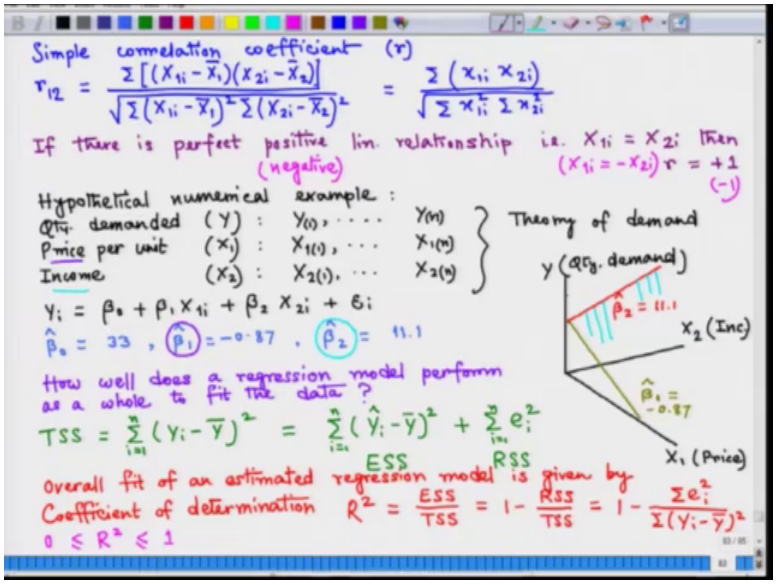
Microeconomics: Theory & Applications
Prof. Deep Mukherjee
Department of Economic Sciences
Indian Institute of Technology, Kanpur

Lecture – 60
Linear Regression (Prat-3)

Hi. Welcome back to the lecture series on Microeconomics. We have been discussing basic econometric theory and practice which is bread and butter for and applied macroeconomist. So, we have studied the ordinary d square estimators, first we will finish the discussion on the wireless estimators, and then we will move to another related topic which is known as hypothesis testing which is considered to be the heart and soul of econometrics.

(Refer Slide Time: 00:45)



Utilize a graph to show you what is exactly happening then probably that will be more effective.

So, let me draw a 3D diagram where I measure my Y or quantity demanded and my price, and the household income or consumer's income. So, this is price, this is my income and Y is quantity demanded ok. So now, let us have the relationship plotted. So, let us first you know look at the relationship between quantity demand and price. So, we expect negative lease slope straight line right. So, let us first draw that ok.

So, this downward slopping straight line actually has the slope value minus 087 and that is precisely the value of the partial regression coefficient for a price variable right ok. And we expect positive lease slope straight line for the X 2 Y plane because as the income increases keeping the price fixed or constant then quantity demanded shall rise. So, we expect something like this and here the slope of this red colored straight line is actually this estimated value of beta 2 parameter, beta 2 hat equals to 11.1, ok.

So, basically we are talking about a regression plane shaded by this area ok. So, the next topic that we are going to study is we know to access the goodness of fit of a regression. So, basically the idea of regression analysis is to pass straight line you know among, you know in between scatter plot right. So, if you have some data a regression line should pass through the center of the data and you know it should give a good fit to you data. So, by fit we mean that you know it should be able to explain the variation in Y, well.

So, how do we go about evaluate evaluation of regression equation? So, we actually through some explanatory variables in the regression model to explain the variation in Y, but if you think about the regression model, then you will see that there is an intersect parameter beta naught as well. So, what does that mean? So; that means, that if we do not throw any explanatory variable in the regression equation to explain variation in Y then the variation will be explained by only a constant number. And you know an intercept term and that is basically the sample mean of the Y. So, the most naive explanatory variable is basically the constant term and that constant term is the sample mean of the dependent variable Y.

So, after adding explanatory variables in the model whether i c is some improvement in the explanatory power of my model to explain the variation in Y ok. So, now we have to look at some measures or concepts what we have already introduced and that is basically the notion of total sum of squares, explains sum of squares and residual sum of squares. So basically, let us first express the squared variations of Y around its mean because that is the most naive explanatory variable, it is a measure of the amount of variation to be explained by the sample mean only. So, let us first write down total sum of square ok.

Now, we have seen earlier that this can be broken down into two components and the first component is called the explained sum of squares and the second component is known as the residual sum of squares right ok. Then, the overall fit of the estimated

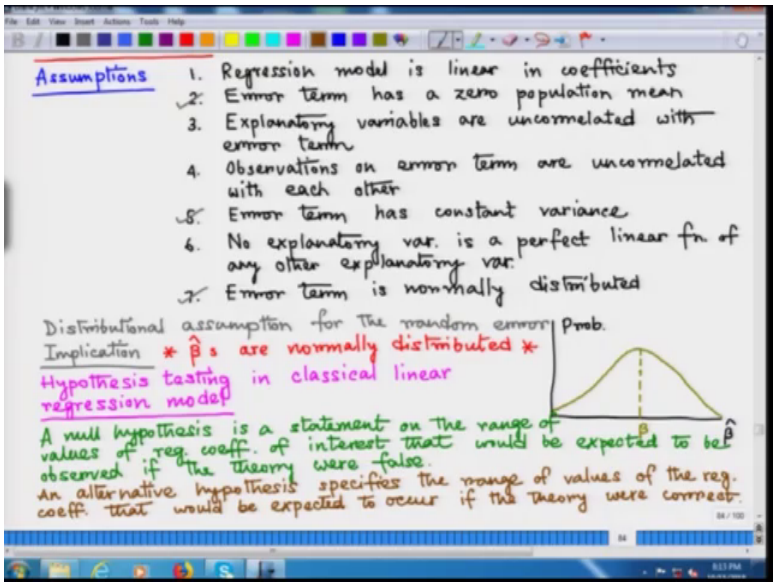
regression model is given by a concept called coefficient of determination. And this coefficient of determination is represented by symbol R square notation R square and this is basically the ratio of the explained sum of squares to the total sum of squares. This is not an econometric scores; so I really do not want to get into deeper details of these concepts because that will consume a lot of time. But there are many text books and online sources from where you can study yourself, I am just you know going to give you the basic results what are required to do some applied microeconomic data analysis.

So, if you see some software output then you will understand what do this measures actually mean ok. So now, note that this R square takes value in a range. So, my TSS RSS and ESS are all nonnegative numbers right because there are the square deviations and the explanatory some of squares that is basically ESS is less than or equal to TSS the total some of squares. So, by these two facts I can safely write that my coefficient term determination actually lies between 0 and 1.

So, if my regression model fails to analyze the variation in y all together then we will see the R square value will be 0 or very close to 0, it will be a very low fraction number, but if the regression model performs well to explain the variations in Y get implies that we have chosen probably the perfect explanatory variables to explain the variations in Y. Then we will observe a very high R square and at and extreme if the explanatory variables are able to explain all variations in my dependent variable Y. Then we will see R square value equal to 1, but that is just a theoretical possibility in practical purpose you will probably not fine R square value equal to 1.

So, now we are going to study what is known as classical linear regression model. So, this model is basically not to be confused with classical economics or classical micro economics. Classical here means, basically the basic or simplest possible linear regression model.

(Refer Slide Time: 12:16)



So, let us start with a set up assumptions which are required to the formulation and working of classical linear regression model. So, we will list down assumptions. So, the first assumption is that the regression model is.

So, we have listed down all the assumptions now let us look at the explanation for some of the assumptions in detail because they will help us in our analysis. So, assumption number 2, assumption number 5 and assumption number 7 basically talks about distribution for the random error or stochastic noise in the regression model and we assume that you know this has a normal distribution. So, I hope that you know you are aware of the normal distribution. So, I will not get into details of normal distribution ok.

But what is the implication of these assumptions ok. So, the normality of error terms actually imply that this partial regression coefficients, the betas the estimated values of these betas which are basically beta hats are themselves normally distributed. Because one of the property of the normal distribution says that any linear function of normally distributed variables is itself normally distributed.

So, we can write down this implication or very interesting result that my regression coefficients are normally distributed. One may ask why this is important to assume? Soon we will find that we need to make some distributional assumption for betas because that will help us in testing economic theories ok.

So, we mean that my beta hat for any particular explanatory variable has a distribution and the mean of the distribution for beta hat will be beta, the population regression parameter right. And betas can take various values, it can be say 0, it can be 1, it can be minus 2 so all set of possibilities are there. So, what do we measure along the y axis in this diagram. So, just to complete the diagram let me have the vertical axis here. So, actually we measure probability right.

So, now let us turn to the topic of hypothesis testing which is the heart and soul of econometric theory. So, let me remind you about econometrics, what does an econometrician do actually? So, an econometrician uses a statistical techniques and economic theory to quantify the relationships between economic variables using sample data. But you know does the journey of an econometrician end right there after getting the estimates of population parameters beta naught, beta 1 and so on? The answer is no, there is more to it.

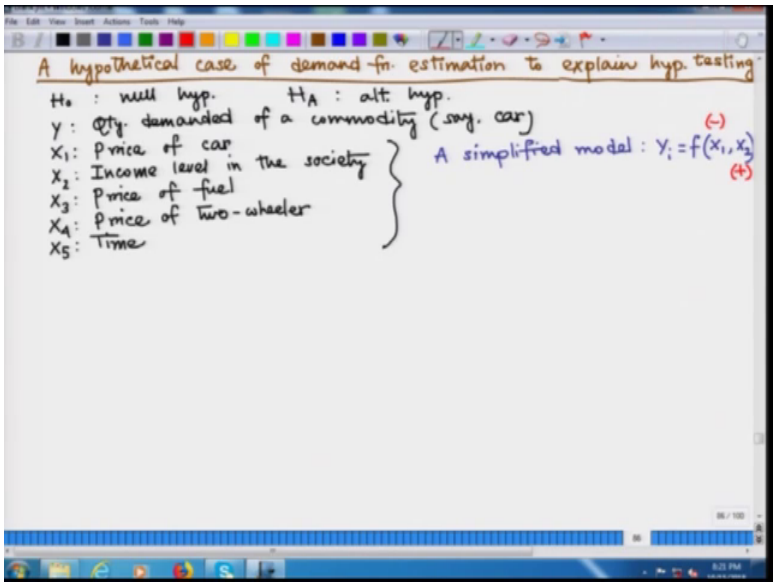
So, now once the equation, regression equation has been established parting and question can be what do we learn from this regression equation, what do we conclude? So, whatever fact we are observing from this sample regression equation does it match with the expectation from micro economic theory or any other economic theory? So, that is to be done through the lens of hypothesis testing which is originally borrowed from statistical methods.

So, to do hypothesis testing we have to set proper hypothesis right. So, there could be two types of hypothesis that we use in practical hypothesis testing and they are known as null hypothesis and alternative hypothesis. So, we will start with defining these two concepts. So, a null hypothesis is a statement on the range of values of regression coefficient of interest that would be expected to be observed, if the theory were incorrect right.

So, basically the word null means 0 as well, null hypothesis can also be thought as the hypothesis that the researcher does not believe in ok, there is also a concept of alternative hypothesis that we are going to start. So, an alternative hypothesis uses specifies the range of values of the regression coefficient of interest or under consideration that would be expected to occur. If the opposite happens it implies that the theory were indeed correct ok.

So, hypothesis testing theory is a vast area in statistics, we do not have time and also we do not require to know bit each and every corner of this vast area, but I will just you know educate you on the basics of hypothesis testing so that you can practice it in real life. So, we will first start with the case of the demand function estimation problem and see how hypothesis testing can be useful in order to test microeconomic theory. So, before we jump on to the demand function example.

(Refer Slide Time: 25:45)



So, the notation used to refer a null hypothesis is H naught and the alternative hypothesis is denoted usually by this notation fine.

So, now let us look at an example of making null and alternating hypothesis using the demand function estimation problem, let us assume Y is the quantity demanded of a commodity say a car right. So, a researcher can think about three or four variables which could explain the variation in Y, the first one is a of course, the price of car then X 2 could be the income level in the society and X 3 could be the price of fuel ok.

Then X 3 4 could be price of some alternative commodity like two wheeler ok. So, there could be another one like the role of time because tests and preferences change and economic conditions do change as well. So, in a nutshell there could be a host of X factors which could explain variation in Y, but let me just work with only two of them as of now ok. So, a simplified model could be defined as Y i, i is basically the number of

observations that we have. So, that is basically the size of the sample and I assume that its a two variable regression model.

Now, note that as X X 1 is price of car my microeconomic theory says that there is law of demand which actually dictates the relationship between X 1 and Y and that is negative right and we also know from the theory of consumer behavior that the relationship between Y and X 2 is expected to be positive right. So, this is basically coming from my theory. So, this is basically my hypothesis that I would like to test whether indeed the relationships between these variables are following the theoretical result or not.

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