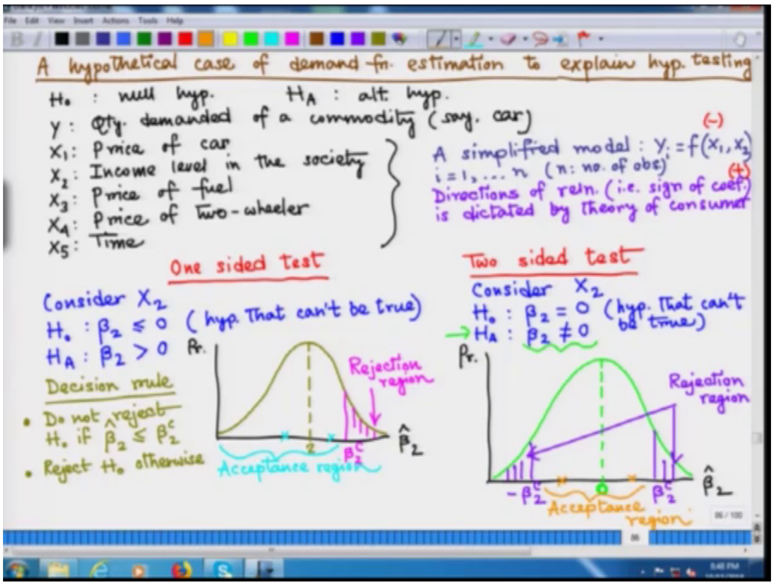
Microeconomics: Theory & Applications
Prof. Deep Mukherjee
Department of Economic Sciences
Indian Institute of Technology, Kanpur

Lecture – 61
Linear Regression (Part-4)

So let us start the discussion with hypothesis testing. So, hypothesis testing could be of two types: one sided test and two sided test. So, let us go through this individually one by one.

(Refer Slide Time: 00:28)



So, in the case of one sided test let me take the example of the variable X 2 which is income level and I assume that my microeconomic theory tells me that the relationship between Y and X 2 is positive right ok.

So, in this case we can write the null hypothesis for one sided test as beta 2 less than or equal to 0. So, this is basically cannot be true or the researcher does not believe right ok. So, what could be the alternative hypothesis then? So, it should be the exactly opposite and then in that case, we write beta 2 is positive right. So, that is basically the alternative hypothesis.

Now, let us move on to the two sided test and again we consider X 2 variable. So, in this case the alternative sorry the null hypothesis could be written as beta 2. So, beta 2 equals

0 and the alternative hypothesis could be written as beta 2 is not equal to 0 right. So, again here note the same thing these null hypothesis, which states beta 2 equal to 0 is basically hypothesis that cannot be true. So, as I stated previously I do not want to enter the first theory, statistical theory on hypothesis testing, but I would like to briefly discuss some decision rules to conduct hypothesis testing in practice.

So, let us note down some basic fundamental facts about the theory and practice of hypothesis testing in linear regression context. So, while we are testing a hypothesis, we have to first compute sample statistic from the sample data and then you know we have to find whether the null hypothesis is accepted or rejected depending upon the sample statistic magnitude. So, if you remember the previous discussion, there we have stated that beta hats are basically not one particular number it has some sampling distribution. As you change the sample you get new set of observations, you are expected to see a new value of beta hat the estimated value of the population parameter. So, there is a distribution and generally we assume normal distribution for the beta hats.

So, if that is the case then basically the way to proceed with this hypothesis testing is to first break these or divide this range of potential beta hats into two regions, one is called the acceptance region and the other one is rejection region. So, if we observe a statistic value which falls within the acceptance region, then we say that our null hypothesis is accepted. But of course, you can see here that we have different null hypothesis and alternative hypothesis depending upon whether we are conducting a one sided test or a two sided test. So, just to explain these things simply I would take aid of two diagrams, I will not get into statistical details that much.

So, first let me concentrate on the case of one sided test. So, here let me say that we measure different values of beta 2 hat along the horizontal axis and as there could be several possible values for beta 2 hat we expect sampling distribution and it follows normal distribution closely from the statistical theory. So, then basically we can say that suppose and this is arbitrary of course, suppose we say that the mean of this sampling distribution for beta 2 hat is found at some number some positive number say 2 fine. So, now, basically what we are saying? We are saying that we now divide this area under the curve which is basically a probability curve. So, let me complete the diagram the vertical axis, where I measure the probability.

So, now I say that my statistical theory suggests that there is some critical value of beta hat, which I need to figure out. So, if my beta hat critical value denoted by say beta 2 C; C stands for critical if my observed beta hat 2 falls in this shaded region; that means, that my beta 2 hat is greater than beta C 2 then I will reject my null hypothesis because I am too sure that beta hat 2 is indeed positive right. So, this is basically the rejection region and the rest area is known as the acceptance region. So, if I observe any particular value of beta 2 hat which say lies here or say here, then I can say that I fail to reject my null hypothesis and that is it. So, what would be my decision rule in this case, decision rule ok.

So, now let me move on with the case of two sided test. Here also I will take the help of a diagram to explain the concept of hypothesis testing; as usual I measure the possible values of beta 2 hat along the horizontal axis. And, the probability of these various possible values of beta 2 hat along the vertical axis. And, as there is a sampling distribution related to this variable beta 2 hat, random variable beta 2 hat I assume that you know there is normality.

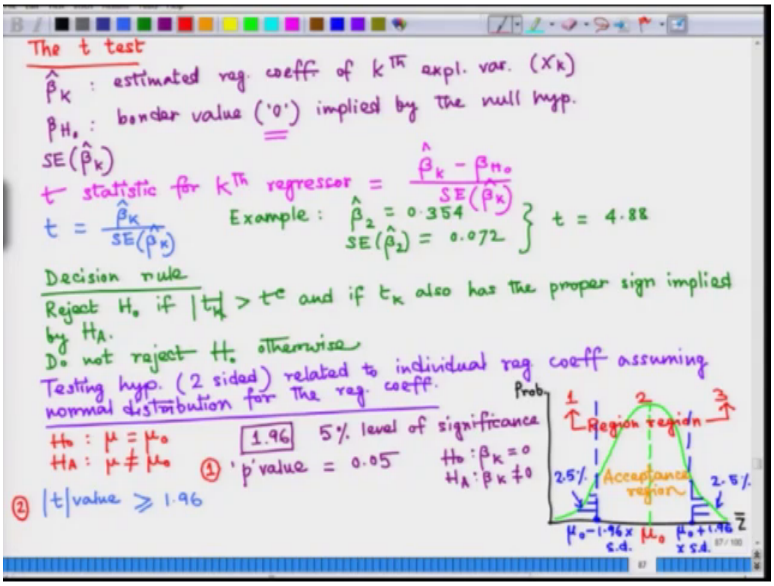
So, again I can assume a mean value. Now the mean could be anything, but note that here the problem is it will be different because now, you know if we concentrate on the alternative hypothesis, then the alternative hypothesis actually permits both positive and negative numbers for beta 2. So, in order to divide the range of beta 2 values we have to now think about two critical values. So, one on the one side of the distribution and one on the another side or the other tail of the distribution and let me call them beta C 2 and this should be minus of beta C 2 just the mirror image if we assume that we have 0 here as the mean.

So, now if we observe a value of beta hat two which is above the beta C 2 value, then we can say that I can safely reject my null hypothesis and that is why this region is known as rejection region ok. Now alternatively we can observe beta C 2 values such that that we are here in this region. So, basically we reject the null hypothesis here in this case also. So, there could be some interim values of beta hat 2 say right here or right there. So, in that case we can say that we fail to reject, it implies we accept the null hypothesis right. So, this is indeed my acceptance region right.

So, we have covered the basics of hypothesis testing, now let us talk about practice. So, in a regression context we are interested to know whether a particular variable whether it is x 1 or x 2 or it can be when the intercept term, whether they play some kind of significant role in explaining the variation in y. So, that could be judged by the hypothesis testing.

So, in other words we basically would like to say whether we can reject the null hypothesis that beta corresponding beta has a value equal to 0. A beta equal to 0 means, that this particular variable of interest actually fails to explain any variation in y. So, they are not statistically associated at all ok. So, the simplest test that we use in linear regression analysis is known as t test. So, t test is used to do hypothesis testing for individual regression coefficient. So, let us now study t test in somewhat detail.

(Refer Slide Time: 13:50)



Now, let me introduce some notations first. So, suppose we are working with multiple regression model. So, beta hat K is basically the estimated regression coefficient of kth explanatory variable it implies X k alright ok. Now, let us also introduce a new notation beta H naught. So, that is basically the border value, usually it is 0 which is implied by the null hypothesis, which is said to test microeconomic theory and finally, let me also introduce the estimated standard error for this estimate beta hat K fine.

So, now we define the concept which is a statistic which is to be computed from the sample the observed data and that is known as t statistic, and it is defined as beta hat K minus beta h naught divided by standard error of beta hat K fine. So, since most of the regression hypothesis tests whether a particular explanatory variables coefficient is significantly different from 0 or not, we can safely assume that the border value beta H naught is 0 right; we have already mentioned that here. So, the simplified form which is used in practice is basically this right.

So, this is basically the test statistic that we talk about while do hypothesis testing in linear regression context, in order to check whether a particular explanatory variable has some statistical association ship with the dependent variable or not. So, now, let us look at the critical value of this t statistic, which is used in practice to determine the outcome of a hypothesis testing. Before we set some ground rules or some decision rule, let me give some example here so, that you can understand. So, suppose you know I found the. So, suppose you know we go back to our demand function analysis exercise. So, let me give some example actually here with utilizing the demand function problem.

So, there beta 2, we were interested in to start our discussion that was basically the regression coefficient for the income variable I expected positive regression coefficient right. So, suppose after computing you know these estimators and we get some estimated values for these estimated this regression coefficients, and software can be also used to get these numbers. But suppose you know whatever way we have adopted we get some value beta 2 hat equal to say 0.354 some arbitrary number ok.

And we also can compute or statistical software can also get us this value, which is standard error of this beta 2, and that is some arbitrary number again say 0.072 something like that. So, now, basically I can utilize this information to find out the t statistic value, which is 4.88 ok. You do not have to worry about computing t value if you are using a statistical software. These days statistical softwares routinely compute these numbers and report ok.

So, now what would be the decision rule? So, note that here the kind of problem we are handling with t statistic, actually is related to the concept of two sided test because we are taking a border value 0 right. So, if that is the case then we have to have a new decision rule to be applied, when we test a single regression coefficient and let us now write down. So, this null hypothesis is basically associated with the two sided test; as it is two sided we have to use the mod value right of the test statistic t and it has to be greater than the critical value of the t statistic. So, this is for kth variable. So, let me put the superscript k here ok.

So, now let us talk about some general decision rule which is used in econometric applications, while conducting linear regression analysis. So, here we assume that our estimated beta hat values follow a normal distribution and then basically we can borrow some statistical theory, which talks about testing the population parameter in a normal distribution case. Or in otherwise we can test the parameter, we can test for a particular value of the parameter for a variable which follows normal distribution.

We assume that we are conducting a two sided test and let me measure that random variable, let me measure a sample statistic which is denoted by Z bar measured along the horizontal axis and as the sample statistic, actually follows a distribution you know we measure the probabilities for different values of the sample statistic Z bar along the vertical axis. So, here the sample statistic could be thought about you know the t value also. So, you can say that t value is a sample statistic which we are measuring along the horizontal axis.

So, now you know let us plot a normal distribution curve a bell shaped curve right and then we know some result from statistical theory that we are going to use now. So, let me assume the mean of the random variable is here, which is basically mu naught right. So, that is basically the null hypothesis value that we are testing right. So, we are testing against the alternative hypothesis which is stated by this

So, now there is a result that we can adopt, which states that there is some critical value which is given by mu naught minus 1.96 times the standard deviation of this random variable. So, let me say that this value is right here and the mirror image of this point on this line would be somewhere here and that would be mu naught plus 1.96.

So, then basically I can use these critical values which is supported by statistical theory, I am not going to statistical theory here I am just going to talk about the usefulness of that theory or that theoretical result. So, now, we can draw some perpendicular lines on these two values and let me denote these regions as 1 2 and 3. So, my regions 1 and 2 are known as rejection region, and my region number 2 is known as my acceptance region ok. So, that is why we know that these value 1.96.

Now, what is the use of this number 1.96? So, this is as of now is arbitrary, but let me tell you that if we apply this number then what happens. So, this area that I am shading currently gives 2.5 percent of the probability mass and as this second part of the rejection region is a mirror image, this also gives 2.5 percent of the probability mass. So, in total I can say that the rejection region actually has an area about 5 percent of the total probability and rest 95 percent is basically the acceptance region.

So, we now see that this number 1.96 is a magical number because then we know the area below the probability curve, and we say that 1.96 this number is basically associated with 5 percent level of significance. So, sometimes you also see in the statistical software report on regression analysis, they report p value. So, if we notice p value is equal to 0.05 that is basically is giving a threshold for 5 percent level of significance while, judging the regression coefficient against null hypothesis H naught equal to beta K equal to 0 against the alternative hypothesis it is not equal to 0 ok.

So, if p value is less than 0.01 then; that means, that probability is less than 1 percent which actually means, that the null hypothesis now would be rejected at 1 percent level ok. And if a p value takes if a p value lies between 0.01 and 0.05; that means, that the null hypothesis would now be rejected at 5 percent, but not at the 1 percent level. Alternatively we can also look at the t value which is also reported in statistical software report, after you conduct a regression analysis in it. And, if you find the more t value is greater than or equal to 1.96, then also you can say that you can reject the null hypothesis.

So, basically there are two different approaches we have studied, one is basically looking at the p value whether it is small enough or not and the second approach that we have studied is basically by looking at the t value we can also take a decision. There is a close relationship between the t value and the p value. So, if we observe a t value higher than 1.96, then basically we can certainly say that p value is less than 0.05 and if we observe a t value less than 1.96, then we can say that our p value is higher than 0.05.

So, in practical work that is why the applied econometrics or statisticians put enormous emphasis on these two values right. So, by looking at these numbers now you can do testing hypothesis testing for individual regression coefficients, in order to check whether they are statistically different from 0 or not.

So, this brings an end to our discussion on regression analysis.