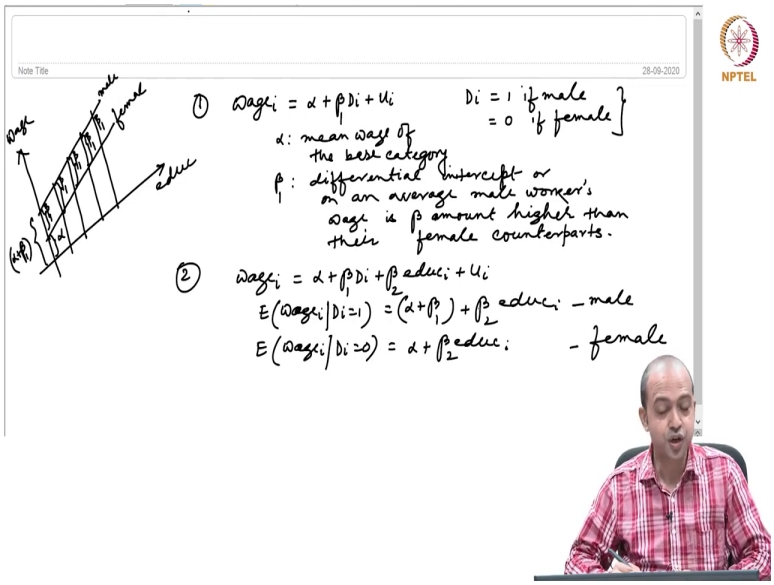
Introduction to Economics
Professor Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology Madras
Dummy Variable Analysis and Application of Difference- in Difference for Impact Evaluation Part-3

(Refer Slide Time: 00:16)



So, we were discussing dummy variable model in our last class. And we have discussed two types of, actually three types of dummy variable models, ANOVA model wherein we have only dummy variable or qualitative variable as explanatory variable. Then in the second case we have also introduced a quantitative variable as well in the right hand side and we call that model as ANCOVA model.

Then in that ANCOVA model itself we have got another case which was the third case of our discussion where we were interacting the dummy variable with the quantitative variable. So, we will quickly recap all those models once again, before we start about our new models of dummy variable in today's discussion.

So, the first case was this, this is the first case what we were discussing, this is wage of the ith individual was alpha plus beta 1 Di plus Ui and this model, we said that the ANOVA model, since we have only dummy variable model in the right hand side. This model is not, it is very, it is a rare case wherein we have only qualitative variable in the right hand side, rather we have

qualitative and a mix of qualitative and quantitative variables. Anyway, for simplicity's sake we have assumed this is our first case.

And how was our Di defined? Di was equals to 1 if male and 0 if female, these are the two cases, two categories of the gender variable that we were discussing. And then we say that when you do not have any other quantitative variable in the right hand side, the intercept, what was the interpretation of the intercept alpha?

Alpha basically indicates the mean salary or mean wage of the base category and what is the base category here? Here base category is female since we have not introduced an entirely for the female category. And then what does beta 1 indicate? Beta 1 basically indicate the differential intercept.

So, that means if alpha denotes the intercept of the base category and beta 1 indicates differential intercept, so you can easily understand that beta 1 indicates on an average male's salary or wage is beta 1 times, beta 1 amount higher than the female, very easy. So, alpha is the intercept for the female category or the base category and alpha plus beta 1 is the intercept for the male category that means beta 1 indicates the differential intercept.

So, that means we can say that the interpretation of beta 1 is the differential intercept or alternatively you can say that on an average male workers wage is beta 1 amount higher than their female counterpart. This is the interpretation. And what was the second case we were discussing? Second case we introduced one quantitative variable also, and that was education.

So, wage equals to alpha plus beta 1 Di plus beta 2 let us say education i plus Ui. Now, when we have education variable, that means one quantitative variable involved in the model, the interpretation of alpha is not basically the mean salary of the base category, because you have education also involved. So, in that case, what would be the interpretation? You can say that alpha is basically the intercept.

And what is beta 1 in this case? And in this case, so, beta 1 would be again the differential intercept. So, that means, in this model quickly if you take expectation, expectation given Di equals to 1, so this would become alpha plus beta 1, alpha plus beta 1 plus beta 2 education,

education and if you take expectation of wage i given Di equals to 0, so, you will get alpha plus beta 2 education.

So, that means, here for the female category intercept is alpha, for the male category it is alpha plus beta 1, this is for male and this is for female. So, for female the intercept is alpha and for male it is alpha plus beta 1. So, you can quickly say that beta 1 indicates the differential intercept. And how will you interpret in terms of a, what would be the practical interpretation?

You can say that for every level of education that means, male workers, they earned beta 1 amount higher, higher wage than their female counterpart. So, this particular, the second case if you represent in terms of a diagram, here it is wage and here it is education, so let us say this is male and this is female. So, males intercept is how much? Alpha plus beta 1 and this is simply alpha which is for the female.

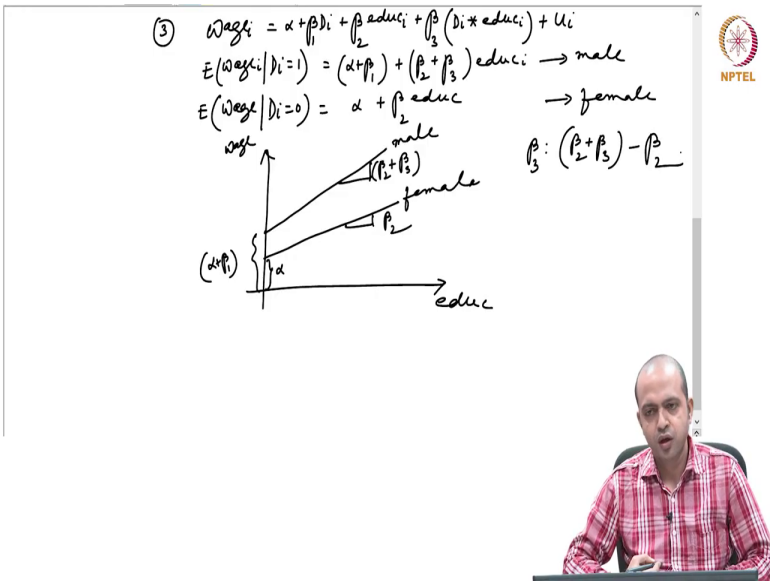
So, beta 1 is basically the differential intercept. So, that means, you can easily understand for, so that means, these differences basically the beta 1. So, for every level of education male workers they are beta 1 amount higher wage than their male counterpart. And the slope is same. So, that means these are the two parallel lines.

That is why the difference at every point it is only beta 1. For each and every level of education, the male workers they are beta 1 amount higher wage than their female counterpart. That is why we say that beta 1 is the differential intercept and the practical interpretation is male workers on an average on beta 1 amount higher wage than their female counterpart.

This was our second case. And in the third case, in this case, what is the implicit assumption we said if you recall, since the slope is same that means responsiveness of wage, with respect to education is same for the male and female workers, That means, if the male worker goes for 1 extra years of education, then the increment what the worker will experience in their wage is same as if the female workers also go for 1 extra year of education.

So, that means both male and female workers, they will enjoy beta 2 amount of increment in their wage, if they go for 1 extra years of education. Since both male and female workers wage function have the same slope as beta 2 which is the slope that is the assumption. If we relax this assumption, then we will get our third case.

(Refer Slide Time: 11:09)



So, our third case was, wage i equals to alpha plus beta 1 Di plus beta 2 education and we have 1/3 variable also where we are interacting Di with education plus Ui. So, in this case what we said, that if you quickly take again expectation of wage, expectation of wage given Di equals to 1, what will happen? This would become alpha plus beta 1 plus beta 2 plus beta 3 education.

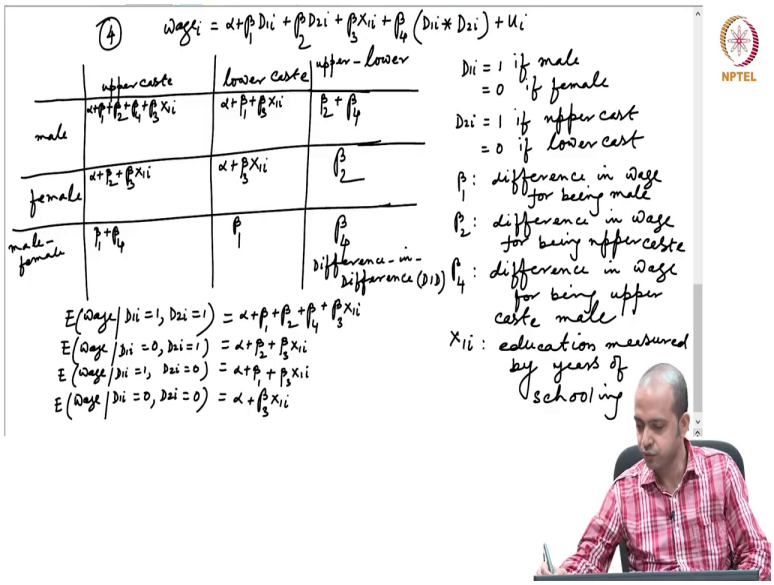
And if you take expectation of wage given Di equals to 0, what you will get? You will get alpha plus beta 2 education. And that means this is male and this is for female. So, that means from this what you can say that beta 1 still indicates differential intercept. But, beta 3 indicates a differential slope as well.

So, that means for one extra year of schooling or one extra year of education while the male workers enjoy beta 2 plus beta 3 amount of increment, female workers enjoy only beta 2 amount, because the slope of the female workers is only beta 2. So, that means in terms of a diagram, if you represent this is education and this is wage.

So, let us say this is the male categories wage function, where the slope is alpha plus beta 1 and female workers wage is alpha and slope is beta 2, assuming alpha is positive beta 2, beta 3 all the coefficients are positive. So, what will happen, this would become like this, this is female. So, this intercept is only alpha and their slope for the male is beta 2 plus beta 3 while the slope for the female is only beta 2.

So, from this diagram, we can easily understand that for extra years of education, male workers enjoy beta 3 amount of extra increment. While both male and female will enjoy increment, male workers increment is beta 3 amount higher than the female counterpart, because how do you derive beta 3? As I said the interpretation is derived, it is basically beta 2 plus beta 3 minus beta 2. So, these are the 3 cases we have, we have discussed yesterday in our previous class.

(Refer Slide Time: 15:37)



Now, what we will do, we will discuss about, discuss about the fourth case. We will introduce one more qualitative variable, apart from the gender, let us say that is castee. So, that means our wage function is wage equals to alpha plus beta 1, let us say this is D1i plus beta2 D2i plus beta 3 education, education I am denoting by x1i plus beta 4 I am interacting these two, interacting these two dummies, D1i multiplied by D2i plus Ui.

And once again, I assume that increment is same, that is why I am not interacting the x with D1 or D2 there is no interaction between D1 and D 2 with the quantitative variable that is again for the simplicity's sake. How we have defined D1i and D2i? D1i as usual equals to 1 if male and 0 if female and D2i equals to 1 if upper castee and 0 if lower castee.

We assume that there are only two castees available, which is upper castee and lower castee and two categories for the gender also male and female. That means here why we have introduced this type of model? we have introduced this type of model because our assumption is that, the

gender discrimination for the upper caste is significantly different from the gender discrimination with the lower caste.

And what is the rational for this? Maybe we are hypothesizing that the male, female difference in their wage is a little lower for the upper caste, because in upper caste, maybe female workers because of their empowerment or higher bargaining power, they could reduce the gap between the male and female labor force.

While for the lower caste, female in the lower caste, because of their lower empowerment or lower level of bargaining power, they are not able to do so. That is why we assume that male, female gap, the gender gap for the upper caste is significantly lower than the gender gap that prevails in the lower caste.

That is our hypothesis may or may not be true, may or may not be true, that we have to, we have to estimate using a data and look at their significance side. So, now what we will do to derive the interpretation of beta 4, so that means beta 1, what we can say that beta 1 indicates difference in wage for being male. And what is beta 2?

Beta 2 indicates difference in wage, in wage for being upper castee. Then what is beta 3? Beta 3 indicates difference in wage, difference in wage for being upper castee male, upper castee male. And we are going to sorry, this is not beta 3, this is beta 4, this is beta 4, and beta 3 is basically the responsiveness of wage with respect to education because x1i is education and education is measured by years of schooling.

Now, what we will do? We will nicely derive the interpretation of beta 4 and this is very very important case. So, please concentrate and think about this case carefully. So, what I will do? I will put here let us say upper castee. This is lower castee, then here it is male and here it is female. So, that means I will get several cases.

So, what we will do? For upper castee and male, what would be the mean salary? that means, you have to take expectation of wage i given D1i equals to 1. Because, male is defined as that, an upper castee for getting upper castee you have to get D2i also equals to 1 and if you do so, the 0 you will get in this column, in this sale for upper castee male you will get. So, that means I will get several cases.

So, you have to put D1i equals D2i equals to 1. So, you will get alpha plus beta 1 plus beta 2. And here also you are putting 1 and 1. So, plus beta 4 plus beta 3 x i or x 1 i. This is the average salary or wage for the upper castee male. So, the same thing I will write here, alpha plus beta 1 plus beta 2 plus beta 4 plus beta 3 x1 i.

Similarly, for the female category what will happen? Expectation of wage given D1 i equals to 0 but D2i is still 1 because I am considering upper caste female, what will happen? So, this would become alpha, then beta 1 will vanish, beta 2 would be there, beta 2 would be there plus beta 3x i, x 1 i, and here also since I am putting D 1 i equals to 0, the beta 4 will also vanish. So, this will become alpha plus alpha plus beta 2, beta 2 plus beta 3, beta 3 x 1 i.

Then for the male and lower castee, what will happen? This category, this particular case expectation of wage, given D1i equals to 1, because it is still male, but D2i equals to 0 because I am considering lower castee. So, that means, this would become alpha plus beta1 would be there, but beta 2 will vanish. So, beta 1 would be there, because D1 i equals to 1 alpha plus beta 1, D2 i is 0 means beta 2 will vanish, plus beta 3 x i x 1 i and this will also vanish.

If I put D2 i equals to 0. So, that means this column for male lower caste this will become alpha plus beta 1 plus beta 3 x1 i. And for female and lower caste that means expectation of wage, when do you want i equals to 0 and D2 i is also 0 that would become alpha plus beta 1 will vanish, beta 2 will vanish then I will get alpha plus beta 3 x1 i and beta 4 will also vanish because I will put D1 I, both equals to 0.

So, this will become alpha plus beta 3 x1 I, x1 i. Now, when I am putting D1 i equals to 0, D 2 i also equals to 0 that means, please keep one thing in mind clearly, when you have two dummies, two qualitative variable, what is your base category or benchmark category? To identify the base category or benchmark category, you have to assign 0 for both the dummies.

So, that means when I am putting D1 i equals to 0, D2 i is also equals to 0 that is now becoming my base category. So, what is my base category then? Female lower caste. So, lower caste female is my base category. So, now what I will do, since my objective is to get the gender difference, gender difference in upper and lower caste.

So, this I will get as male minus female, male minus female for the upper class. So, from these minus, this if you do what will happen? Alpha, alpha will cancel beta, beta will cancel, beta 2 and beta 2 will get cancelled out. So, I will get this minus this equals to alpha plus beta 2 will vanish alpha plus beta 2. So, you will have beta 1 plus beta 4, beta 4 plus beta 3 sorry, beta 3 will also get canceled. This minus this. So, I will get alpha plus beta 4.

For lower castee, if I take male, female gender difference, then alpha will get cancelled and beta 3 x1 i will get cancelled, I will get beta 1, beta 1. Now, if I take the gender difference of upper caste and gender difference of lower caste, then what I will get? This minus this equals to beta four. Please try to understand beta 1 plus beta 4 indicates the gender difference because this is male minus female for the upper caste.

Beta 1 indicates gender difference for the lower caste, if I take these differences in difference, that would become my beta 4. And what is the interpretation of this? The interpretation is beta 4 indicates what is the difference in gender gap between upper caste and lower caste? Both (())(30:36) for both upper caste and lower caste, we assume there is a difference in gap between male and female, but we are trying to hypothesize whether the gap, gender gap is significantly different for the upper caste compared to their lower caste counterpart.

You may get another input for interpretation of beta 4. So, that means, here basically beta 4 is difference in difference, because we already got one difference here, this is male minus female, this is also male minus female, again I am taking difference. So, that is why this is called, this is called difference in difference. Difference in difference estimate, which is a very very important, interesting and powerful tool in econometrics, you have to carefully understand this difference in difference estimate, or in short DiD.

Similarly, if you want to understand the caste difference between male and female, so we hypothesize, there is a caste difference for male, there is a caste difference for female then we can test whether the caste difference is significantly higher for male, than their female counterpart. So, this is your upper minus lower. So, you, if you take difference for upper and lower for the male what will happen, alpha and alpha will get cancelled, beta will get cancelled, beta 3 will get cancelled. So, you will get beta 2 plus beta 4.

So, this is beta 2 plus beta 4. Similarly, if you take the caste difference for the female category, this minus this alpha will get cancelled, beta 3 will get cancelled, then you will get beta 2. And then if you take again the difference in difference this also, so that means, column wise if you take difference still you will get beta 4. So, that means here beta 4 basically gives you two types of interpretations, what is the two types of interpret, what are the two types of interpretation.

So, the first interpretation of beta 4 is basically if you take row wise difference, then that gives you what is the, what is the row wise interpretation because I have taken male minus female. Which is basically the gender gap. So, gender gap for the upper caste and gender gap for the lower caste and then again I am taking difference. So, that means, I am trying to understand difference in gender gap for the male and female category.

Difference in gender gap, we are hypothesizing that gender gap is for the upper caste is quite different from the gender gap the lower caste. That is one interpretation. And what is the second interpretation I said? That caste difference for the male, caste difference for the female and then you are taking (())(34:04) difference. So, difference in caste. So, caste difference, difference in caste difference. So, caste gap between male and female.

These are the two ways by which you can actually interpret the beta for coefficient and this is known as difference in difference estimate, difference in difference estimate. This is very, very powerful. This is very, very powerful in policy evaluation kind of studies, policy evaluation kind of studies. So, what I will, what I will do? I will, I will give you another example of this case, to show how this difference in difference estimate can actually be used for impact evaluation or impact of policy evaluation. This is called impact evaluation.