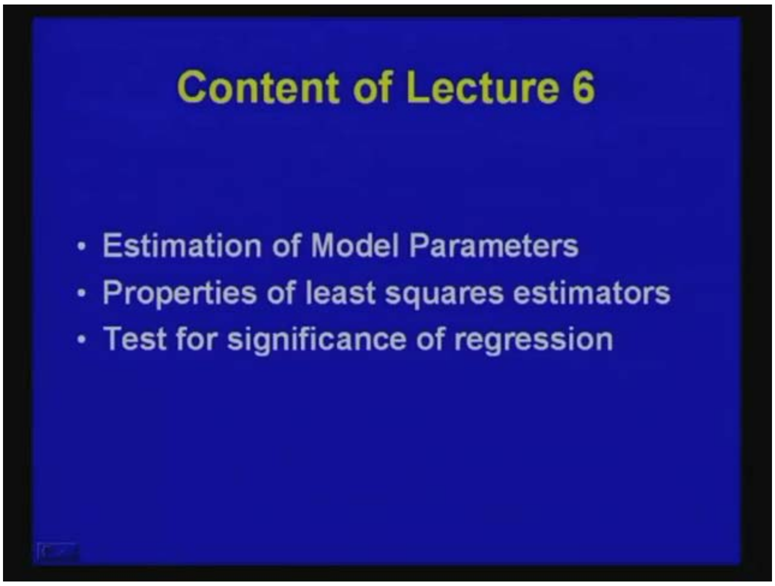
Regression Analysis
Prof. Soumen Maity
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture - 6
Multiple Linear Regression

Hi, we shall be talking on Multiple Linear Regression, so this is my first lecture in multiple linear regression.

(Refer Slide Time: 01:04)

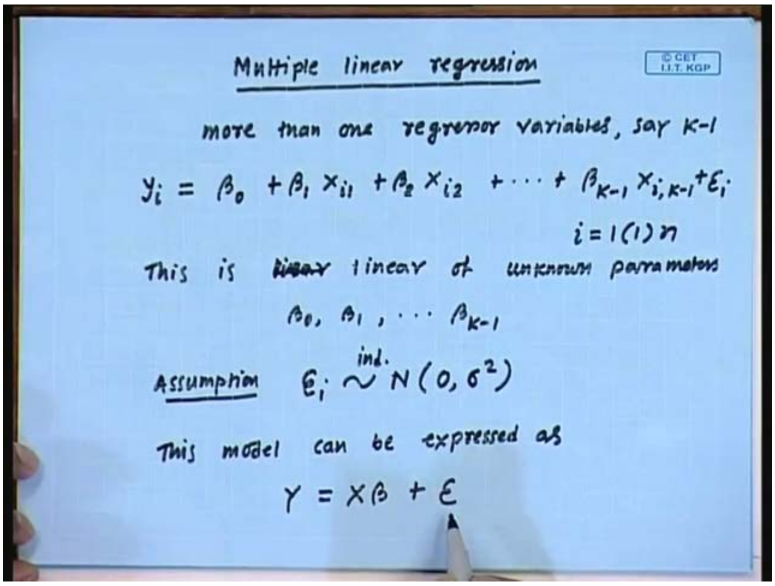


And the content of today's lecture is estimation of a model parameters in multiple linear regression and properties of least squares estimators, and then we will be talking about once the model has been fitted, will be talking about testing for the significance of regression. So, let me recall the disney toy problem, there we had only 1 regressor variables that is the amount of money spent on advertisement.

Well we have observed that the regression variable, there that means the amount of money is spent on advertisement that explained 80 percent of total the variability in response variable, that is the variability in the sales amount. And the 20 percent of the variability, in the response variable that remained on explained, so that we say that is you know the S S residual part. Now there could be one more the regress able variable, which can explain the part of that unexplained variability in this response variable, that means the part of that 20 percent of the variability, which remain on explained in that case.

And the one important regressive variable could be you know the number of cells parts in you employee, so also in the in most of the cases in practice. There you will have more than one regressive variable; and in that case, we need to move for multiple linear regression, let me explain the multiple linear, I mean multiple linear regression model well.

(Refer Slide Time: 03:53)

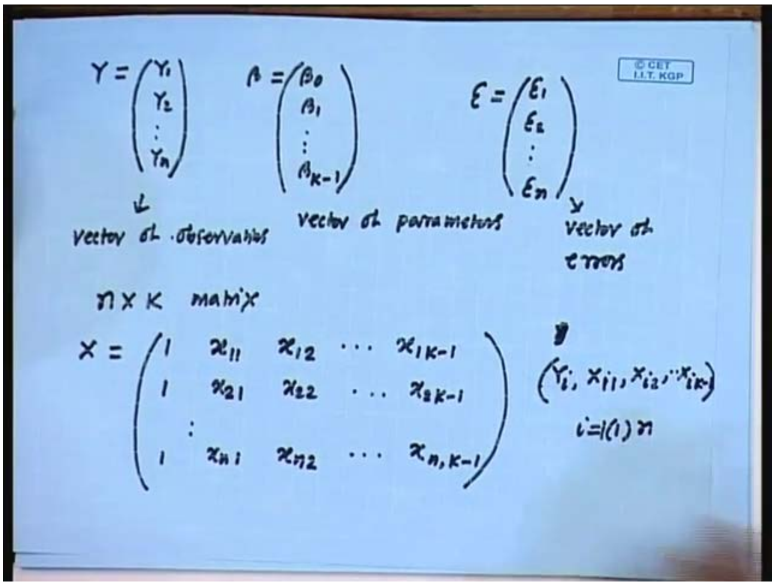


So, the situation here is that instead of 1 regressor here, we have more than 1 regressor variable, say we have K minus 1 regressor variable and the deniled form of multiple linear regression is y i equal to beta naught plus beta 1 X i 1. So, this one is the first regressor variable plus beta 2 X i 2 up to beta K minus 1 X i K minus 1 plus epsilon and this model is the you know basically, for the i th observation, so i runs from 1 to n.

So, since we have no the more than 1 regressor variable, then that is why, we call it a multiple regression and since the model is linear that is why, it is called multiple linear regression. But, one should be careful you know, this is a linear function, when linear means, it is a the linear function of the unknown parameters, hear the unknown parameters are beta naught beta 1 beta 2 and beta there are K unknown parameters. So, this one is this model is linear of unknown parameters beta naught beta 1 up to beta K minus 1.

So, it is not, I mean if the model is linear in unknown parameter then only called you know linear model well and we make the assumption that, the error this is the i th error, which follows normal distribution with I mean 0 and the variance sigma square. And they are also independent all the epsilon eyes are independent, so now will defined some matrices.

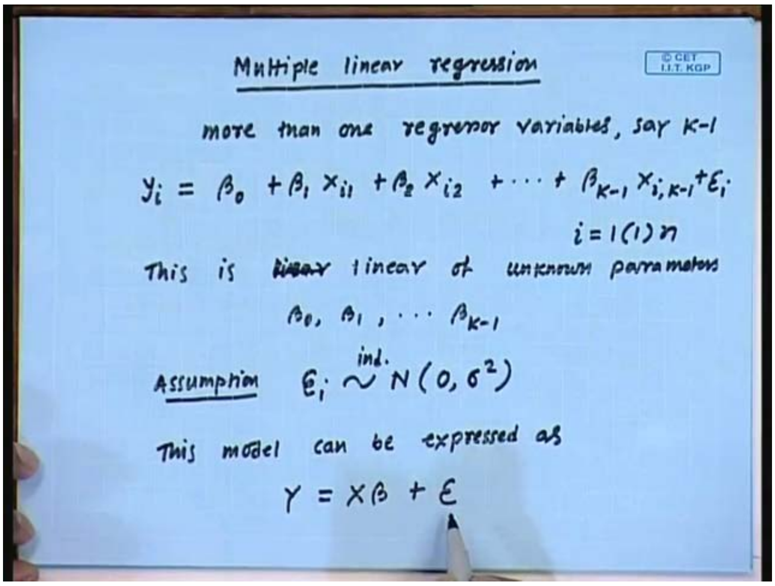
(Refer Slide Time: 08:07)



Y equals to Y 1 Y 2 Y n, these are the observations n observations, beta equal to beta naught beta 1 K minus 1. So, this is the K cross 1 vector, this is the vector of parameters and this is the vector of observation and epsilon equal to epsilon 1 epsilon 2 up to epsilon n. So, this one is the vector of errors and also it defined n cross K metrics, which is equal to X that is 1 x 1 1 x 1 2 x 1 K minus 1, so these are basically observation, the first observation is on regerssor 1, this is the first observation regressor 2, this is the first observation on regressor K minus 1.

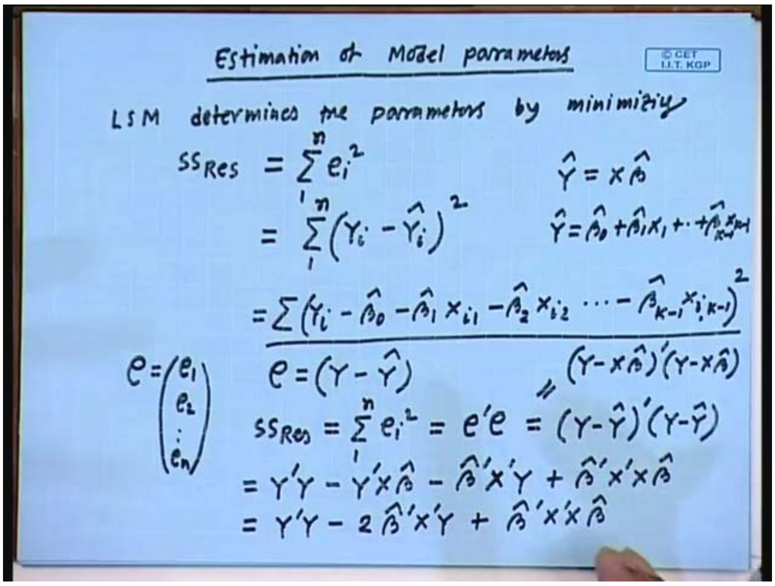
So, 1 x 2 1 x 2 2 x 2 K minus 1 and similarly, 1 x n 1 x n 2, this is first 1 to the n i th observation x n K minus 1, so this is you know, symmetric of known form, because all the value is unknown well we well, we have the data like, we have the data of this form Y i x i 1 x i 2 x i K minus 1. So, we have this data for i equal to 1 2 n and we have to using this state of observations.

(Refer Slide Time: 11:58)



We have to feet a model like this, it is a multiple linear regression model and this model can be, now using the metrics rotation, this model can be expressed as you know, Y equal to x beta plus epsilon well. This is the vector of observations, vectors of parameters, vectors of errors well, this is the model, we have 2 fit and this is the model in metrics form. We are giving the data of this form and using this data, we have to find, we have 2 fit the model, that means, the basically, we have 2 estimate the parameters well.

(Refer Slide Time: 13:28)



Now, will be talking about the estimation of model parameters, I mean in the multiple linear regression, there is almost know a new concept every all the concept, we have already talked about in the simple linear regression. So, like simple linear regression, here also, you know the estimating, we will be estimating parameters using the least squares method, so the parameters are determined by minimizing the S S residual well.

So, the least square method determines the parameters, so hear instead of you know the only beta naught and beta o n, we have basically, K unknown parameters that is the only difference. So, the least squares method determines, the parameter by minimizing by minimizing S residuals residual, so what is S S residual, So, S S residual is the basically, it is e i square from 1 to n, which is again nothing but Y i minus Y i hat square 1 to n right.

Now, suppose my fitted model is beta naught hat plus beta 1 hat X 1 plus beta K minus 1 hat X K minus 1, so this quantity is equal to, so S S residual is equal to Y i minus beta naught hat minus beta 1 hat X i 1, because you know, I am talking about the i th fitted value beta 2 hat X i 2 like this beta K minus 1 hat X i K minus 1 hole square. Now, you know, the least square method determines, the parameter by minimizing this S S residual, what will do here is that I mean will also represent this S residual in metrics form, for that will defined the residual effecter e, which 1 is basically e 1 e 2 e n.

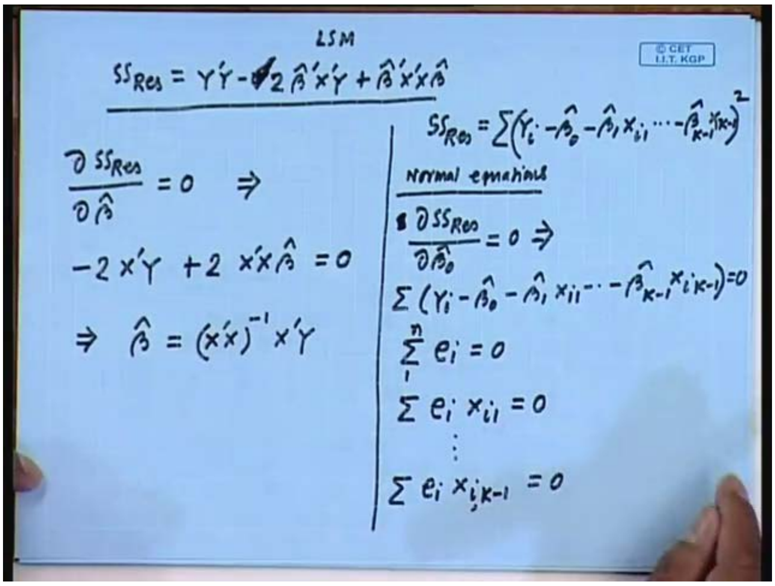
So, e i is the i th residual, so e can be written as e is Y minus Y hat, so this is Y is the vector, I mean vector of observations and the vector of observations, for the repeated value well, this is my e. Now, S S residual is equal to summation e i square 1 to n, so we are basically, you know talking about, another I mean, how to express this thing in terms of the metrics rotation. So, this can be written as e prime e right, if you yeah now this one is equal to equal to Y minus Y hat prime Y minus Y hat and this can be written as Y prime Y minus Y prime X beta hat minus beta hat prime X prime Y plus beta hat prime X prime X beta hat.

I just missed one step in between, the basically you know I am replacing Y hat by Y hat by this expression, so metrics rotation in this nothing but Y hat equal to x beta hat. So, you replace Y hat by Y minus X beta hat prime Y minus X beta hat. So, and then you have this expression here, you know, this is you can take that, this is 1 cross 1 metrics, that means, it is a scalar quantity similarly, this one is also 1 cross 1 matrices. So, the basically, it is a scalar, so everything is scalar here.

So, this 2 quantity, this 2 are same, so this can be written as Y prime Y minus 2 times, I am tacking this form beta prime, X prime Y plus beta hat prime X prime X beta hat. This is my S S residual metrics form, but if you do not understand this one, here is your S S residual here is your S S residual, which is very similar to the simple linear regression only, we have this additional terms, because of the additional regressor variable and the same thing is represented here in metrics form well.

So, we have 2 different representation of the S S residual and now, we need to defined said this S S residual with expect to the unknown parameters. So, there are you know, there are K unknown parameters, so we have to define said this is S S residual with expect to each unknown parameter and that will give you, normally questions. So, then you will be having K normally equations and k unknown, so using this K normal independent, normal equation, you can find out get the estimator, for the unknown parameters, K unknown parameters.

(Refer Slide Time: 23:04)



So, here is the you know process, you know well, least squares method well, so what we have that, we have 2 Y, I am explaining both the things, if we do not understand the metrics representation. So, here the S S residual is of the form Y prime Y minus Y prime minus 2 times beta hat prime X prime Y plus beta hat prime X prime X beta hat. So, this is the metrics representation of the S S residual and another way to represent same thing is that, like the usual technique S S residual is equal to summation Y i minus beta naught hat minus beta 1 hat minus beta 1 hat X i 1.

And similarly go up to beta k minus 1 hat X K minus 1 sorry, you to put, I here that is all, so this is my S S residual. Now, what I will D Y S that, I will differentiate to get the normal equations, I will defined said this S S residual with respect to beta naught fast. So, S S what, I have 2 do that is I will defined said this S S residual with expect to beta naught hat and equal to 0, this is the normal equation, which implies or which gives you define said this with expect to beta naught hat that, will give you summation Y i minus beta naught hat minus beta 1 hat X 1 X i 1 minus beta K minus 1 hat X i K minus 1 equal to 0.

So, this is my first normal equation and you know this term is nothing but e i Y i minus Y i hat, so this can be also return, this first normal equation can be also return has summation, e i i from 1 2 n equal to 0, so this is my first normal equation. Similarly next you defines, it is this quantity, I mean you defines, it is residual with expect to beta 1 hat that, will give you the normal equation, summation e i X i 1 equal to 0, so this is very similar to the simple linear equation.

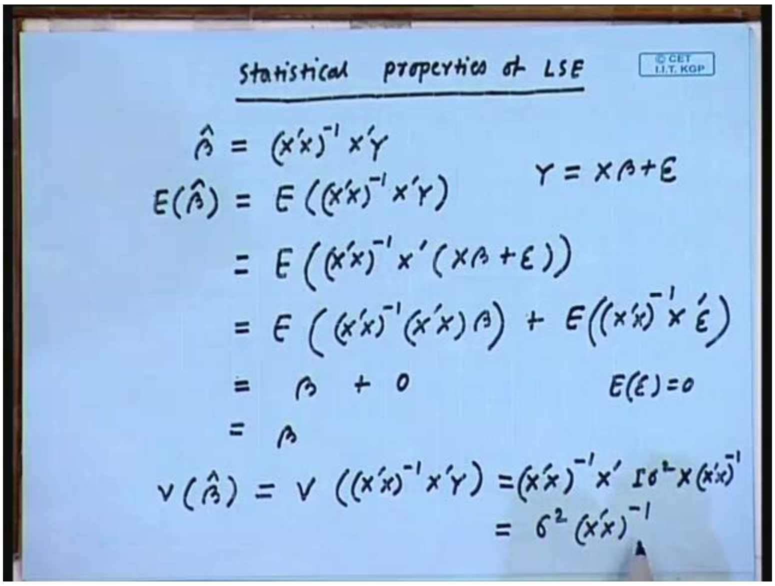
And similarly, you defines it with respect to beta 2 hat and you go up to beta K minus 1 hat and the final normal equation is summation e i X i K minus 1 equal to 0, so here you have K normal equations and you have K unknown parameters and all this normal equations are independent, So, solving this K normal equations will give you K unknown parameters beta naught, beta 1 beta up to beta K minus 1, this is the usual tech, I mean form, what we have used in the case of simple liner regression.

Now, I mean you know, I will go for the metrics representation of the same thing, what I will do is that I just define said that S S residual, which has been explained you know, which as been expressed in terms of metrics rotation and I will defined and said that with respect to beta hat well. This is my S S residual with expect to in terms of this is presented in terms of metrics rotation.

Now let me define said this one S S residual with respect to beta hat, defined said in respect to beta hat equals to 0, which implies or which gives you know defines side, this with respect to a beta hat that will give you minus 2 X prime Y. So, I am define setting, so this is independent of beta, so while depending this term, it is 0, now you defined set the second term. So, that will give you minus 2 X prime by when you define it third term that will give you plus 2 times X prime X beta hat and you equate, this equal to 0, I mean you can write down the metrics forms in detail and define state, you will get this one.

So, from here you know, hear it is combine to since the same thing written here and here, now finding the beta hat, in this metrics representation is easy from is this normal equation. So, this is in fact, you know it consist of K normal equations, this K normal equations, so from hear, we get this implies beta hat is equal to X prime X inverse X prime Y. So, here is the least square estimator of the unknown parameters beta naught beta 1 up to beta K minus 1. And if you solve this K normal equations, you will be getting the same thing, now will be talking about the statistical property of this least square estimator.

(Refer Slide Time: 30:59)



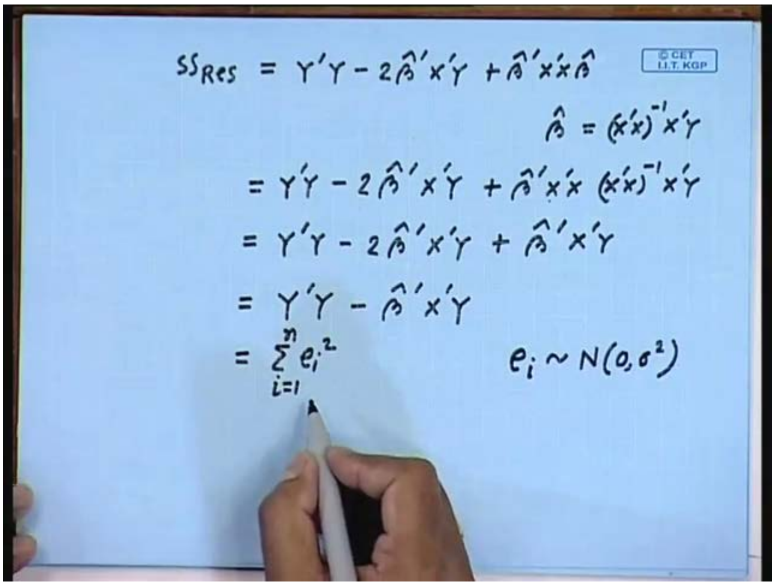
So, of least square estimator, so what I am going to do is that, I am just going to prove that, whatever estimator, we have obtained that means, beta hat, which is equal to X prime X inverse X prime Y. This is an unbiased estimator of beta, let me prove that, unbiased means to prove that expectation of beta hat is equal to beta. So, expectation of this one is expectation of X prime X inverse X prime Y, now what is Y is in metrics rotation equal to X beta plus epsilon, so this one is basically equal to E X prime X inverse X prime X beta plus epsilon right.

So, this can be written as expectation of X prime X inverse X prime X beta plus expectation of X prime X inverse X prime epsilon well, this quantity, this is going to be identity. So, expectation of and this one is equal to beta only plus expectation of this term or this random variable here, you know this epsilon is random variable, which follows normal distribution with expectation 0. And variance sigma square with the expectation of epsilon, we know that expectation of epsilon is equal to 0, sure that is equal to 0, which means this is equal to beta.

So, here prove that expectation of beta hat is equal to beta, that means, beta hat the estimator the least square estimator, we have obtained that is an unbiased estimator of beta. So, next we are going to derive the variance of this estimator, so the variance of beta hat is equal to the variance of X prime X inverse X prime Y right. So, this one is going to be, we know the variance of Y is equal to sigma square well, the variance covariates metrics, this Y is basically vector and observation vector.

So, the variance of the whole thing and this one is independent of, I mean this is constant on with as not inform any random variable. So, this one is going to be X prime, X inverse, X prime i sigma square into X prime X inverse, so this ones you know, this can be finally, return as sigma square into X prime X inverse. Because, X prime X and X prime X inverse will cancel out, so it is sigma square into X prime X inverse well. So, next will be talking about, the different representation of S S residual.

(Refer Slide Time: 36:48)



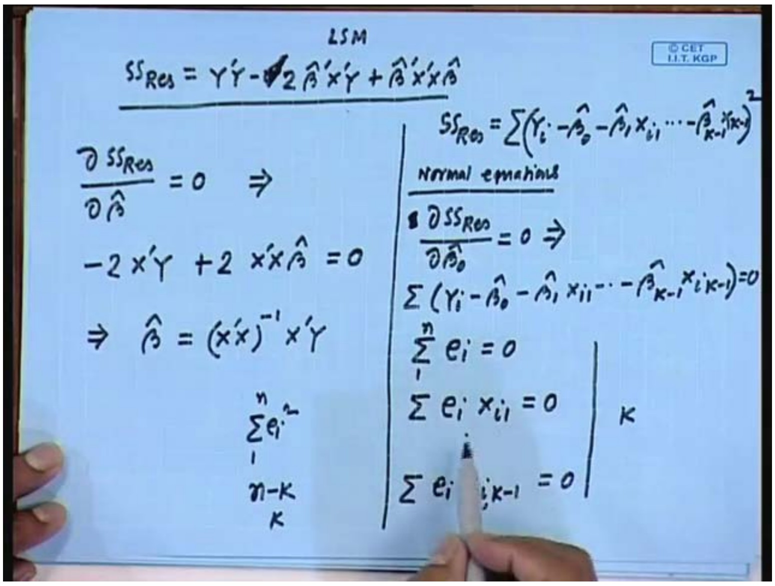
S S residual in metrics rotation, this is we observed, we derived that this is equal to Y prime Y minus 2 beta hat prime X prime Y pulse beta hat prime X prime X beta hat right. Now, you know, we know that beta hat is equal to X prime X inverse X prime Y, so I am going to put value here, just to simplify this expression, this is equal to Y prime Y minus 2 times beta hat prime X prime Y plus beta hat prime X prime X and now will plug this beta hat hear, this is going to be X prime X inverse X prime Y.

So, just I replaced this by beta hat by this expression, so this quantity is now, Y prime Y minus 2 beta hat prime X prime Y plus beta had prime X prime Y, because this is identity well. So, simplified form is Y prime Y minus beta hat prime X prime Y, because this is identity well. So, the simplified form is Y prime Y minus beta hat prime X prime Y, see the this same thing, you know this one's nothing but summation e i square and here is the metrics representation of the summation e i square well.

What we know is that, we that the summation e i follows, normal distribution with means, 0 and variance sigma square, now let me talk about, what is the degree of freedom of S S residual well, i equal to 1 2 n. So, we know that, the summation S S residual is summation e i square from 1 to n and e i follows, normal with mean 0 and variance sigma square.

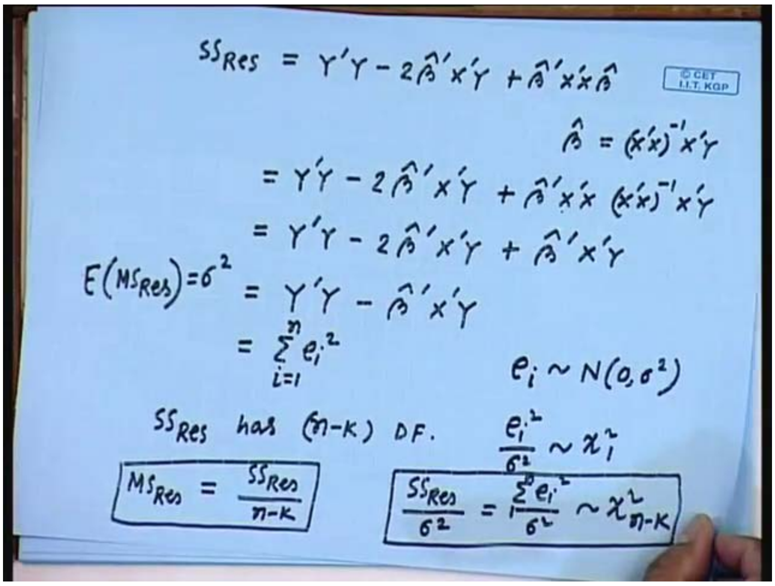
Now, I want to talk about the degree of freedom for this S S residual S S residual is some of in e i square, but just now we have derived that, you know this e i is they satisfy K constant, that means, there are K normal equations involving e i. So, here all the e i are I mean, you do not have freedom of choosing, all the e i is in e i independently, you can choose n minus K of them. You have the freedom of choosing n minus on n minus K of the n e i is and the remaining K have to be chosen in such a way that, they satisfy those K constants well. So, in the case of simple linear regression, we had 2 constants, on e i that is why you had the freedom of choosing n minus 2 e i is independently then the remaining 2, we have chosen, we have chosen such a way that, they satisfy the constant, those 2 constant.

(Refer Slide Time: 41:55)



And here instead of 2 constant on e i, we have basically, n constant and here, at the these n constants, you know sorry, K constant, these are the K constant we have. So, you cannot this e i square here, I mean you cannot choose n of them, you do not have the freedom of choosing all the n e is you can choose, you have the freedom of choosing n minus K e i is independently and then the remaining K have to be chosen in such a way that, they satisfy this K constant. So, basically, while losing K degree of freedom, because of this K K constant, on the residuals well.

(Refer Slide Time: 42:37)

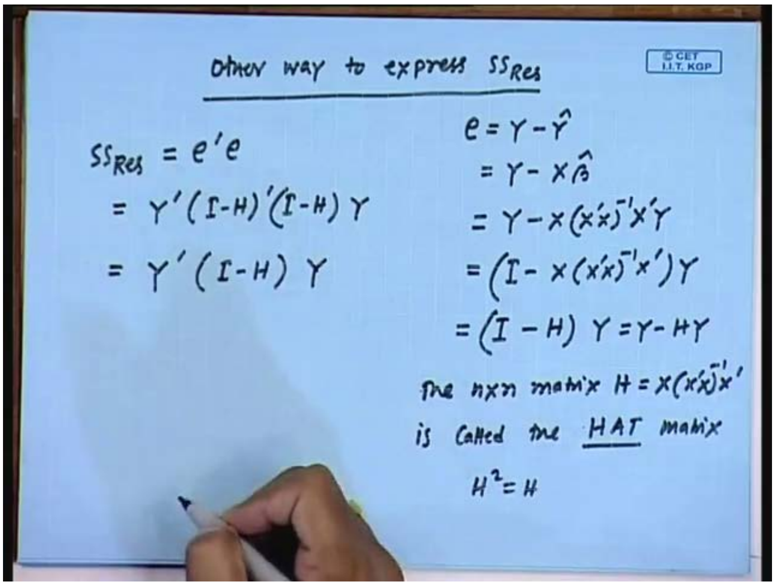


So, that explain that the S S residual here, S S residual has n minus K degree of freedom right, now we know that, this follows from here, you can say that e i square by sigma square follows, chi square 1. And from here, you can say that S S residual S S residual by sigma square, which is nothing but summation e i square by sigma square, this follows 1 to n, this follows chi square n minus K naught n, because of those K constant well and we have this result.

And also you can defined the mean square, residual mean square that is M S residual, which is obtained by dividing the S S residual by degree of freedom n minus k. So, and we know that, it is not difficult to prove that, this M S residual is an unbiased estimator of sigma square that means, we can it is easy to prove that, expected value of M S residual is equal to sigma square. So, we have an unbiased estimator for sigma square as well before moving to the statistical significance of the regressive model, I just want to give another representation of S S residual.

So, the S S residual can be represented in several ways, you know just simply, you can write summation e i square i equal to 1 2 n then we had the metrics representation of S S residual and now, I am going to give another representation of the S S residual, which is in terms of the hat metrics right. Now, I do not have any use of this expression in future maybe will be using this expression, let me give another just another representation of the S S residual, using the hat metrics.

(Refer Slide Time: 45:40)

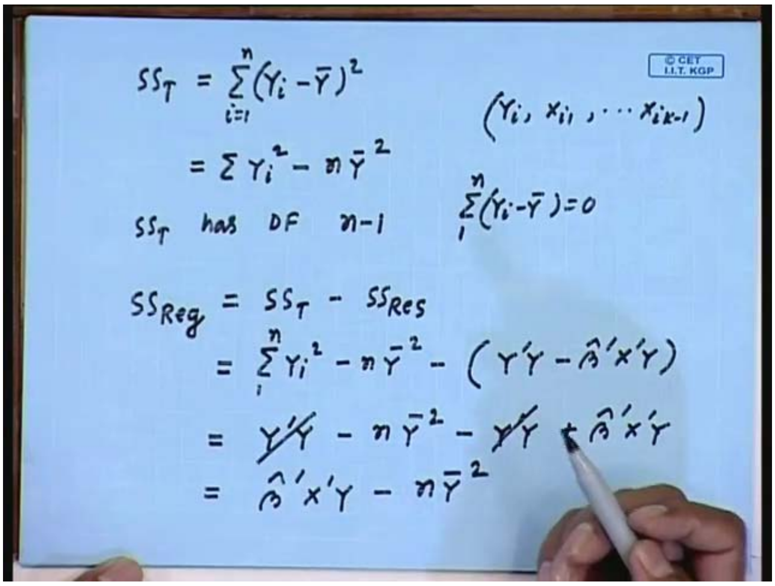


So, you say that, this is other way to express S S residual, so well what we know is that, we know that, e equal to in matrix rotation e equal to Y minus Y hat. So, Y this is basically the observation vector and this one is going to be Y minus, what is why hat is nothing but X beta hat right. Now, this one is going to be Y minus X, now will replace this beta hat by it is estimator X prime X inverse X prime Y right. So, what I got is that, this is equal to I minus X prime X inverse X prime right, this one is using the notation of H metrics, this is i minus H into Y.

So, this this is an n cross metrics, the n cross n metrics H, which is equal to Y t S into X prime X inverse, X prime is called the hat metrics here, this is called the hat metrics, because you know ultimately, what we had is that, here it is equal to Y minus H Y. So, this is called hat metrics, because it this H metrics terms from Y to, so this one is H Y nothing but Y hat, so this metrics comes from Y to Y hat that is why, it is called hat metrics. And now you know, you can you can prove that, you know H square equal to H well, so this is the specialty of this metrics.

Now ,S S residual can be returned as S S residual is equal to e prime e, which is equal to Y prime i minus H prime i minus H Y and you can check that, this i minus H prime i minus H nothing but i minus H, so this can be returned as Y prime i minus H Y. So, this is know, the another way to express the S S residual right and as I said, at I am not going to use this expression of S S residual, in terms of hat metrics at this movement, I will be using in future well. Next, I will be moving to the sort of you know and I approach to test the statistical significance of the regressive model, for that I will be preparing with I will first, I will talk about S S total and then the S S regression well.

(Refer Slide Time: 50:03)



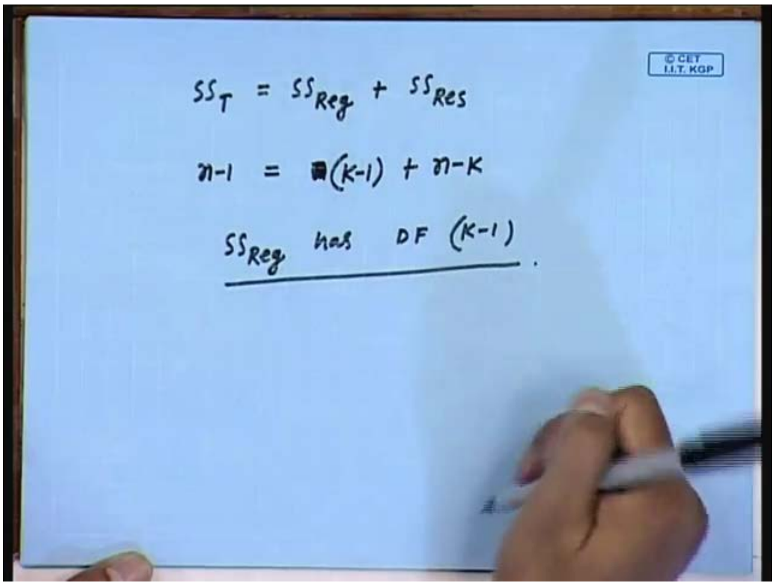
So, well what is S S total here, this is S S total is nothing but the variation in the observation or variation in the data, which is nothing but Y i minus Y bar whole square, i equal to 1 to n. So, we have n observations of the form Y i and then X i 1 and then X K minus 1, so this S S T is nothing but the variation in the response variable well. So, this can be returned as summation Y i squire minus n Y bar square, so this is not difficult to check well.

What is the degree of freedom of this S S total has degree of freedom some of n terms, but of course, it satisfy the constant that, Y i minus Y bar, this is equal to 0. So, you do not have the freedom to choose all the terms, I mean Y 1 minus Y bar Y 2 minus Y bar up to Y n minus Y bar. So, you can choose n minus of you have the freedom of choosing n minus 1 of them and then the n th one as to chosen in such a way that, it is satisfy this constant, so S S has degree of freedom n minus 1.

Now, what is S S regression S S regression is equal to S S total minus S S residual right well, so S S total equal to, we know that is equal to summation Y i square 1 2 n minus n Y bar squires minus S S residual, if you can recall, it is Y prime Y minus beta hat prime X prime Y in metrics rotation. Now, I can also, you know slowly, I mean this can be replaced in, I mean this can also returned as Y prime Y minus n Y bar square. So, this Y bar is nothing but the mean of the observations minus Y prime Y plus beta hat prime X prime Y.

So, this 2 will cancel out and left with beta hat prime X prime Y minus n Y bar square, so we have the expression for S S regression, we have the expression for S S total, we have the expression for S S regression and just we left with the degree of freedom for S S regression.

(Refer Slide Time: 54:12)



What we know is that S S total is equal to S S regression plus S S residual well, so let me say again that, this is the total variability in the response variable and that variability is partisian into 2 parts. One is I mean, how much of the variability in the response variable is explained by the model that is S S residual and the part, which is not being explained by the regression model is called the S S residual well. We want to the model to be, such that we wanted the model to maximize S S regression and then obviously, minimizing S S residual.

So, S S total as degree of freedom n minus 1, we know that S S residual has the degree of freedom n minus K, then the degree of freedom for S S regression is n minus sorry, is equal to K minus 1, so here is the degree of freedom for. So, S S regression as degree of freedom K minus 1 well, so in the next class, I will be talking about the statistical significance of the regression model, in case of multiple linear regression.

Thank you very much.