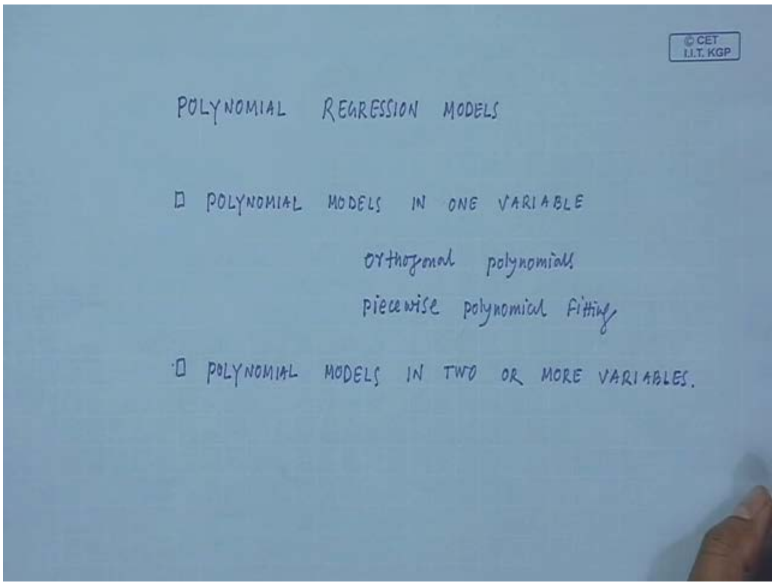
Regression Analysis
Prof. Soumen Maity
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture - 27
Polynomial Regression Models.

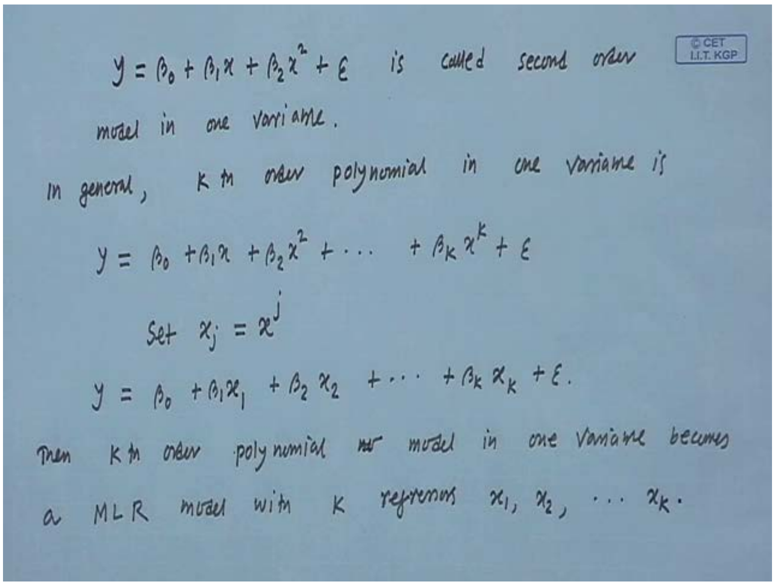
Hi, this is my 1st lecture polynomial regression models and here is the content of this topic.

(Refer Slide Time: 00:31)



So, we will be talking on polynomial models in one variables and orthogonal polynomials, piecewise polynomial fitting and also we will be talking about polynomial models in two or more variables. Well so, polynomial models are used in regression analysis, when the response variable is nonlinear. That means given a set of data x i y i or i equal to 1 to n. 1st you prepare the scat of plot and the when the scatter plot indicates that, relationship between the response variable the regressor variable is nonlinear. Then we need to go for polynomial model.

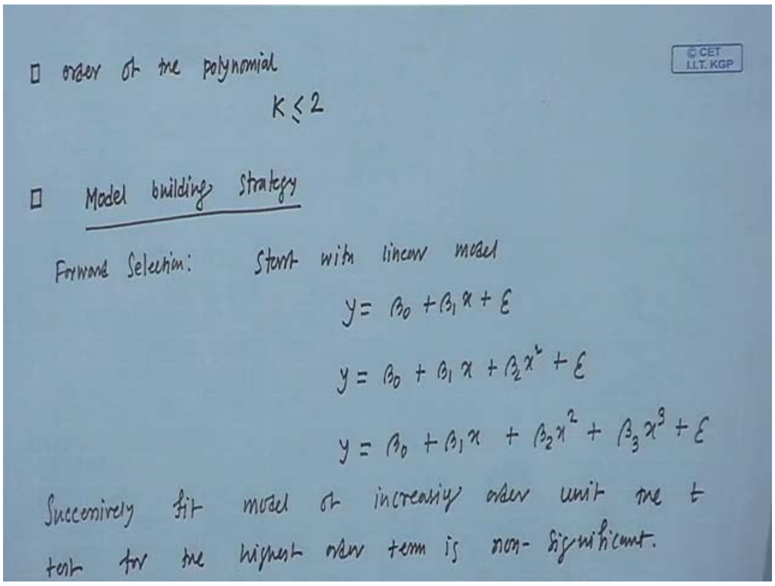
(Refer Slide Time: 01:54)



So, here y equal to beta naught plus beta 1 x plus beta 2 x square plus epsilon is called 2nd order model in one variable. So in general, k-th order polynomial in one variable is y equal to beta naught plus beta 1 x plus beta 2 x square plus beta k x to the power of k plus epsilon. So now, if you put say for example, x set x j equal to x to the power of j then, this can be rewritten as y equal to beta naught plus beta 1 x 1 plus beta 2 x 2 plus beta k x k plus epsilon. So, this one is nothing but, a multiple linear regression model involving k regressors; so, then k-th order polynomial model in one variable becomes a multiple linear regression model with k regresses x 1, x 2, x k ok.

So, there is I mean fitting a k-th order polynomial is same as fitting a multiple linear regression model involving k regressors. But, there are several important consideration while fitting a multiple linear regression model; the 1st one is what would be the order of the polynomial because, we talking about fitting a k-th order polynomial so, we need to decide about the order of the polynomial. So, here the suggestion is that we would like to keep the order of the polynomial as low as possible, so when the response variable is nonlinear that means when the scatter plot indicates that is a non-linear relationship between the response and the regresses variable. 1st you try for some transformation to make the model linear if that fails then, you can you go for a 2nd order polynomial ok. So, we do not recommend polynomial fitting of very higher degree usually, the order of the polynomial is less than or equal to 2.

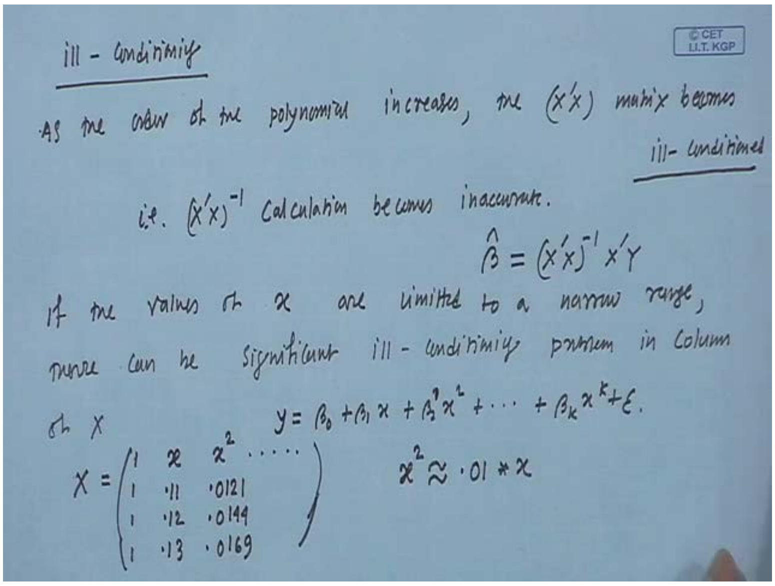
(Refer Slide Time: 06:59)



The next issue is, you know that, is called the model building strategy, so the 1st one was order of the order of the polynomial and here, we sort of decided that, you know its recommended that k is usually less than or equal to 2. So, 2nd one so, this is the order polynomial of k the 2nd one is model building strategy; model building strategy ok. So, this is also I mean degree of the polynomial, sorry the order of the polynomial this is called forward selection: so, what this forward selection suggest that, you start with linear model, start with linear model. That means you start with y equal to beta naught plus beta 1 takes plus epsilon and then, you go for the 2nd order polynomial say y equal to beta naught plus beta 1 x plus beta 2 x square plus epsilon.

Then after fitting this model the 2nd order model, you need to test the significant of the highest or a term that is beta 2 here. If beta 2 is significant then, you go for a 3rd order model say y equal to beta naught plus beta 1 x plus beta 2 x square plus beta 3 x cube plus epsilon. But, if you see that beta 2 is not significant, then you can stop here; so you will stop in the 2nd order model. So, this is what the algorithm say the general so, ultimately its successively fit model of increasing order until the t test for the highest order term is non significant. So, this is what the model building strategy is.

(Refer Slide Time: 10:38)

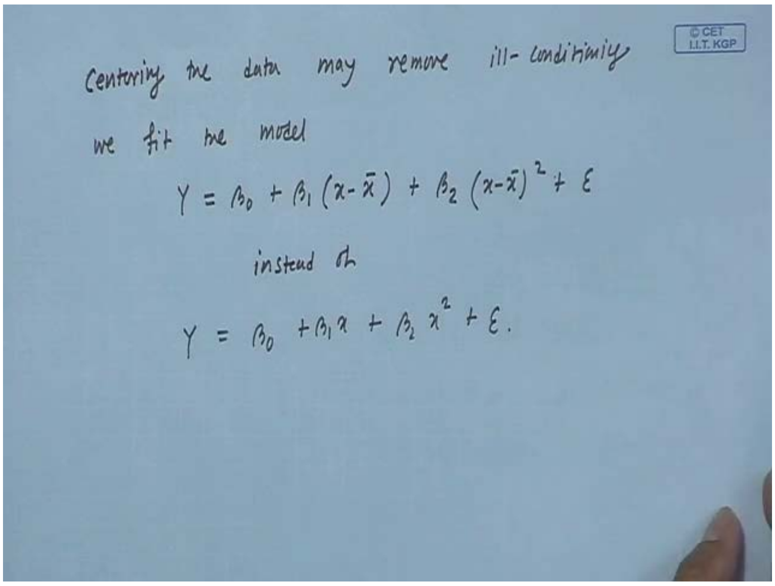


And next another condition that is called the ill conditioning ok, so here, as the order of the polynomial increases the X prime, X matrix becomes ill condition. So, what is the meaning of this it is become ill conditions? Is that, the X prime, X matrix is becomes the near singular that means, that is same as X prime, X inverse calculation becomes inaccurate ok. Because, we need to compute this one so because, the estimation of regression coefficient beta hat is equal to X prime X inverse X prime Y; so we need to compute this inverse. But, as the order of the polynomial increases this

X prime X matrix become near singular, so the computation of inwards becomes inaccurate ok. The specific case if, the values of x are limited to a narrow range there can be significant ill conditions; ill conditioning problem in column of X. Let me give an example of this one, you must have understood that, we are talking about the polynomial y equal to beta naught plus beta 1 x plus beta 2 x square and beta k x to the power of k plus epsilon. So, here is the X the coefficient matrix X is, the 1st column is 1 the 2nd column corresponds to x values, the 3rd column is corresponds to x square values like this right. So, if you have say the x value very a limited narrow range, suppose the x values are like: 0.11, 0.12, 0.13 these are anyway 1. Then, the x square value is 0.0121, 0.0144, 0.0169.

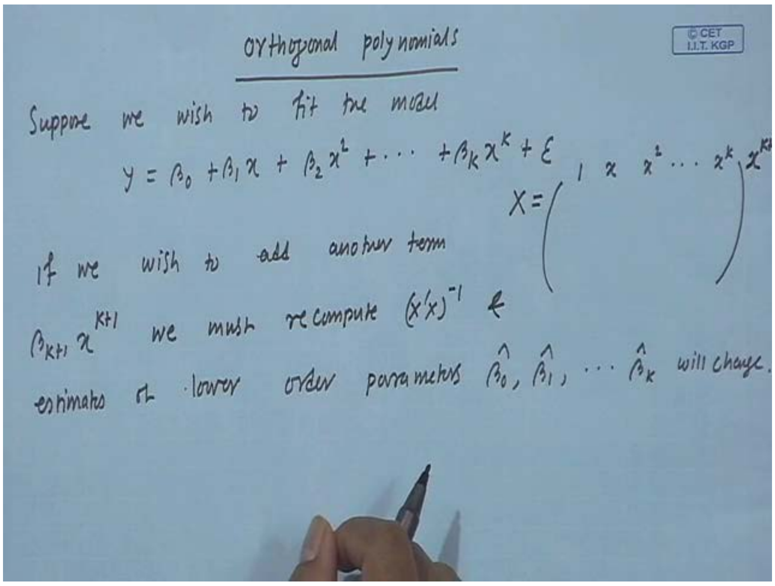
So here, you can see let me write then, the x square column this is approximately equal to 0.01 time x column. So, here you can see there is near dependency between these two columns, so that means the matrix become near singular ok. So that is why it says that value of x are limited to a narrow range; if, the x values from narrow range there could be significant ill conditioning problem in the column of x ok.

(Refer Slide Time: 16:22)



And, how to remove this ill conditioning problem is that, you know one way is to do that centering the data may remove ill conditioning. That means, we fit the model say Y equal to beta naught plus beta 1 x minus x bar plus beta 2 x minus x bar whole square plus epsilon. You fit this model for the data at centered instead of Y equal to beta naught plus beta 1 x plus beta 2 x square plus epsilon. So, this is you know one way to remove ill conditioning problem ok.

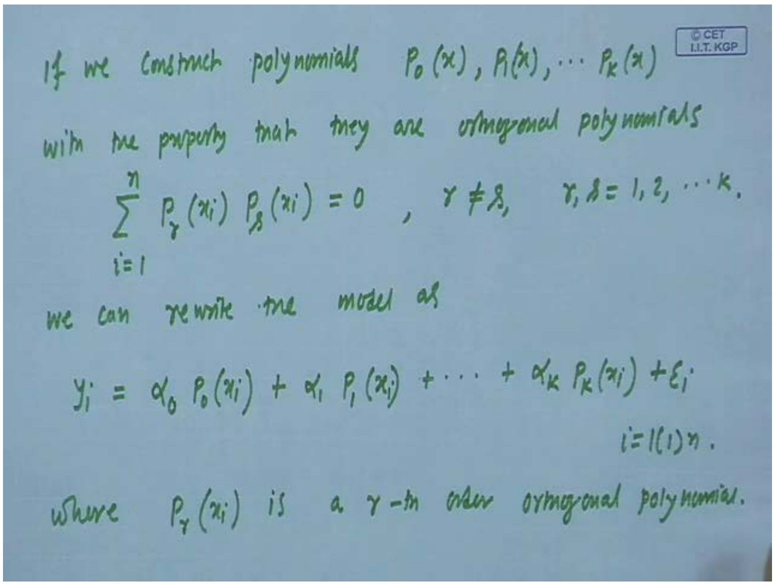
(Refer Slide Time: 17:49)



So, next we talk about orthogonal polynomial, suppose, we wish to fit the model: y equal to beta naught plus beta 1 x plus beta 2 x square beta k x to the power of k plus epsilon. And here you have observed that the X the coefficient matrix X is sort of 1 x x square up to x to the power of k ok. So now, if we wish to add another term like beta k plus 1 x to the power of k plus 1 then, we must recompute X prime X inverse. Because, once you add this term in the polynomial you have to add the one more column x to the power of k plus 1. So, you have to recompute the new X prime X inverse and also estimates of lower order parameters beta naught hat, beta 1 hat, beta k hat; this thing will change.

Once you add 1 higher order term in the polynomial model. So, how to that means you have suppose, you start with the 2nd order model and then you compute beta naught, beta 1, beta 2 now, if you add say 3rd order term like beta 3 x to the power of 3 in the model then, again you have to re compute X prime X inwards and the lower order parameters also.

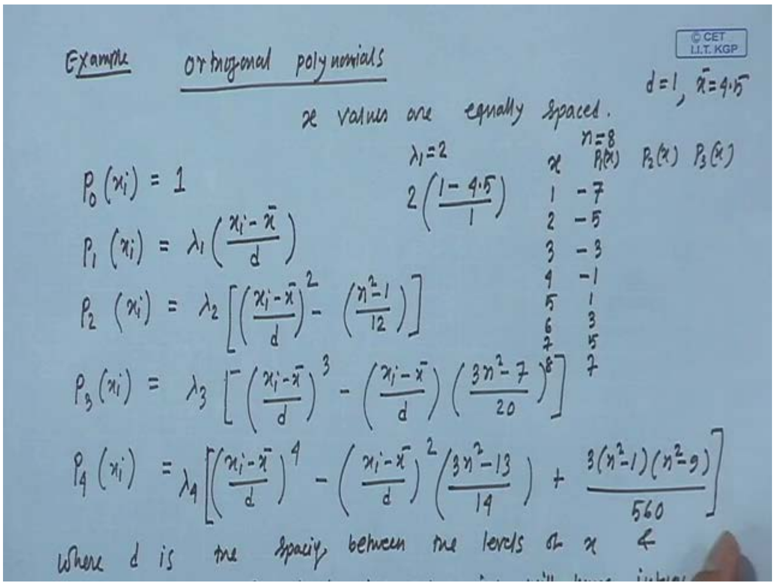
(Refer Slide Time: 20:30)



So how to avoid this problem; so, one way to do this use orthogonal polynomial. So, here so, if we construct polynomials P naught x; so P naught x is a polynomial of degree order 0, P 1 x of order 1 P k x with the property that, they are orthogonal polynomial that means, summation P r x i into say P s x i is equal to 0 for i equal to 1 to n for r is not equal to s and r s there from 1, 2 up to k. So, if you can find polynomial like this you know, they are called orthogonal polynomial. Then we can, rewrite the model as y i equal to alpha naught P naught x i plus alpha 1 P 1 x, so this is a orthogonal polynomial of order or degree 1 plus alpha k P k x i.

So, we replacing x by P 1 x and x to the power of k by P k x i ok, so that means, this is a polynomial orthogonal polynomial of degree or order k plus epsilon i for i equal to 1 to n. So where, P r x i is r-th order orthogonal polynomial so, instead of fitting the model alpha naught plus alpha 1 x alpha 2 x square plus alpha k x to the power of k plus epsilon. We are fitting the model p naught sorry alpha naught p naught x i alpha 1 P 1 x i plus alpha k p k x i. And these are equivalent problem and these are orthogonal polynomial ok. Let me just before we you know learned how to compute or how to estimate this regression coefficients; let me give example of orthogonal polynomial, to make this I mean to get better idea about this polynomial, orthogonal polynomials ok.

(Refer Slide Time: 25:56)



So, here is the example of orthogonal polynomial, so here the condition is that, the x values are x values are equally spaced. So here the zeroth order polynomial P naught x i is equal to 1, P 1 x i is equal to lambda 1 x i minus x bar by d, I will explain now why they are orthogonal polynomial. P 2 x i is equal to lambda 2 x i minus x bar by d minus n square minus 1 by 12; this is of order 2 so, this is the 2nd order orthogonal polynomial. And then, P 3 x i equal to lambda 3 x i minus x bar by d to the power of 3 minus x i minus x bar by d into 3 n square minus 7 by 12.

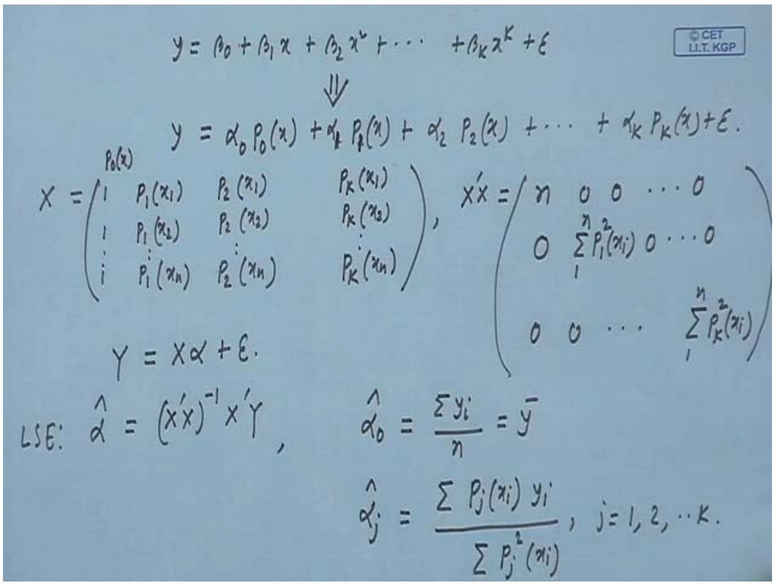
And, let me write one more P 4 square x i is equal to x i minus x bar by d to the power of 4 minus x i minus x bar by d squared 3 n square minus 13 by 14 plus 3 n square minus 1 n square minus 9 by 560. I am sorry, you do not need to remember all these thing, so given a problem you will be given the orthogonal polynomials now, you do not need to memorize this thing, lambda 4. Let me define some terms here, I have used where d is the spacing between: the levels of x and lambda j are chosen so that, the polynomial will have integer values ok.

These are the orthogonal polynomials let me just give what I mean by d lambda 1 say it is for P 1. Suppose, you are given a data with n equal to say 8; you are given 8 observations and you want to find the orthogonal polynomial, for that observation. And it does not matter what is what, are the values of x because, you need to you know that this x values are equally spaced. So you can you can say the x values are just like 1, 2, 3, 4, 5,6,7 and 8; because there are 8 observation.

So, here the d is the spacing between the level of x, so here, d is equal to; in this example d is equal to 1 and the x bar is of course, for this particular case x bar is equal to 4.5 you can check that. Then, what is P 1 x 1? P 1 x 1 is 1 minus 4.5 by 1 into lambda 1 ok, so this is minus 3 point 5 and it says that lambda are chosen so, that the polynomial will have a integer value. So, to make it integer value you take lambda 1 is equal to 2, so 2 into this is minus 7 right. So similarly, if you put 2 here x is equal to you will get minus 5, so this one, this is what my P 1 x and if you put 3 here; you will get minus 3, if you put 4 here then, it is minus 1, if you put 5 then it is 1357.

So this is the how you have to for different n you will have different orthogonal; the values will be different I mean the same orthogonal polynomial of course. So, you can compute P 2 x, P 3 x all these things, so, you know what is d looking at the value of these you can decide about lambda 2 right.

(Refer Slide Time: 33:16)



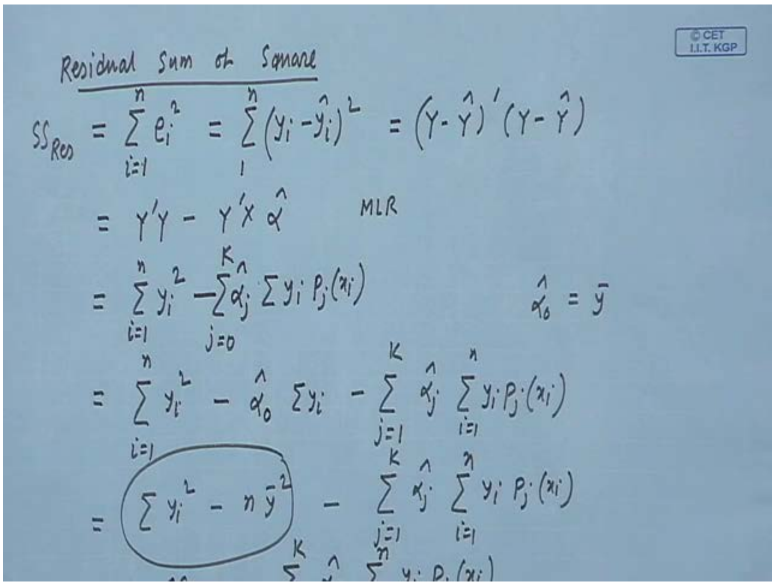
So, my aim not to talk more about this orthogonal polynomial, what I wanted to do is that I had a model like I started with a model y equal to beta naught plus beta 1 x plus beta 2 x square plus beta k x to the power of k plus epsilon. And then, there is some problem some consideration of this model instead of fitting this model I wanted to; I want to fit the model y equal to alpha naught P naught x plus alpha 2 P 2 x sorry P 1 sorry plus alpha 1 P 1 x plus alpha 2 P 2 x plus alpha k P k x plus epsilon. So I want to fit this, I want to find the value of alpha naught, alpha 1, alpha 2, alpha k. So, how do I do that? This is a multiple linear regression model I will write down what is my x the coefficient matrix is 1 1 1 ok the 1st column. And the 2nd column is because P naught x is equal to 1 for all x. Now P 1 x so, this is my P 1 x 1 the 1st observation P 1x 2, for the 2nd observation and P 1 x n for the n-th observation. And similarly, my 3rd column might be P 2 x 1, P 2 x 2 and then P 2 x n and my k-th column that, is P k x 1, P k x 2 and P k x n ok.

So, this is my x matrix now, we will realize the advantages of this orthogonal polynomial and these are orthogonal polynomials right. So then, what is X prime X? X prime X is n 0 0 0 and then the 2nd row is, see this is nothing but, my P naught x that is P naught x 1, p naught x 2, p naught x n. So, this column into this column is since, there are orthogonal that is this term is equal to 0 and the 2nd diagonal element is P 1 x i square of course, 1 to n. And all other elements, so it is become a diagonal matrix right; last 1 is P k x i square 1 to n. So this is my X prime X matrix which is the diagonal matrix and I can write down this one as matrix form Y equal to X alpha plus epsilon.

So, alpha hat; so the least square estimate of alpha hat is equal to X prime X inverse X prime Y. You know X prime X you know of course y, y is nothing but, y 1, y 2, y n so, you can compute alpha hat. So, let me write down, what is alpha naught hat? Alpha naught hat is first you compute X prime Y. So X prime Y that is summation y into X prime X inverse that means, 1 by n so that is nothing but, y bar. And similarly, for other parameters say alpha j hat is equal to you can check that, you take the j-th column here and then that is P j x i into y i.

And here the j-th diagonal element 1 by just P j square x i, I am sure you understand is so, this is for j equal to 1, 2 up to k. So, this is how you can estimate the regression coefficients. Now here the advantage is that, you should observe this now, we add say one more this term alpha k plus 1 P k x plus 1. This things does not change so, you do not need to recomputed X prime X and the value of the lower parameters also does not also change. So, this is the advantage of using orthogonal polynomial.

(Refer Slide Time: 40:26)



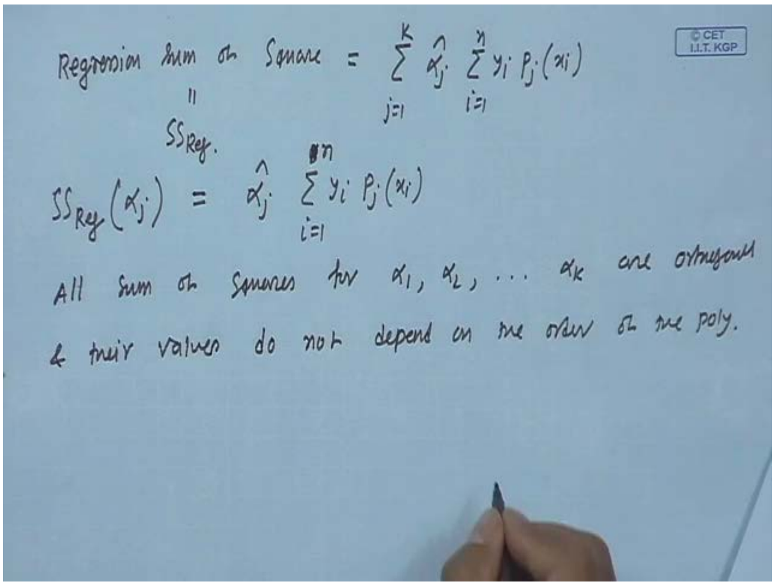
Now, let me write down the Anova table for this model. Here, for this fit what is the residual some of square? Residual some of squares, that is called S S residual ok, so s s residual we know that this is nothing but, e i square as e i is the i-th residual for i equal to 1to n. And, this 1 is again nothing but, y i observed value by minus the estimated value whole square 1 to n. Now, you can write this one to in terms of matrix form Y minus Y hat prime into Y minus Y hat right. So, this one is same as Y prime Y minus Y prime Y prime X alpha hat.

So, this you can check you know why this one is equal to this, from your from the second topic on multiple linear regression. So, we talked about this one before now, Y prime Y is nothing but, y i square and then you 1st compute Y prime X and that is nothing but, and you can check that, the whole thing is y i P j x i. That is the j-th element in X prime Y row; it is a row now (Refer Time 42:43) yeah and then, while you multiple with this vector alpha hat this become alpha j hat sum over j equal to 0 to k, it is not difficult to check this one. You just write down the matrix and check this ok.

Now, this 1 is equal to from i equal to 1 to n i equal to 1 to n y i square, now the zeroth for j is equal to 0 I will separate it out, that is alpha 0 hat and for j equal to 0 P j is x 1. So this one is nothing but, summation y i and I will keep the other terms j equal to 1 to k here. Alpha j hat minus y i P j x i right, now you know that this alpha naught hat, alpha naught hat this is nothing but, y bar. Then this one is summation y i square minus n y bar square minus j equal to 1 to k alpha j hat sum over this is i equal to 1 to n, y i p j, x i i equal to 1 to n.

Now you know that, this thing is nothing but, S S T, so S S T minus j equal to 1 to k alpha j hat i equal to 1 to n y i P j x i. So, S S residual is equal to S S T minus something and this one is nothing but, S S regression right. So, regression sum of square we can write this is nothing but, this part the 2nd term is S S regression ok.

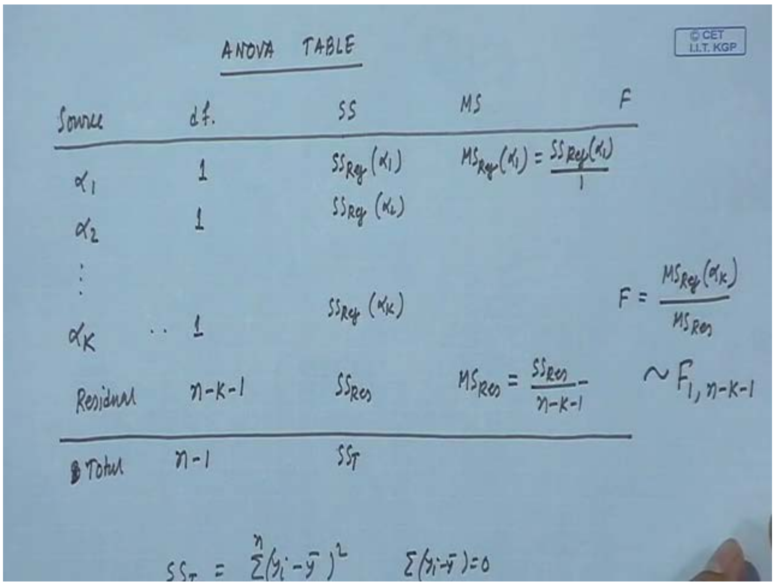
(Refer Slide Time: 45:29)



So, the regression sum of square is equal to alpha j hat summation y i P j x i i equal to 1 n j equal to 1 to k. Now, what I want to say here is that so, this is the total I mean regression sum of square; this S S this one is nothing but, S S regression. Now, what is S S regression due to the j-th term that the notation for that is S S regression due to the due to the alpha j the j-xth term alpha j P j x i. That is nothing but, the j term here that is nothing but, alpha j hat summation y i P j x i equal to 1 to n. Similarly, for S S regression due to alpha 1 is just replace this h j by 1.

So you will get S S regression due to every regression coefficients separately and here, it is very important that you know and also useful that all sum of square for the coefficient say alpha 1 alpha, 2 alpha k; they are orthogonal and their value, their values do not depend on the order of the polynomial. So, if you have say 2 degree polynomial then, the S S regression due to alpha 1 and the S S regression due to alpha 2 you have. And now say you make this polynomial to say 5 degree polynomial then, there you will again have you know S S regression due to every regression coefficient alpha 1, alpha 2, alpha 3, alpha 4 and alpha 5. But, this alpha 1 and alpha 2 they does they do not change, they remain the same even if you go for the high module. So, this is in the beauty of this orthogonal polynomial.

(Refer Slide Time: 48:53)



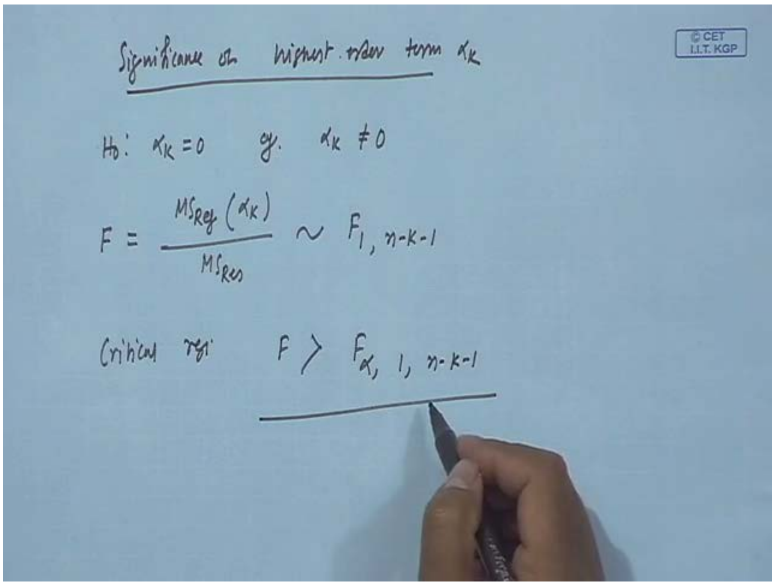
Now, let me just write down the Anova table for this one. S o, Anova table ok, so source, degree of freedom, sum of square M S and finally F. So the sources are so, S S regression again S S regression due to alpha, 1 S S regression due to alpha 2 and similarly, S S regression due to the k-th term. And you also have the total variation in the response to variables that is S S, sorry I should write just total and the part which is not explained by this regression model or this terms is called the residual ok.

Well, now total degree of freedom, we know that s s total is y i minus y i bar square, there is the variation in this response variable and this has degree of freedom n minus 1 because, of the constant that they satisfied the constant that y i minus y bar is equal to 0. So you know that, so the degree of freedom is n minus 1 now, alpha 1 has degree of freedom 1 alpha 2 1, I hope you understand all these thing. So you have k coefficients and the residual degree of freedom is n minus k minus 1.

So, other way to explain this is that, other way to explain this degree of freedom is that, there are n observations so n residuals. But, there are k plus 1; there are k plus 1 constrain on the residual because there are k plus 1 coefficients like including: alpha 0 alpha 0, alpha 1 and alpha k. So, there will be k plus 1 constraint on residual so, the residual degree of freedom is n minus k minus 1. So, this one is: S S regression due to alpha 1, S S regression due to alpha 2, S S regression due to alpha k and you know all these things. So, you know what is this S S regression alpha 1 that is nothing but, (Refer Slide Time: 45:29) you put just j equal to 1 here to get that. So, you know how to compute S S regression due to the coefficients and we know what is S S residual that is, S S T minus S S regression, so this called S S residual. And of course, the M S are same as S S because the degree of freedom is vocal is they are one, so M S regression due to alpha 1 is same as S S regression due to alpha 1 and that is by 1; so that is same thing.

Now, only the MS residual is equal to S S residual by n minus k minus 1 ok. And the F value suppose, you want to test the significance of say the highest product of alpha k, so the test statistics for the that is F equal to M S regression due to alpha k by M S residual right. And this, F follows F 1 n minus k minus 1, so, this is the Anova table for this one. Now, see in module buildings strategy I told that you know you start from lower order model say 1st order model and then whether you need 2nd model to test that you test the significance of alpha 2 the highest product term.

(Refer Slide Time: 54:17)



So similarly, here the significance of highest order term to check that is alpha k, to check that you have to test the hypothesis alpha k equal to 0 against alpha k not equal to 0. And you know the test statistics F is equal to MS regression due to alpha k by MS residual and this follows F 1 n minus k minus 1. And hence, the critical region is F greater than F alpha 1 n minus k minus 1 ok.

So, if the observe F is greater than this tabulated F then, we reject the number null hypothesis that means the k-th order term is significant. So, the k-th order term is significant you can consider the k-th degree polynomial and then, you have to check for k plus 1 half degree polynomial. And if you see the k plus 1-th degree polynomial is not significant, then stop there otherwise if it is significant again you have go for the higher order polynomial. So, in a next class I will given example to illustrate this orthogonal polynomial, today we have to stop now.

Thank you.