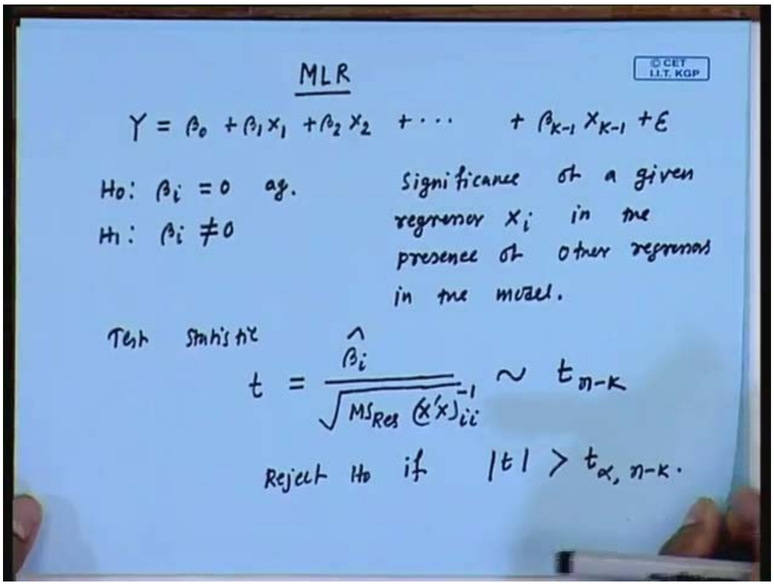
Regression Analysis
Prof. Soumen Maity
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture - 12
Selecting the BEST Regression Model (Contd.)

This is my lecture on Selecting Best Regression Model. In the previous lecture we have learned how all possible model involving at most k minus 1 regressors is evaluated, using according's to some criteria and the best model has been selected well. So, today we will be discussing on sequential selection, so here instead of you know evaluating all possible subset models, we I mean we do not evaluate all possible subset regression models here. Here the best model is found by adding all or removing one regression in each step. So, there are three basic three algorithms of these type no sequential selection the algorithms are backward elimination, and next one is forward selection, and the third one is you know stepwise selection. So, before I start talking on a sequential selection I want to recall partial test, which we discussed in multiple linear regression setup.

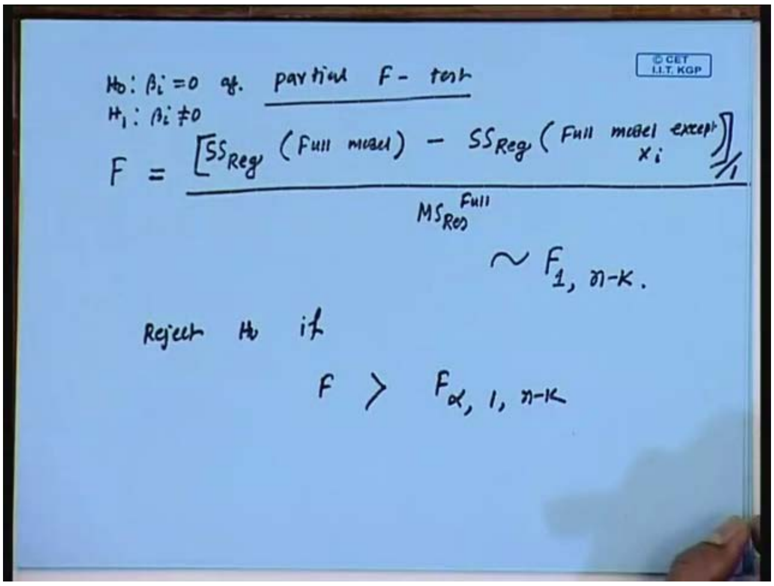
(Refer Slide Time: 03:02)



So, by considering the multiple linear regression setup and we have a model, with k minus 1 regressions. So, y equal to beta naught plus beta 1 x 1 plus beta 2 x 2 plus beta k minus 1 x k minus 1 plus epsilon. And by testing this hypothesis like H naught which says say beta i equal to 0 against the alternative hypothesis H 1 which is beta i not equal to 0. It is the meaning of this hypothesis is that basically by testing this hypothesis we test, the significance of a given regressor say x i in the presence of other regressors in the model.

So, we know how to test this hypothesis we can either use the t statistic to test this hypothesis or also we can, use the extra sum of square technique to test this hypothesis. Well just let me recall that thing, the first one is we can test this hypothesis using the test statistic t, which is equal to beta i hat by MS residual x time x inverse i i. So, this the i'ith element of x and x inverse, and this follows we know that this follows t distribution with degree of freedom n minus k. And we reject the knot hypothesis, we reject H naught if mod t is greater than t alpha n minus k. So, this is the test using t test statistic, now this can be also tested using the f statistic that is the extra sum square approach.

(Refer Slide Time: 07:01)

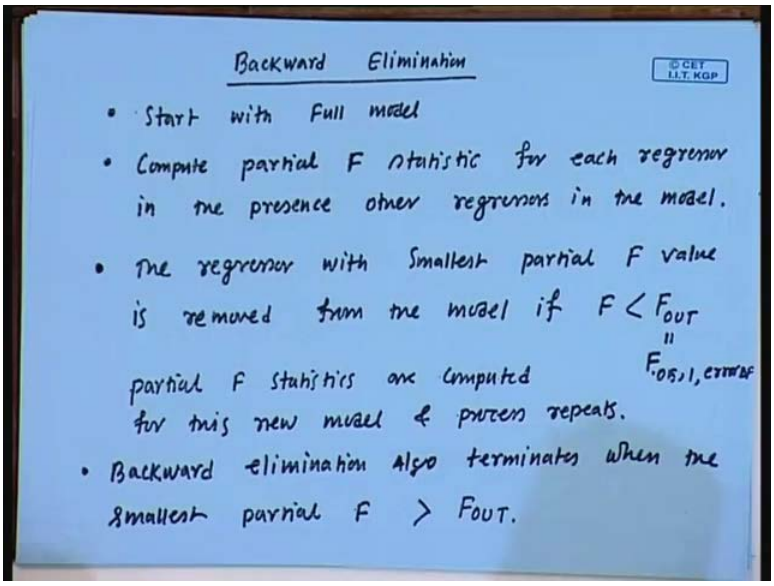


Let me talk about that, let us call we will say it partial f test and we will be using this partial f test repeatedly today. So, what we do is that the test that is tic for to test H naught beta i equal to 0 against the alternative hypothesis H 1 that beta i naught t equal to 0; the test statistic is F which is equal to SS regression for the full model minus SS regression for the full model except x i. So, here you have all the regressors x 1 to x k minus 1, here you have all the regressors except the regressor x i.

So, this difference will give you the extra sum of square due to the regressor x i or more precisely we can say that this difference is the extra regression sum of square due to the regressor x i well. So, this has degree of freedom once we divide it by 1 and by MS residual that is for the full model, this follows F distribution with degree of freedom 1 and n minus k. And here this is called the partial F test statistic and we reject H naught, if this F value is small sorry we reject H naught if F is greater than the tabulated value F alpha 1 n minus k. So, rejecting H naught means, accepting this one; that means, x i the i'th regressor has significant contribution to explain the variability in y, in the presence of other regressors in the model.

So, if this difference is large; that means, the i'th regressor has significant contribution. So, if F is large the difference is large then F is large, F is large then we are going to reject the now hypothesis and accept the alternate hypothesis, so this is all things we talked before also.

(Refer Slide Time: 10:40)



Let me talk about the first technique that is called backward elimination, so it says that the basic idea behind the backward elimination is that, see we are looking for the best subset regression model. The basic idea is that, you start from the full model and then in each step you reject the regressor which is less significant, so you if it is a problem four regression variables, then you start with the model which is having four regressors. And then the first step, you remove the regressor which is less significant to explain the variability in the response variable.

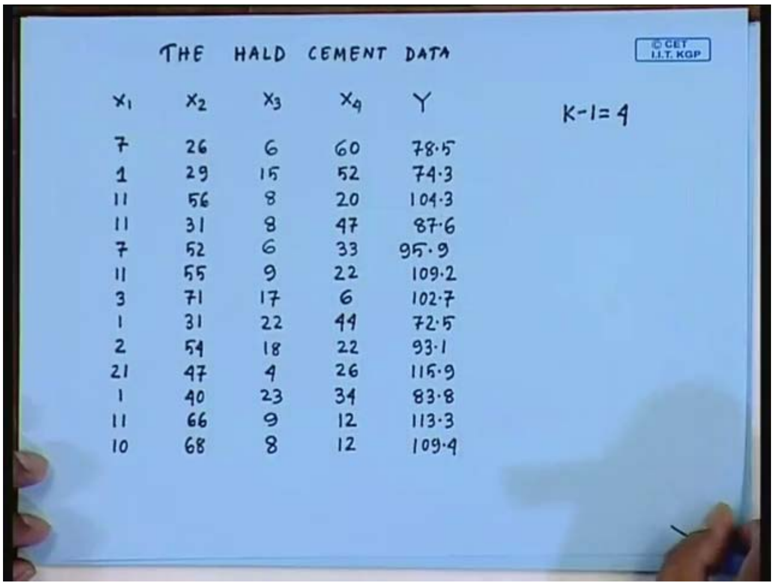
So, this is the basic idea behind the backward elimination, and there is some stopping criteria of course, we will be talking about that well. So, it says that you start with the your first thing is that you start with full model, and then you compute partial F statistics statistic for each regressor in the presence of other regressor in the model. I will explain all these things, let me just write down the algorithm first, next is see you are computing the partial F statistic for each regressor in the presence of other regressor.

So, if the F value is very small for some particular regressor; that means, that is not significant. So, we look for the smallest partial F value the regressor with smallest partial F value is removed from the model, if the value is less than F out well sometime this F out is p specified or this F out is equal to the tabulated value of F 051 and this error degree of freedom right. And then partial F statistic computed for this new model, by new model I mean the model which has been obtained by removing the first regressor for this new model and the process repeats.

And here is the stopping criteria, backward elimination algorithm terminates when the smallest partial F value is greater than F out. So, let me just explain little bit and then illustrate this algorithm using one example, we start with the full model and then here we compute the partial F statistics. So, this partial F statistics say for example, for the random for the regressor x i that will give the partial F statistics associated with the regressor x i, will give you the significant or significance of the regressor x i in the presence of others regressors.

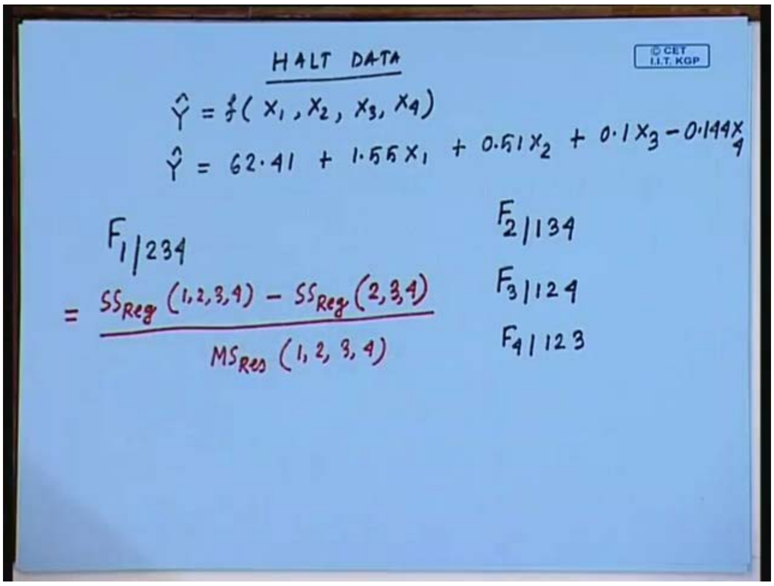
So, this way we compute the partial F statistics for all the regressors, so lower the value of F statistic indicates that the associated regressor is less significant. So, next we look at the smallest partial F value, and if the partial F value is you know less that F out or less than the tabulated value then of course, we reject the we remove that regressor from the model. And again we repeat the same process, until there is no partial or there is until a step or stage where the smallest partial F statistic is greater than the F out. That means, the smallest one is also significant, so we cannot remove anymore regressor from the model well. Let me explain illustrate this backward elimination using one example.

(Refer Slide Time: 18:32)



So, I will considering the same data that is the HALD cement data, so we have four regressors and one response variable, and to we will see which one is the best subset model according to the backward elimination technique well.

(Refer Slide Time: 19:02)



So, what you do is that we are considering the HALT cement data, so we start with the full model. So, we will first fit Y as a function of all the four regressor variable X 1, X 2, X 3, X 4; that means, you fit the full multiple linear regression model, and the fitted model is Y hat which is equal to 62.41 plus 1.55 X 1 plus 0.51 X 2 plus 0.1 X 3 minus

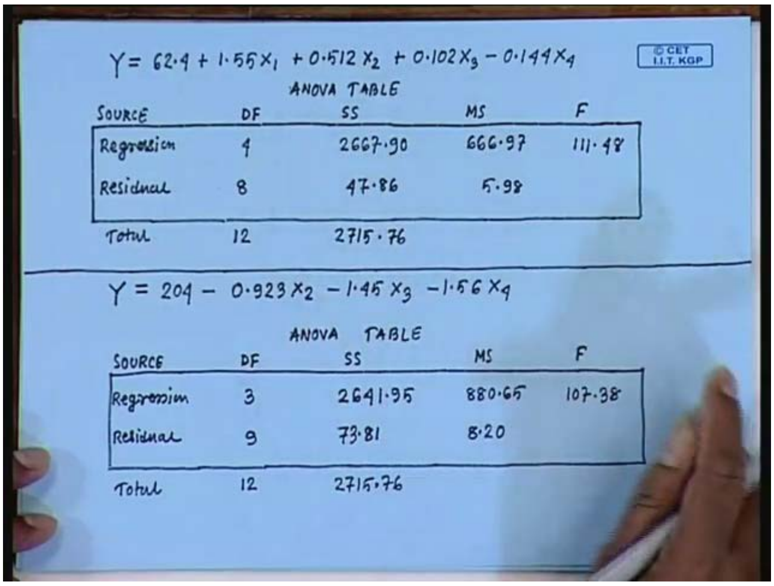
0.144 X 4. So, this is the fitted model and now what we want is that we want we look for the less significant regressor in this model, for that what we do we compute the partial F statistic associated with X 1.

We, compute the partial X statistic associated with the regressor X 2 similarly for X 3 and X 4. So, this is the notation F 1, 2, 3, 4 this is called the partial F statistic associated with the regressor X 1 in the presence of X 2, X 3, X 4 similarly also we will compute the partial F statistic associated with X 2 in the presence of X 1, X 3 and X 4 in the model. And also we will compute for all the regressors, so we will compute F 3 in the presence of 1, 2, 4 we will compute F 4 in the presence of 1, 2, 3.

Now, let me explain how to you know just now we discussed now how to compute this partial F statistic value. So, this can be computed using I mean this is the value of this statistic, you compute SS regression for the full model, so I will use some other notation like 1, 2, 3, 4 that is my full model. Because, there are only four regressors in the problem and by 1, 2, 3, 4 I mean that all the four regressors X 1, X 2, X 3 and X 4 are in the model.

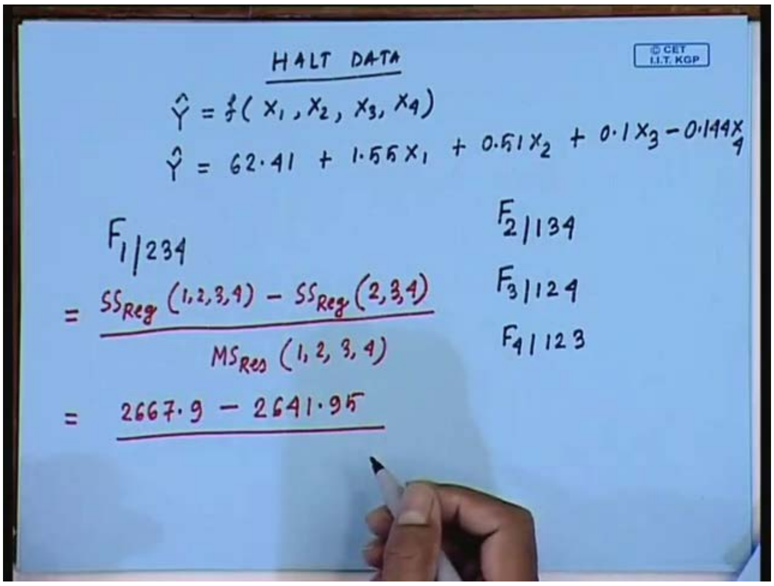
So, this is the SS regression for the full model minus SS regression for the model with random variable 2 X 2, X 3 and X 4 also I mean instead of 1, 2, 3, 4. You can write X 1, X 2, X 3, X 4, X 2, X 3, X 4 maybe next time I will do that by MS residual for the model; that means, 1, 2, 3, 4 all the four regressors at there in that model. I mean this is the MS residual when all the four regressors are present in the model. well. So, what you have to do is that we need to get this one we need an ova table for the model involving four regressors variables.

(Refer Slide Time: 23:42)



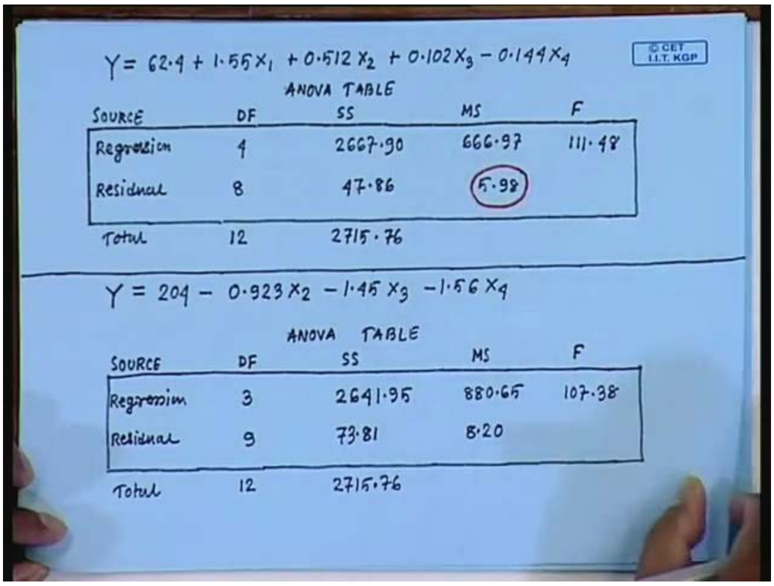
So, here is the fitted model involving four regressors X 1, X 2, X 3 and X 4 and here is the an ova table for this model, for this full model. Now, the SS regression is 2667.9, right.

(Refer Slide Time: 24:14)



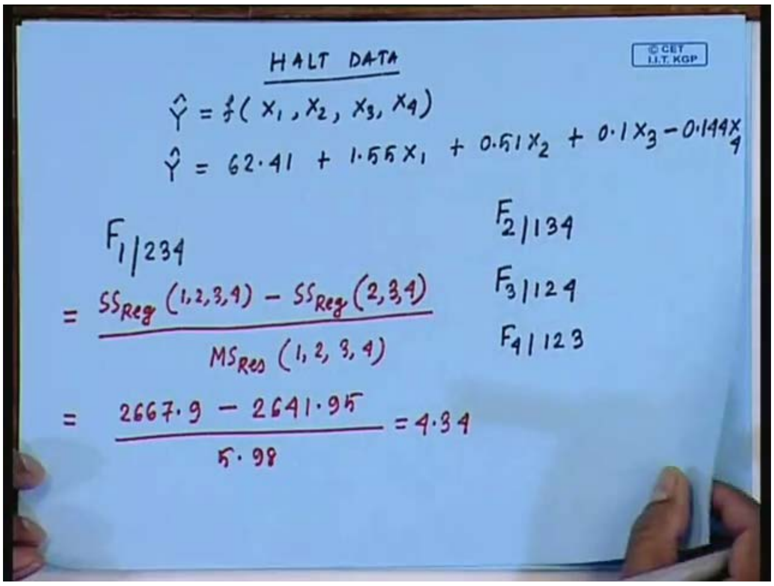
So, I will use this value this SS regression for the full model is 2667.9 minus the SS regression for the model involving X 2, X 3, X 4. So, again I need to fit this model I need fit the model involving X 2, X 3 and X 4 and here is the associated I mean the an ova table for this model, and my SS regression for this model is 2641. So, I will use this value 2641.95. Now, this difference you know this difference is the extra sum of square due to X 1.

(Refer Slide Time: 25:20)



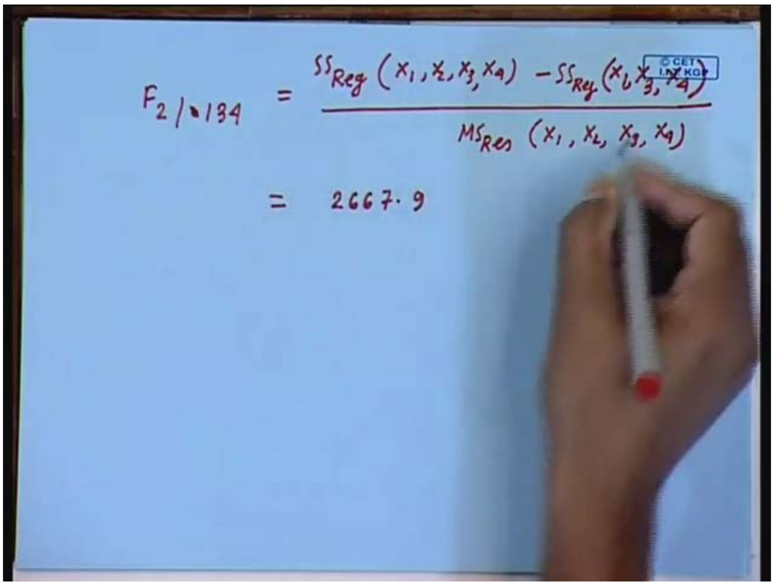
Now, the MS residual value for the full model, again require I mean, so this the full model and the MS residual value is 47.86 and look at the degree of freedom here you know we have 13 data. So, the degree of total degree of freedom is 2 and the residual has degree of freedom 8 because of the fact that you have 1, 2, 3, 4, 5 five angles, so you will be getting 5 what is called normal equations. That means, the five constant known residual. So, 13 minus 5 is equal to 8, so the residual has degree of freedom 8, and hence the regression has degree of freedom 4. Now, what I want to say is that I will be using this MS residual that is the MS residual for the full model.

(Refer Slide Time: 26:20)



So, 5.98 this value is equal to 4.34 right, now, we need to compute the other values also may be I will just explain this one also if 2 given 134.

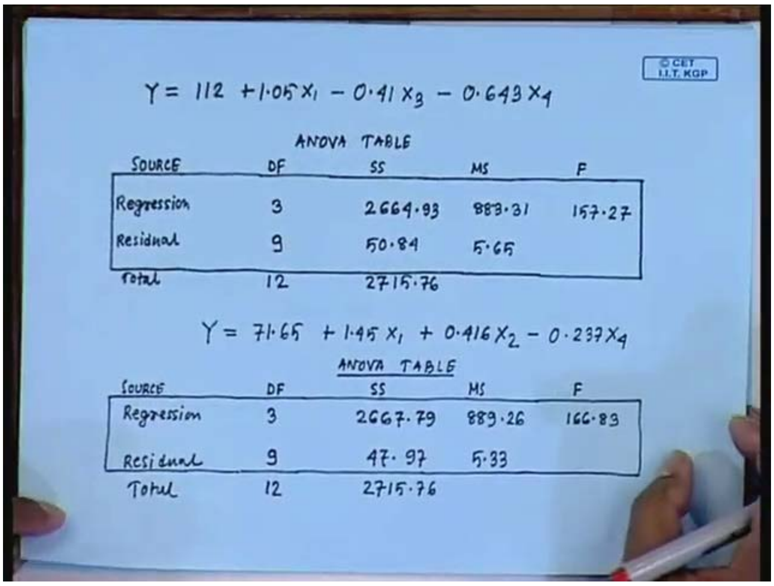
(Refer Slide Time: 26:52)



So, if 2 in the presence partial F statistics for X 2 in the presence of X 1, X 3 and X 4 in the model, so this is regression SS regression for the full model. That means, X 1, X 2, X 3, X 4 minus SS regression for the model involving X 1, X 3 and X 4 alright by MS residual for the full model; that means, X 1, X 2, X 3, X 4 well. So, let me check whether

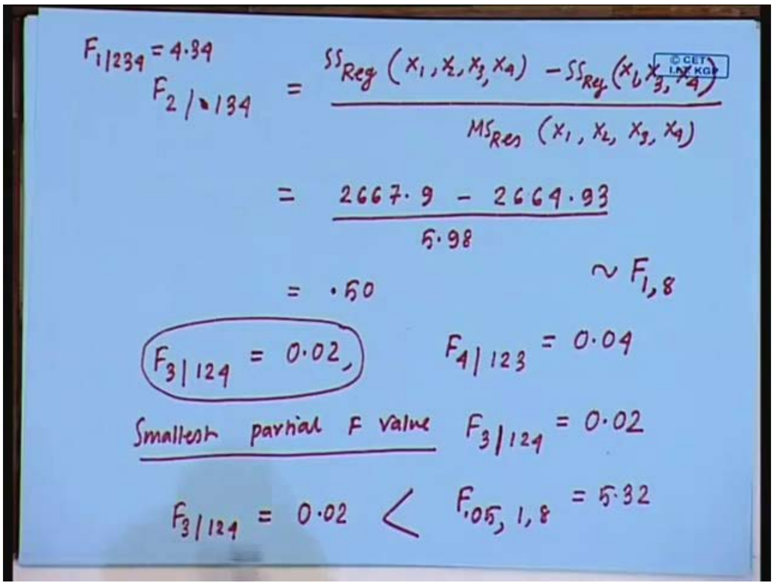
I have an ova table for this one yes, I have look we know that this SS regression is 2667.9. Now, I need the model involving X 1, X 3, X 4.

(Refer Slide Time: 28:08)



So, here is the fitted model Y equal to X 1, X 3, X 4 and the an ova table here. So, the SS regression is 2664.94 alright.

(Refer Slide Time: 28:26)



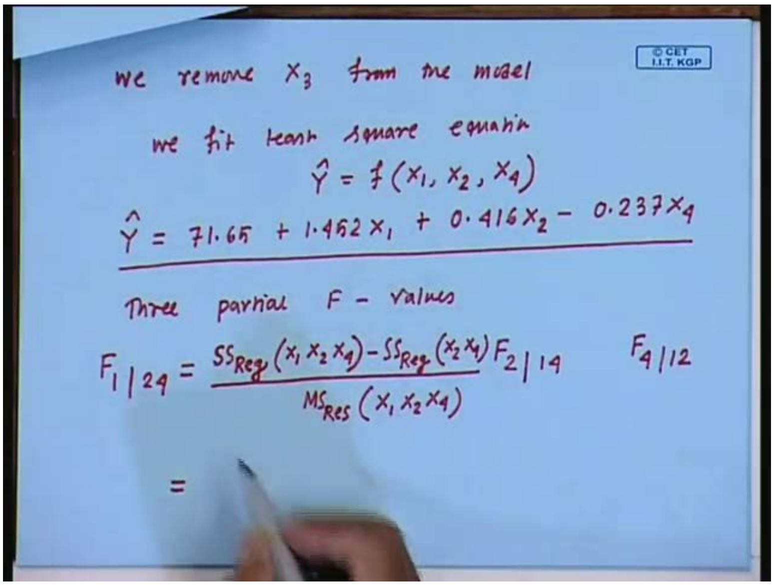
So, minus 2664.93 by the MS residual for the full model that is 5.98 and this value you can check that this close to 0.50. Similarly, you compute I am not going through detail for the other calculation, so if 3 given 1, 2, 4 you can check that this is equal to 0.02 and

F 4 given 1, 2, 3 is equal to 0.04. So, we have all the four partial F values, the first one was F 1, 2, 3, 4 that value was 4.34. Now, we have to see which one is the less significant, so the smallest partial F value now the smallest because we want to remove one regressor, the less significant regressor from the model.

So, the smallest as we look for the smallest partial F value is which one is the smallest this one is the smallest is F 3, 1, 2, 4 which is equal to 0.02. Now, we can remove this, so the associated random variable is X 3, we can remove the regressor X 3 from the model provided that this observed partial F value is less than the tabulated value.

So, the tabulated value or if F out here is F 0.05, so it took alpha equal to 05 and the degree of freedom of this partial F statistics you know it follows F distribution with degree of freedom 1 and the error degree of freedom for the full model the error degree of freedom is 8. So, it follows F 18, so we will find the value of F 105, 1, 8 which is equal to 5.32. Now, this one is the smallest partial F value which is F 3 associated with X 3 1, 2, 4 this one is less than the tabulated value. So, X 3 is insignificant we can remove x 3 from the model.

(Refer Slide Time: 32:10)



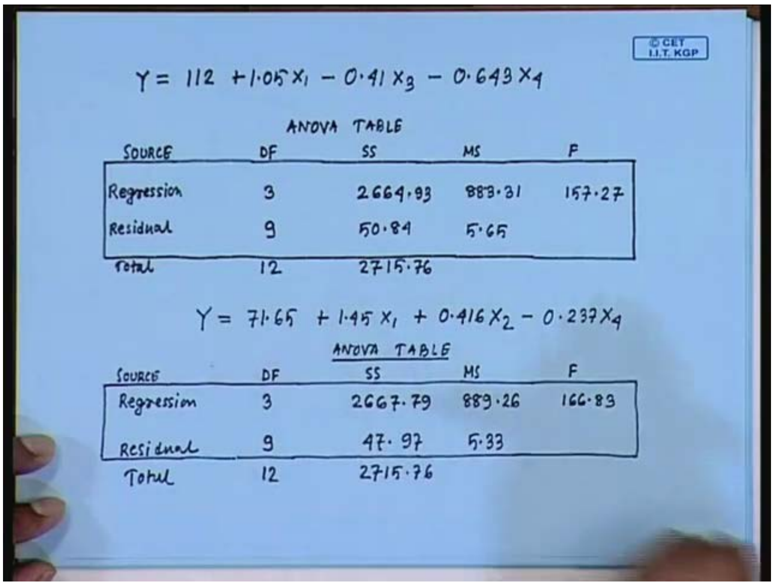
So, we remove X 3 from the model, now, what we do is that once X 3 is removed from the model, we are left with X 1, X 2 and X 4. So, we fit model between the response variable and X 1, X 2 and X 4 right, so we fit least square equation on model between Y and X 1, X 2, X 4 right. And the fitted model is Y hat equal to you can check that you have the data, so know how to you know how to fit multiple linear regression between Y and X 1, X 2, X 4 that is 71.65 plus 1.452 X 1 plus 0.416 X 2 minus 0.237 X 4 alright.

So, this is the model after removing X 3 from the full model well, now again what we do is that we will try to see whether we have a regressor, which is less significance in the presence of other regressors. So, what we do is that we compute three partial F values and their F 1 what is the significant I mean we check whether X 1 is significant in the presence of X 2 and X 4, we compute this partial F statistics 1 2, 4 right. We also compute F 2 in the presence of 1, 4 and F 4 in the presence of 1, 2.

And we again we look for the smallest partial F value, and which is and that is again you know less than the F out, then we remove that regression from the model. So, we will keep on doing that well, so what is this value, this value again no I will like to explain in detail this is obtained by computing. So, there is a little difference you know I that is why I want to discuss this one in detail.

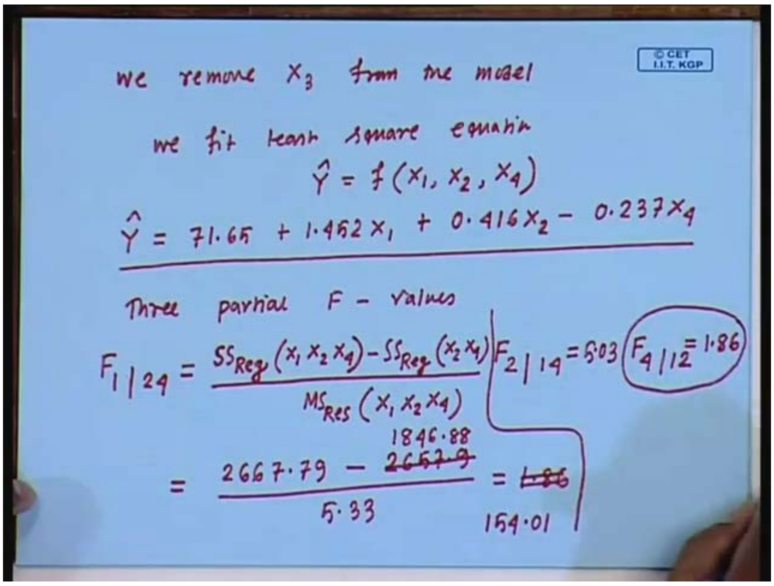
So, here you compute SS regression for the model X 1, X 2, X 4, so this is the SS regression when the mode involves X 1, X 2 and X 4 minus SS regression, when the model involve only X 2 and X 4. So, this difference will give you the contribution the extra sum of square or extra regression sum of square due to X 1 and this divided by MS residual sorry yes MS residual for the model X 1, X 2, X 4. So, see now this MS residual is not any more for the full model, it now at this moment our full model is involving three regressors X 1, X 2 and X 4 right.

(Refer Slide Time: 37:05)



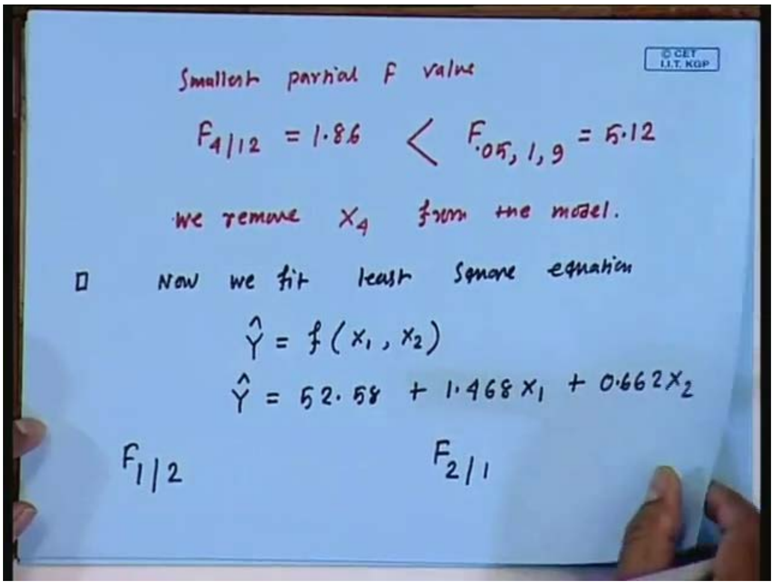
So, to get this value again X 1, X 2 and X 4 is I have an ova table for that, so here is the fitted equation X 1, X 2, X 4 and here is the an ova table, now you can check that the total degree of freedom is 12 this is same. Now, the residual has degree of freedom 9 because we have 1, 2, 3, 4 unknown parameters, so 13 minus 4 is equal to 9 right. So, this is the SS regression for this model and this is the MS residual, we are going to use this MS residual now right we are not going to use this MS residual that is for the model involving all the four regressors right.

(Refer Slide Time: 37:54)



So, at this moment my full model is this one, so SS regression is 2667.79 I do not know the value of this one, you can check that you fit a model a between y and X 2 and X 4 you find out the SS regression value that will be 2657.9. And the MS residual value is 5.33, it is not 5.98 this one is equal to 1.86, no I did mistake here this one is not 2657 it is 1846.88 the final value here is equal to 154.01 right. Now, you can check that the value of this partial F statistic is equal to 1.86 value of this partial F statistic is equal to 5.03. So, the smallest partial F value is this one; that means, if this is F is you know less than the tabulated value, then we are going to remove X 4 from the model because this partial F statistics is associated with X 4 right.

(Refer Slide Time: 40:06)

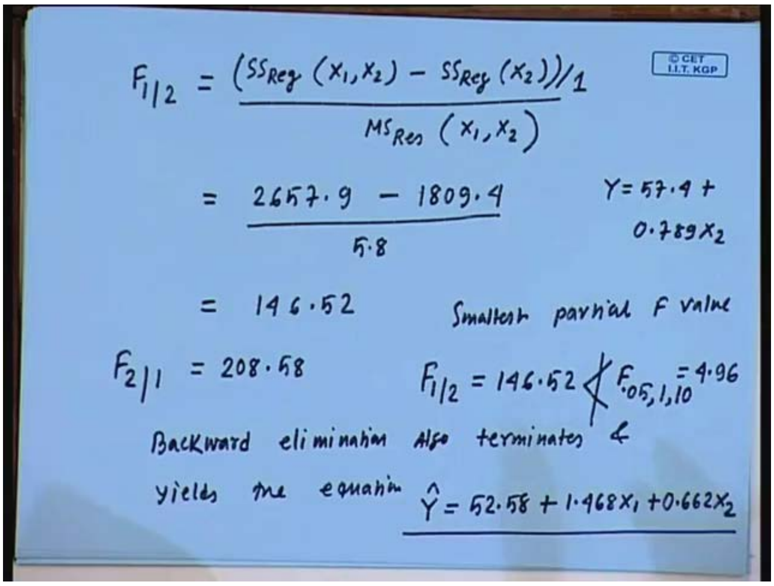


So, let me write down the smallest partial F value F 4 1, 2 is equal to 1.86, and the tabulated value is F 0.05 with degree of freedom of 1 and here now the error degree of freedom is not 8 it is 9 because see this our full model at this moment and error degree of freedom is oh sorry residual degree of freedom is 9. The residual degree of freedom is 9 and this value is equal to 5.12 and this is the F out, now again the partial F value associated with X 4 is less than this one.

So, what you do is that we remove we remove X 4 from the model, the next step what we do is that see we will we have initially we started with X 1, X 2, X 3, X 4. Then the four step we have removed X 3 from the model, in the second step we have removed X 4 from the model, now we are left X 1 and X 2. So, we now we fit least square equation Y hat equal to f X 1, X 2 and this Y hat is going to be equal to so; that means, we I am trying to I mean we will fit a model involving X 1 and X 2 only and the fitted model is Y hat equal to 52.58 plus 1468 X 1 plus 0.662 X 2.

So, you know how to fit you know how to find the unknown parameters, now again see we are having a model with involving X 1 and X 2. Now, we need to check whether again we can remove any more regressors from this model; that means, we have to check the significance of X 1 in the presence of X 2, and also we have to check the significance of X 2 in the presence of X 1. So, we need to compute two partial F statistics, one is F partial F statistics for associated with X 1 in the presence of X 2, and the other one is partial F statistics associated with X 2 in the presence of X 1 right this can be done.

(Refer Slide Time: 43:59)



Like F 1, 2 this one is equal to this is regression involving X 1 and X 2 in the model minus SS regression involving X 2 in the model. And this one you know basically this divided by the is 1 by MS residual for the model involving X 1 and X 2 well, I do not think that I have an ova table for this one. But, you can check that you fit a model between Y and X 1, X 2 we have the fitted model is it just now we have fitted it. So, once you have the fitted model you can compute the SS residual, and then you know the SS total is same, so then you can get the SS regression value right.

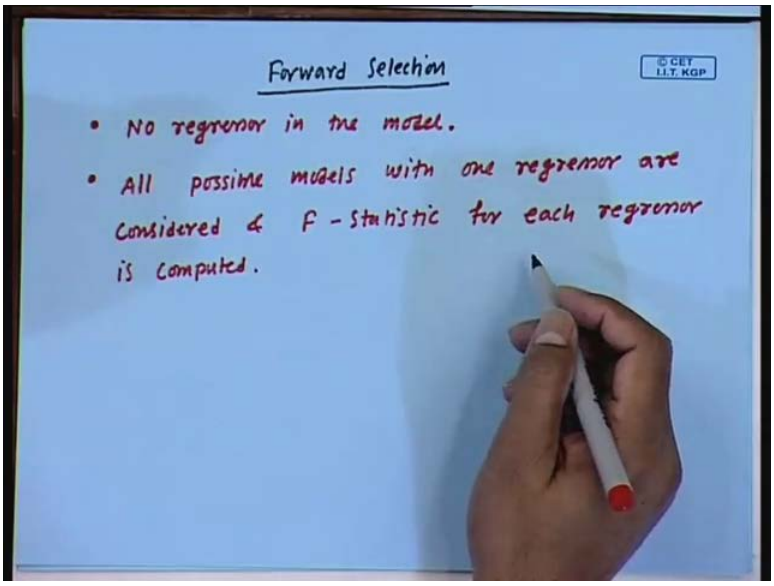
So, the SS regression you can check that this is equal to 2657.9 minus you fit a model between Y and X 2 the fitted model is going to be Y equal to 57.4 plus 0.789 X 2. So, I am giving you the fitted model, now you can check that SS regression value is equal to 1809.4, and the MS residual for the model involving X 1 and X 2 is 5.8 that you can check. So, the partial F value is going to be 146.52 and similarly, you can check that if 2, 1, so the partial F statistics associated with the random associated with X 2 in the presence of X 1 is equal to 208.58.

Now, these partial F values are large also still we will check you know the smallest partial F value is which one is smallest F 1 2, which is equal to 146.52. And we compare this value with the tabulated value that is F 0.05 with degree of freedom 1, and if there are two regressors in the model. Then the number of unknown parameters is equal to 3, so the residua degree of freedom will be 13 minus 3. So, that is 10 and this value is equal to 4.96, but this one is not, smaller than this one.

So; that means, the model we have at this moment involving X 1 and X 2 both are significant, I mean both X 1 is significant in the presence of X 2, and similarly X 2 is also significant in the presence of X 1 that is why their partial F statistic value are large and we cannot remove any of them from the model. So, the backward elimination algorithm terminates here, so here we say that the backward elimination algorithm terminates and yields the equation the final equation is Y hat equal to 52.58.

That means, that model involving X 1 and X 2 is the best subset regression model 1.468 X 1 plus 0.662 X 2. So, the output of the backward elimination algorithm for the HALD cement data is that, the subset model involving X 1 and X 2 is the best, according to the backward elimination algorithm. Next we will be talking about the forward selection.

(Refer Slide Time: 49:44)



So, what is the basic motivation behind the forward selection is that, instead of you know in the backward elimination we started from the full model, and then in each step we tried to eliminate the less significant regressor from the model. And in the forward selection what we do is that, we start from a model having no regression variable, so initially there is no regressor in the model. And then every in every step we tried to find the most significant regressor for the response variable, and we add every step we add one regressor to the model. So, that is the no just opposite of backward elimination.

Let me just write down the algorithm now, so the first step is that no regressor in the model, and then you know all possible models. The first step is that the all possible models with one regressors are considered and F statistic for each regressor is computed; that means, if you have four regressor variable we will start with the we will conside, the model one regression model involving X 1, and then the regression model involving X 2 only, and then the regression model involving X 3 only, and then the regression model involving X 4 only. And then we compute the F statistic value for each of them, and the regression having the highest F statistic value is added to the model. Anyway I do not have you know time today, so better I will start this forward selection in the, I will talk about this forward selection in the next class.

Thank you for your attention.