Statistics for Experimentalists

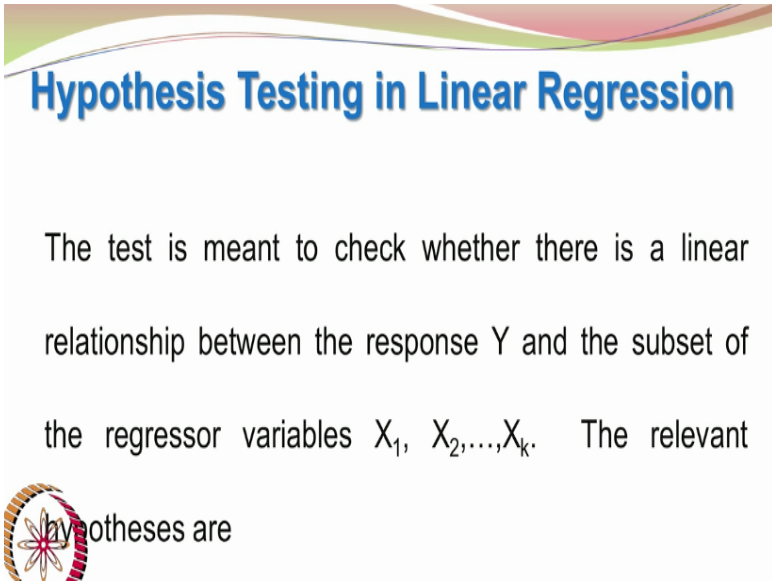
Prof. Kannan A
Department of Chemical Engineering
Indian Institute of Technology – Madras

Lecture – 39
Hypothesis testing in Linear Regression

Hello, welcome back. In today's class, we will be looking at hypothesis testing in linear regression. So, what is the motivation for doing this test? When we develop a linear regression model, we want to see which of the variables we had taken are considered in the experiment are really important. At the beginning, when we are not having prior knowledge or experience, we really do not know which variables are important, which variables are not important?

So we would like to include as many variables as possible in our experimental program and we perform the experiment and we get the data. Now we want to analyze the data and identify which of the variables are really significant and influence the experiments strongly. So how do we go about it that is what we are going to see in today's lecture? So as the slide indicates.

(Refer Slide Time: 01:43)

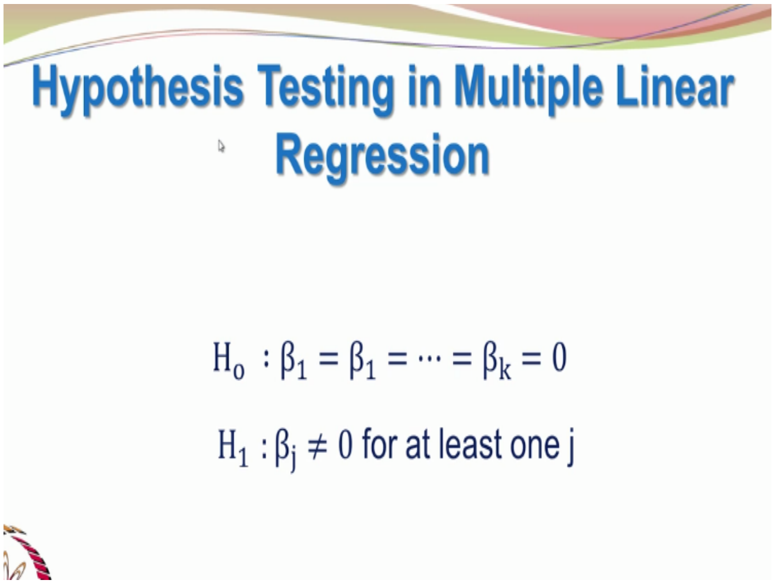


Hypothesis Testing in Linear Regression

The test is meant to check whether there is a linear relationship between the response Y and the subset of the regressor variables X_1, X_2,…,X_k. The relevant hypotheses are

The test is meant to check whether there is a linear relationship between the response Y and the subset of the regressor variables X1, X2 so on to Xk. So these regressor variables are actually the variables we are investigating in the experiment.

(Refer Slide Time: 02:09)



Hypothesis Testing in Multiple Linear Regression

$$H_0 : \beta_1 = \beta_1 = \cdots = \beta_k = 0$$

$$H_1 : \beta_j \neq 0 \text{ for at least one } j$$

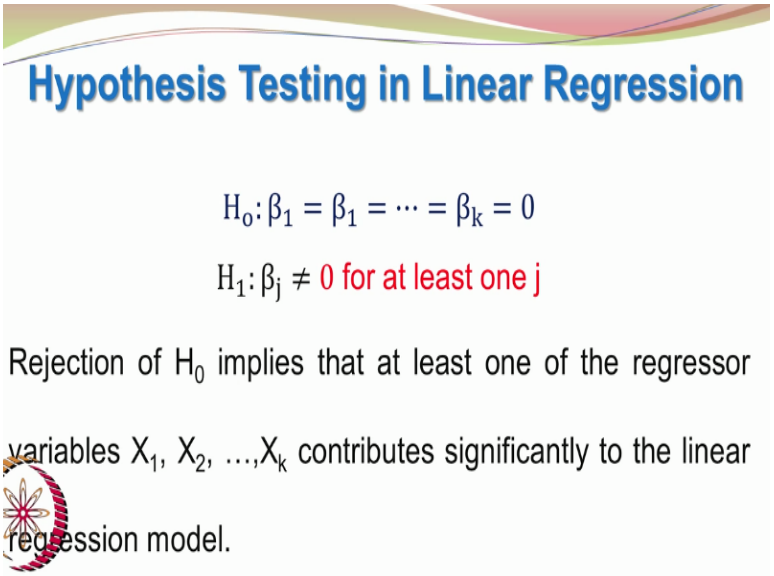
So we carry out the hypothesis testing in multiple linear regression our null hypothesis is beta1 = beta2 so on to beta k = 0. I will make a small correction here. So we have beta1 = beta2 so on to beta k = 0 that is a null hypothesis. So what it really means is that none of the regression coefficients are having numbers that are significantly different from 0. Look the beta1, beta2 so on to beta k may take either negative values or positive values.

If they take positive values, it means that they are positively affecting the response. For example, the yield of a chemic reaction may increase with increasing temperature. On the other hand, if you have a negative value for beta j then it means that when the variable increases, it actually has a negative effect on the response. For example, when pressure increases, the volume may decrease. So it depends upon the experiment we are looking at.

When the beta j that is one of the regression coefficients become 0, then beta j xj will be 0. This means that whatever maybe the value taken by xj, the effect of that particular variable on the experiment is insignificant. So this is what we are trying to test. The null hypothesis says that all the regression coefficients are 0 that means none of the variables are really affecting the process.

This is the most skeptical point of view a person may take at the beginning of the experiment, but as experimenters we should be really skeptical and not have some preconceived notions. See the alternate hypothesis says that beta j != 0 for at least 1 j. This means that among all the regression coefficients at least one of them is nonzero. In other words, there is at least 1 variable in the experiment which is actually affecting the process response.

(Refer Slide Time: 05:02)



Hypothesis Testing in Linear Regression

$$H_o: \beta_1 = \beta_1 = \cdots = \beta_k = 0$$

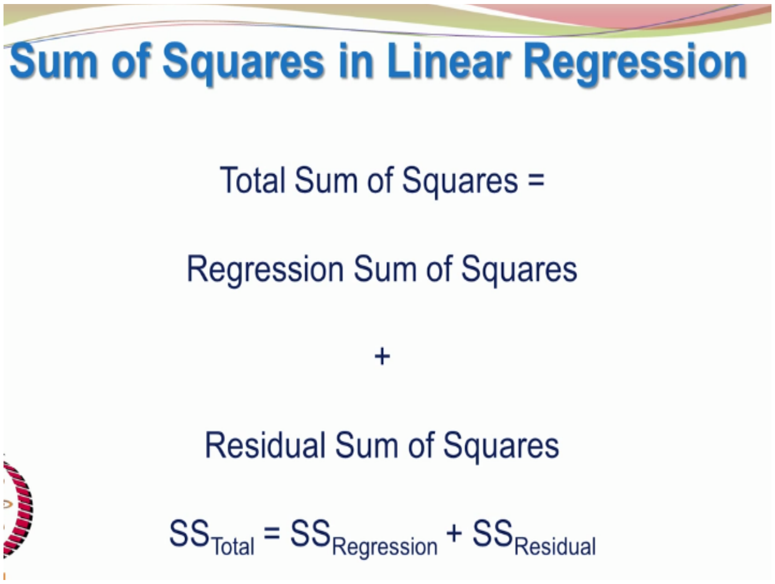
$$H_1: \beta_j \neq 0 \text{ for at least one j}$$

Rejection of H_0 implies that at least one of the regressor variables X_1, X_2, …,X_k contributes significantly to the linear regression model.

So when we accept the null hypothesis, we agree that none of the regression coefficients are taking a value other than 0. So we say that H0 be null hypothesis is beta1 = beta2 so on to beta k = 0. If we agree with this, then none of the variables are really affecting the response. The alternate hypothesis is for at least 1j, beta1 or beta2 or so on to beta k, at least one of them is nonzero. It may be negative or it may be positive.

So the rejection of H0 implies that at least one of the regressor variables X1, X2 so on to Xk contributes significantly to the linear regression model. So here we are having k independent variables X1, X2 so on to Xk and these are the regression coefficients which are attached to these regressor variables and when then say that beta j takes a value 0, beta j xj will be = 0 and there will not be any effect of that particular regressor variable xj on the process response. So how to carry out this hypothesis testing?

(Refer Slide Time: 06:38)



Sum of Squares in Linear Regression

Total Sum of Squares =

Regression Sum of Squares

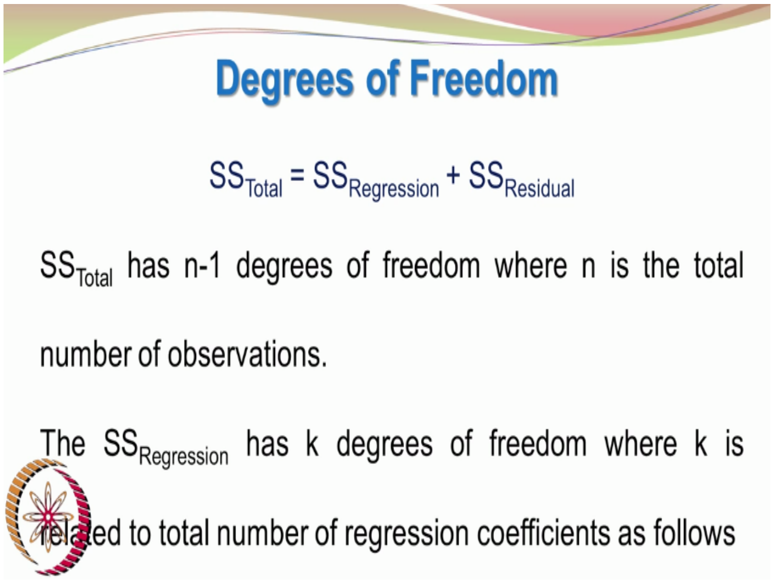
+

Residual Sum of Squares

$SS_{Total} = SS_{Regression} + SS_{Residual}$

So we have the experimental data with us and we first find the total sum of squares and then we split it into regression sum of squares and residual sum of squares. So I have indicated this briefly here. Sum of squares total = sum of squares regression + sum of squares residual.

(Refer Slide Time: 07:19)



Degrees of Freedom

$SS_{Total} = SS_{Regression} + SS_{Residual}$

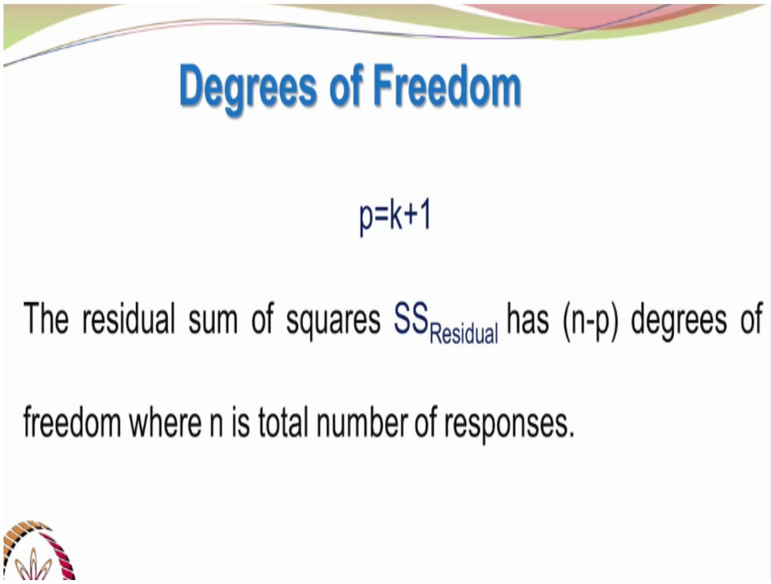
SS_{Total} has n-1 degrees of freedom where n is the total number of observations.

The $SS_{Regression}$ has k degrees of freedom where k is related to total number of regression coefficients as follows

So whenever we compute the squares we also have to find the degrees of freedom. Whenever we want to compute the variance not only we find the deviation from the mean, but we also divided by n - 1 where n is the total number of observations. So the sum of squares is actually divided by a certain value which is related to the data size.

In our present analysis also whenever we are considering linear regression we have the total sum of squares and we have to scale it by the appropriate or associated degrees of freedom. So the total sum of squares has n - 1 degrees of freedom, where n is the total number of observations. The sum of squares of regression has k degrees of freedom where k is related to the total number of regression coefficients in the following manner.

(Refer Slide Time: 08:30)



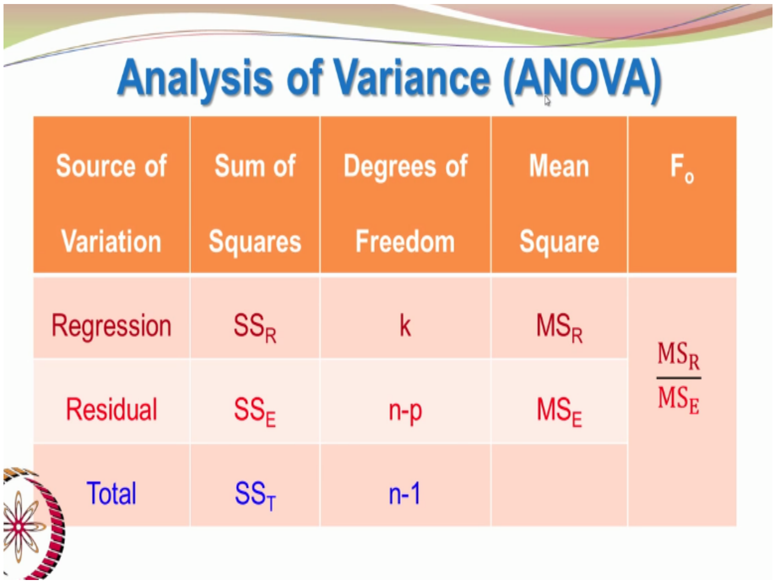
Degrees of Freedom

$p=k+1$

The residual sum of squares $SS_{Residual}$ has (n-p) degrees of freedom where n is total number of responses.

So we have p = k + 1. I think we have already come across this earlier to re-iterate p is the total number of parameters and that includes the parameter beta 0. The so called intersect of the regression model. The residual sum of squares will then have n - p degrees of freedom, where n is the total number of responses.

(Refer Slide Time: 09:13)



Analysis of Variance (ANOVA)

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_o
Regression	SS_R	k	MS_R	$\frac{MS_R}{MS_E}$
Residual	SS_E	n-p	MS_E	
Total	SS_T	n-1		

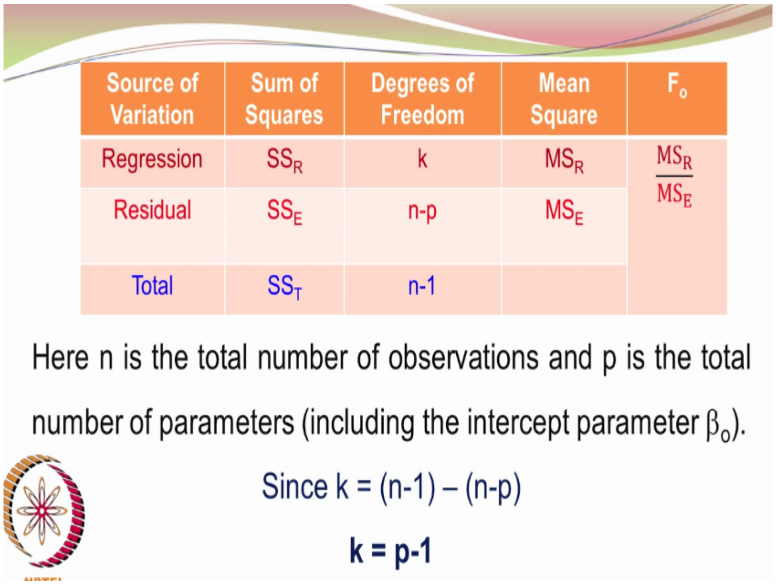
So now we can go to the analysis of variance table. We have in the ANOVA table the usual entities. The source of variation, the sum of squares associated with the source of variation, the degrees of freedom and we divide the sum of squares with the associated degrees of freedom to get the mean square. So k = p - 1 that means the total number of parameters - 1. Here we are not considering the intercept.

We are only considering the regression coefficients beta1, beta2, so on to beta k and when we divide the sum of squares of regression with the k degrees of freedom, we get the mean square regression and then we also have the residual sum of the squares. This is a very important aspect in regression analysis because only by looking at the residuals and the pattern of the residuals we can really judge about the quality of the fit.

So we have the residual sum of squares as sum of squares of E again instead of rather writing residuals I have used the subscript E residuals may also be associated with error because it is a difference between the experimental value and the model prediction. So the residual is defined as the difference between the experimental value in the model prediction and so we have the error with respect to the model prediction.

And the sum of squares associated with the residuals is given by SSE and the degrees of freedom associated with it is n - p so the mean square would be sum of squares of the residuals divided by n - p that will give you mean square residuals. So we take the ration of mean square regression to the mean square residua to get the F0 value. So we also have the total sum of squares SST which is having n - 1 degrees of freedom. So it will look like n - p + k and k is nothing but p -1. So n - p + p - 1 will give us n - 1 degrees of freedom. So when you add up these 2 you get the total degrees of freedom as n – 1

(Refer Slide Time: 11:54)



Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_o
Regression	SS_R	k	MS_R	$\frac{MS_R}{MS_E}$
Residual	SS_E	n-p	MS_E	
Total	SS_T	n-1		

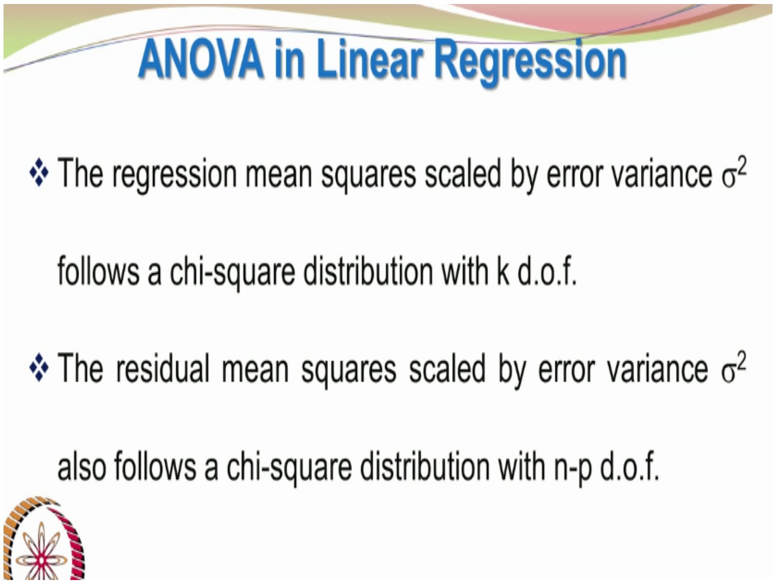
Here n is the total number of observations and p is the total number of parameters (including the intercept parameter β_0).

Since k = (n-1) – (n-p)

k = p-1

So that is what this slide also tells. Repeating, n is the total number of observations and p is the total number of parameters including the intercept parameter beta 0 and so we have k = n - 1 - n - p and we get k = p- 1. What I am doing is we saw that k and n - p add up to give n - 1. This is telling the same thing in a different way we just subtract n - p from n - 1 and we get p - 1 k = p - 1. So we can take whatever route we want.

(Refer Slide Time: 12:36)



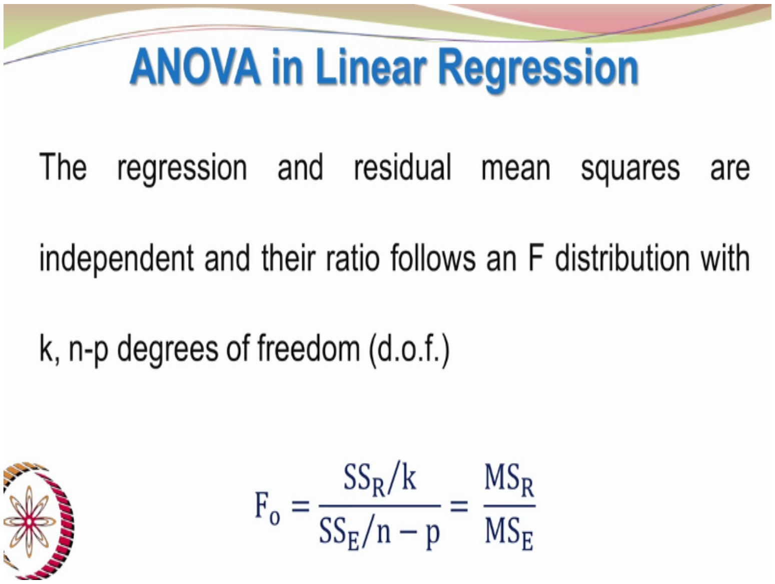
ANOVA in Linear Regression

- The regression mean squares scaled by error variance σ^2 follows a chi-square distribution with k d.o.f.
- The residual mean squares scaled by error variance σ^2 also follows a chi-square distribution with n-p d.o.f.

The regression mean square scaled by error variance sigma square follows a chi-square distribution with k degrees of freedom. We have to make a proper judgment regarding the observed mean squares. So we have mean square regression. We have mean square error or mean square residuals. So the ratio of the 2 we consider and we have to test it against a suitable distribution. What is that suitable distribution?

We also know that the mean square regression and the mean square residual are independent and we divide both of them by sigma square and we then have the mean square regression/sigma square to form the chi-square distribution with the k degrees of freedom. So when we divide MSR/sigma square it leads to a chi-square distribution with k degrees of freedom and when you divide MSR/sigma square you have to divide MSE also by sigma square and you get another chi-square distribution with n - p degrees of freedom and the ratio of the 2 chi-square distributions is the F distribution.

(Refer Slide Time: 14:13)



ANOVA in Linear Regression

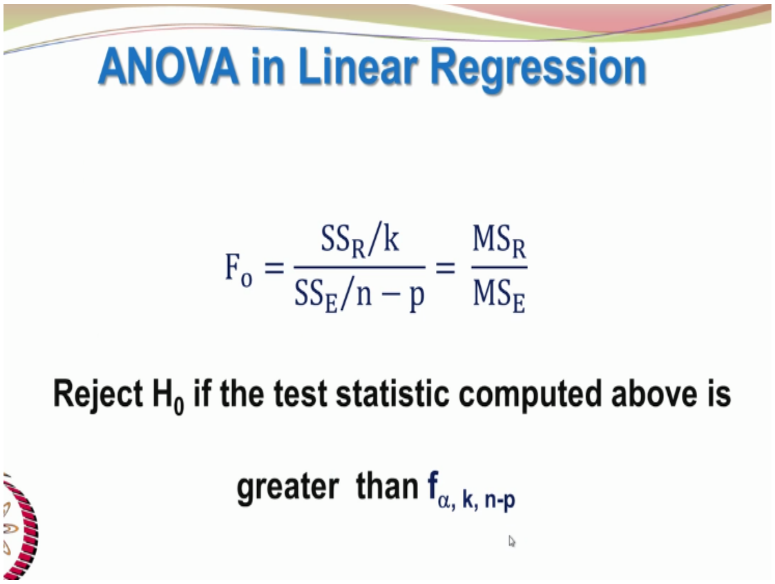
The regression and residual mean squares are independent and their ratio follows an F distribution with k, n-p degrees of freedom (d.o.f.)

$$F_0 = \frac{SS_R/k}{SS_E/n-p} = \frac{MS_R}{MS_E}$$

So the regression and residual mean squares are independent and the ratio follows an F distribution with k numerator and n - p denominator degrees of freedom. So you can see that we have F0 here and that the sum of squares of regression/k and this is the sum of squares of the residuals or the sum of square of the error/n - p. The k degrees of freedom are in the numerator and n - p degrees of freedom are in the denominator.

What I am trying to say here is the k degrees of freedom are associated with the sum of squares in the numerator and the n - p degrees of freedom are associated with the sum of squares of the residuals which is in the denominator. So we have F0 = mean square regression by mean square error. The sigma square actually cancels out. So we can simply take F0 as MSR/MSE.

(Refer Slide Time: 15:21)



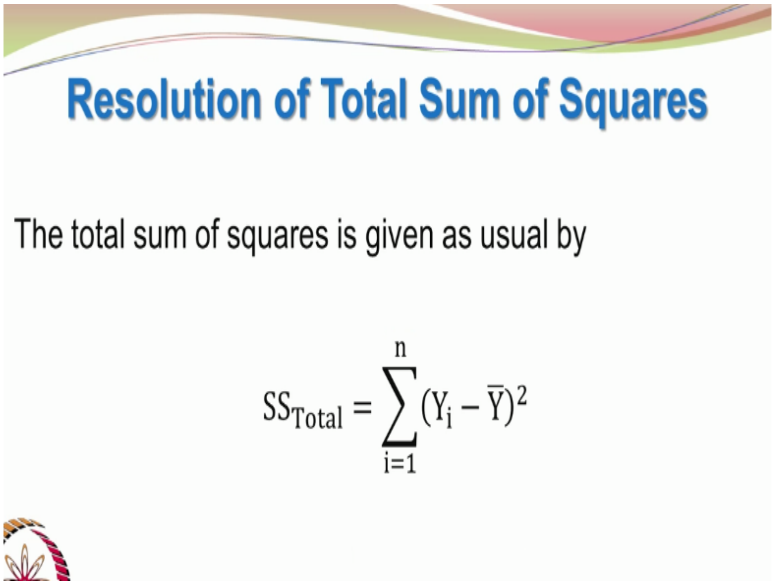
ANOVA in Linear Regression

$$F_0 = \frac{SS_R/k}{SS_E/n-p} = \frac{MS_R}{MS_E}$$

Reject H_0 if the test statistic computed above is greater than $f_{\alpha, k, n-p}$

We do the usual F test by now you should be familiar with the implementation of the F test. We have also done some practice problems or example sets earlier. So I request you to go through those problems and refresh your memory. So we also know that we reject the null hypothesis H0 if the test statistics computed above is > f alpha, k, n - p. Alpha is the significance level usually taken as 0.05. So if it lies in the critical region or in the rejection region, then we reject the null hypothesis.

(Refer Slide Time: 16:08)



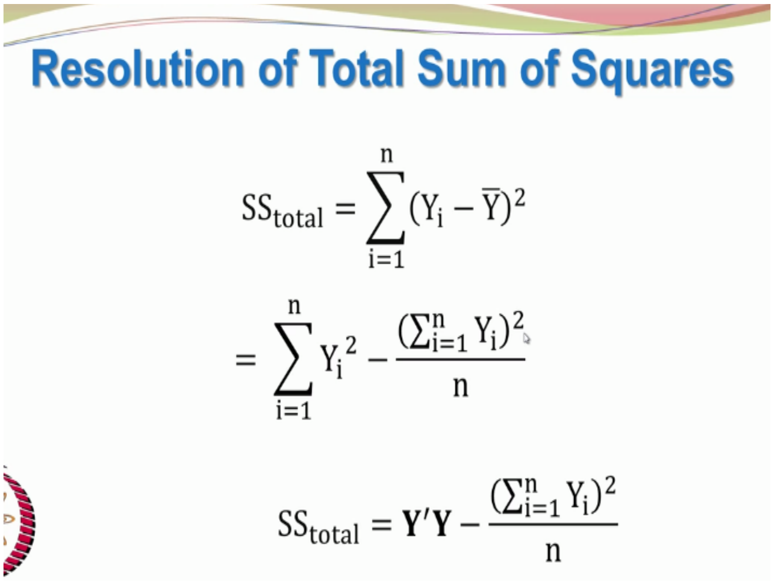
Resolution of Total Sum of Squares

The total sum of squares is given as usual by

$$SS_{Total} = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

So continuing with our discussion on the resolution of the total sum of squares, the total sum of squares is given by sigma I = 1 to n yi - y bar whole squared where yi is the actual ith experimental data recorded by the experimenter and y bar is the average of all the n experimental observations.

(Refer Slide Time: 16:42)



Resolution of Total Sum of Squares

$$SS_{total} = \sum_{i=1}^{n} (Y_i - \overline{Y})^2$$

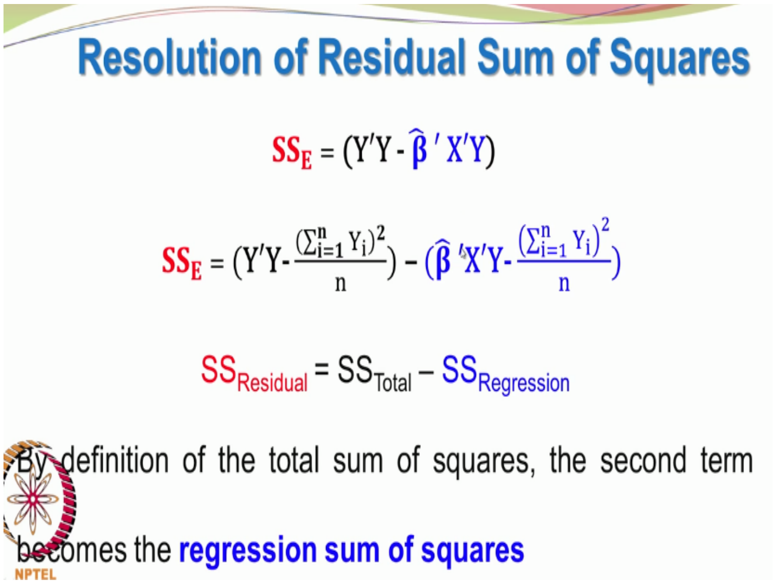
$$= \sum_{i=1}^{n} Y_i^{\,2} - \frac{(\sum_{i=1}^{n} Y_i)^2}{n}$$

$$SS_{total} = \mathbf{Y'Y} - \frac{(\sum_{i=1}^{n} Y_i)^2}{n}$$

So you have this relationship given here. This may be expanded and simplified the derivation is fairly straight forward and you get sigma I = 1 to n yi square - sigma I = 1 to n yi whole square/n. This indicates the sum of the square of all the responses and this is the sum of the observations is squared. So please do not confuse this with this term. Here the individual observation is squared. Similarly, all the other observations are also squared. Then the sum is taken.

Here, first the sum is taken and then it is squared. So this may be represented by y prime y. Sigma I = 1 to n yi squared is nothing but y prime y. So you have the column vector of the responses and a transpose is taken for the column vector and then it is multiplied with the actual response column vector and when you do that you will get the sum of the square of all the observations and then you also have the sum of the observations squared/n. Here n is the total number of responses. So this is the total sum of squares.

(Refer Slide Time: 18:42)



Resolution of Residual Sum of Squares

$$SS_E = (Y'Y - \hat{\beta}'X'Y)$$

$$SS_E = \left(Y'Y - \frac{(\sum_{i=1}^{n} Y_i)^2}{n}\right) - \left(\hat{\beta}'X'Y - \frac{(\sum_{i=1}^{n} Y_i)^2}{n}\right)$$

$$SS_{Residual} = SS_{Total} - SS_{Regression}$$

By definition of the total sum of squares, the second term becomes the **regression sum of squares**

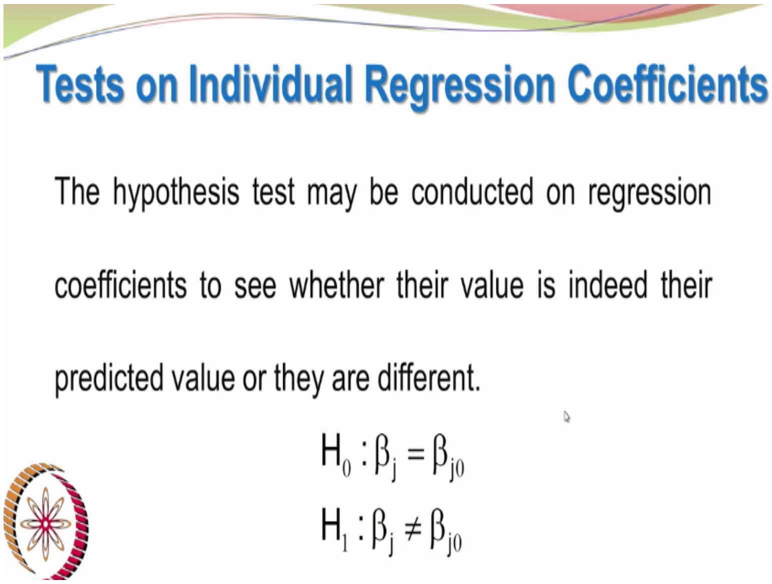
And as you can see we are gradually moving on to the representation of various sum of squares using matrix notation. The matrix method is quite convenient and it helps us to do the calculations which are otherwise tedious in a very efficient manner. So we have the sum of squares of the residuals as y prime y - beta hat prime x prime y. So the sum of squares of the residuals may be written as y prime y - sigma = 1 to n yi whole square/n - beta hat prime x prime y - sigma = 1 to n yi whole square/n.

What I am doing here is I am subtracting and adding this term sigma = 1 to n yi whole squared/n and that leads to by definition the sum of squares of the total here and we also have this term as a sum of squares of regression. We started off by saying that sum of squares total = sum of squares regression + sum of squares residual. So we have the expression for the sum of squares of the residuals.

And we also have the expression for the sum of squares total and so when we subtract the sum of squares of the residuals from the total sum of squares, we get the regression sum of squares. So this blue term here represents the sum of squares of regression. So when you have the linear regression parameters estimated and you have the x matrix and you have the y column vector you can get the regression sum of squares by considering a beta hat prime x prime y - sigma = 1 to n yi whole square/n.

Beta hat is nothing but the vector of the estimated regression parameters including the intercept beta hat 0 and then x is the matrix x matrix so x prime would be the transpose of the x matrix. We have already seen how to set up the x matrix in one of the earlier lectures and then y is the vector of observations. So we have sum of squares of residual as sum of squares of total - sum of squares of regression and we have the sum of squares of error given by his relation and then the sum of squares of regression is given by this relation.

(Refer Slide Time: 21:40)



Tests on Individual Regression Coefficients

The hypothesis test may be conducted on regression coefficients to see whether their value is indeed their predicted value or they are different.

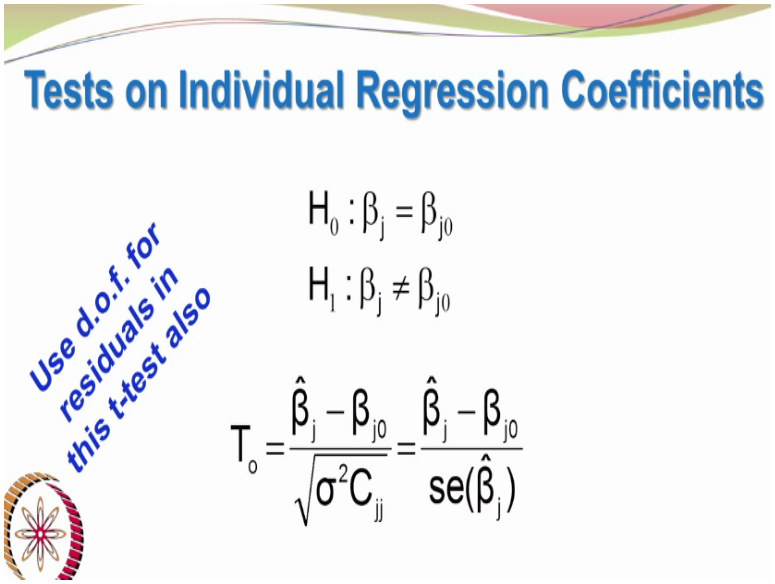
$$H_0 : \beta_j = \beta_{j0}$$
$$H_1 : \beta_j \neq \beta_{j0}$$

So, now coming back to the hypothesis tests on individual regression coefficients. So we have to see whether a particular regression coefficient beta j is actually taking up a particular value beta j0 or it is not taking that particular value. So now we are concentrating on the individual regression coefficients and whether they take up a value or not. So you can put 0 here and say that pretty much the regression coefficient is insignificant and does not affect the model or it does not affect the response in fact and then the alternate hypothesis is the value is != 0.

But it may be < 0 or > 0. so to be more general in suffixing the value to be 0 all the time instead of fixing beta j0 to be 0 all the time, we can fix it to some other value 100 for example, so it need not be always 0. You can also hypothesize on a particular value taken by the regression parameter. Instead of looking at the whole bunch of regression parameters now we are concentrating on a single regression parameter.

It may be a good idea for you to not proceed with the lecture as of now just pause a bit and then think yourself how you will carry out the test for this particular case. We have already come across this earlier and I would like you to think about it and then write down on a notebook you must be carrying with you as to how you would proceed. So I hope you have at least made an attempt and let us see how to do it.

(Refer Slide Time: 23:45)



Tests on Individual Regression Coefficients

$$H_0 : \beta_j = \beta_{j0}$$

$$H_1 : \beta_j \neq \beta_{j0}$$

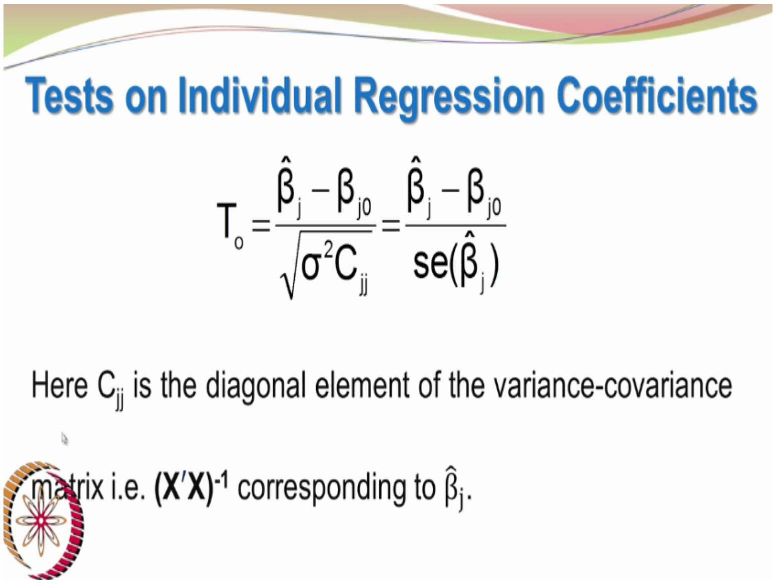
$$T_o = \frac{\hat{\beta}_j - \beta_{j0}}{\sqrt{\sigma^2 C_{jj}}} = \frac{\hat{\beta}_j - \beta_{j0}}{se(\hat{\beta}_j)}$$

Use d.o.f. for residuals in this t-test also

So we have the tests on individual regression coefficients so the H0 is beta j = beta j0 and H1 is beta j != beta j0 and then as you can see here we carry out a T test. So you must recollect the T test now if you are unable to remember I request you to just go back and refresh your memory. So T0 = beta j hat - beta j0/square root of sigma square Cjj that is beta j hat - beta j0 by standard error of beta hat j. Now we know that the T test is associated with the certain degrees of freedom and what degrees of freedom we should use in the T test.

Very interesting result is use the degrees of freedom which you had used for the residual sum of squares in the T test also and you should also by now be familiar with what is meant by the standard error of beta j hat and you should recollect that it is sigma squared Cjj where Cjj is the diagonal jjth element of the variance, co-variance matrix and sigma square is the error variance. Unfortunately, we do not know the error variance that true value of the error variance so what we do is we use the standard error instead.

(Refer Slide Time: 25:45)



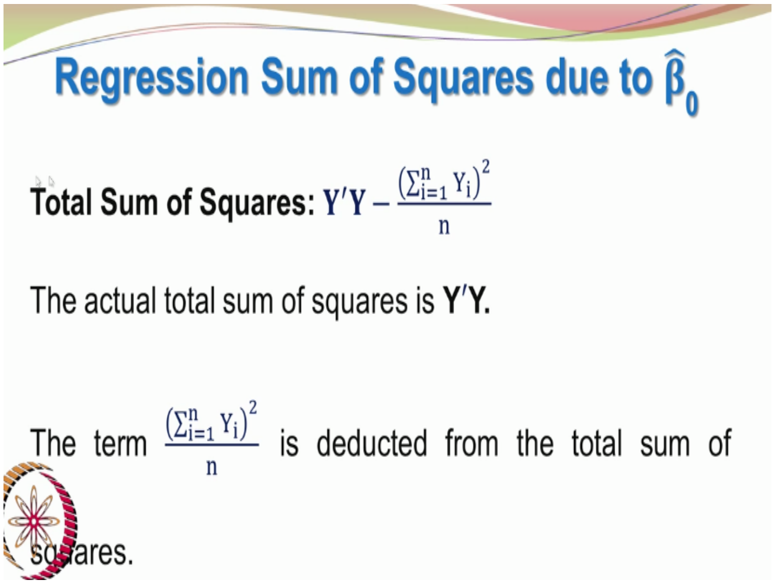
Tests on Individual Regression Coefficients

$$T_o = \frac{\hat{\beta}_j - \beta_{j0}}{\sqrt{\sigma^2 C_{jj}}} = \frac{\hat{\beta}_j - \beta_{j0}}{se(\hat{\beta}_j)}$$

Here C_{jj} is the diagonal element of the variance-covariance matrix i.e. $\mathbf{(X'X)^{-1}}$ corresponding to $\hat{\beta}_j$.

So as I said earlier just now Cjj is the diagonal element of the variance-covariance matrix and the variance-covariance matrix is given by x prime x inverse corresponding to beta hat j.

(Refer Slide Time: 25:59)



Regression Sum of Squares due to $\hat{\beta}_0$

Total Sum of Squares: $Y'Y - \frac{(\sum_{i=1}^{n} Y_i)^2}{n}$

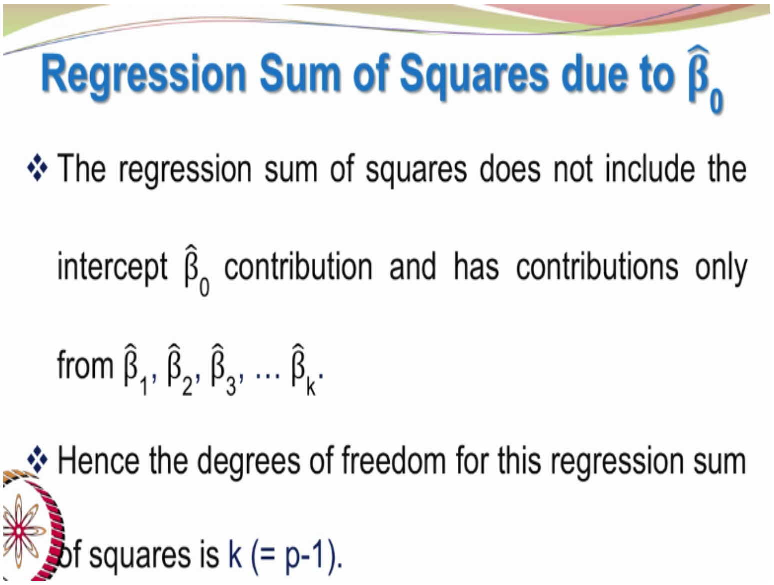
The actual total sum of squares is $\mathbf{Y'Y}$.

The term $\frac{(\sum_{i=1}^{n} Y_i)^2}{n}$ is deducted from the total sum of squares.

Now let us see a regression sum of squares due to the intersect beta hat 0. This is a very interesting thing and in some places it may be skipped and that may lead to some loss of clarity in understanding the concept of linear regression. In some ANOA tables you would find the sum of squares corrected for beta hat 0 or in some tables of ANOA you will find uncorrected total sum of squares.

So what is really the correction all about? So it depends on whether we consider the intercept or not. So we know the total sum of squares is y prime y - sigma is = 1 to n yi whole square/n. The actual total sum of squares based on the responses is y prime y. You simply square each response in total it up and that gives you the actual total sum of squares. So your detecting some portion from the actual total sum of squares that is you are detecting I = 1 to n sigma yi whole square/n. So this is the correction you are doing to the total sum of squares.

(Refer Slide Time: 27:28)

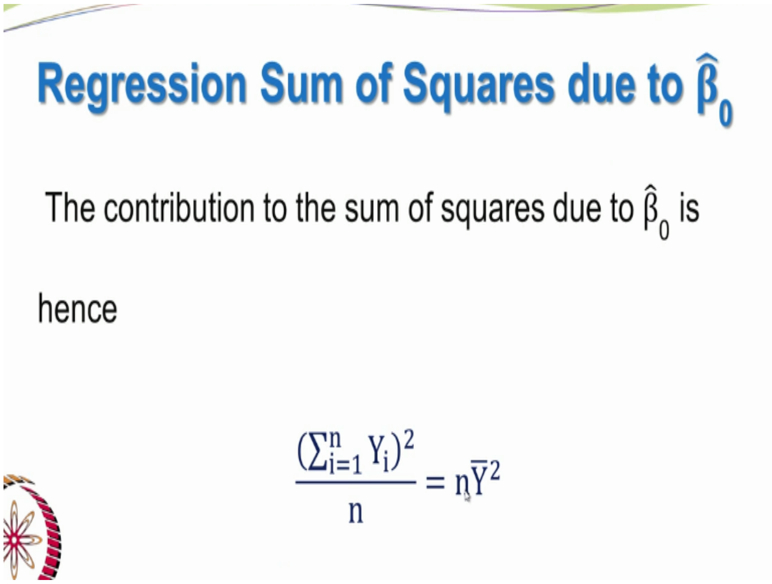


Regression Sum of Squares due to $\hat{\beta}_0$

- The regression sum of squares does not include the intercept $\hat{\beta}_0$ contribution and has contributions only from $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$, … $\hat{\beta}_k$.
- Hence the degrees of freedom for this regression sum of squares is k (= p-1).

The regression sum of squares does not include the intercept beta hat 0 contribution and has contributions only from beta hat 1, beta hat 2, beta hat 3 so on to beta hat k. So that is the reason why since you are having these 1 to k which is k independent regression parameters you have k degrees of freedom.

(Refer Slide Time: 27:57)



Regression Sum of Squares due to $\hat{\beta}_0$

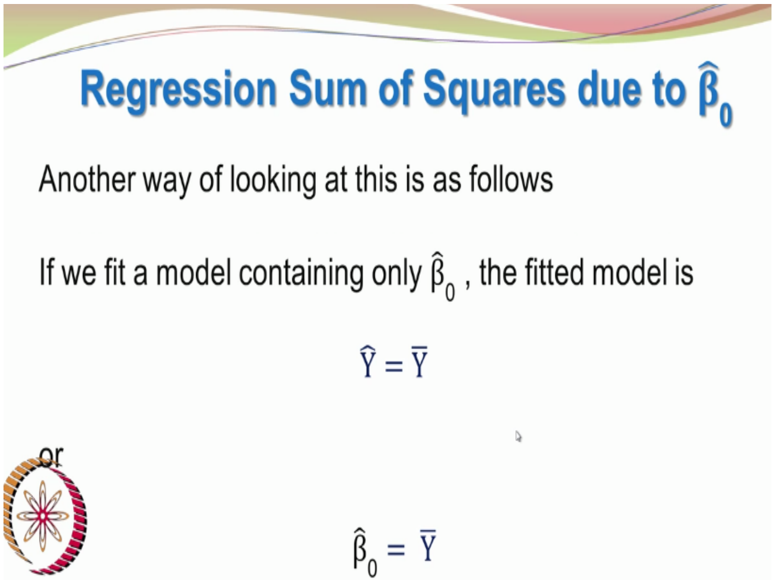
The contribution to the sum of squares due to $\hat{\beta}_0$ is

hence

$$\frac{(\sum_{i=1}^{n} Y_i)^2}{n} = n\bar{Y}^2$$

So what actually happens is the contribution to the sum of squares due to beta hat 0 is sigma = 1 to n yi whole squared by n. So we are removing the contribution to the total sum of squares that is y prime y with the subtraction by n y bar square. Sigma = 1 to n yi whole square/n is n y bar square and what we are doing is we are subtracting from the total sum of squares n y bar square. We call that as the contribution by the intercept parameter beta hat 0.

(Refer Slide Time: 28:55)



Regression Sum of Squares due to $\hat{\beta}_0$

Another way of looking at this is as follows

If we fit a model containing only $\hat{\beta}_0$, the fitted model is

$$\hat{Y} = \bar{Y}$$

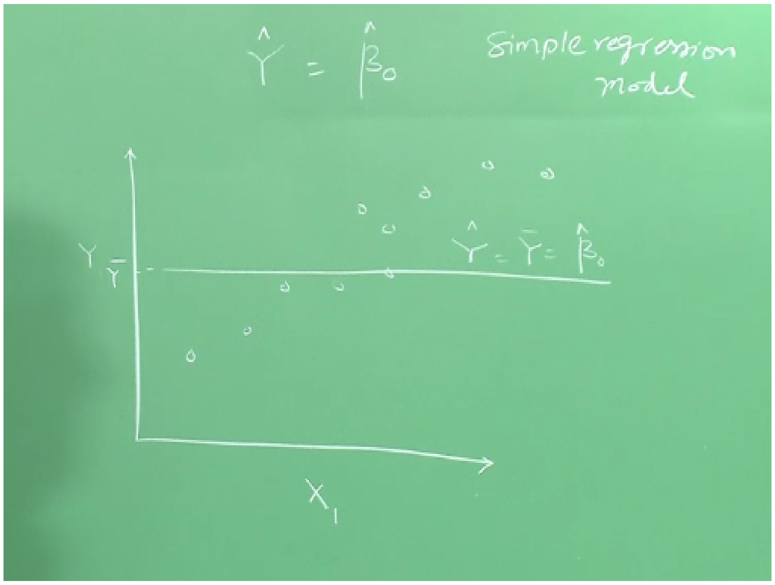
or

$$\hat{\beta}_0 = \bar{Y}$$

Why should it by n y bar square? So this is a very simple explanation to this it is quite nice actually. So when you consider no other parameter except beta hat 0 in your regression model then the regression parameter beta hat 0 would be simply the average of the experimental data points. What does it mean? Suppose we are very lazy to fit a regression model considering the variables. We say that y predicted = beta hat 0 only.

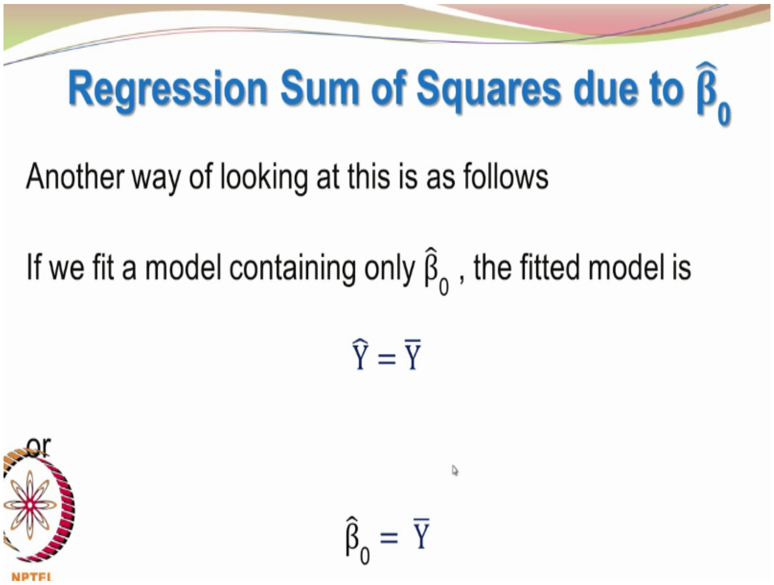
So then what will happen is if you carry out the regression analysis we will find that the estimated beta 0 parameter would be only y bar the average of all the responses. So when you have scattered data then let us say that we are having only one regressor variable x1. So we are having y1 as a function of x1 and when you plot the data on the graph sheet you will find that you will have scattered data and when you are fitting only a simple mode then the model will be nothing but y hat = y bar where y bar is the average of the responses and you will have one horizontal line passing through the data points. Let me illustrate this on the board.

(Refer Slide Time: 30:49)



So what we have here is the experimental data we are plotting y as a function of x1 obviously there is a effect of x1 on the response that is why you are finding that when x1 increases the data also increases, but if we take up a regression model saying that y hat = beta 0. This is our very simple regression model. Then all we are doing is fitting a straight line which is nothing but the average of all the responses and so we get a horizontal or a straight line parallel to the x axis and that straight line is nothing but the average value y bar.

(Refer Slide Time: 32:28)



Regression Sum of Squares due to $\hat{\beta}_0$

Another way of looking at this is as follows

If we fit a model containing only $\hat{\beta}_0$, the fitted model is

$$\hat{Y} = \bar{Y}$$

or

$$\hat{\beta}_0 = \bar{Y}$$

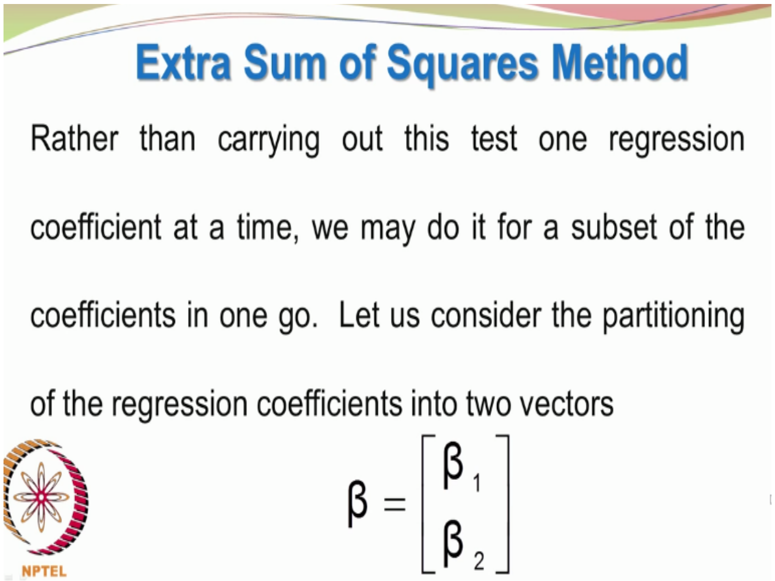
So since beta hat 0 = y bar, the sum of squares contribution from beta hat 0 will be y bar square + y bar square for n experimental data points and you will get n y bar squared. So this is the contribution to the sum of squares by the parameter beta hat 0. So if you want to correct your sum of squares and the regression sum of squares with the contribution from beta hat 0 then you subtract it with n y bar square and that is what we are doing.

Now let us go back to the resolution of the residual sum of squares and even before that we looked at the resolution of the total sum of squares you can see that the sum of squares of total we have subtracted the contribution by the parameter beta hat 0 and that is y prime y - n y bar square. So the n y bar square represents the contribution from the intercept beta hat 0. So when you are subtracting n y bar square from the total sum of squares you should also subtract n y bar square on the other side of the equality so that you maintain the balance.

So we see that this is the total sum of squares and this is the regression sum of squares y prime y is the actual total sum of squares beta hat prime x prime y is the regression sum of squares including all the regression coefficients and we are subtracting here n y bar square and then we are also subtracting n y bar square so this is the total sum of squares corrected for beta hat 0 and this is the regression sum of squares excluding the parameter beta hat 0.

So I hope now you have understood why we subtract n y bar square from the total sum of squares and from the regression sum of squares. Then we looked at the contribution from beta hat 0 and the regression sum of squares when it is subtracted by n y bar square does not include the contribution from beta hat 0. So the number of degrees of freedom is reduced by 1 because we are removing beta hat 0 from the list of P parameters so p - 1 will be = k.

(Refer Slide Time: 35:15)



Extra Sum of Squares Method

Rather than carrying out this test one regression coefficient at a time, we may do it for a subset of the coefficients in one go. Let us consider the partitioning of the regression coefficients into two vectors

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

Now let us look at the extra sum of squares method. This is a very interesting issue. What we saw earlier was looking at individual regression coefficients so we can keep doing it for all the regression variables or regression coefficients. We can start with the beta 1 hat then we can look at beta 2 hat and so on to beta k hat. So that is the somewhat tedious process and sometimes you may also have an existing model and when you report the existing model to your supervisor.

He may say that you have considered only a model with 2 variables why do not you consider or build a model with 5 variables? So what I am trying to say is we can use the matrix linear algebra concepts to do this pretty efficiently rather than do 1 variable or 1 regression variable at a time which is a somewhat tedious process. We can first analyze the model with a certain bunch of variables and that would be an existing model.

And then we can also see the impact of adding another bunch of variables to the already existing model and we can then decide whether adding the additional bunch of variables also has any impact or value addition to the regression model. So normally the simpler the model, the less number of variables the model has. It is elegant and it is convenient to use and it is also efficient. So you have done lot of work and then reduced complicated process by describing its dependence with only a few selected variables and when you present this model to let us say to the management, the people there may be a bit disappointed.

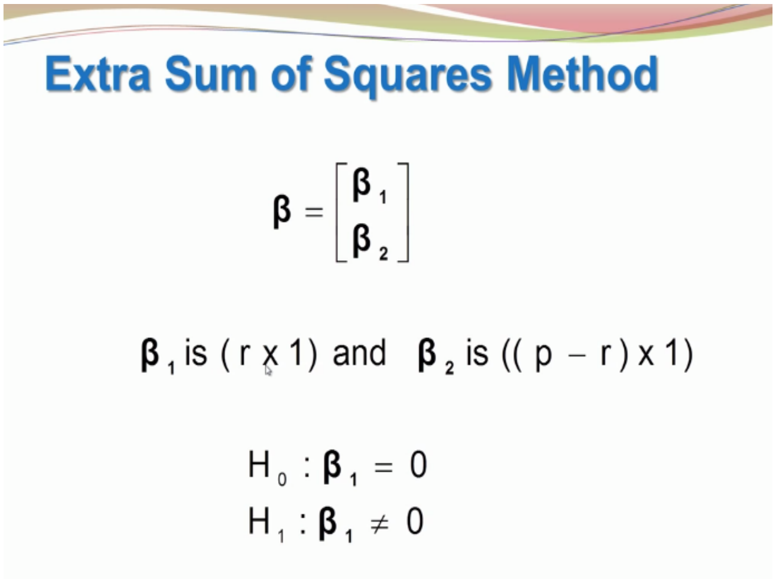
We thought it is such a complicated process why do you have only few variables describing the response. It looks like other parameters or other regressor variables also might influence the experiment. So why do not you go back and check your model. So what we can do is instead of adding 1 regressor variable by considering the effect of 1 regression coefficient at a time we can take a whole bunch of regressor variables with their associated regression coefficients and use a method called as extra sum of squares approach to see the impact on the process response.

So what we are doing is, we are going to conduct a hypothesis test to see whether the new bunch of regression coefficients are indeed valuable and if the test says that none of the new added regression coefficients are significant all of them may be pretty much taken to be 0 then you may go to the management and say look my original model was in fact adequate. There was absolutely very negligible impact of considering the effect of additional variables.

So what we do here is something which may be a bit difficult for people who are not familiar with linear algebra, but actually it is very simple. So let us look at the beta column vector which is comprised of 2 sub vectors if you can call it like that beta1 and beta 2. So beta 2 is column vector and beta 1 is also another column vector. When you put them one below another it leads to the complete column vector beta.

So beta 2 is a preexisting or model, which is already exists and beta 1 is the set of regression coefficients which you want to add to an already existing model.

(Refer Slide Time: 40:07)



Extra Sum of Squares Method

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix}$$

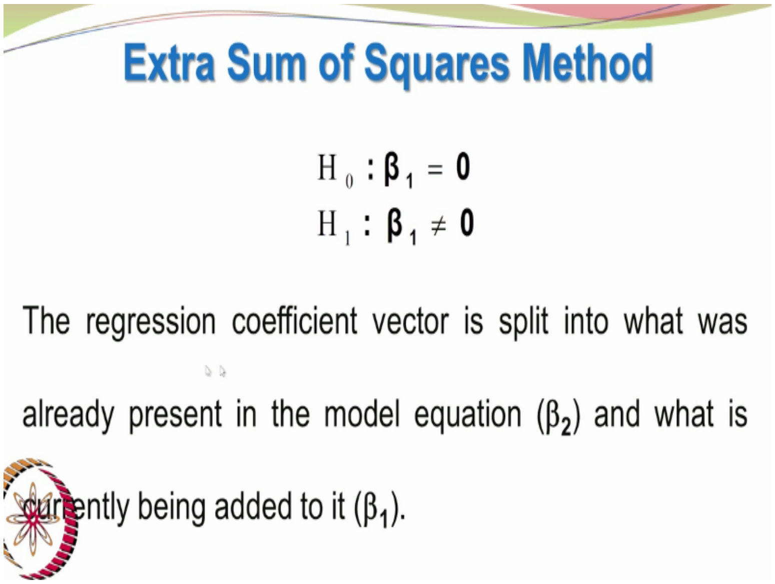
$\boldsymbol{\beta}_1$ is (r x 1) and $\boldsymbol{\beta}_2$ is ((p − r) x 1)

$$H_0 : \boldsymbol{\beta}_1 = 0$$
$$H_1 : \boldsymbol{\beta}_1 \neq 0$$

So let us say that beta 1 comprises of r regression coefficients and beta 2 is comprising of p - r regression coefficients. So we say that H0 beta1 = 0 and H1 beta 1 != 0. There is a small difference here from what we have done earlier. Earlier we were looking at scalars or just single values beta 1, but now I am putting beta 1 in bold that means, it is a vector comprising of r regression coefficients beta 1 hat, beta 2 hat so on to beta r hat.

So we are saying that the entire bunch of entities in that beta1 column vector = 0 and the alternate hypothesis says that beta1 != 0 and so you are having the new model represented by beta1 and the already existing model by beta2.

(Refer Slide Time: 41:21)



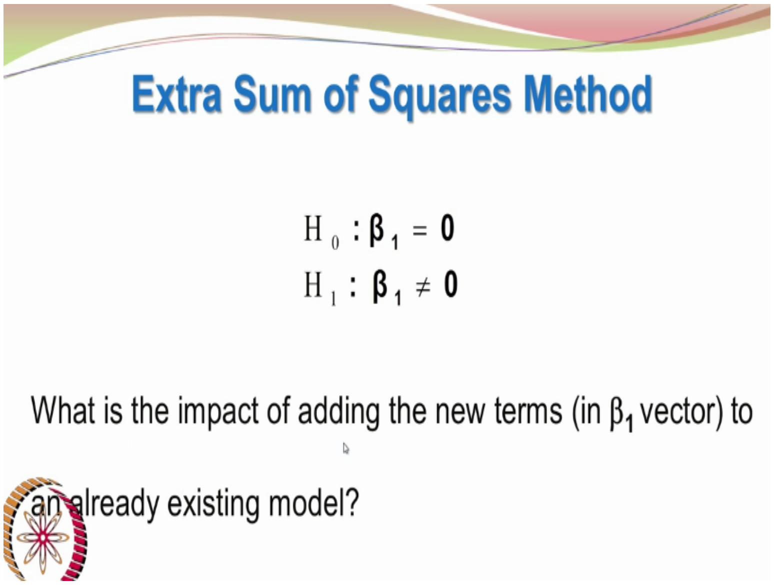
Extra Sum of Squares Method

$$H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$$
$$H_1 : \boldsymbol{\beta}_1 \neq \mathbf{0}$$

The regression coefficient vector is split into what was already present in the model equation (β_2) and what is currently being added to it (β_1).

So the regression coefficient vector beta is split into what was already present in the model equation beta 2 and what is currently being added to it which is beta1.

(Refer Slide Time: 41:36)



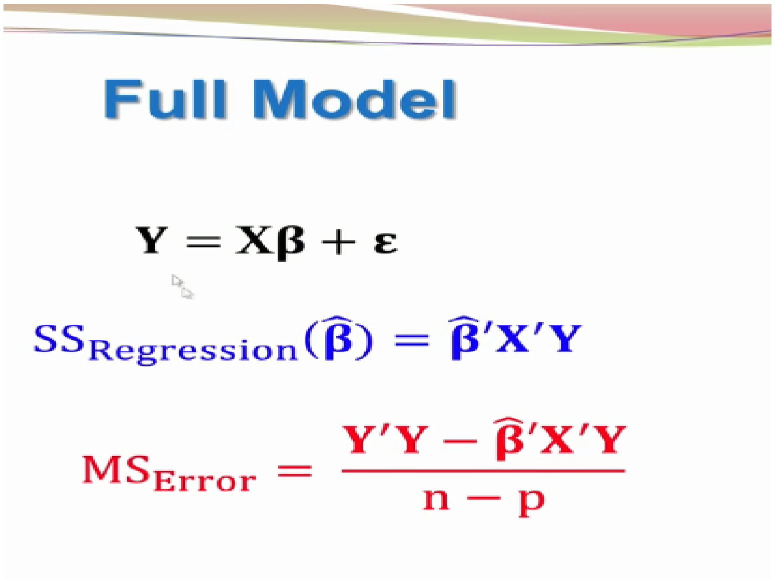
Extra Sum of Squares Method

$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$

What is the impact of adding the new terms (in β_1 vector) to an already existing model?

So we want to see what is the impact of adding the new terms in beta 1 vector to an already existing model?

(Refer Slide Time: 41:44)



Full Model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$SS_{Regression}(\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y}$$

$$MS_{Error} = \frac{\mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y}}{n - p}$$

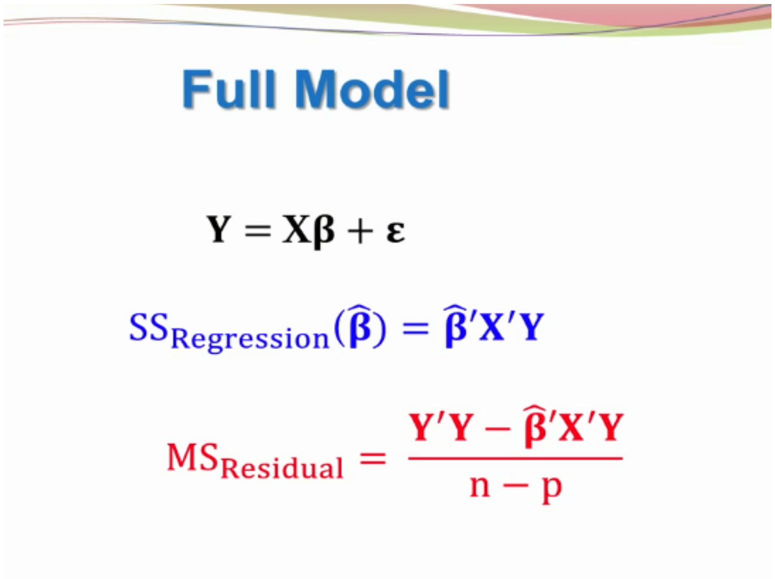
So what we do here is we first look at the full model so that is what we have to do first. We know the sum of squares of regression including the parameter beta hat 0 as beta hat prime x prime y. Let me sort of revise. When you include the intercept also we have beta hat prime x prime y, but if you want to exclude the parameter intercept beta hat 0 then you have to subtract n y bar square, but you are doing it here.

You are considering all the parameters including the intercept beta hat 0 that is why you have sum of squares of regression as beta hat prime, x prime y and then you have the mean square error as y prime y - beta hat prime x prime y and that you scale it by the n - p degrees of freedom and also another thing you have to notice whether you consider beta hat 0 the mean square error does not really care because the n y bar square you subtract it from y prime y you are also subtracting from beta hat prime x prime y so that the n y bar square actually cancels out.

So whether you consider ny bar square or not consider ny bar square it does not really matter to mean square error because you are subtracting consistently n y bar square from y prime y and also from beta hat prime x prime y so that thing actually cancels out and so this mean square does not really bother about it. In other words, it does not really care whether you are considering the model with the intercept of without the intercept.

So we have the mean square residual or the mean square error here and let me sort of make a correction here to be consistent with what I had written earlier I will change this mean square error to mean square residual since both of them have the same starting alphabet r we use MSE, but we use mean square residual when we use the full form.

(Refer Slide Time: 44:33)



Full Model

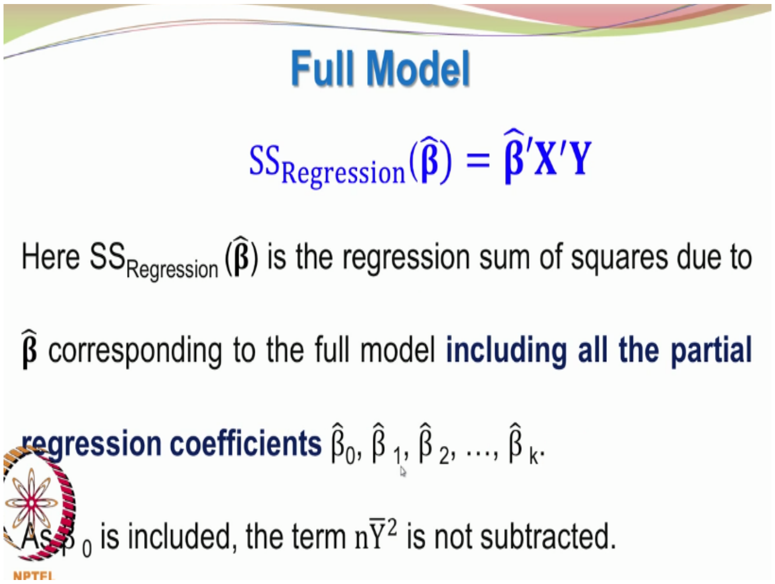
$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$SS_{Regression}(\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y}$$

$$MS_{Residual} = \frac{\mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y}}{n - p}$$

So, I would like to conclude by saying that the mean square residual does not really depend upon whether you have considered the actual total sum of squares or the corrected sum of squares. Be it total sum of squares or regression sum of squares. If you are considering the corrected sum of squares the n y bar square will consistently cancel out here and here, but if you are using it well and good. No problem we are considering the parameter beta hat 0 and the mean square residual value will be unchanged.

(Refer Slide Time: 45:13)



Full Model

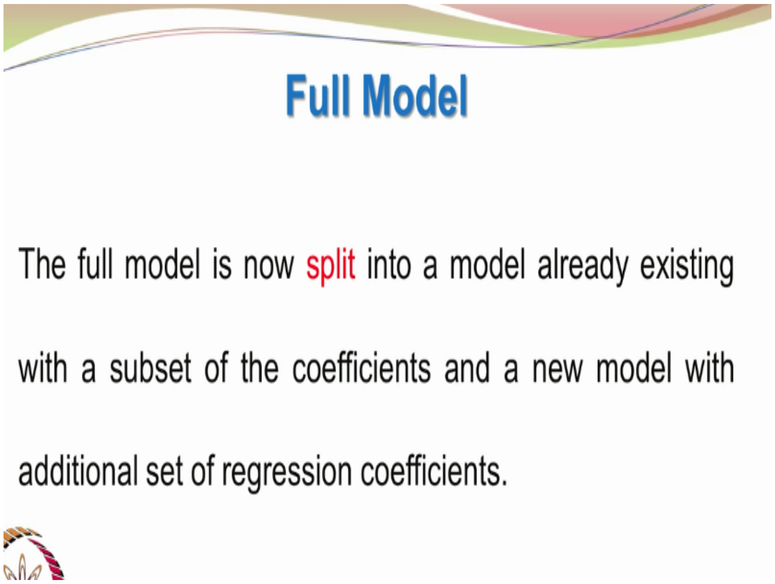
$$SS_{Regression}(\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y}$$

Here $SS_{Regression}(\hat{\boldsymbol{\beta}})$ is the regression sum of squares due to $\hat{\boldsymbol{\beta}}$ corresponding to the full model **including all the partial regression coefficients** $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k$.

As $\hat{\beta}_0$ is included, the term $n\bar{Y}^2$ is not subtracted.

So the sum of squares of regression beta hat 0 is a regression sum of squares due to beta hat not corresponding to the full model including all the partial regression coefficients beta hat 0, beta hat 1, so on to beta hat k. So as beta hat 0 is included, the term n y bar square is not subtracted.

(Refer Slide Time: 45:33)

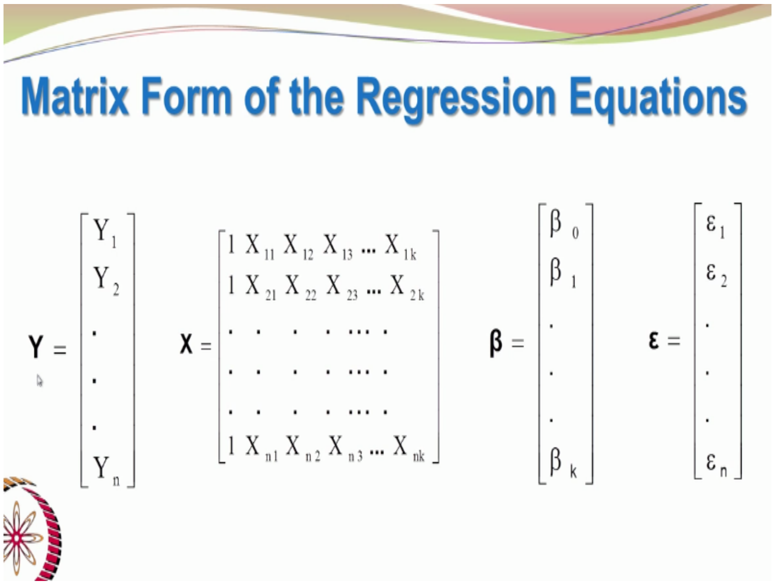


Full Model

The full model is now split into a model already existing with a subset of the coefficients and a new model with additional set of regression coefficients.

So there are full model this is what we have being considering until now is now split into a model already existing with a subset of the coefficients and a new model with additional set of regression coefficients.

(Refer Slide Time: 45:49)



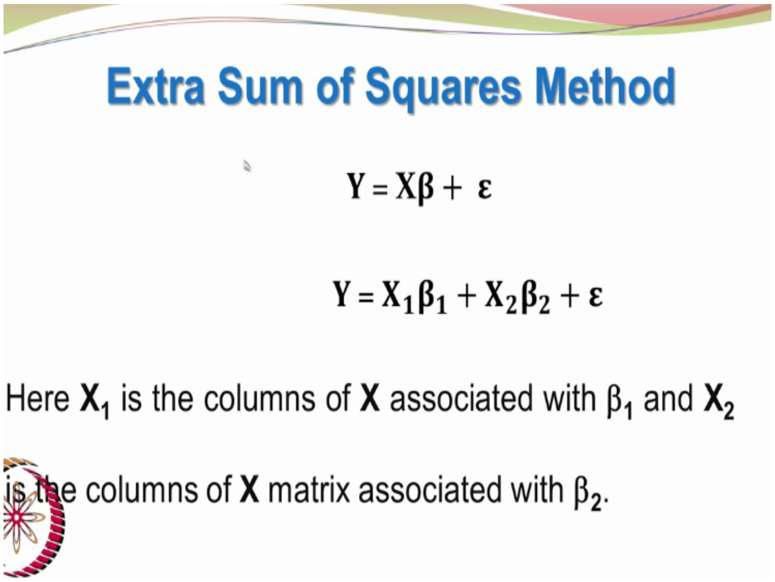
Matrix Form of the Regression Equations

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_n \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & X_{11} & X_{12} & X_{13} & \dots & X_{1k} \\ 1 & X_{21} & X_{22} & X_{23} & \dots & X_{2k} \\ . & . & . & . & \dots & . \\ . & . & . & . & \dots & . \\ . & . & . & . & \dots & . \\ 1 & X_{n1} & X_{n2} & X_{n3} & \dots & X_{nk} \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ . \\ \beta_k \end{bmatrix} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_n \end{bmatrix}$$

Let us look at the full model. This is the vector of responses. This is the x matrix. This is the beta column vector which is the full set of regression coefficients and then you also have the error column vector you might not probably for whatever reason you might not have considered beta 0 and beta 1. You might have started your model with beta 2, beta 3, so on to beta k only that is your existing model, but then your boss would say what happened to the intercept what happened to the variable 1.

They also look important to me from an (()) (46:31) point of you why do not you include it. So then new model would be adding beta 0 and beta 1.

(Refer Slide Time: 46:42)



Extra Sum of Squares Method

$$Y = X\beta + \varepsilon$$

$$Y = X_1\beta_1 + X_2\beta_2 + \varepsilon$$

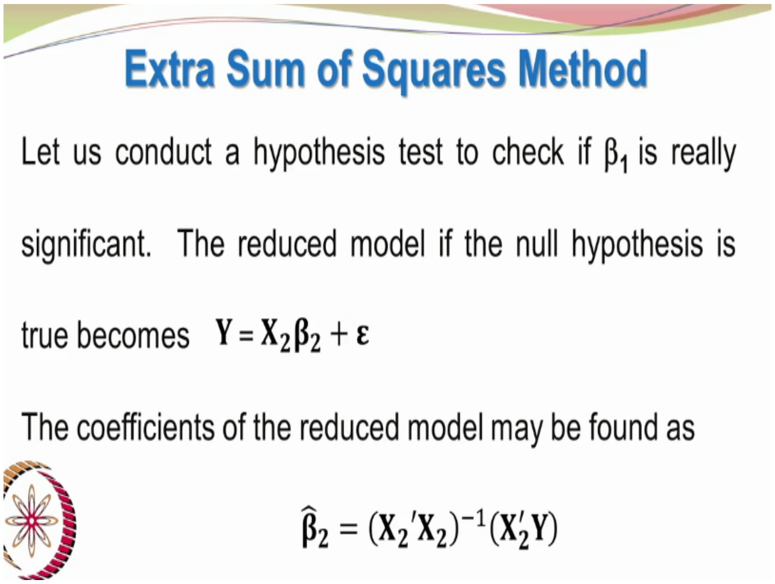
Here $\mathbf{X}_1$ is the columns of **X** associated with β_1 and $\mathbf{X}_2$ is the columns of **X** matrix associated with β_2.

So what we do is this is the full model. We split this into x1 beta1 + x2 beta2 + epsilon. Actually this is not simple algebraic addition. This is involving matrices. This is the overall response vector y and then you have the x1 beta1 + x2 beta2, beta1 is the new model. It is a new vector comprising of the new regression coefficients and beta2 is the column vector comprising of the old regression coefficients and x1 is again a submatrix of x which are dealing with the regressor variables corresponding to beta 1.

So x1 is the columns of x associated with the beta1 and x2 is the columns of x matrix associated with beta 2. So just let us go back for example I told you that the new model was based on beta0 and beta1 based on the boss's recommendation. So then the x1 matrix will be the submatrix obtained by taking the first 2 columns that is what we are looking at. The intercept will be associated with just 1 and then the beta1 would be associated with x11, x21, so on to xn1.

For example, you will have beta 0 + beta1 x11 and then you will have beta 0 + beta 1 x21. So you are considering the effect of the intercept and you are also considering the effect of the first regressor variable x1. So this is how we do it. The old model already had these regressor variables starting from x2, x3, so on to xk. So x2 would be associated with beta2, x3 would be associated to beta3 and xk would be associated with beta k.

(Refer Slide Time: 49:26)



Extra Sum of Squares Method

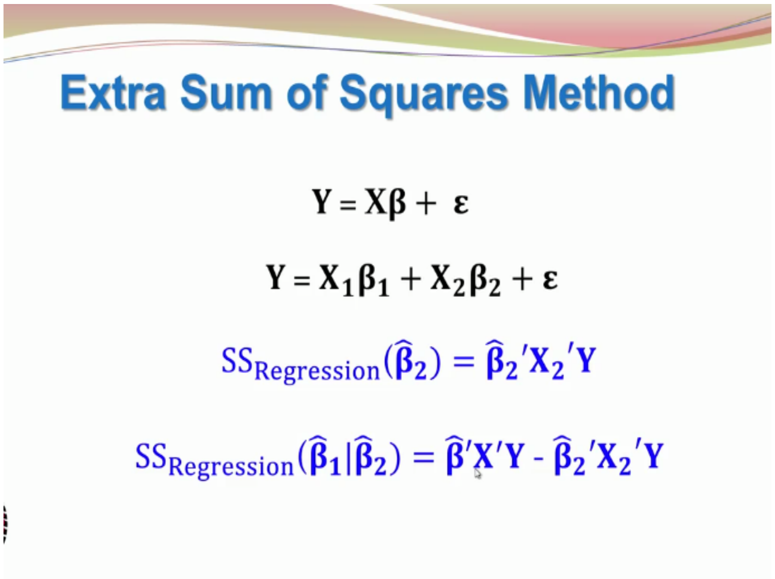
Let us conduct a hypothesis test to check if β_1 is really significant. The reduced model if the null hypothesis is true becomes $\mathbf{Y} = \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$

The coefficients of the reduced model may be found as

$$\hat{\boldsymbol{\beta}}_2 = (\mathbf{X}_2{}'\mathbf{X}_2)^{-1}(\mathbf{X}_2'\mathbf{Y})$$

So we have to consider hypothesis test to check if beta1 is really significant. The reduced model if the null hypothesis is true becomes y = x2 beta2 + epsilon. Null hypothesis means that there is no value addition on adding the elements in beta 1 so you are okay with this model y = x2 beta2 + epsilon. So the coefficients of the reduced model can be found in the usual way by x2 prime, x2 inverse, x2 prime y.

(Refer Slide Time: 50:05)



Extra Sum of Squares Method

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\mathbf{Y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$$

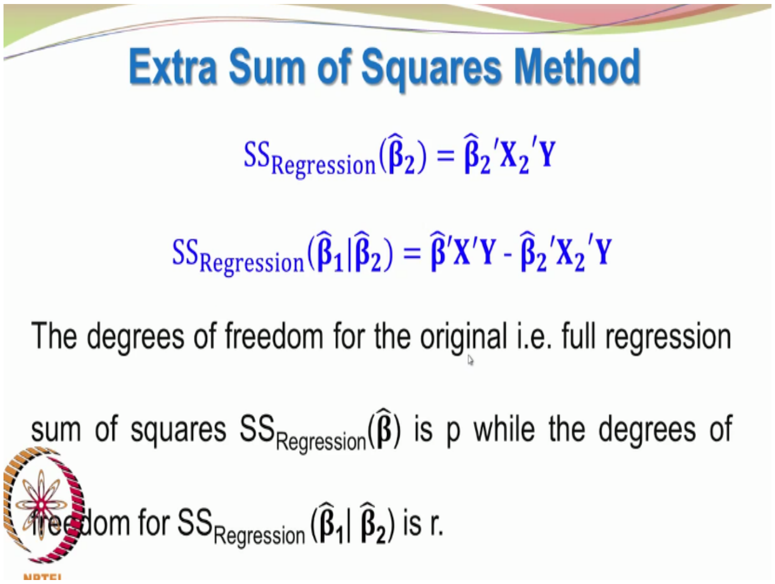
$$SS_{Regression}(\hat{\boldsymbol{\beta}}_2) = \hat{\boldsymbol{\beta}}_2{}'\mathbf{X}_2{}'\mathbf{Y}$$

$$SS_{Regression}(\hat{\boldsymbol{\beta}}_1|\hat{\boldsymbol{\beta}}_2) = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} - \hat{\boldsymbol{\beta}}_2{}'\mathbf{X}_2{}'\mathbf{Y}$$

So we have the extra sum of squares method. This is the full model. This is the model split into contributions from beta 1 and then beta2 or in fact it may come again. It is the contributions from beta 2 and then from beta 1 and the sum of squares of regression due to beta2 hat alone is pretty straight forward. It is beta hat2 prime x2 prime and so the sum of squares of regression due to beta hat1 given beta2 hat already present in the model is the beta hat prime x prime y - beta hat2 prime x2 prime y.

In order to find the regression contribution by the new model we take the full model first the regression sum of squares from the full model and then from that we subtract the regression sum of squares from the already existing model so that the difference will give you the contribution to the regression sum of squares from the new model.

(Refer Slide Time: 51:23)



Extra Sum of Squares Method

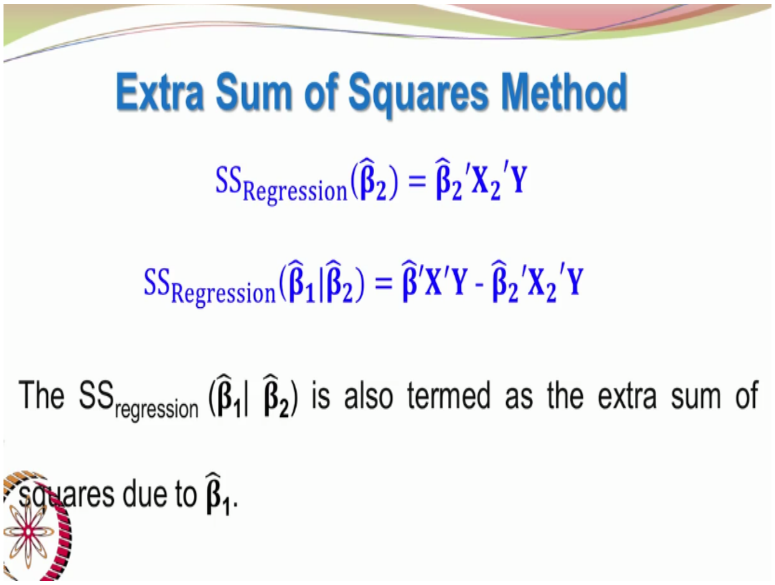
$$SS_{Regression}(\hat{\beta}_2) = \hat{\beta}_2'X_2'Y$$

$$SS_{Regression}(\hat{\beta}_1|\hat{\beta}_2) = \hat{\beta}'X'Y - \hat{\beta}_2'X_2'Y$$

The degrees of freedom for the original i.e. full regression sum of squares $SS_{Regression}(\hat{\beta})$ is p while the degrees of freedom for $SS_{Regression}(\hat{\beta}_1|\hat{\beta}_2)$ is r.

And the degrees of freedom for the original or the full regression sum of squares is p. It includes all the parameters while the degrees of freedom for the sum of square of regression beta hat1 given beta hat2 is r, because if you recollect we split the beta column vector into 2 parts into 2 column vectors, the first column vector was of size r and the second column vector was of size p - r. So the full model is having degrees of freedom of p and the new model is having degrees of freedom of r.

(Refer Slide Time: 52:03)



Extra Sum of Squares Method

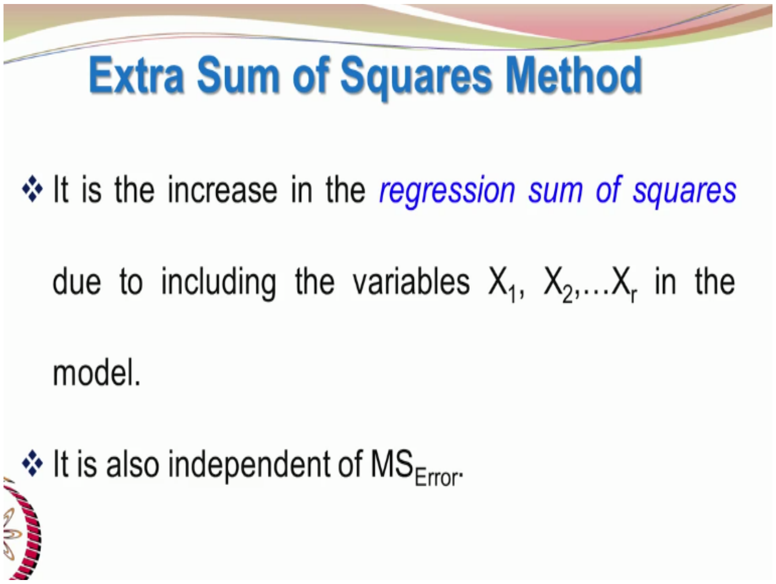
$$SS_{Regression}(\hat{\beta}_2) = \hat{\beta}_2'X_2'Y$$

$$SS_{Regression}(\hat{\beta}_1|\hat{\beta}_2) = \hat{\beta}'X'Y - \hat{\beta}_2'X_2'Y$$

The $SS_{regression}(\hat{\beta}_1|\ \hat{\beta}_2)$ is also termed as the extra sum of squares due to $\hat{\beta}_1$.

So the sum of squares of regression beta hat1 given beta hat2 is also termed as the extra sum of squares due to beta hat 1. So what is the extra regression sum of squares brought in by the new set of regression coefficients.

(Refer Slide Time: 52:25)



Extra Sum of Squares Method

- It is the increase in the *regression sum of squares* due to including the variables X_1, X_2,…X_r in the model.
- It is also independent of MS_{Error}.

So it is the increase in the regression sum of squares due to including the variables x1, x2, so on to xr in the model and it is also independent of mean square error. So this concludes our lecture on the hypothesis testing in linear regression. It is quite elegant and you can see that whatever we did in our earlier phase or the first phase of the design of experiments namely the hypothesis testing is also playing a very valuable role here. Thanks for your attention.