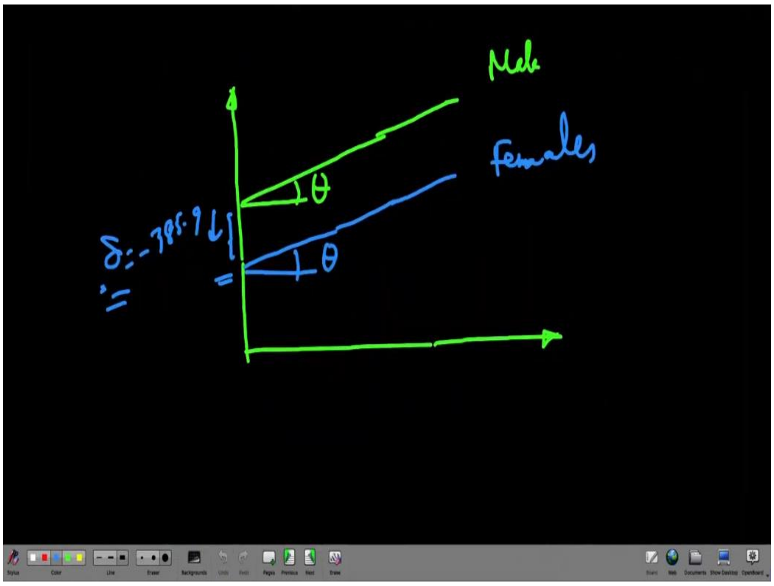
Applied Econometrics
Prof. Tutan Ahmed
Vinod Gupta School of Management
Indian Institute of Technology - Kharagpur

Module - 8
Lecture - 63
Dummy Variable (Contd.)

Hello and welcome back to the lecture on Applied Econometrics. So, we have been talking about dummy variable, and in the previous lecture, we have seen the reduced form equation. And so far, we have been talking about interactive dummy. And in interactive dummy, we have seen how that looks like.

(Refer Slide Time: 00:43)

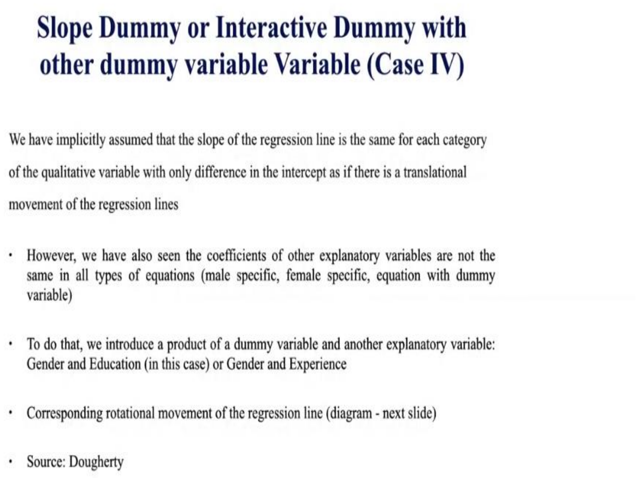


It only impacts on the intercept; nothing to do with the slope. Now, if I want to have the slopes changing, so, how we are going to deal with the dummy variable? That is the topic of this lecture. Now, we have seen that one assumption we have made all throughout is that the slopes for all the different equations which are clubbed together in the reduced form equation; so far, what we have seen, the slopes are constant.

So, the slopes are same for; in the previous equation we have seen; for be it female, be it male, be it a person a part of union or non-union, be the person is white or non-white; for all the cases, the slopes are same. Or in the previous equation we have seen, for males and females if we run regression equation separately using national sample survey data in India; we have seen that.

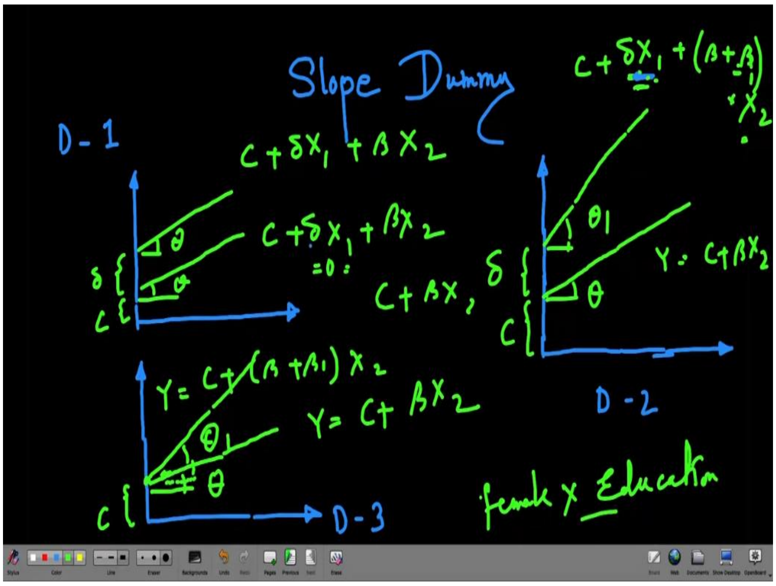
So, we have seen that the coefficient of education or experience, they actually change. So, what we get for the full sample data, the regression coefficients, they are quite different from, when we use the sub-sample of male and female separately. So, that kind of gives us an idea that, well, if we keep the slope constant, perhaps that is somewhat a big assumption we are making, and we need to look into that. And to address that issue, we essentially bring in the concept of slope dummy. And let me actually show you how it looks like if I use a slope dummy.

(Refer Slide Time: 02:09)



So, it is essentially, there are different types of slope dummy we can have.

(Refer Slide Time: 02:21)



So, I will draw different cases here. So, I have in this case, the first case, let us say, I do not have any slope dummy here, only intercept dummy. And if I have intercept dummy, so, this is the same equation like we had previously. It is, the 2 lines are parallel, the angle between them is, the slope is basically same for both the lines, and only difference is the intercept delta.

Now, if I have this equation is equal to, let us say, some constant term plus some delta into X 1 and some beta X 2. So, this is the equation of this line. And here, since this is the reference category, delta is equal to 0. So, what I will get? I will get C plus beta X 2 for this equation. Whereas for the other line, we will get, because of the delta value, we will have this equation C plus delta X 1 plus beta X 2. Now, that is the usual case.

So, there is no change in the slope. Now, when I introduce a slope dummy, things are little different. And let me actually show you how it would work. So, let us say I have drawn the same line here, C plus beta X 2, the same line as the reference dummy category, Y is equal to C plus beta X 2. And I have drawn the other line; but this time, I am actually allowing a change in the slope.

So, here I will have theta and theta 1, let us say. And the change in the slope will come from the fact that I have allowed the X 1 to interact with X 2. So, essentially, the equation is going to be, let us say, C plus delta X 1; and there is a coefficient, let us say, beta plus beta 1, another beta, beta 1 into; this is beta 1; into X 2. So, what I have done here? I have basically included this variable X 1 and I have allowed this variable X 1 to interact with X 2.

And because they are interacting, I am actually getting a slope coefficient added, so, which is beta 1. So, earlier I had only this coefficient delta. And now, because the 2 variables are interacting, I am getting another component here; beta 1 X 2, which is actually rotating theta, which is actually adding the slope component here. It can go either way, but this is what is happening. So, because I am allowing the interaction here, so, I get something like beta 1 X 2.

Now, I can have another category. So, let me actually name the diagram. So, let us say this is diagram 1, this is diagram 2 and this is diagram 3. Now, in diagram 2, what we have done? We have actually allowed both the; this is the intercept dummy, just like the previous case. And this value is still going to be the same value which is delta; but in this case, if I want, I can simply avoid the intercept component.

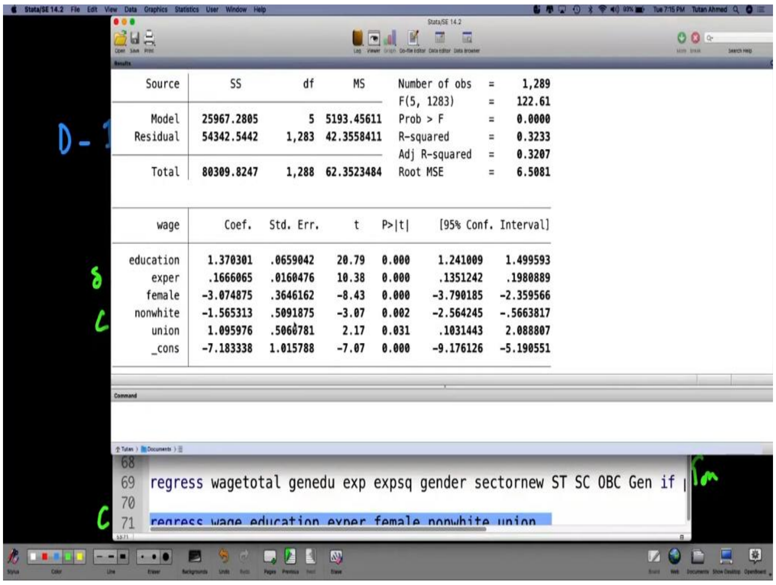
What I can do instead, I can just introduce the slope component here, just the slope component. So, how will it be? So, let me try to draw it. Let us say the same line here of the reference category. This is theta. And my equation is as it is, Y equal to C plus beta X 2. And in the next equation, I do not allow the; so, if I do not allow the delta component to come in, essentially, both the lines actually pass through the same point, because the intercept is basically the C for both the lines.

So, it is the same as in all the cases. The intercept remains C. But what I am allowing here? I am allowing the interaction term to be present. And that is why the slope is becoming theta 1 here. So, this is the slope; this is theta 1, just like before. And the equation is going to be Y equal to C plus beta plus beta 1 into X 2. Essentially, what we have done? We simply omitted the intercept dummy.

So, essentially, these 2 cases, in D 2 and D 3, what we have seen is that, we have allowed the slope to change. We have allowed the interaction to happen. And to give you an example, like it could be for example, in our previous case, we had different dummy variables like female; we had union, non-union; we had white, non-white. Now, if I try to understand, let us say, the impact of female education, so, which should be an interaction between female and education, what would be the impact of female education on the wage?

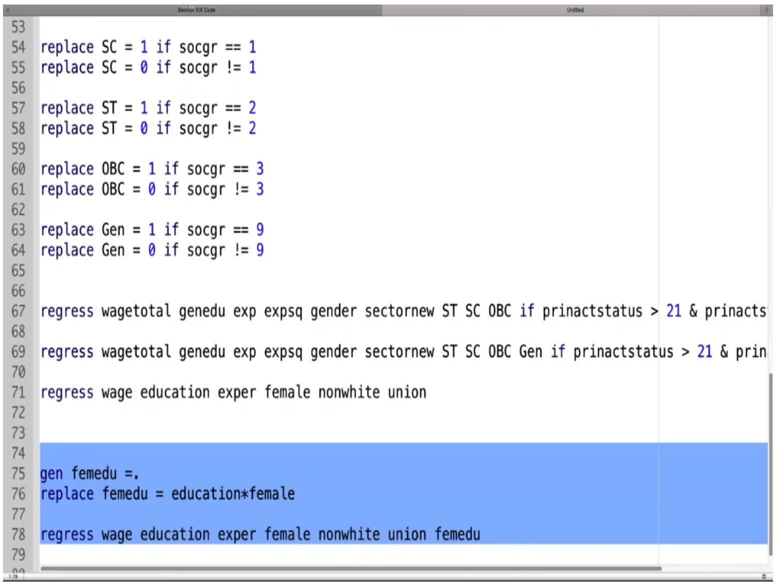
So, in that case, I will allow something like, let us say, my dummy variable female into education. So, this is how we will explain the interaction dummy. And let me actually show you in code, how it actually will look like.

(Refer Slide Time: 09:37)



So, we had this equation here, right? Now, what I am going to do?

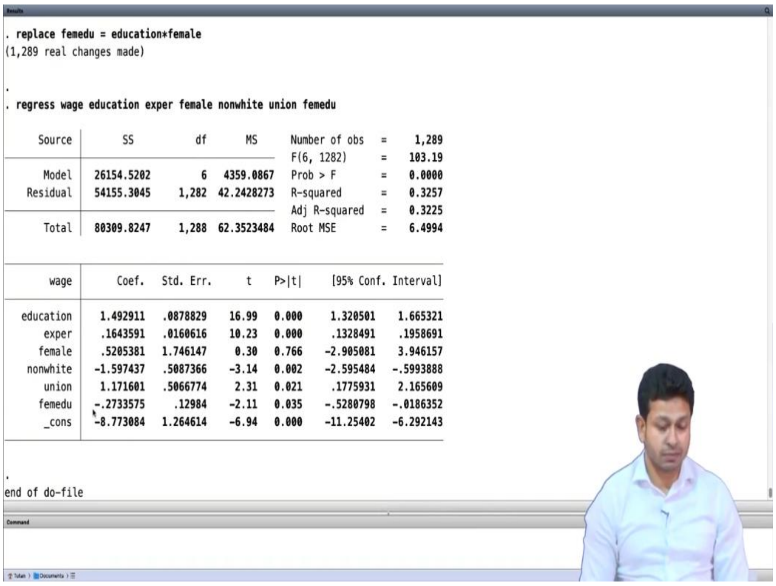
(Refer Slide Time: 09:43)



I am actually going to write another line here. So, this was my code. Now, here I am going to do gen, let us say female education is equal to dot; replace is equal to, let us say, my education variable and my female variable. So, when my female is equal to 1, so, it will give some value; but if my female is equal to 0, it will basically give a 0 value. So, female education could have some value for female education if the female is equal to 1.

So, only in that condition, you will get some value; otherwise, it will be 0. If it is male, it really does not look into that. So, let me actually run it and let me actually add all these variables in my regression. And here we will have, say this new variable, fem edu. So, let me run the whole, all the 3 equations. Let us see how it behaves. I have run it, and let us see the result here.

(Refer Slide Time: 10:56)

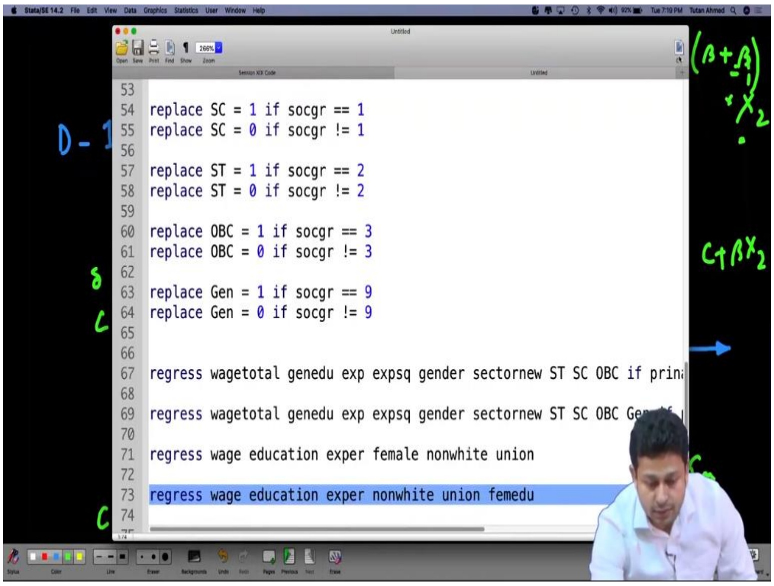


We got some results. So, let us try to interpret the result. So, here you see, we have, when we included this new variable, female education, so, we got this value, certain value for female education. And here, we see that this is a value of minus of 0.27, and which is, P value is 0.035. So, that is, it is less than my alpha equal to 5%. So, it is going to be significant. So, it shows that, if a female is having education or the impact of female education is going to be 0.27 unit less than the corresponding reference category; so, basically, the male education; basically, male is the reference category here.

Now, so that is basically the impact of female education vis-a-vis impact of male education. And this is how we should interpret. Now, what has happened here is that, because of this new coefficient term, this new variable, female education, your slope of the new regression line will change. So, this is how we will understand the intercept dummy. So, we have the contribution for female here.

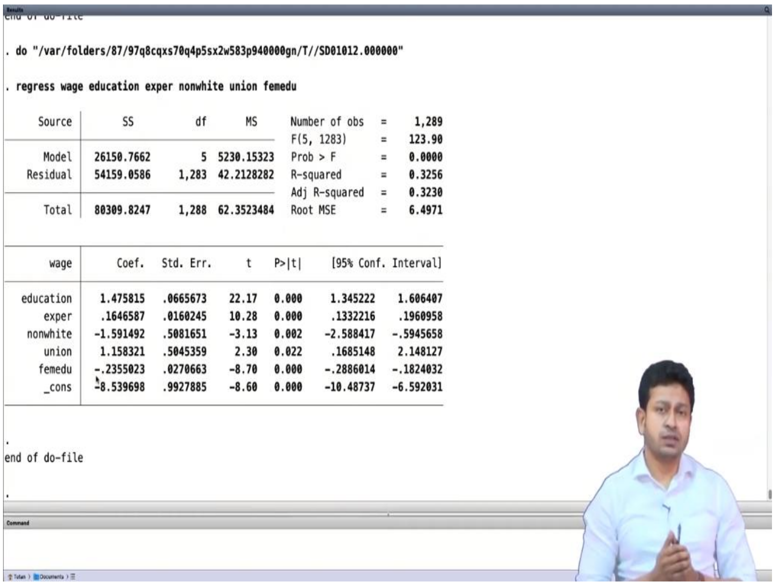
It is as it is, like the intercept term, but we have allowed this rotation to happen, this rotation of the line to happen because of the female education coefficient. So, this is how we will understand the concept of slope dummy. Now, I could have very well omitted, say female altogether.

(Refer Slide Time: 12:50)



I could have simply written this, in which case, what I would have had is, let us say, I am just deleting this female part and I am just including female education. And if I run it, so, what I get is this equation.

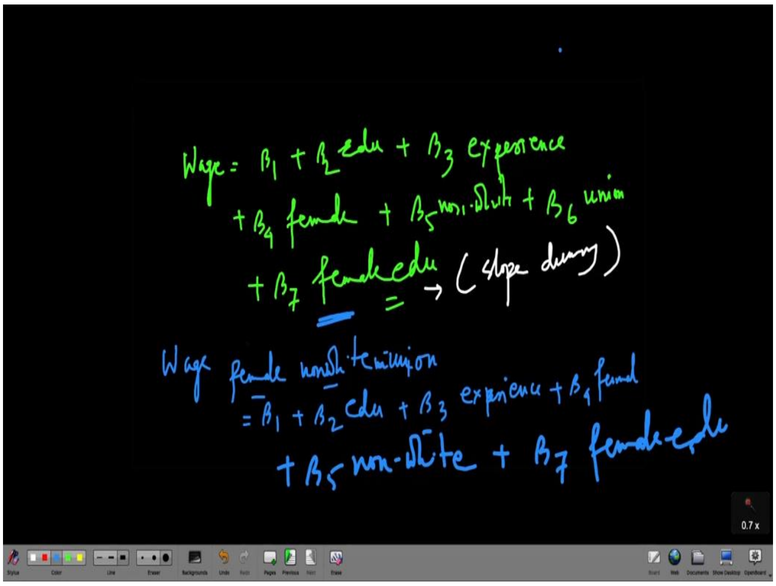
(Refer Slide Time: 13:09)



So, in this case, what I have done is, I have actually removed the intercept dummy, I have only kept the slope dummy. So, this is the case D 3; this is basically same as the D 3; and the previous one is basically same as the D 2. Now, we got the concept of slope dummy. Now, if I want to understand the impact of slope dummy on reduced form equation; so, previously we have seen that; we had all these different variables, different dummy variables and like putting value is equal to 0 or 1, we get the coefficient terms, and we add that with the intercept term, and that is how we get many regression equations for different categories.

And that is only for the intercept dummy. Now, in case of slope dummy, if I have a slope dummy, how do I really interpret my reduced form equation. How do I really unbox the reduced form equation and get many other regression equations? And let us actually see that. So, let me actually go back to my whiteboard and let me create this equation.

(Refer Slide Time: 14:26)



So, here, what my regression equation is; in the previous example, my wage is equal to, let us say there was some constant term, beta 1 plus beta 2 into education. Then I had beta 3 into experience. Then I had say beta 4 into female. Then I had beta 5 into non-white. It was union. And then, I have this another variable that I have added, let us say beta 7 into female education. So, this is the product of female education. This is a slope dummy here; this one is a slope dummy.

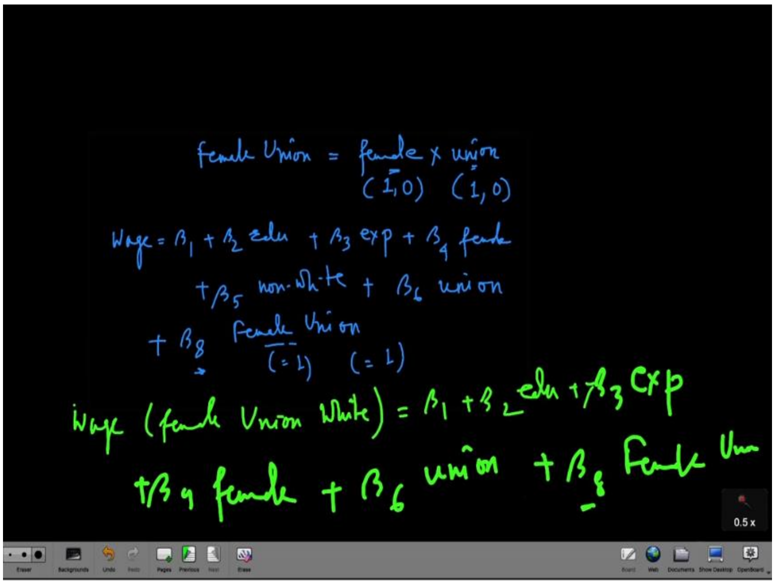
So, let me actually identify, because this is a new element we have included; this is a slope dummy. Now, the point is, when I am talking, when I am actually trying to get the reduced form equation, and when I am trying to get all the different equations from the reduced form equation, so, how do I really interpret the new equation? So, let us say I want to understand for female non-white union. So, how we are going to include the beta 7 here?

So, what will happen is that, I will have a value for; I will have beta 1; then I will have beta 2 into education as it was; then I will have beta 3 into experience as it was; then of course, I will have beta 4 into female; then I have, because I have included female, then I have included non-white, so, I will have; actually I could have written non-union; I can exclude one variable. And I will have one coefficient for non-white; so, this beta 5 non-white.

And then, I have to include; beta 6 is omitted, because it is non-union; so, I have to include beta 7 into female education. This is because of the fact that I am talking about the females. So, I have to include beta 7 here. So, that is going to give me the specific regression equation for female non-white non-union category. Now, if I look at only males, then, beta 7 is going to be omitted.

So, in this case, if I am actually talking about males, then the beta 7 is going to be omitted. So, this is how we will understand the reduced reduced from of the equation here. So, here, we have actually seen the interaction between a dummy variable with a numerical variable. Now, I can also have an interaction between 2 dummy variables, 2 categorical variables. And let us see how that will work.

(Refer Slide Time: 18:14)



So, let us say my interaction term here; let us say I want to include a slope dummy for female who are in union. So, then it will be female cross union. So, that means, only if female; female can take 0 and 1; so, only when it is female, it will be 1. And union can take 0 and 1. So, that will mean, only if the person is a part of union, that is going to be 1. So, essentially, if we multiply, it will have only one value of 1 when both are correct, female and union; in all other cases, it is going to be 0.

Now, if I include this in the regression equation, so, my regression equation, let us say, wage is going to be; let us come back; so, beta 1, plus beta 2 into edu, beta 3 into experience, beta 4 into female; female is going to be there, because we are talking about female; beta 5; note that we are actually keeping beta 5; let me see, it was non-white; beta 5 for non-white; beta 6 into union. So, we are keeping both female and union dummy variable as it is.

So, we have not removed them, but we are including one more dummy, let us say it is beta 8, because we have used one, this beta 7 in the previous equation, we can write whatever you want; beta 8 female union. Now, you will include beta 8 only if I am asking for the, someone is interested to know the impact of both female and union. So, if and only if you are interested in the impact of both female and union, you will include beta 8, but in all other cases, if it is male, beta 8 is going to be 0.

So, you basically do not include anything. Or if it is female non-union, this is going to be 0. Or if it is male union, it is going to be 0. So, all other categories are going to be 0. Only if female equal to 1 and union equal to 1, you are going to include beta 8. So, if I finally want a wage equation for; let me use a different colour; wage for female union white. So, I will have beta 1 plus beta 2 edu, beta 3 experience; let me reduce the size a little bit; then of course, beta 4 female; and because it is white, so, beta 5 is going to be omitted; beta 6, of course is going to be there, because we are interested in union; and then, beta 8 female union.

But had it been male, so, this term would have been 0; I would not have included. Had it been female non-union, I would have excluded beta 8. So, this is how we will understand the reduced form equation when we are including the slope dummy. So, one thing that we have to keep in mind in this case is that; so, here I have only used, say 1 or 2 slope dummies; so, female education, female union and so forth; but if I actually allow all the possible combinations to be present for all the interaction terms, so, then, it is going to be something like what we have seen for subpopulation regression.

So, in that case, the equation for the male population and the female population as we have seen in the national sample survey data, that is going to be same as the same dummy variable equations if we managed to introduce all the different slope dummies. So, this is basically the idea of slope dummy. And we are going to see that, actually, how to introduce slope dummy and how many slope dummies to be introduce and how it is actually going to become like our normal regression of subpopulation regression.

That we will see in the next lecture where we are going to talk about the test, different kinds of tests and to understand, basically the best model when you use a dummy variable. So, with this, we end this lecture here. Thank you.