plus beta 1 and what is the slope? Beta 2 plus beta 3. This is intercept and this is slope for the mean wage function for the male category.

Now, if you derive expectation of wage given Di equals to 0 what you will get? Beta 0 plus beta 2 x1i and this would become 0, this also becomes 0. So, this is beta 0 plus beta 2 xi this is mean wage for female category. And what is the intercept? This is the intercept and this is the slope. So, that means beta 0 is the intercept and beta 2 is slope.

Now, you can easily understand one thing that how do you derive beta 3 then? Beta 3 is basically beta 2 plus beta 3 minus beta 2, as I have told already, that interpretation of the coefficients in dummy variable model is always derived. So, the moment you think about what is going to be the interpretation of beta 3, you have to really derive beta 3 and how I am deriving?

It is very simple, beta 3 equals 2 beta 2 plus beta 3 minus beta 2. What is beta 2 plus beta 3? That is the slope of the mean wage function for the male category and beta 2 is actually the slope of the mean wage function for the female category that means, if you take the difference of beta 2 plus beta 3 minus beta 2 it tells you the differentials slope.

So, that means beta 3 basically indicates the differential slope that means, when education increases by one unit increment for the male category is beta 3 unit higher than that of their female counterpart. For the unit change in education, average wage for the male category increases by beta 2 plus beta 3 unit, while for the unit change in education, the mean wage for the female category increases by only beta 2 unit.

That means, there is a change by which mean wage responds towards education for the male and the female category. That is why beta 3 has now become a differential slope coefficient in this model, is this clear? how do you derive beta 1 now?

Beta 1 it is again same see, beta 1 is nothing but beta 0 plus beta 1 minus beta 0. So, that means this indicates differential intercept. So, this interpretation we have already derived. So, the interaction, the term, the coefficient which is attached with the interaction between the qualitative and quantitative variable actually indicates differential slope.

So, when the dummy variable interacts with the covariates, the corresponding coefficient indicates differential slope, while when the dummy variable enters into the model additively, without any interaction with the covariate, then we say that the coefficient attached to that dummy indicates the differential intercept.

So, that means, male categories wage function will have intercept which is beta 1 amount higher than the female counterpart, and slope would be higher than beta 3 unit. So, if you now try to draw the diagram once again between male and female in the x axis you have education and on the y axis you have wage.

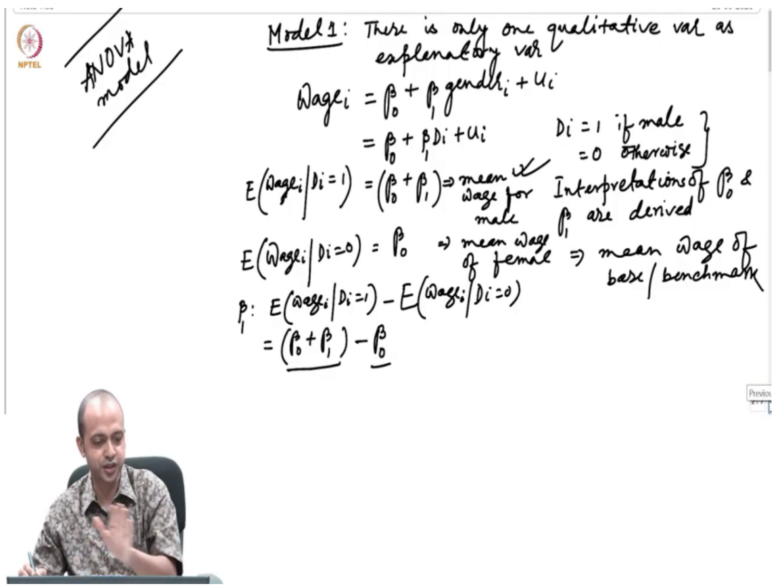
So, this would become, let us say this is my wage function for the male category, where the slope is, what is the slope? beta 2 plus beta 3 and intercept is actually beta 0 plus beta 1, and female category slope would be little. So, female category is slope is this. So, it is only beta 2 and intercept is actually beta 0. So, a female category is this, which shows lower amount of intercept and lower slope as well.

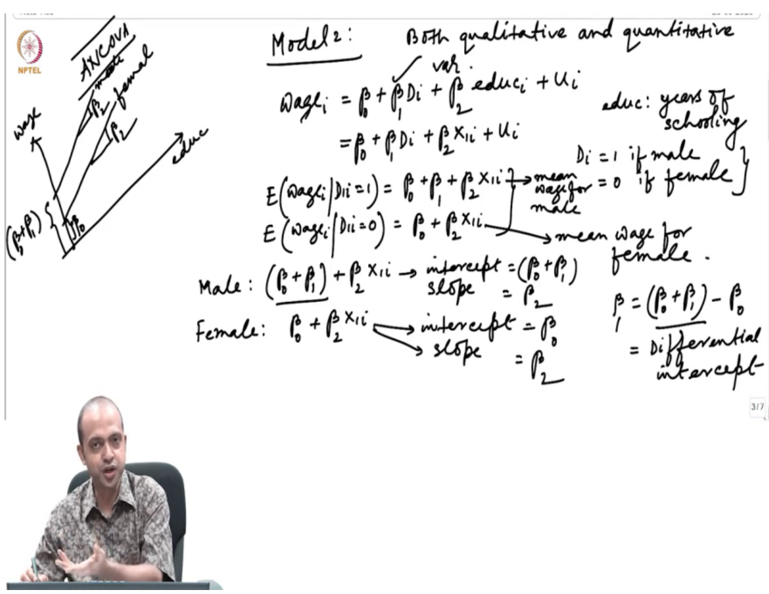
So, that means for every additional unit of education, female laborers receive less increment compared to their male counterpart. One thing you have to keep in mind that the both of these male and female workers they will get increment, it is not that for extra education, females are not getting increment, both of them are getting increment.

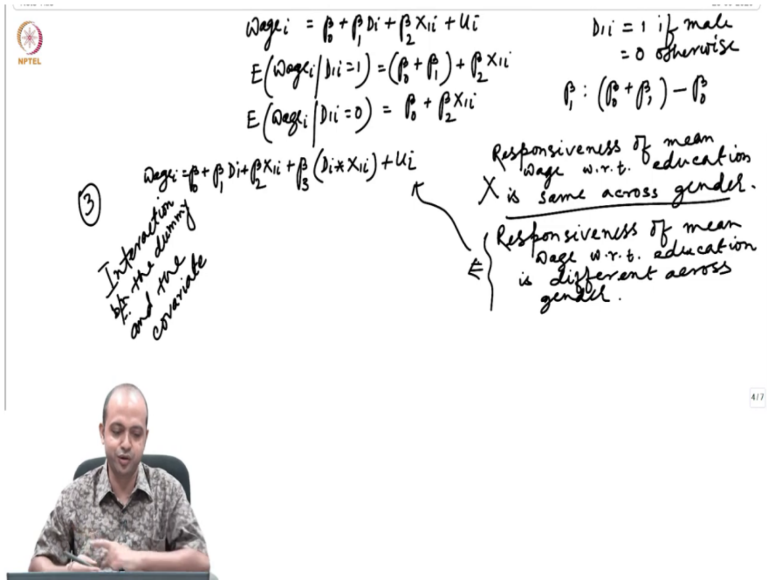
But what I am saying here male labor force get more increment compared to their female counterpart, for every additional unit of extra education. If male laborers, they get 100 rupees increment, probably the female counterparts would be getting less than 100, 80, 90, 75 whatever might be the amount, that is what I am saying. Increment is there for both, but the amount of increment is more for the male category than their female counterpart.

That is why I say that in this model, we assume responsiveness of mean wage is different, with respect to education is different across the two categories male and female. So, this is the model. So, that means how many models we have discussed today. So, we have discussed only two models, model two, where we have both qualitative and quantitative model information. No, actually we discussed three models.

(Refer Slide Time: 19:34)







Then in model one, we said there is only one qualitative variable gender, there is no covariate that is called ANOVA model. And then in model two we have both qualitative and quantitative, that means while Di indicates the gender, then education indicates a quantitative variable which is covariant, that is model two.

And Model three is this. So, that means, this is our model 3. So, model three, it basically makes interaction between the dummy and the quantitative variable or the covariate, that is model three. So, three models we have discussed. This is how we have to design your variable.

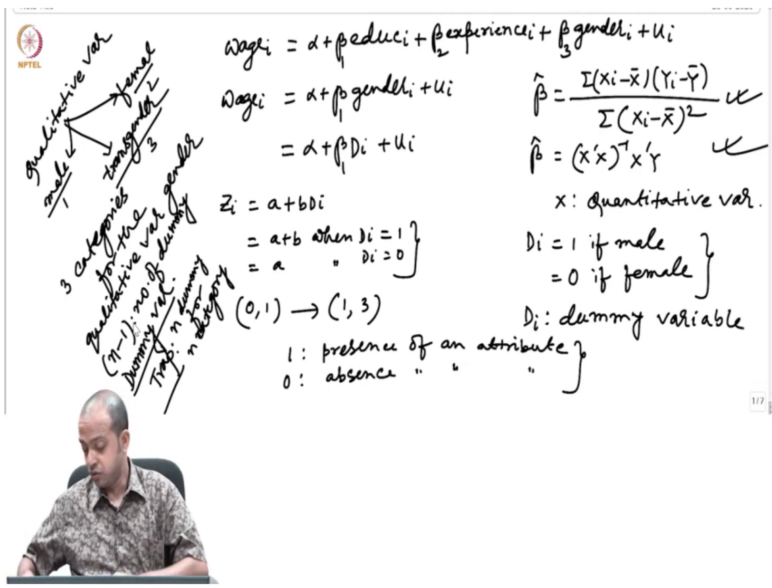
So, that means, before estimation itself, you have to formulate your dummy variable, which is very interesting and very very important for empirical application. But only thing is that unlike the standard econometric model, you have to first formulate the variable which is little involved, I will not say difficult, little involved than the standard econometric model.

And after formulation of the model, you have to first derive their interpretation, what is the meaning of the intercept, what is the meaning of the slope, what is the meaning of the coefficient what you are getting by interacting the dummy with the quantitative variable, you will just derive the interpretation then you estimate the model.

So, that after estimation you will know what exactly you have estimated and as I said the interpretation of the coefficients are all derived, there is no standard interpretation. So, it changes depending on what type of model you are estimating. We have also said that, if you have n number of categories for a qualitative variable, the number of dummy variables should be n minus 1.

That means, in our case, we have assumed two categories- male and female for the qualitative variable gender. that is why we have introduced only one dummy Di, 2 minus 1 equals to 1. If you introduce two dummies for two categories, then you will have some problem that is called dummy variable trap.

(Refer Slide Time: 22:42)



So, here, see, I have two categories male and female. That is why I have introduced one dummy. So, that means number of dummies should be n minus 1 where n is actually number of dummys. And if you introduce n dummy for n category, then you will have a problem situation which is called dummy variable trap. What is the implication of dummy variable trap?

That we will discuss in a later part of our discussion. This is called dummy variable trap. What does that mean? n dummy for n category. Where the rule says actually the number of damage should be n minus 1. With this, we are closing our discussion today. In our next class what we will do? We will have two qualitative variables instead of one. And we will see what would be the interpretation of the coefficients.

And we will see what would be the interpretation of the coefficient when both the dummies they interact with each other. Here we saw the interaction between dummy and the quantitative variable. But there we will see the interpretation of the coefficient when the two dummies they themselves interact with each other. With this, we are closing our discussion today.