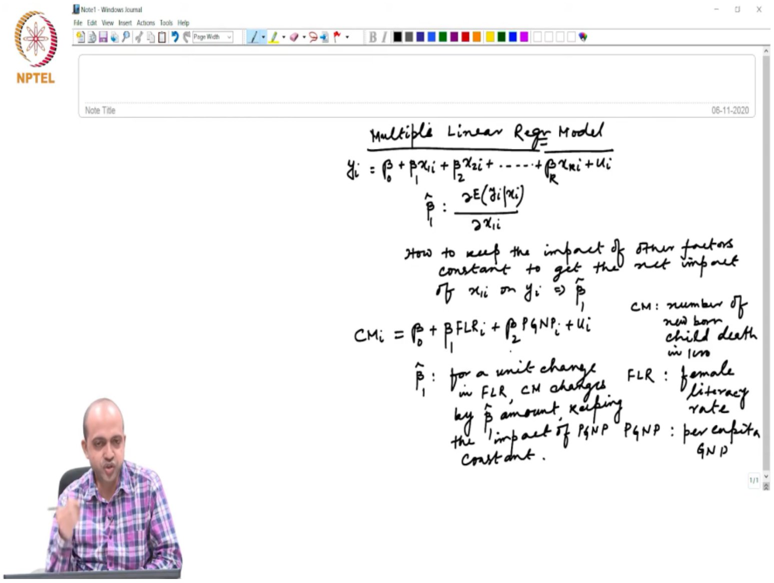
Introduction to Econometrics
Professor Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology Madras
Lecture 20
Application of STATA for hypothesis testing and introduction to multiple linear regression model Part - 4

(Refer Slide Time: 00:14)



Multiple Linear Regr Model

$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + u_i$

$\hat{\beta}_1 : \frac{\partial E(y_i|x_i)}{\partial x_{1i}}$

How to keep the impact of other factors constant to get the net impact of x_{1i} on $y_i \Rightarrow \hat{\beta}_1$

$CM_i = \beta_0 + \beta_1 FLR_i + \beta_2 PGNP_i + u_i$

CM : number of new born child death in 1000

$\hat{\beta}_1$: for a unit change in FLR, CM changes by $\hat{\beta}_1$ amount, keeping the impact of PGNP constant.

FLR : female literacy rate

PGNP : per capita GNP

Welcome. So we were discussing about multiple linear regression model and our model was something like this, yi equals to beta 0 plus beta 1 x1i plus beta2 x2i plus dot dot dot beta k xki plus ui. Once we estimate the model, we will get beta 1 hat, beta 2 hat, like that and the interpretation of beta 1 hat, we were trying to derive from this: delta of expectation yi given xi divided by delta x1i.

So, that means for a unit change in x1i, yi changes by beta1 hat amount on an average, keeping the impact of other factors constant. That means, in economics, many a time we have heard about the term called ceteris paribus condition meaning keeping the impact of other factors constant. And today, we are going to learn how to keep the other factors constant to get the net impact of x1i on yi which is basically beta 1 hat.

And this, we will explain with a dataset and before we deal with the data set we will take one example. We are taking this example where our dependent variable is child mortality rate equals to beta0 plus beta1 female literacy rate plus beta2 per capita GNP gross national product plus ui.

So, this data is basically is a state level data, our dependent variable is child mortality rate. CM is called child mortality rate of i^{th} state. And how child mortality rate is defined? Is it defined as the number of new born child death in 1000. And then, our first dependent variable is FLR which is female literacy rate. And then PGNP is per capita GNP. So, this is what we are trying to understand from this particular data set. We will first try to understand why these particular variables are included in the model. Why we are trying to explain the child mortality rate, and there, as a determinants of the child mortality rate, we have included female literacy rate and per capita GNP. What is the logic?

Now, as mothers become literate, they become more aware of how to give proper care to their child during the pregnancy as well as post that. What to eat, what type of specific food to eat, how much time to give for rest and what are the proper medicines to take, so on and so forth. So, these type of factors, so that means the care, how much care you take during and after, during pregnancy and after delivery, that will basically determine the mortality rate.

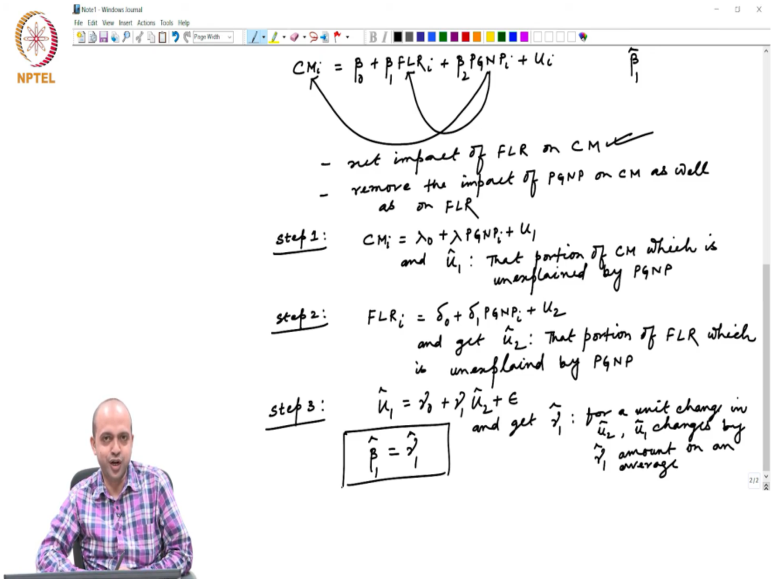
If proper care is given, then the mortality rate will come down. And we believe that as the mothers, they become more literate, they become more aware of this type of things, they take proper care to their child and as a result of which, the child mortality rate comes down. So, that is the reason female literacy rate is included in the model. This is a state level data, so we will be measuring how many women are literate at, out of 100 or 1000.

Then per capita GNP, state level. It is some kind of indicator of per capita income actually. We believe that as income increases, the female literacy rate will also increase because female literacy rate has some connection with the state's income and also higher the state's per capita income, we believe the standard of living of that particular state would be little, would become higher and at higher level of income, mothers will be able to take proper care in terms of food, medicine, so on and so forth during pregnancy or after delivery.

Because of this, in our model, we have hypothesized that there might be many other factors but we have hypothesized these are the two factors in determining the child mortality rate. And that is the reason we have included these two factors in the model. Now, here our objective is to get beta1 hat and what is the beta1 hat? What is the interpretation? For a unit change in FLR, child mortality rate changes by beta1 hat amount keeping the impact of PGNP constant. That is an interpretation.

So, for a unit change in FLR, CM (child mortality rate) changes by beta1 hat amount on an average keeping the impact of PGNP constant. This is the interpretation. And what would be the interpretation of beta2 hat then? For a unit change in per capita GNP, child mortality rate changes by beta2 hat amount keeping the impact of FLR constant. This is how we can interpret the coefficients in multiple linear regression model. And now we will see step by step, what is the procedure that we must follow to keep the impact of other factors constant.

(Refer Slide Time: 10:49)



$CM_i = \beta_0 + \beta_1 FLR_i + \beta_2 PGNP_i + u_i$ $\hat{\beta}_1$

- net impact of FLR on CM
- remove the impact of PGNP on CM as well as on FLR

step 1: $CM_i = \lambda_0 + \lambda PGNP_i + u_1$
and $\hat{u}_1$: That portion of CM which is unexplained by PGNP

step 2: $FLR_i = \delta_0 + \delta_1 PGNP_i + u_2$
and get $\hat{u}_2$: That portion of FLR which is unexplaind by PGNP

step 3: $\hat{u}_1 = \gamma_0 + \gamma_1 \hat{u}_2 + \epsilon$
and get $\hat{\gamma}_1$: for a unit chang in $\hat{u}_2$, $\hat{u}_1$ changes by $\hat{\gamma}_1$ amount on an average

$$\boxed{\hat{\beta}_1 = \hat{\gamma}_1}$$

So, we will write the model once again here. Child mortality rate equals to beta0 plus beta1 FLR on the ith state plus beta2 PGNP for the ith state. So, first we will try to get the impact of beta1 hat from this model. That means we will try to keep the impact of PGNP constant. Now, if we think logically, this PGNP will have some impact on FLR per capita GNP and then, PGNP will also have direct impact on child mortality rate, since it is an explanatory variable.

So, if we want to get the net impact of FLR on CM, then what should we do logically? Logically, if we think, then we can understand, we have to remove the impact of PGNP from both CM as well as FLR which is quite simple to understand. We want to get the net impact of FLR on CM. What is our objective? Objective is to get net impact of FLR on CM.

And to achieve that objective, what we should do? We should then remove the impact of PGNP on CM as well as on FLR. We want net impact, so we have to remove the impact of per capita

GNP from both this. If we remove, then this child mortality rate would be purified from the impact of PGNP, FLR will also be purified from the PGNP and then if we regress CM on FLR, we will get the net impact. So, this is our objective, net impact of FLR on CM, that is what we want to achieve. And to achieve that, we need to remove the impact of PGNP on CM as well as on FLR.

How to do this? So, step 1, to remove the impact of PGNP on CM, what we actually do? We will regress CM on FLR and we will run this type of regression. Let us say, lambda0 plus lambda1 PGNP plus let us say u1. We will run this regression and get u1 hat. Now, once we regress CM on PGNP and collect the residual from that regression, what does this u1 hat indicate basically? If we recall, error term captures the impact of other factors.

So, that means when we run this regression CM on PGNP and then collect u1 hat, that is basically, that portion of CM which is unexplained by PGNP. So, that means u1 hat indicates the unexplained portion, that portion of CM which is unexplained by PGNP. So, that means I can say that u1 hat is nothing but the purified value of CM. It is also CM, child mortality rate only, but only after removing the impact of PGNP. PGNP will certainly explain certain portion of child mortality rate since we have included this variable in the model.

So, you remove this by regressing, how to do that? Get the u1 hat which is basically the residual, which is that portion of CM which is unexplained by PGNP. Then in step 2, what we should do? We will run another regression where FLR is regressed on PGNP plus let us say, u2. And from this regression, what we will get? We will get u2 hat. Now, what does this u2 hat indicate? From the same logic we will say that u2 hat is that portion of FLR which is unexplained by PGNP.

So, u2 hat basically captures that portion of FLR which is unexplained by PGNP. So, that means this u1 hat and u2 hat, they are basically indicating the purified value of CM and FLR respectively. So, that means what we can say? That means u1 hat is free from PGNP, u2 hat is also free from PGNP.

And then, in step 3, what we should do? We will run another regression of u1 hat equals to let us say gamma 0 plus gamma1 u2 hat plus let us say epsilon. And from this regression what we will get? We will get, let us gamma1 hat. Now, what would be the interpretation of this gamma1 hat? Gamma1 hat basically indicates for a unit change in u2 hat, on an average, u1 hat changes by gamma1 hat amount. That means the interpretation is for a unit change in u2 hat, u1 hat changes by gamma 1 hat amount on an average. But then what is u2 hat? u2 hat is basically the unexplained portion of FLR which is free from PGNP. And what is u1 hat? u1 hat is basically unexplained portion of CM which is free from PGNP. So, that means, I can say the interpretation as, for a unit change in u2 hat, I will say that, for a unit change in FLR, purified value of FLR, what is the change in purified value of CM? That is nothing but gamma 1 hat and that is nothing but our beta 1 hat.

So, what we can say? That actually, beta 1 hat equals to gamma 1 hat. So, this is how we can get the net impact of FLR on CM following these three steps. So, the idea is, since we want to get net impact of FLR, we need to remove the impact of PGNP on FLR as well as CM. And how will you remove? Simply just regress both CM and FLR on PGNP and then we need to collect the error term, predicted value of the error term.

Since the predicted value of the error term captures the impact of omitted variable, that means we can say that at least this portion is not explained by PGNP. We are doing this and then we are regressing one error term on the other so as to get the net impact of FLR on CM which is nothing but beta 1 hat.

But in step 3, if you look at the equation carefully, we made a mistake while specifying our equation. Look at step 3; u1 hat equals to gamma 0 plus gamma 1 u2 hat plus epsilon. This equation, we made a mistake. What is the mistake can you think of? In step 3, is there any mistake I have committed? I deliberately made a mistake for your own understanding and now I am asking you to think what is the mistake? Is this equation in step 3 theoretically correct?