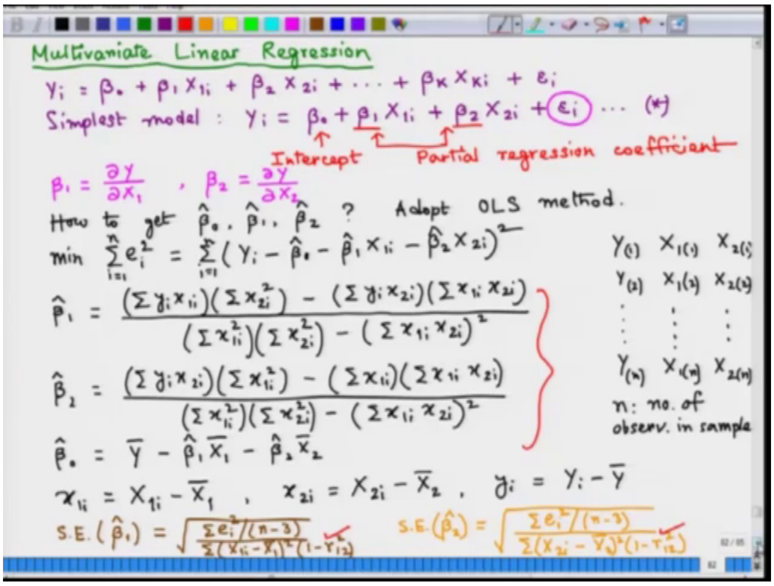
Microeconomics: Theory & Applications
Prof. Deep Mukherjee
Department of Economic Sciences
Indian Institute of Technology, Kanpur

Lecture – 59
Linear Regression (Part-2)

Hi, welcome back to the lecture series on Microeconomics. Let us now continue our discussion on Linear Regression analysis. Last time we have seen simple linear regression analysis with one explanatory variable; now we are going to extend that simple model to more than one explanatory variables. So, if we have more than one x, then that kind of model is called multiple linear regression model or multivariate linear regression model. So, let us first start with a model expression.

(Refer Slide Time: 00:51)



So, multi radiate linear regression model could be expressed as a linear function of this sort where we have more than one explanatory variables right, now for simplicity sake we can assume k equal to just 2. So, the simplest possible linear regression model would involve only two explanatory variables, and then we have a model with a smaller dimension right.

So, for this lecture we are going to work with this two explanatory variable multiple regression model, but needless to say that all the results that we are going to derive and see can be easily generalized to k variable case. So, first thing that I would like to note here is the following. What would be the interpretation of these coefficients in a multivariate linear regression model. So, of course, this beta naught is going to be my intercept as usual, but we have slightly different intercept, slightly different interpretation for the slope parameters or coefficients beta 1 and beta 2. So, how can we interpret these coefficients?

So, these are known as partial regression coefficients. So, what do we mean by partial regression coefficient? So, this term actually has some relation to calculus where we take partial differentiation right. So, this multivariate regression coefficients actually are partial derivatives of this variable Y i with respect to the corresponding explanatory variable. So, we can actually write that beta 1 is del Y i del X 1 i and beta 2 is del Y i del X 2 i or in fact, we can remove this Y notation this i to make it simpler expression ok.

So, these betas are basically indicating the change in the dependent variable associated with 1 unit increase in the explanatory independent variable, but simultaneously holding all other independent explanatory variables constant in the equation. So, the coefficient beta 1 measures the impact of 1 unit increase in the variable X 1 on the dependent variable Y, holding the X 2 variable constant, but note that here we are not holding any other omitted variable which are part of this term epsilon i constant right. So, omitted variables can vary, but those variables which are in the regression equation, you have to keep them constant while you are changing one explanatory variable at a time ok.

Let us now concentrate on a statistical method to get estimates of these parameters right. We are again going to adopt the wireless method that we have explained before right. So, brief recap we know that in the wireless method we have to minimize the sum of squared residuals alright and here this will look like the following right ok. So, this is the expression that we need to minimize with respect to the unknowns which are basically beta naught hat, beta 1 hat and beta 2 hat.

So, I am skipping the steps you know what to do you have to take derivatives of this sum of squared residuals with respect to these unknown variables set them equal to 0, you have to then solve those normal equations, 3 of them to find the solution for these 3 unknown variables. So, if you do all this, then actually you get this nice looking formula ok.

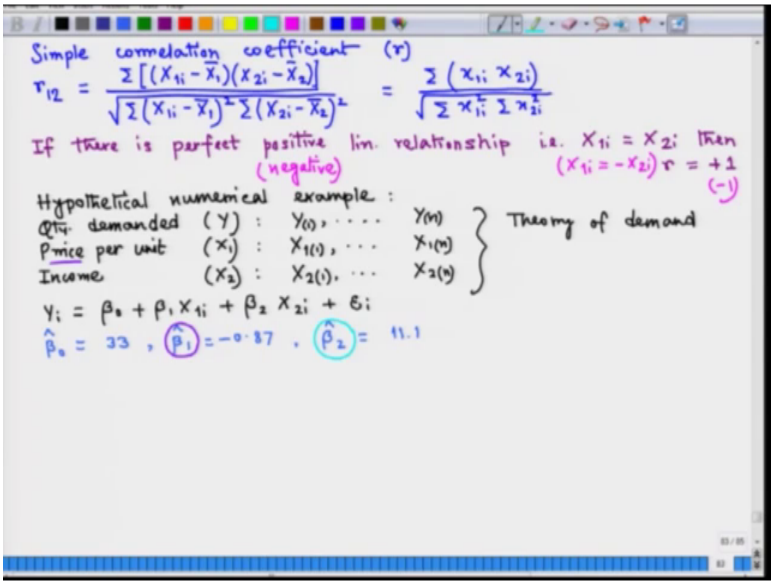
So, here the lowercase variables actually indicate the deviations from the mean. So, if you have observed some data sets like this, where n is basically the number of observations in the sample, then x 1 i is equal to X 1i minus the arithmetic mean of the variable X 1. Similarly, we can write x lowercase 2 i equals mean variable X 2 minus the arithmetic mean of the variable X 2 and similarly y i is also centered around the arithmetic mean of the dependent variable.

So, here in these expressions that we obtained for our regression coefficients, the sum all the some ranges from i equal to 1 to n ok. Now as we have discussed earlier that these coefficient values beta 1 hat beta 2 hat and beta naught hat will vary from sample to sample, because if the data changes of course, the formula will result into a different number. So, there is some kind of dispersion in the data see you we change our sample we get different betas.

So, there is some kind of variation in the beta. So, of course for that reason there is some standard error that will emerge we have looked at that before as well. Here in this case let me also give the formula for standard error for the slope parameters or the partial regression coefficients, because these are the most important variables when an applied econometrician uses this regression tool. So, again the sum ranges from 1 to n, I am skipping this sum range to save space and produce less clumsy expression. I am not showing you the derivation of this, because this is tedious and that serves no purpose for this course as well. So, let me just give you the final expression ok.

So, these are the tedious expressions for standard error, we will see the usage of these standard errors later on. Now note that we have introduced two new notations here, sorry I forgot to write 1 2 here. So, actually one new notation and that is basically r 12. So, what is it? So, that is the thing that we are going to study next. So, the new symbol that we have introduced r 1 2, actually is a measure called simple correlation coefficient which measures the strength and direction of linear relationship between 2 variables. So, here r 1 2 measures the linear associationship between two explanatory variables x 1 and x 2. So, let us have a formula for these expression r 1 2 so, that you can compute it from the data given.

(Refer Slide Time: 17:55)



Again the sum ranges from 1 to n. So, I skip these to make it less cluttered makes the expression nice looking ok. So, you can actually see that this expression can be simplified further and this will be in the deviation term. So, we are talking about the centered data now. So, if the original data is given on capital Y, capital X 1 and capital X 2. Now we can center all these variables with respect to their arithmetic mean and then we can use the lowercase variables to have a simplified expression, but basically they are the same thing ok.

So, now note that this r can take two extreme values. So, if there is a perfect positive linear relationship between x 1 and x 2, it implies X 1 i is exactly equal to X 2 i, then r takes value plus 1. And, if we have a negative linear relationship which is also perfect in that case, we mean X 1 i is equal to minus of X 2 i, then the expression that we have above on r 1 2 results into a value of minus 1 for r. Well we have discussed so, far on the multivariate linear regression model in a theoretical setup.

Now, let us you know look into the philosophy of this model and let us look at the interpretation of these coefficients through a simple graphical exercise, a simple numerical exercise you can say ok. So, let me talk about some hypothetical numerical example ok. So, suppose an econometrician or an applied economist is given data on quantity demanded and quantity demanded of a commodity and let me denote this by Y, that is going to be my dependent variable and I observe n number of data points ok. Then

I got two explanatory variables one is price per unit denote this by X 1. So, I have wait a minute so, 1 n ok.

And, we have another explanatory variable which is per capita income of the household, you remember the theory of consumer behavior what we have studied earlier; there we have seen that not only the price of a particular commodity has impact on the quantity demanded, but consumers income also plays a big role in determining the demand. So, we have this simple hypothetical model. So, basically this is coming close to our theory of demand right ok.

So, given the data set we have if we apply the formula, that we have jotted down so far, we can get the estimates for our partial our regression model and we can find the coefficients of our linear regression model, which is written as Y i equals beta naught plus beta 1 X 1 i plus beta 2 X 2 i plus the random or stochastic noise epsilon i ok. So, do not be scared to the complicated formulas because, these days the statistical softwares are available and they can help you to find out these numbers quite quickly within a second. But, you know it is important to know what goes behind the screen right; what actually is happening behind the curtain when you ask a computer to compute these things for you.

So, given the data suppose you have done calculations yourself manually or a computer produces some numbers for these coefficients and measures, and let us assume some simple numbers purely hypothetical just you know for the purpose of illustration ok. So, suppose these are my numbers fine. So, what do we get to know from these numbers? So, first we have to concentrate on the partial regression coefficients beta 1 hat and beta 2, because they have sorry let me put a negative sign in front of this number 0.87 because the reason it will be clear to you soon.

So, here let us first concentrate on beta 1 hat. So, what does that mean? It means that if there is an unit increase in the price of the commodity then what will be the impact on the quantity demanded. So, we know that the law of demand states that there is an inverse relationship between price of a commodity and quantity demanded of that commodity. So, we expect a negative number for this beta 1 hat coefficient, and that is why you know I placed this negative sign in front of the number. So, this basically satisfies that law of demand right, that every other thing remaining the same if price increases quantity demanded falls.

Now, let us look at the other partial regression coefficient beta 2 hat. So, what does that mean? It says that if the price of the commodity does not change and if the income of the consumer changes by 1 unit, then what will be the impact on Y? So, by how much the quantity demanded will change. So, if you remember our discussion on the theory of demand and income increase actually relaxes the budget constant of a consumer and the consumer tends to purchase more to find the equilibrium right. So, there is a positive income effect on the quantity demand and purchase right. So, that is what we exactly see here as well ok.

So, we will continue with this discussion in the next lecture.