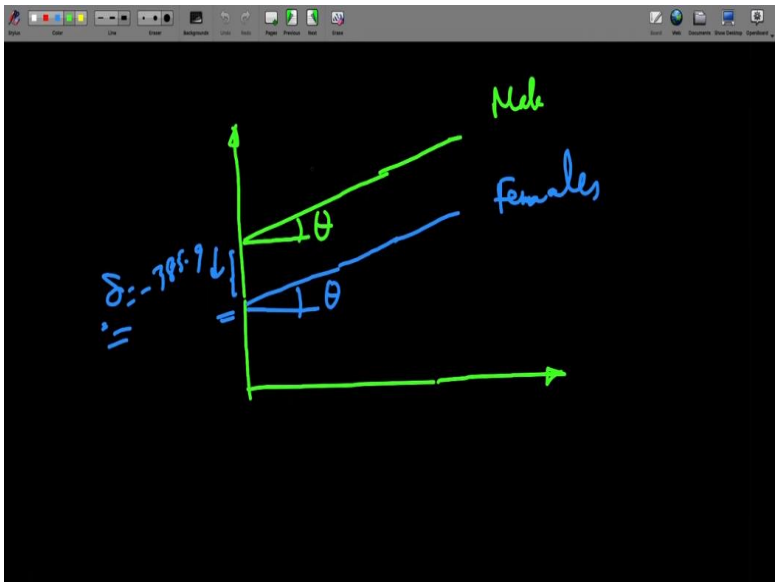
Applied Econometrics
Prof. Tutan Ahmed
Vinod Gupta School of Management
Indian Institute of Technology - Kharagpur

Module - 8
Lecture - 62
Dummy Variable (Contd.)

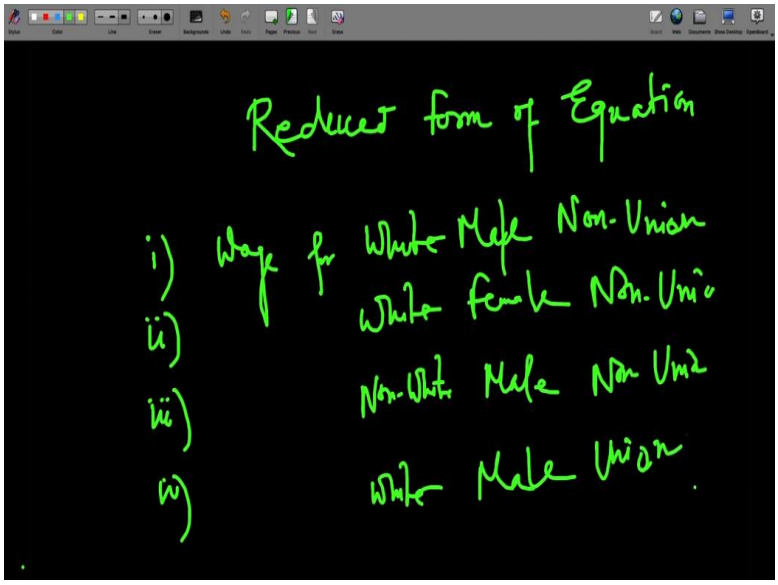
Hello and welcome back to the lecture on Applied Econometrics, and we have been talking about dummy variable. Now, see, till now, we have spoken about intercept dummy.

(Refer Slide Time: 00:40)



And previously, we have spoken about the equations for male, female; and if we have dummy variables, we can actually have equations for different categories of dummy variable. And these equations we have seen, we can actually call them reduced form of equations.

(Refer Slide Time: 00:55)

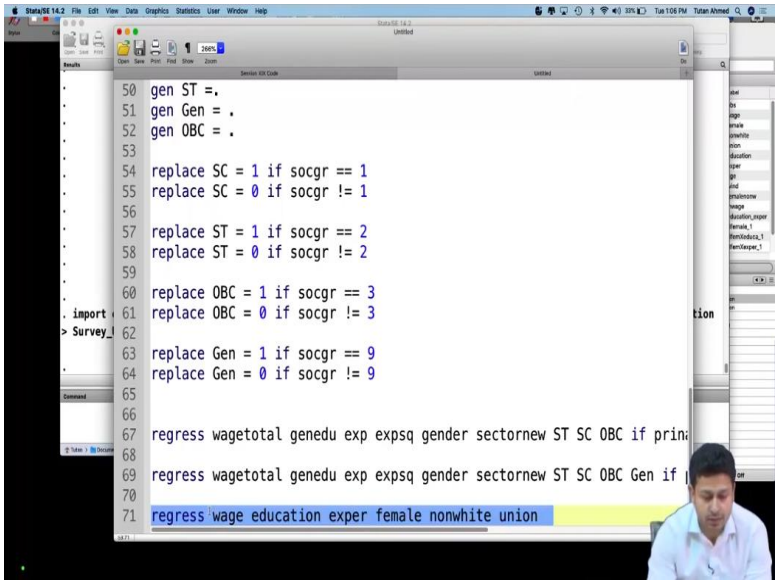


So, essentially, what is happening here is that, in 1 equation, if I have, say male, female, I am boxing them in 1 equation. So, in 1 equation, I have a full population equation, so, like full sample equation, where I have both male and female. I can actually unbox it; I can have an equation for female. I can unbox it and I can have an equation for male. Now, if I have another dummy, for example, sector; so, it could be rural and urban.

And if I unbox it, I can have an equation for the people who are in the rural area, and I can also have equation for people who are in the urban area. And with that, if I further include, let us say social group, so, I can have male, female; I can have rural, urban; I can have the equation for SC, ST, OBC and General. So, in 1 equation, if you bring all these dummy variables together, you can actually include all these different equations; of course, you are making some assumptions about the slope.

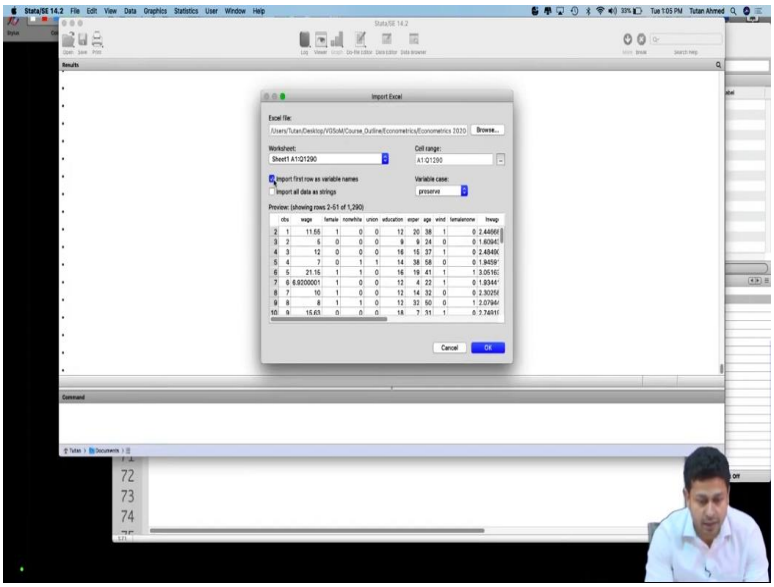
So, in all the equations, you are assuming the slopes are constant; but then, at that cost, what you are getting is, within 1 equation, you can actually get so many different equations. And because you are able to capture all these different possibilities in 1 equation, we often call, by running a dummy variable, we often call reduced form of equation. Now, let us try to explain the reduced form of equation a little further. And we already have some notion about how it works.

(Refer Slide Time: 02:36)



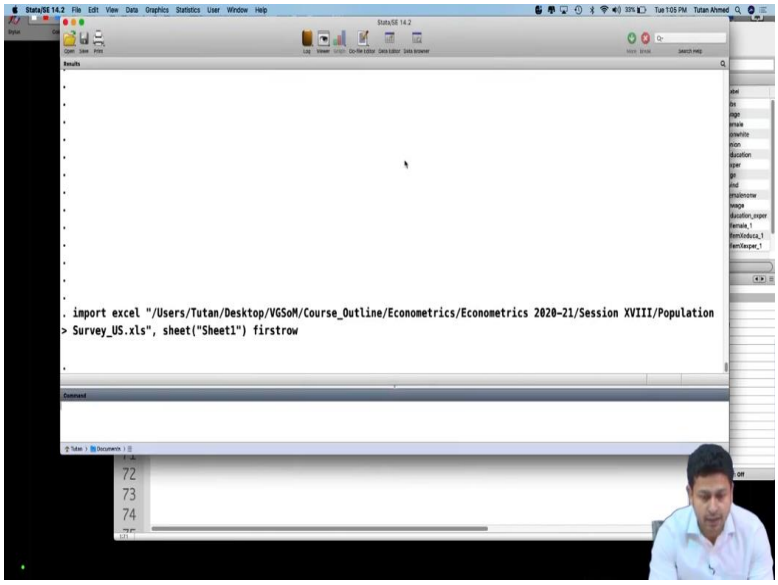
So, let us just open a new data set to see how we can really get the reduced form of equation.

(Refer Slide Time: 02:44)



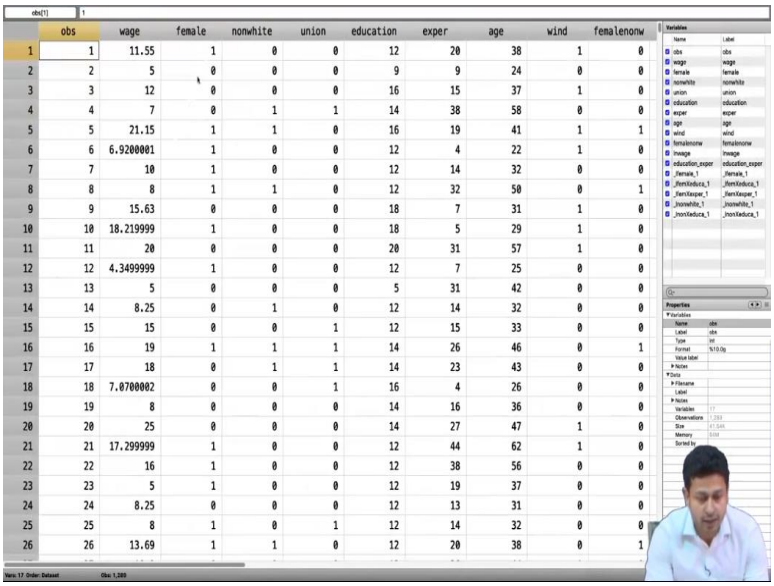
So, let us say I want to open another data file that is on U.S. Population Survey, and I already have this data here. And I will have import first row as variable name.

(Refer Slide Time: 03:00)



So, I get the data set here, and I can see the data set here.

(Refer Slide Time: 03:02)



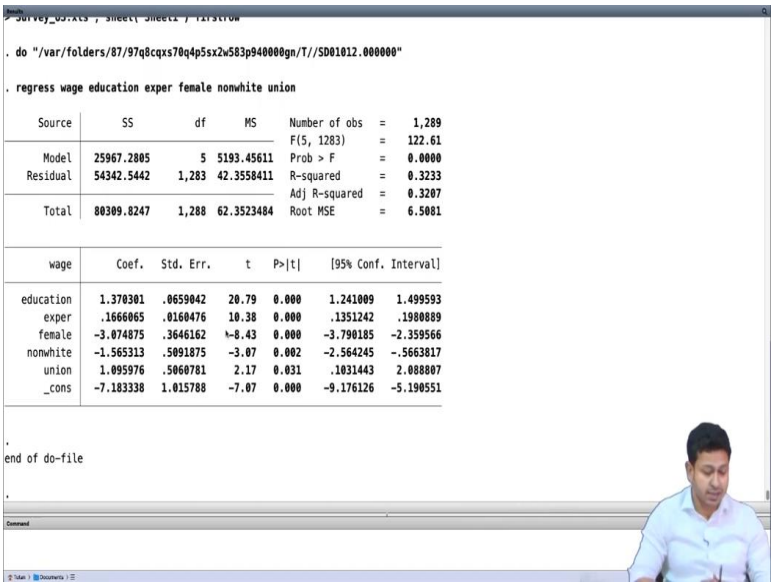
	obs	wage	female	nonwhite	union	education	exper	age	wind	femalenonw
1	1	11.55	1	0	0	12	20	38	1	0
2	2	5	0	0	0	9	9	24	0	0
3	3	12	0	0	0	16	15	37	1	0
4	4	7	0	1	1	14	38	58	0	0
5	5	21.15	1	1	0	16	19	41	1	1
6	6	6.9200001	1	0	0	12	4	22	1	0
7	7	10	1	0	0	12	14	32	0	0
8	8	8	1	1	0	12	32	50	0	1
9	9	15.63	0	0	0	18	7	31	1	0
10	10	18.219999	1	0	0	18	5	29	1	0
11	11	20	0	0	0	20	31	57	1	0
12	12	4.3499999	1	0	0	12	7	25	0	0
13	13	5	0	0	0	5	31	42	0	0
14	14	8.25	0	1	0	12	14	32	0	0
15	15	15	0	0	1	12	15	33	0	0
16	16	19	1	1	1	14	26	46	0	1
17	17	18	0	1	1	14	23	43	0	0
18	18	7.0700002	0	0	1	16	4	26	0	0
19	19	8	0	0	0	14	16	36	0	0
20	20	25	0	0	0	14	27	47	1	0
21	21	17.299999	1	0	0	12	44	62	1	0
22	22	16	1	0	0	12	38	56	0	0
23	23	5	1	0	0	12	19	37	0	0
24	24	8.25	0	0	0	12	13	31	0	0
25	25	8	1	0	1	12	14	32	0	0
26	26	13.69	1	1	0	12	20	38	0	1

So, I have like this data set where I have wage; I have a dummy variable called female; I have a dummy variable called non-white. So, already that the values of the dummy variables are assigned as 0 and 1, so, we do not have to do anything here. And if someone is non-white, it is going to be 1; someone is white, it is going to be 0; and someone is a union member, it is going to be 1; someone is not a union member, it is going to be 0.

Then we have a quantitative variable, education, experience and age, let us say. And there are some other variables we do not need to bother really. So, let us first run a regression equation with all these different variables. And here is the regression equation. I have to regress wage and our independent variables are female, non-white, union, education and experience. And if

I run this regression equation; I have already imported the data, so, I can run the regression equation.

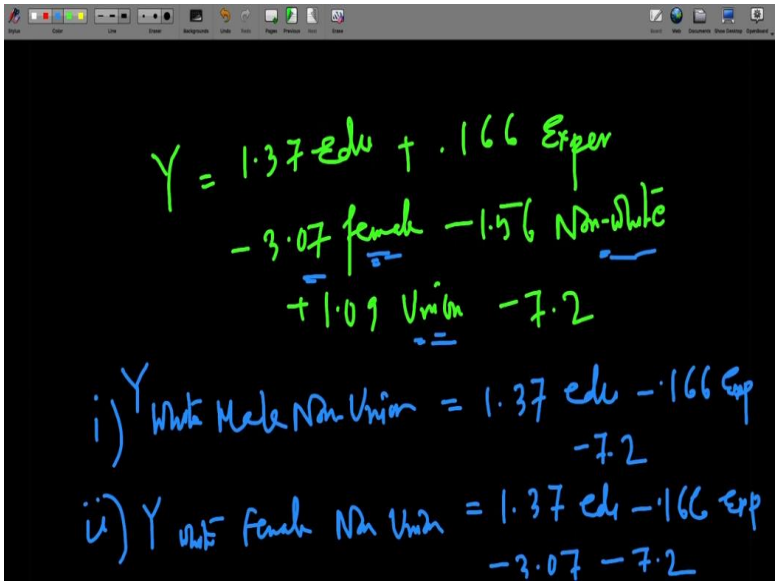
(Refer Slide Time: 04:02)



And if I run it, so, I get this regression table. So, I have all these different variables here. Now, how do I really interpret? So, how do I really now unbox multiple equations from this one reduced form of equation. So, let us say, if I want to get the wage for white male who are not in the union; non-union. So, I will just create different cases; white male non-union. Or I want to get, let us say, white female non-union.

Then I may want to get the case for, let us say, non-white male non-union. Or I may want to get a case where I want white male union. So, all these possible cases, I want to understand using this 1 equation. And how do I do that? I will do that, let us say, I will get the equation here. So, what is my equation going to be? The reduced form equation is going to be 1.37 into education.

(Refer Slide Time: 05:53)



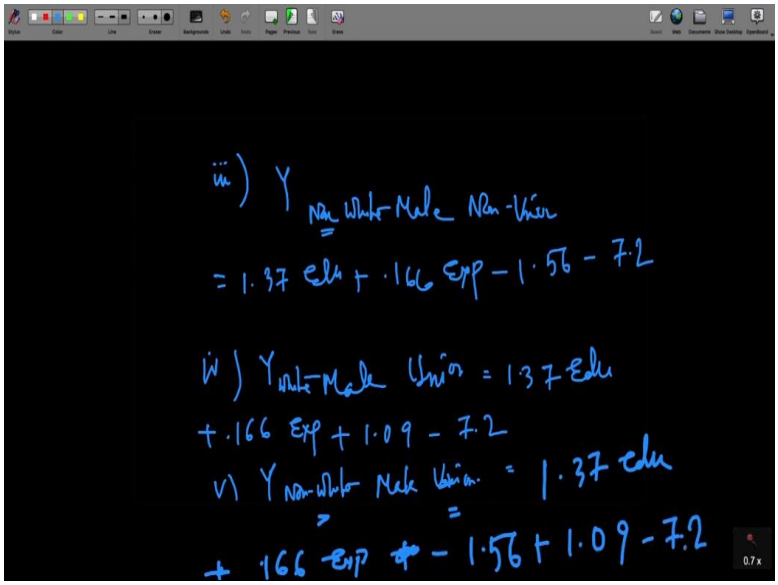
Y = 1.37 into edu, 0.166 experience; and then you have female is -3.07; -3.07 into female. And then, I have non-white; minus of 1.56 non-white. And then, I have 1.09 union; plus of 1.09 union. And there is a constant term, minus of 7.2. Now, let us go back to our first case. Our first case was that I want the wage for white male non-union. So, we will use a different colour; it is too boring to have one colour.

So, for case 1, white male non-union; so, if it is a white, so, my non-white is equal to 0; so, this is going to be 0. If it is a male, the female is going to be 0; so, this is equal to 0. And non-union; if non-union, union is going to be 0; so, that is again going to be 0. So, Y for white male non-union is going to be only this, 1.37 edu minus 0.166 experience. This is equal to 0; the female is equal to 0; non-white is equal to 0; union is equal to 0.

So, all these terms are going to be 0; minus 7.2. If I want to go for the second case where I have, only change is, from the first equation to second equation, I have from male to female. So, all I have to do is that; when it is female, so, my union is going to be 0, and my non-white is going to be 0; only value I will have for female; so, I will have to put 1 for female. So, let us say Y white female non-union is going to be 1.37 edu minus 0.166 experience.

Now, for female, now I have a value equal to 1. So, that means, I have to add minus of 3.07 here, and minus of 7.2. So, that is how I get my second equation. Now, if I want to get my third equation, I want it to have non-white male non-union.

(Refer Slide Time: 09:05)



So, let us say Y non-white male non-union. So, it is going to be; in the previous equation, I have 1.37 edu, then experience; because it is male, so, female is going to be 0; and because it is non-white, now I have the dummy for non-white; that is going to be 1; so, I will have value for this. So, it is going to be minus of 1.56. And non-union; so, union is going to be 0; so, I do not need to bother about this particular variable.

So, it is going to be 1.37 edu; and then, plus 0.166 experience; then is male; so, female is going to be 0. So, and the next term is non-white; that is going to be 1; minus of 1.56. And then, I will have my union term; because it is non-union, so, it is 0. And then I have the constant term. This is equal to minus of 7.2. So, that is going to be the reduced form equation. What was the last case we had? We had white male union.

So, let us write down, what we want is Y white male union. And that is going to be; so, my white; so, non-white is going to be 0; male; the female is going to be 0; union; so, it is going to be 1. So, then, what I will have is 1.37 edu plus of 0.166 experience. And the only thing that will remain is the union, because it is, we have included white male. So, if it is a white, so, non-white is going to be 0; and female is going to be 0; only union term will be there; so, which is 1.09.

So, if I add, it is going to be 1.09 and the constant term is 7.2. Now, let us say, if I wanted to get another case Y non-white male union; let us say this is, value is equal to 1 for 2 different dummies. So, here, what I will have is, so, I will have the value for non-white and I will have

value for the union, but I will not have any value for female. So, it is going to be 1.37 edu plus 0.166 experience.

And then, I will have, because it is male, so, no value for female; and it is non-white; because it is non-white, so, non-white is going to be 1; so, it is not a plus, it is going to be minus 1.56, because it is a non-white. And then I have a union. So, this is going to be again 1. So, it is going to be 1.09. And then, of course I have the constant term. So, this is how for different cases, you see that there are so many different possible regression equations we can actually derive from just 1 equation where we have used all these different types of dummy variables.

So, this is the idea. This is why we call this regression equation using all those dummy variables; we call it reduced form equation. And from this equation, we can unpack and unbox so many other equations. So, that is the idea of it. Now, having said that, it is really convenient, because we really can get so many different equations from 1 equation; but then, there the cost is that, we have the slopes constant.

Now, can we really have the slope constant? We cannot, and we have seen it previously that if you run a regression for different subpopulations, you are actually going to get different regression equations, so, different with the slopes for other explanatory variables being different. So, we need to understand why it is happening. And in the next lecture, we are going to talk about the slope dummy. We have been talking about intercept dummy, but in the next lecture, we are going to talk about slope dummies. And with this, we end the lecture here. Thank you.