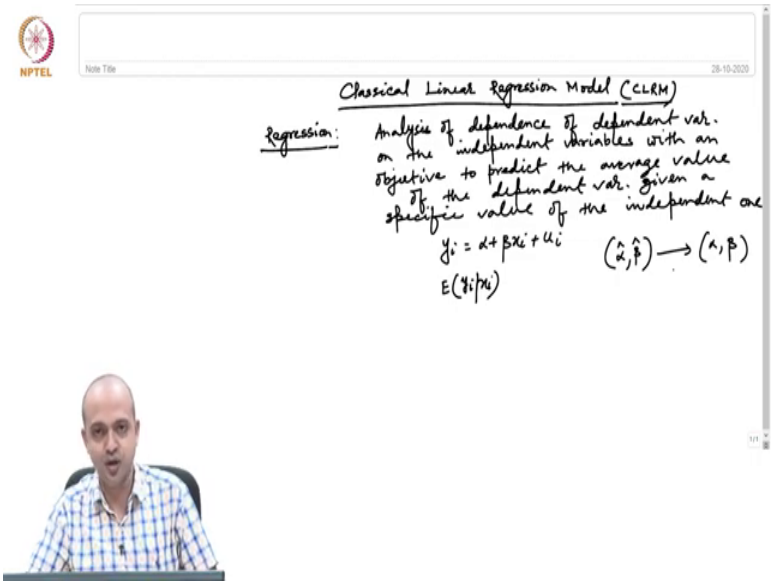
Introduction to Econometrics
Professor Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology, Madras
Lecture 7
Classical Linear Regression Model Part-1

(Refer Slide Time: 00:14)



Welcome to our discussion of econometrics once again and today we will discuss the Classical Linear Regression model. We are going to learn more about regression. This is called classical linear regression model. So, let us first try to understand the meaning of classical. Why this is called classical linear regression model? Do you have any idea about the meaning of classical? See classical like music's and all?

Basically classical indicates the foundation. So that means if you know classical music based on that you can learn many other variants of musics'. Similarly, in the regression analysis-this classical regression is a basic regression model after which econometricians have developed different types of other econometric regression models. So, classical means the basic or foundation. What is the foundation of this regression model that is classical linear regression model or it is also known as CLRM?

And what is regression? The traditional meaning of regression is basically tending towards average. Francis Galton, the statistician and econometrician, first used this term regression in a context where he observed that generally taller parents have taller kids, and shorter parents

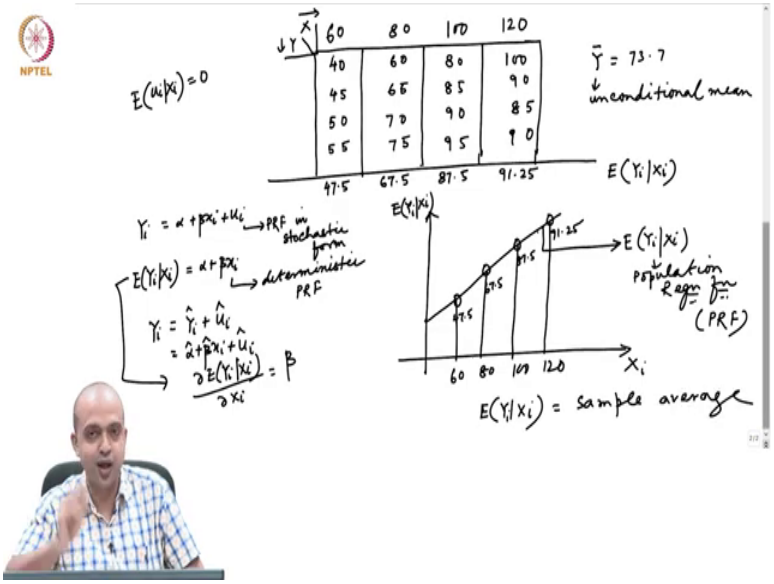
generally have shorter kids, shorter sons or daughter. But, if you have a population of children where different children have parents with different height then we can observe that if we take any particular child's height given his parents height, then that will tend towards the population average. So even though taller parents have taller kids, shorter parents have shorter kids, on an average, a kid's height tends towards the population average. This was the context in which this regression was first introduced by Francis Galton.

But today when you talk about regression it basically indicates regression analysis. It means dependence analysis. So, what does regression mean? It means dependence analysis of the dependent variable on the independent variables with an objective to predict the average value of the dependent variable given a specific value of the independent one. That means when I am writing this type of model where $Y_i = \alpha + \beta X_i + u_i$ in the context of consumption and income, my objective in this regression analysis is to analyze how consumption is dependent on income. That is why we say that it is the analysis of dependence Y on X with an objective to predict the average value of Y. That means to predict the average value of Y given Xi. So, this is called regression analysis. This is the modern meaning of regression analysis and the earlier meaning what was introduced by Francis Galton is actually quite different from this analysis.

Now, we are doing this regression analysis because if you recall from yesterday's class, our entire objective is to draw inference about the true population parameter α and β with the help of $\hat{\alpha}$ and $\hat{\beta}$. $\hat{\alpha}$ and $\hat{\beta}$ is used to draw inference about the true population parameter α and β and we estimate this α and β by regression analysis. So, that is why regression analysis is so important in econometrics and there is only one technical regression to estimate this $\hat{\alpha}$ and $\hat{\beta}$ with an objective to draw inference about the true population parameter α and β. Now, the question is what is exactly the meaning of regression that we need to understand? Because, estimating the model in today's world using the advent of softwares is very simple.

But what is more important is to understand is what exactly we are estimating or what the interpretation of that is. So, for that reason I will make you understand the exact meaning of regression analysis with a simple example in the context of consumption and income. Let us say that we have a data on weekly income and consumption of some 16 families.

(Refer Slide Time: 08:49)



Let us start form 40, 45, 50 then 55 and 60, 65, 70, 75 and then 80, 85, 90, 95 and then 100 and let us say 100, 90, 85 and once again 90 once again 90. This is the data on consumption which is denoted by Y. Now, if I ask you what is the average weekly consumption of these families, you have 16 families' data on their consumption and you can easily calculate the mean and that mean consumption might be let us say, if you take average that would be something around 73.7 or something. This is the average you have calculated by taking the sum of this and dividing by 16.

So, ΣY/n and n equal to 16 will give you some value and that is called mean consumption of this population. Now, let us say that we know some additional information about these households that is their weekly income as well. Let us assume that for this group of household their weekly income is 60, another group have 80, another group have 100 and for another group it is 120.

Now, I have now classified these families into four groups based on their weekly income such that the first group have weekly income of 60 rupees, second group have weekly income of 80 rupees and third group have weekly income of 100 rupees and fourth group have weekly income of 120 rupees. Now, I am asking you to calculate the mean of each group so that mean consumption of the group for whose income is 60 would be something around say 47.5 and for the second group let us say that is something around 67.5 and then this would be something around 87.5 and this would be something around 91.25.

Now, I am calculating mean for each group knowing their weekly income. So that means there are several means- 47.5, 67.5, 87.5 and 92.5. They are quite different from the mean of $\overline{Y}$ while $\overline{Y}$ is called unconditional mean- what we have calculated here this 47.5, 67.5, 87.5 and 91.5. These values are called conditional mean that means expectation of Y_i given X_i.

So, I have derived 4 conditional means. These means are conditional upon their weekly income. Now, in a simple two-dimensional diagram if I plot X_i in the y axis and then Y_i in the Y axis, and then not Y_i rather I would say that this is expectation of Y_i given X_i. So, I am just plotting this conditional mean so this is the first point which indicates 47.5 then second point, third point and fourth point. I have simply plotted the conditional means and if I add this with a line then I would get an upward sloping curve, a upward sloping straight line.

So, this is basically 47.5 and this is 67.5, this is 87.5 and this is 91.25. So when we add the locus of all these conditional means we will get a line which is basically called the regression function or regression line. So, that means I am now getting the regression line which is called expectation of Y_i given X_i which is called the Population Regression Function.

So, that means when I am writing $Y_i = \alpha + \beta X_i + u_i$, that is basically nothing but the population regression function in its stochastic form because I have added one error term also. This is called PRF or PRF in stochastic form. If you take expectation of Y_i given X_i, then that would become $\alpha + \beta Xi$. This is so because we have assumed expectation of u_i given X_i to be equal to 0. Mean value of u_i is equal to 0. That means from the line what we got corresponding to your level of the values 60, 80, 100, 120 that is this line obtained when we know somebody's weekly income, we can easily calculate the conditional average of that group and then we have four such groups. Four such locus of conditional mean and connecting all those conditional means will give you the population regression function. Population regression function is called deterministic when there is no error term. This is called deterministic PRF and the sample counterpart of this PRF would become Y_i equals to actually $\hat{Y}_i + \hat{u}_i$.

So, $\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i + \hat{u}_i$. This is the SRF. That means the sample regression function. Once you estimate the value of the true population parameter alpha and beta then you will get the alpha hat

and beta hat which is called sample statistic. So this is a simple example which basically says the meaning of regression. So regression function is nothing but a straight line connecting the locus of several conditional means that is conditional upon income. This is the meaning of regression analysis.

Now, how do you interpret the coefficient? See our stochastic form where population regression function says $Y_i = \alpha + \beta X_i + u_i$. Now, what is the meaning of α and β? Let us say, I am trying to understand the interpretation of β. β from this function basically indicates for a unit change in X_i what is the change in your Y_i as I said that is nothing but marginal propensity to consume. But, the problem here is since this population regression function involves a stochastic error term I cannot say that interpretation of β is that because the differentiation of Y_i with respect to X_i will give you only β when there is no error term involved and then only you can interpret β in that way as for unit change in income what is the change in consumption.

So, then we need to first take the expectation of Y_i given X_i which is equal to $\alpha + \beta Xi$ and from there if you differentiate this delta expectation of Y_i given X_i with respect to delta X_i that now becomes your β and how do you interpret this for a unit change in X_i? What is the the change in expectation of Y given X_i? What does it mean?

So, that means for a unit change in income, consumption changes by beta amount on an average. On average component is very important because I am able to measure the change only on expectation of Y given X_i and not for any Y because to get the change in Y you need to differentiate this function which is population regression function and that involves a stochastic error term.

So, first of all we need to take the conditional mean or expectation of Y_i given X_i and that becomes $\alpha + \beta Xi$ and if you differentiate that we will get the β value. Now, this is on an average concept is even more important for some other reason. When I am saying that for a unit change in income consumption changes by β amount, natural question that comes to our mind is that for measuring a change, we need to always have a reference point. Without mentioning the reference point the change concept is meaningless.

If I say that my income changes by 10 rupees then immediately you will ask 10 rupees from where is it 40 to 50 or 50 to 60 or 60 to70 or what is the reference point. Unless I tell you then

the change in 10 rupees does not make sense so whenever I am introducing a change concept in econometrics because I am trying to interpret beta, we need to always keep in mind our reference point and the reference point is basically the expectation of Y given X_i. So, that means the reference is from the sample average.