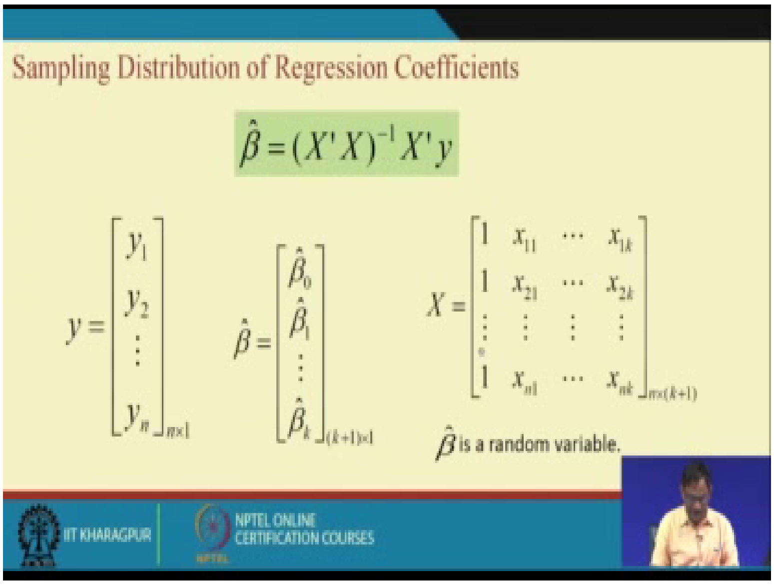
Design and Analysis of Experiments
Prof. Jhareswar Maiti
Department of Industrial and Systems Engineering
Indian Institute of Technology, Kharagpur

Lecture – 21
Sampling Distribution of Regression Co-efficients

(Refer Slide Time: 00:29)

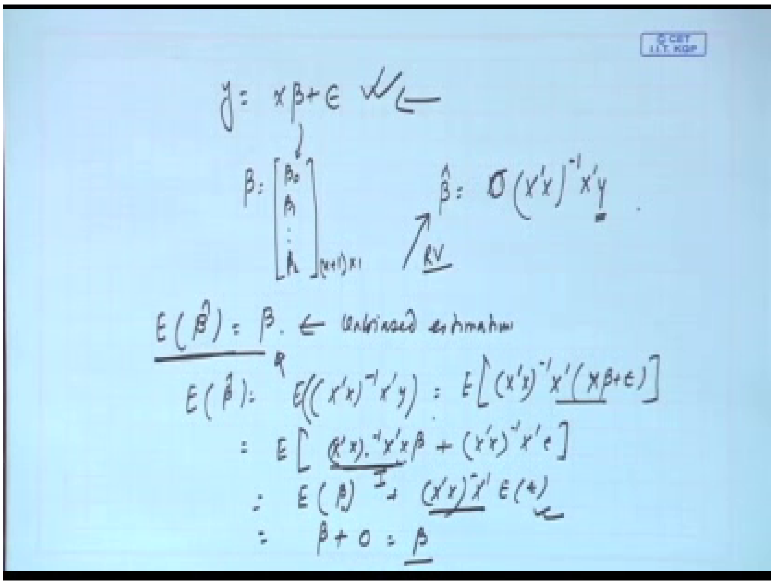
Welcome, today we will continue with linear regression. Today's topic is sampling distribution of regression coefficients. So, what we will cover in this lecture sampling distribution of regression coefficient, test on individual regression coefficient and group of coefficients, confidence interval on the individual regression coefficients.

(Refer Slide Time: 00:48)



So, let us now start with the regression equation.

(Refer Slide Time: 00:54)



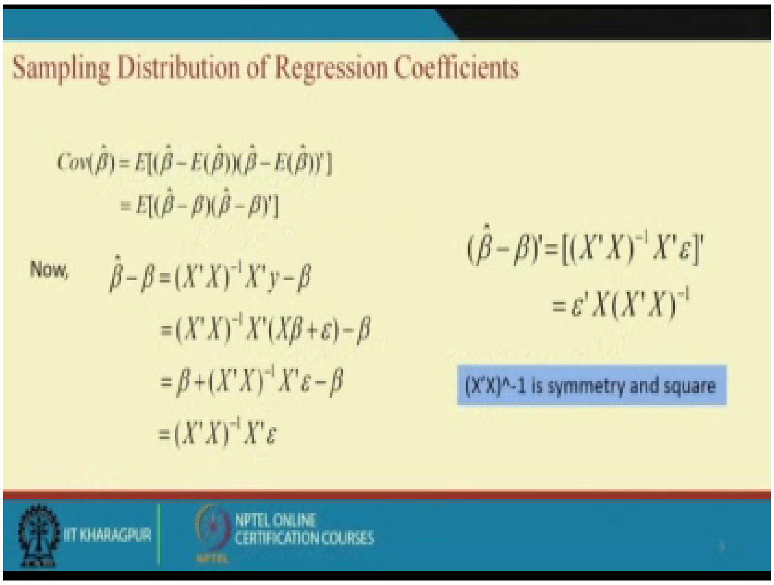
You have already seen that the linear regression equation is y equal to x beta plus epsilon, where beta we have k plus 1 beta values like beta 0 beta 1 to beta k plus 1 into 1. And x is the design matrix epsilon is the error terms. And you have also seen that the estimate of beta is x transpose x transpose x inverse x transpose y that also you have seen where x is the design matrix, so y is data matrix and this part we have completed in last class.

So, if you see the slide, you see that there are n plus 1 observation for y, beta is k plus 1 into one create the parameters of regression and x is the design matrix, where the first column which all having one value. And then these are the data on k number of for various factors otherwise I can say k number of predictor variables x 1 to x k. Essentially x is n into k plus 1 matrix. And from sampling, we know the sampling theory we know that the estimate beta cap this is a random variable. So, it has it expected value and variance component also suppose we want to know; what is the expected value of beta cap. It will be beta and that is; what is unbiased estimation means beta cap is the estimate of beta, beta is the regression coefficient vector, which is from the population point of view. And beta cap is the regression coefficient the estimate of this vector from the sample data.

So, it can be proved also that the expected value of beta cap is beta. So, how can you do, you can write that expected value of beta cap equal to expected value of x transpose x inverse x transpose y this. Then you can write expected value of x transpose x inverse x transpose then y value is nothing but x beta plus epsilon, you are getting from here. And if you multiplied this then you get x transpose x inverse x transpose x beta x transpose x inverse x transpose into x beta plus x transpose x inverse x transpose epsilon.

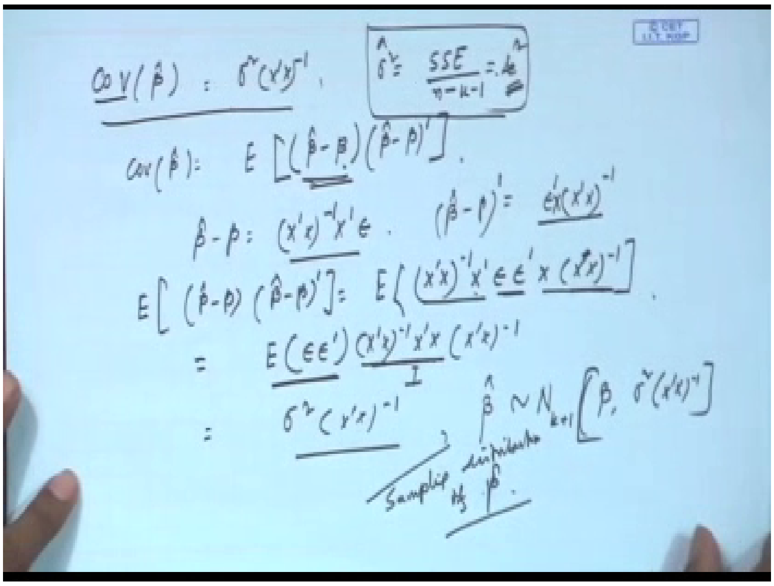
Now, x transpose x inverse x transpose x will be identity matrix I. So, this one will be expected value of beta plus x transpose x inverse x transpose expected value of epsilon because this is coming out of expected value this is the fixed values. So, beta being a constant parameter value, so it will be beta and expected value of this error term by assumption is 0. So, this will be beta.

(Refer Slide Time: 04:53)



So, you may be then what is the variance of beta cap other way we can say covariance of beta cap.

(Refer Slide Time: 04:57)

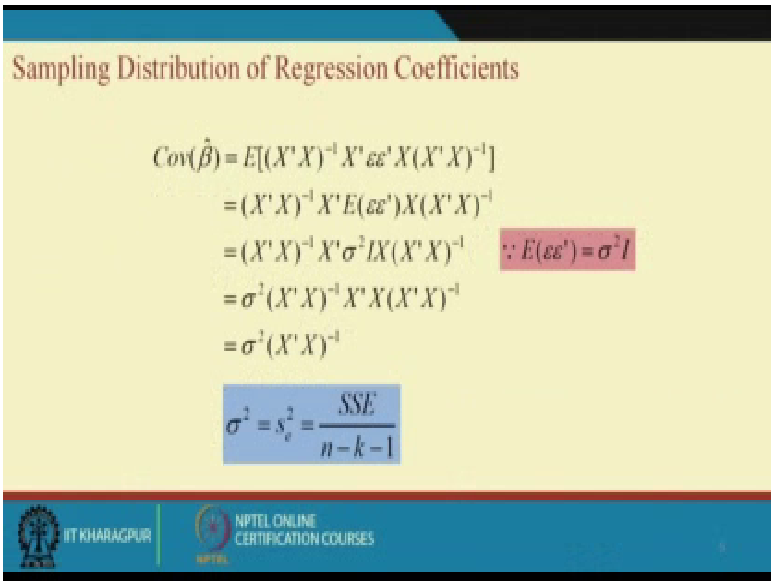


It can be proved there this will be sigma square x transpose x inverse where r sigma square is the sigma square is the SSE estimate of sigma square will be SSE minus n minus k minus 1. You already seen what is SSE sum square error. So, we will discuss about this SSE later on also. So, other way we can write how did it, it can be also proved. So, covariance of beta cap is nothing but expected this is also expected value of your beta cap beta cap minus beta, beta is the expected value of beta cap this into beta cap minus beta transpose.

So, you can see from the slide that that we can prove that beta cap minus beta. So, you are writing this beta cap is x transpose x inverse x transpose y. Now, you are putting the same way when we calculate mean value and that this one and then you found beta plus of this and beta and this beta is cancelled out. So, you are getting you are getting beta cap minus beta is x transpose x inverse x transpose epsilon. Here you assume x transpose this is basically beta, and this one is this now, this beta minus beta is canceled out and this remains.

So, now what will be then that beta cap minus beta transpose, this will be transpose of this, and this quantity will be this transpose x transpose x inverse, and here x will come. This x will come x transpose, transpose is x, epsilon transpose is there x transpose x minus inverse one this because this is a symmetric matrix. So, then your expected value of beta cap minus beta and beta cap minus beta transpose this is nothing but expected value of x transpose x inverse x transpose epsilon epsilon transpose x x x transpose x inverse x x transpose x inverse.

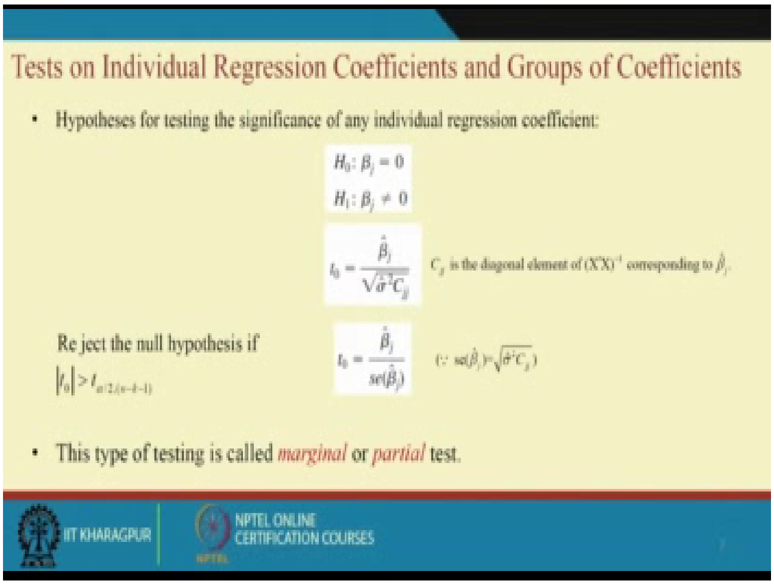
(Refer Slide Time: 07:42)



So, this is a fixed one, this one is also fixed quantity. So, if you taken out all those things then ultimately you will be getting expected value of epsilon epsilon transpose into this x transpose x inverse x transpose x into x transpose x inverse. Now, this quantity become I.

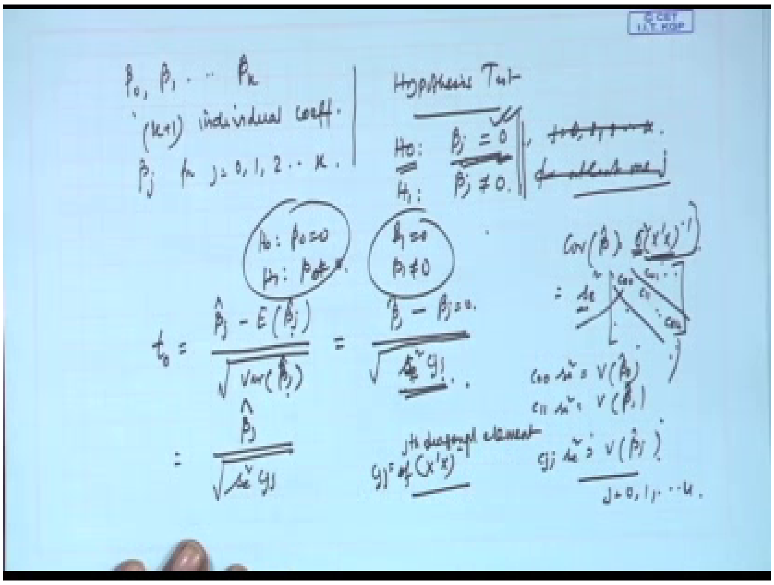
This quantity is what sigma square; this quantity becomes sigma square. So, sigma square x transpose x inverse, so that means, essentially we got beta cap which will be multivariate normal basically whether there are so many. And it will be N there are k n plus 1 number of beta values and then it expected value is beta and covariance matrix is sigma square x transpose x inverse. So, this is known as sampling distribution of beta cap. And I already told you that the estimate of sigma square will be this which is nothing but A C square coming out of the error terms.

(Refer Slide Time: 09:24)



Now, let us see that from sampling distribution two that test of individual regression coefficients. Here took over the concept is individual regression coefficient and group of coefficients. First we will discuss on individual regression coefficients.

(Refer Slide Time: 09:47)



How many coefficients we have we have beta 0, beta 1 to beta k that is k plus 1 individual coefficients. We want to test every coefficient whether they are significant or not that means, we want to test beta j for j equal to 0, 1, 2, to k here test means hypothesis test, test means hypothesis test. So, you have seen earlier that hypothesis test mean there will be null hypothesis, there will be alternate hypothesis. Our null hypothesis is beta j equal to 0, j equal to 1 to we have put 0 to 1, 2 like k; and beta j not equal to 0 is the alternate hypothesis and obviously, for at least 1 j for at least 1 j, but here we are interested to test individual parameter.

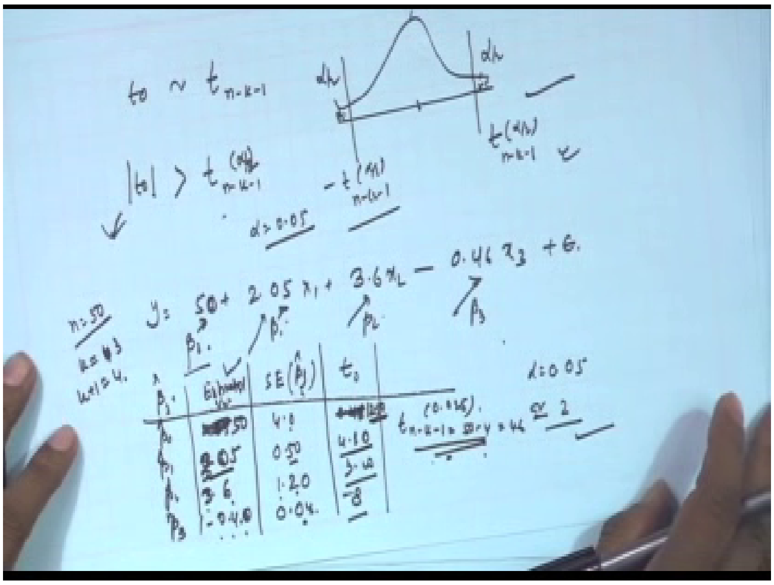
So, this is in total by hypothesis tests for totality that whether this part in under model adequacy test I will discuss again, but in this case suppose if I consider only the j th regression parameter, then this is not applicable, suppose the beta 0. So, we want to test beta 0 then H 0 is beta 0 equal to 0 H 1 is beta 1 not equal beta 0 not equal to 0. Similarly, beta 1 equal to 0, beta 1 not equal to 0, like this. So, individually we are testing here.

So, general term H 0 beta j equal to 0; and beta j not equal to 0 what we would test statistics we will use here we will use a statistics called t statistics here. Suppose we create a statistics like this which is beta j this is a random variable estimate value minus expected value of beta j cap divided by square root of variance of beta j cap. This is t distributed, because of because you have seen earlier that this from standard from central limit theorem, this can be z distributed also provided sample size is large. And as and the variance of beta j these are all estimated one exact not known. So, we will go for the to t distribution. So, this is tested this is true when H 0 is true means beta j beta j equal to 0 or H 0 is true as a result, we say this is t 0. So, this t 0 this follows t distribution when H 0 is true.

Now, this can be written like this beta j cap minus expected value of beta j cap already you have seen this is nothing but beta j. Now, variance part we have already seen that sigma square we have seen that covariance of beta cap equal to sigma square x transpose x inverse. So, this can be written like se square if we assume that the estimate is se square and into this suppose that C 0 0, C 1 1 to C k k and this side also that is C 0 1 like this the off diagonal elements are the variance component. So, this C 0 0 all those things what I am saying this is the component of x transpose x inverse.

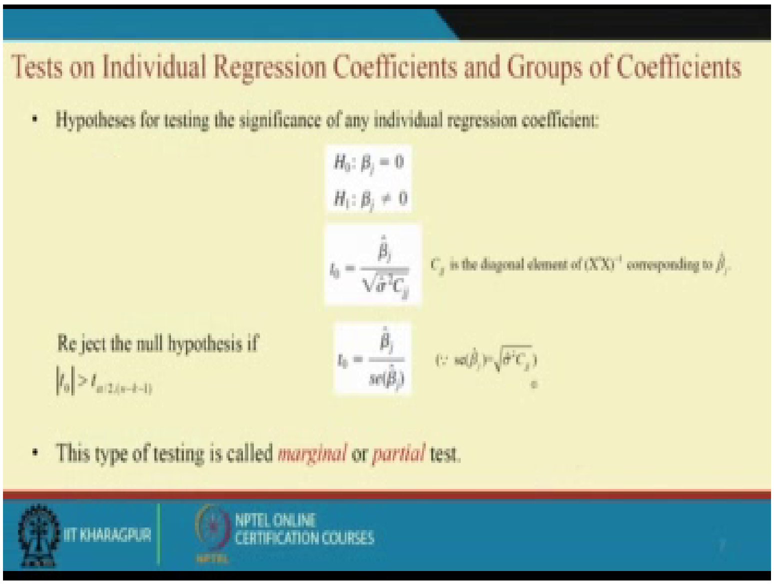
So, C 0 0 sc square C 0 0 se square will be that variance of beta 0 cap. Similarly, C 1 1 se square will be variance of beta 1 cap like the C j j se square will be variance of beta j cap; and obviously, j stands from 0 to k. So, what we like then and then that variance of beta j cap is sigma 0 sigma square in to C j j. Now, the sigma square estimate is se square. So, we are writing this as se square C j j. So, and under H 0 beta j equal to 0, so this quantity becomes 0. So, t 0 become beta j cap divided by square root of se square C j j. Please keep in mind C j j is the jth diagonal element of x transpose x inverse that is C j j j th diagonal element of x transpose x inverse.

(Refer Slide Time: 15:40)



Now, what do you will do you all know that hypothesis testing procedure. So, you calculate t 0, this follows t distribution with certain degrees of freedom that degrees of freedom here is n minus k minus 1 degrees of freedom, because this is the error degrees of freedom n minus k minus 1. So, if you create a t distribution curve like this and suppose that your it should be a two-tailed test, this side alpha by 2, this side alpha by 2, then this is t alpha by 2 n minus k minus 1 and this is minus t alpha by 2 n minus k minus 1. So, if the absolute value of t 0 is greater than t n minus k minus 1 alpha, alpha usually will be 0.05 then we can say the corresponding beta j parameter is significant; it is not 0, it is not 0, it is far away from 0. So, this is the test, this test is also known as you just see, so sorry I have done one mistake that is alpha by 2, alpha by 2, because it is a two-tailed test. Please remember when it is two-tailed alpha by 2, when it is one tailed alpha.

(Refer Slide Time: 17:10)

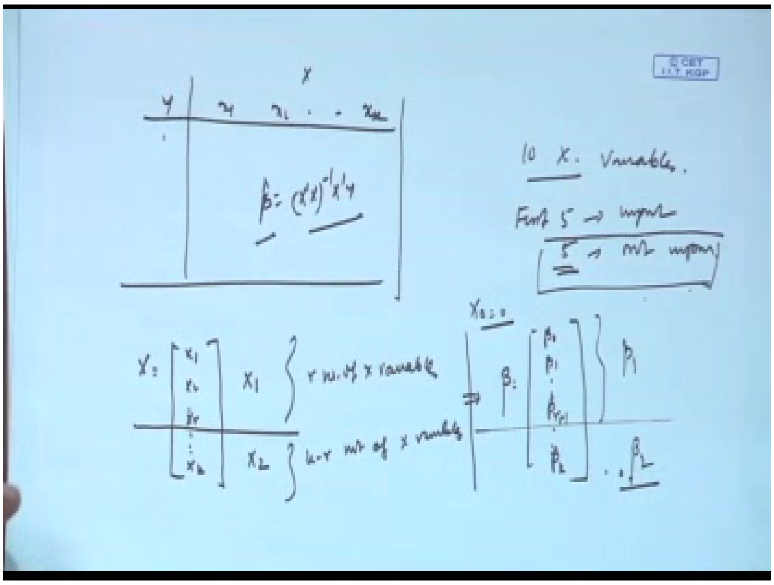


So, see the slide also, so this type of testing is called marginal or partial test, individual regression parameter is tested here. Suppose, I say that y equal to 50 plus 2.05 x 1 plus 3.6 x 2 then maybe minus 0.46 x 3 plus error term then what we are saying beta 0 equal to this, beta 1 equal to this, beta 2 equal to this, beta 3 equal to this. So, by this test we are saying that whether this beta 0 50 is essentially 50 or it is basically nothing more than 0. So, as a result what you are doing you are first finding out the beta values suppose beta 0, beta 1, beta 2, beta 3 here this estimated values these values are nothing but here 2.05, 3.6 minus beta 1 beta 0 beta 1 beta 2, so this is 2.05, 3.6, minus 0.46, these are the estimate value estimated value.

Now, you are finding out the standard error of beta j, let us give it as a cap. So, suppose without when I am doing any calculation suppose we know that the standard error is here 4.0, here 0.50, here 1.20 and here may be 0.04. Then what will be the t value t 0 value, t 0 value will be estimate divided by standard error, so 2.05 by 4 it is almost 0.48 kind of thing then two point oh sorry the beta 0 is 50, I am extremely sorry 50 by 4 50 by 4 minutes it is 12.50. 2.05 by 0.5 minutes 4.10; 3.6 by 1.2 means 12 into 3, it is almost 3; and 0.40 by this I think this is your 100 you multiply then it is 4, and if you 40 then it is almost 80 and this is minus 80.

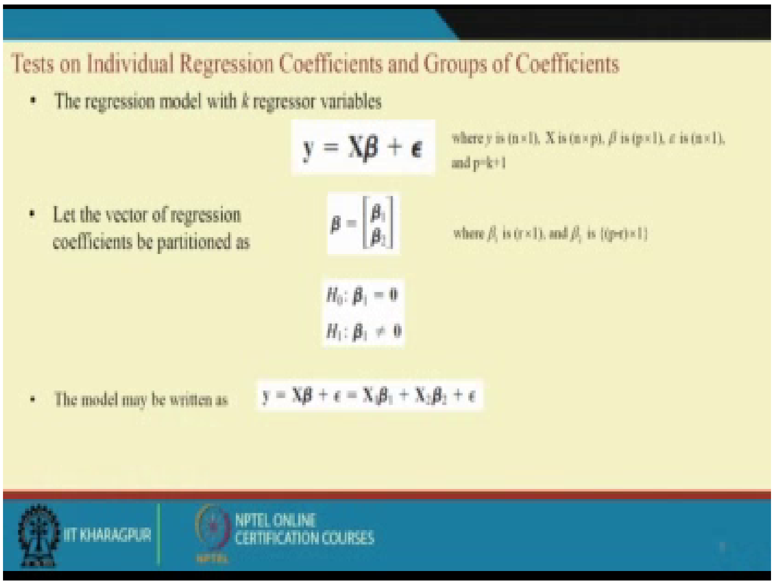
Now, if you compare suppose your alpha is 0.05, then you compare t, suppose your sample size in this example n equal to let it be 50. And what is the k value, k value is that means, 1 to 3 k plus 1 is 4 k equal to 3 k plus 1 is 4, so that mean then n minus k minus 1 this value will be 50 minus 4 that is 46 and alpha by 2 equal to 0.025. So, find out this value from the statistical table. If this value usually I think this value will be may be around 2, this value will be around 2, so see this is 12.50, this is 4.80, this is 3, this is minus 8. So, absolute value of the all these are more than two that mean all the parameters are significant individually significant that is what is the individual parameter test, so that means what I say you have data you have data set y.

(Refer Slide Time: 21:05)



You have data set y, you have x. So, x 1, x 2 to x k, and then here y 1 to y in this data you have. From there using beta cap equal to x transpose x inverse x transpose y calculated these, you calculated the variance and all those things. And then from there you have got the equation this is suppose you got this equation and this type of the step up called computation we have and then this is the result that all the regression coefficients are statistically significant, this is known as individual or marginal test.

(Refer Slide Time: 21:51)



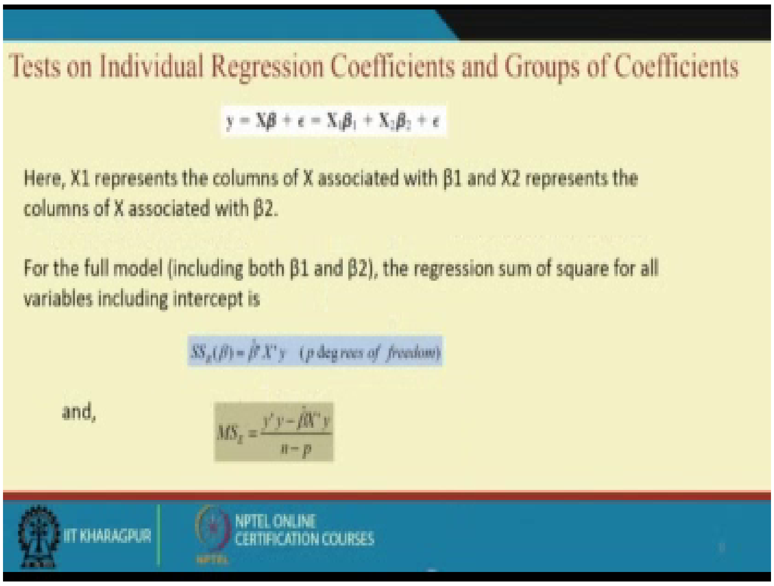
But many a times what happened suppose you have a large number of data variables, suppose you do not want to test individual parameter, you want to test subset of variables. Suppose, I have 10 x variables, 10 x variables. So, maybe I want to test that the first 5 a sub I have arranged them based on certain knowledge that the first five are important and last five not important, but I want to be WCO that yes last five is not contributing.

So, in that case what happened you want to test all those five regression coefficients simultaneously and then accordingly reject the null hypothesis that this set of variables are not contributing or they are contributing. So, essentially the mathematics here is when I have k regressor variables this is my model, and these are the things what is known to you. And then I have dividing into two parts one is that beta, beta 1 and beta 2 where beta 1 is r cross 1, r where number of variables and as I have p number of parameters to be estimated where p is equal to k plus 1. Then p minus r cross 1 this number of variables we are partitioning that means, what you are doing you are partitioning x which is basically x 1 x 2 to x r to x k, you are making two partition here.

This side you are saying x 1 and this one using in here r number of x variables and here then what happened total is k minus r number of x variables are there. Other way actually, but please remember there is there is x 0 also, x 0 which assumes zero all the time. So, this partition from the beta point of view we can write just one will be added only suppose beta is beta 0, beta 1 to beta r to beta k. So, we are taking first r cross 1 and then if it is r cross 1 then this let it be r minus 1, and then remaining p minus r, where is p is equal to k plus 1 minus r, so these two now this side is capital beta 1, this is capital beta 2, so that is what is denoted here.

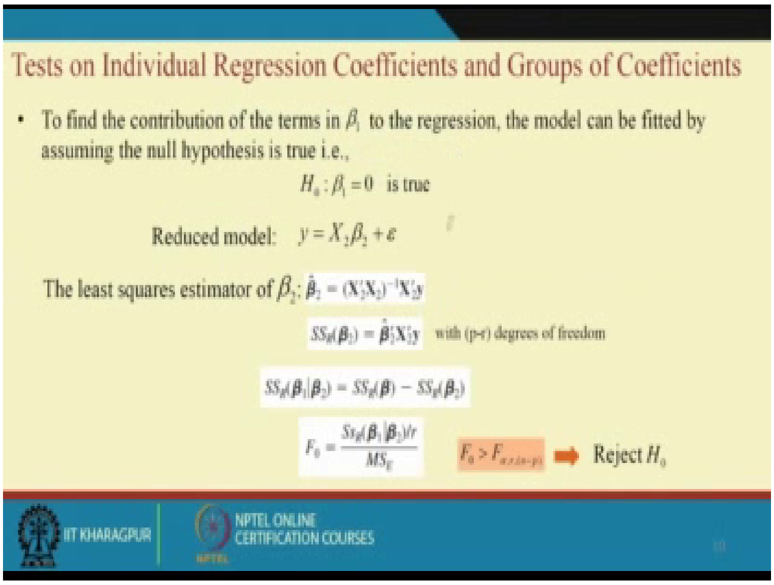
Here you see beta equal to beta 1 beta 2 like this. Now, your test is that this r number variables or r number of degrees and coefficients here considering x 0 I am talking telling r number of regression coefficients are not significant, then H 0 is beta 1 equal to 0 and obviously, alternate hypothesis will be H 1 beta 1 not equal to 0. So, then what you can do you can write down the regression equation in this fashion y equal to x beta plus epsilon is equal to nothing but x 1 beta 1 plus x 2 beta 2 because see this partition will give you this.

(Refer Slide Time: 25:52)



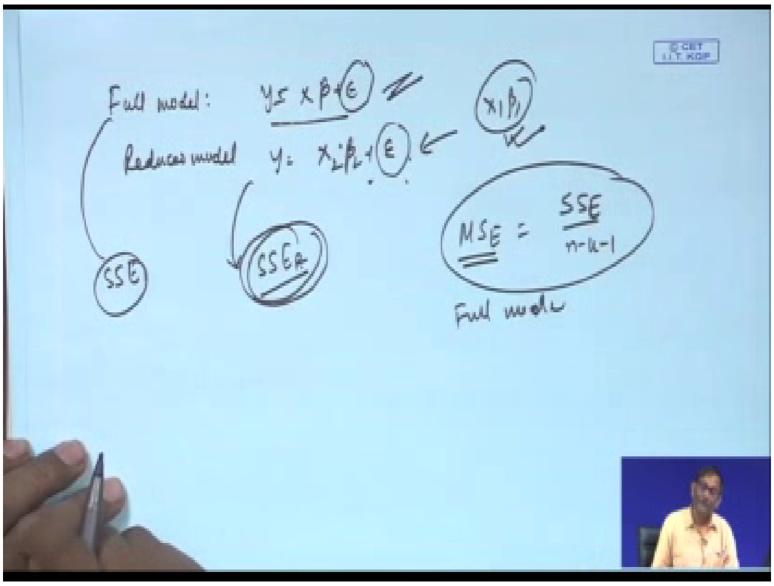
So, as a result what happened, as a result what happened if H 0 is true, suppose H 0 is true then this part become 0; and y will become reduced model will be y equal to x 2 beta 2 plus epsilon. So, that is what is written here reduced model is y equal to x 2 beta 2 plus epsilon.

(Refer Slide Time: 26:13)



So, essentially what happened then we have a full model full model which is y equal to x beta plus epsilon. And we have reduced model where we are writing x 2 beta 2 plus e like this.

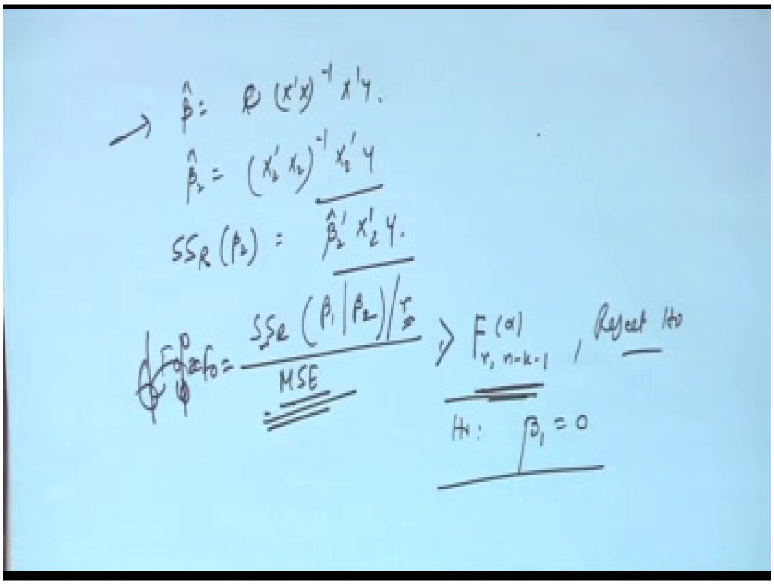
(Refer Slide Time: 26:24)



So, ultimately here in this model as x 1 beta 1 term is not included so that means error term here will be more than error term here, so that different that means, some using this model, what you will you will get SSE or suppose using this model you will get SSE reduced model. Then what will happen this SSE reduce amount will be more than this and the difference is contribution of these to the error and if that 1 is significantly contributing then H 0 will be rejected.

So, what is being done here then first we calculate MSE what is mean square error. And this will be nothing but your SSE divided by n minus k minus 1, and it will be using full model that means, the first model. Now, you have use the second model reduced model and what you are getting you see you are getting beta 2 cap x transpose x inverse x transpose y here x 2 is there only.

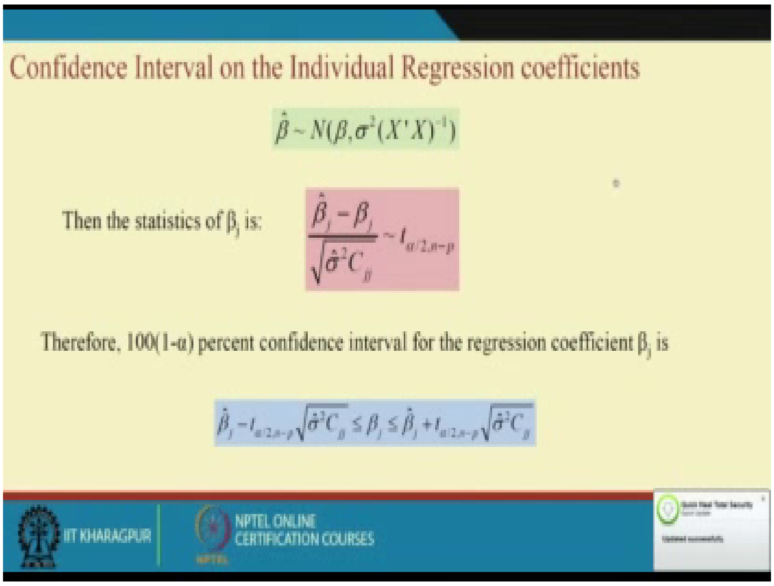
(Refer Slide Time: 28:15)



So, go to slide and see the slide. So, beta cap equal to x trans x transpose x inverse x transpose y in the full model. Now, reduced model beta 2 cap will be x 2 transpose x 2 inverse x 2 stands for y. So, then we will get the SSR beta 2 that one will be beta 2 cap transpose x 2 transpose y, from the previous model we have seen beta test. So, ultimately, so the slides hello, so beta 2 is x 2 this and SSR beta 2 is this. And obviously here with p minus r degrees of freedom, because you have not taken r number of curve this is this one total p r is taken out. And then what is the SSR of beta 1 that is sum square regression beta 1 given beta 2 this will be the full SSR minus the SSR explained by this reduced model, so SSR beta minus SSR beta 2.

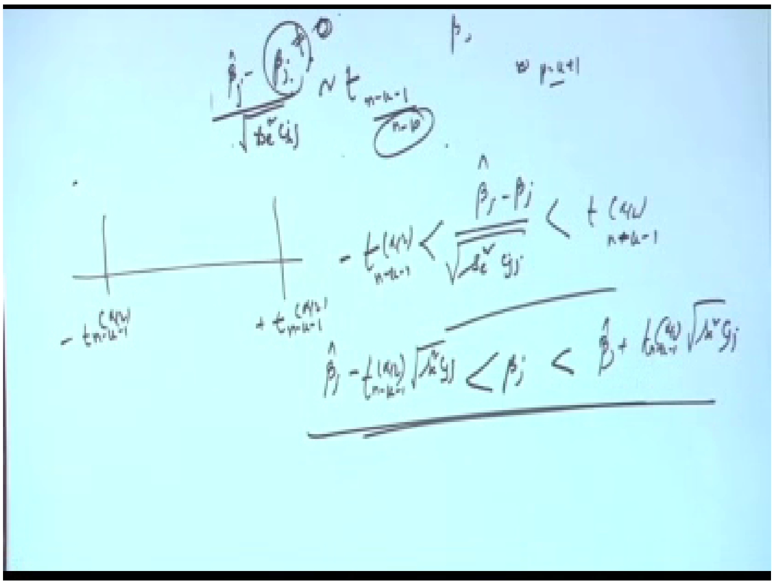
If this one is significant then we will reject H 0 that beta 1 equal to 0 will be rejected this beta 1 equal to 0 will be rejected. So, you are creating f statistics here by saying that F 0 equal to SSR beta 1 given beta 2 divided by degrees of freedom and then your image. So, what will be its degrees of freedom if this one or I can say mod of these I think this it will be always positives. So, if this is greater than f what is the degree of freedom numerator r degrees of freedom in denominator n minus k minus 1 or n minus p and if we create a alpha single significance level, if this is the case that means, F 0 equal to this equal to this greater than this reject H 0. That means, we have considered that big beta 1 equal to 0 that is not correct or I mean those variables are contributing.

(Refer Slide Time: 31:16)



So, I will just brief you a little bit more here related to the individual regression coefficient. I have given you the hypothesis test, but the confidence interval is another important concept.

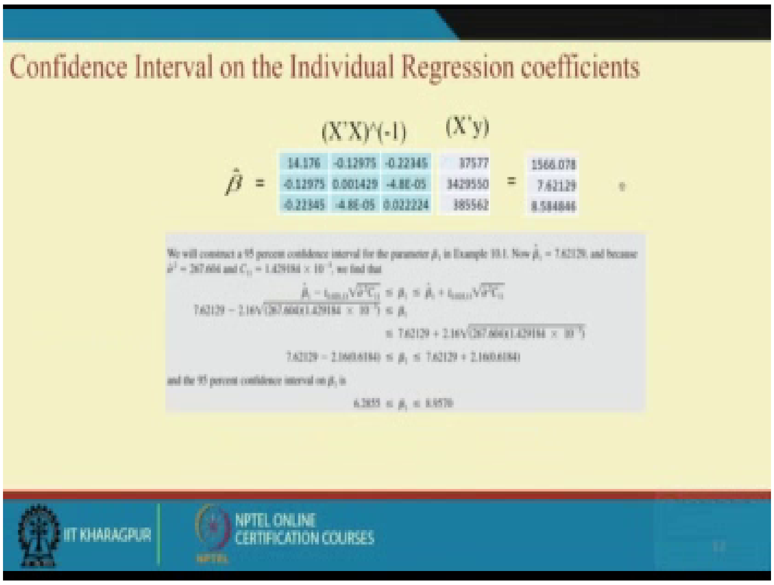
(Refer Slide Time: 31:39)



So, we have seen that that beta cap or you cannot beat a j cap minus beta j by variance that is se square C jj this follows t distribution with n minus k minus 1 degrees of freedom where in k p equal to k plus 1. So, you can write it n minus p also, there is your freedom. And then I have already told you that this is plus t n minus k minus 1 into alpha by 2 and this side will be minus t n minus k minus 1 alpha by 2. Then this quantity will be in between this. Again under H 0 beta j equal to 0, so beta j cap by root over se square C jj this will be a lying in between in between this t minus alpha by 2 n minus k minus 1 t alpha by 2 n minus k minus 1.

So, beta j cap now we want the interval for beta j, so we will not put zero this will not put zero we will keep as it is under H 0 H 0 that is why test we have done. So, we will write this minus beta j understood. We want interval for beta j, then this quantity will become beta j cap minus tn minus k minus 1 alpha by 2 root over se square c jj. And this will be beta j cap plus tn minus k minus 1 alpha by 2 and root over of se square c jj where c jj is the j th diagonal element of the matrix x transpose x inverse.

(Refer Slide Time: 33:48)



So, with this I will show you one example suppose beta is like this from some data you got beta is like this. We will construct a 95 percent confidence interval for the parameter beta 1 in example this is basically a example from a book. And now beta 1 is seven point this and because and sigma square is this, this was also calculated earlier. And see that the C 1 1 that is the x transpose x inverse that this is this. So, we find that that means, the beta 1 cap minus the 2.16 into this one beta 1 cap minus t 0.025113 degrees of freedom. So, 3 plus 1 – 4, 2 plus 1 - 3; that means, 16 minus 3 - the 13 degrees; that means, in sample size is 16 here and this follows like this.

So, beta 1 cap you are writing like 7.621 to 9 t 0 point this is in 2.16. Sigma square is this two sixty point this, C 11 from the x transpose x inverse that with that particular part into 10 to the power minus 3 less than equal to this less than equal to this value and finally, you are getting this type of confidence interval. Now, see that the in the confidence interval there is no if this is positive to positive one sided there is no zero in between. So, beta 1 is statistically significant means it is not 0. And from the hypothesis testing using t value also, you will find that you will reject the null hypothesis that beta 1 equal to 0. So, thank you very much and it is again taken from this Montgomery book, and also my earlier lecture NPTEL lecture on Applied Multivariate Statistical Modeling.

Thank you, I hope that you have understood the concepts.