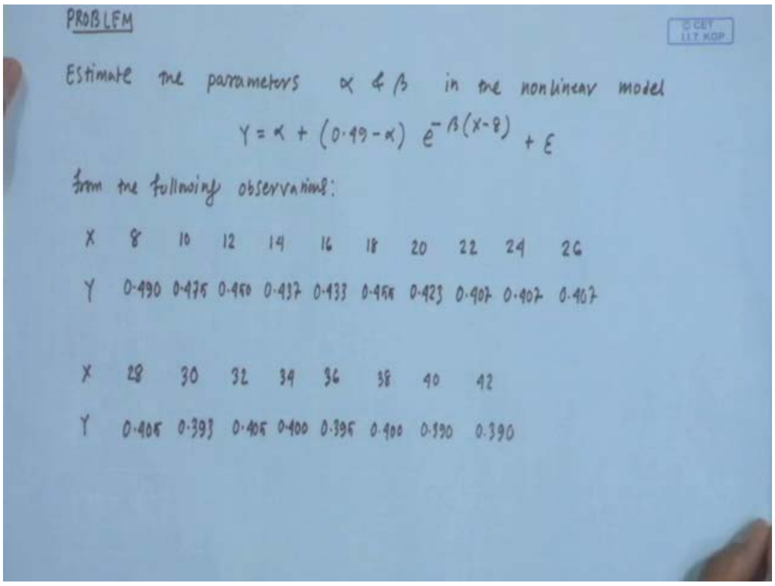
Regression Analysis
Prof. Soumen Maity
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture - 40
Tutorial - 5

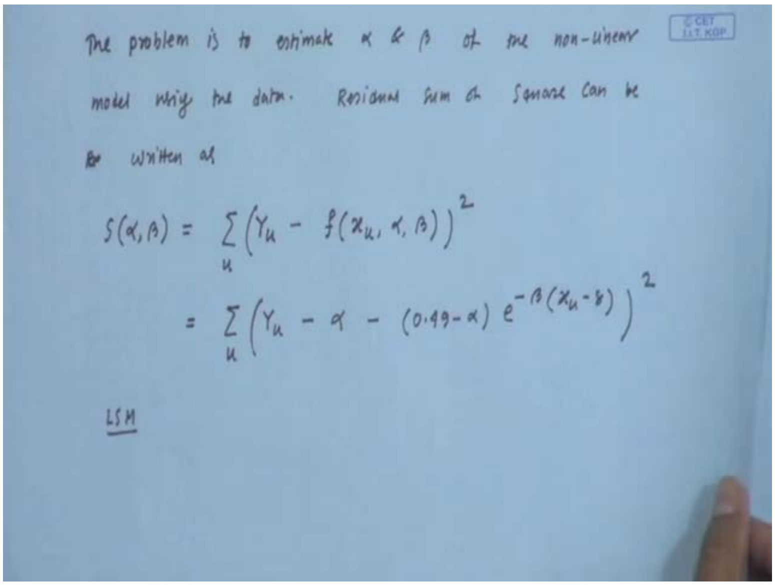
Hi. So, this is my fifth tutorial class. Today, we will be considering problem from non linear estimation, a generalized linear model, dummy variable and also variance stabilizing transformation. So, here is the first problem from non linear estimation.

(Refer Slide Time: 00:48)



So, estimate the parameter alpha and beta in the non linear model, Y equal to alpha plus 0.49 minus alpha in to e to the power of minus beta X minus 8 plus epsilon. So, this is the non linear model we need to fit. That means we have to estimate the parameter alpha and beta from the following observation. So, you are given around 20 observations here. This one is non linear model because of the fact that this is a function; it is a non linear function of the parameters alpha and beta. So, now we will be solving this problem.

(Refer Slide Time: 01:52)

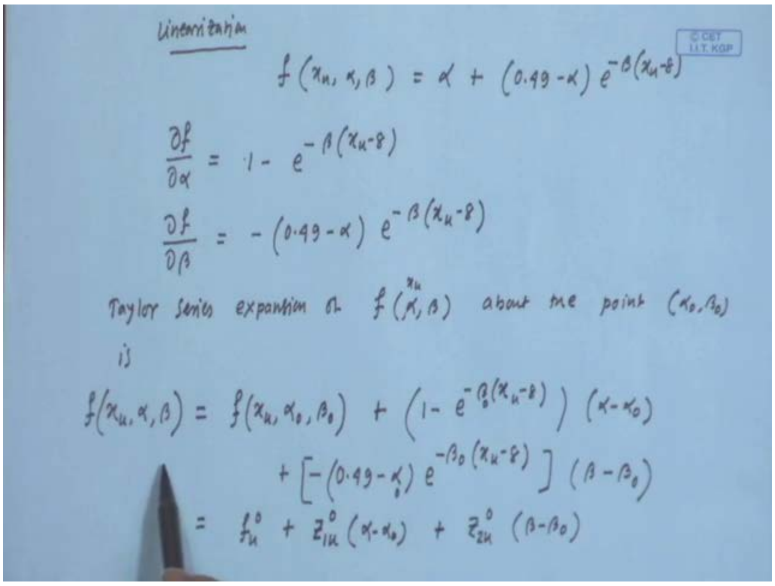


So, the problem is to estimate alpha and beta of the non linear model using the data. So, the residual sum of square can be written as or you can say it is residual sum of square may be the least square function which is alpha, S alpha beta this is equal to Yu. This is the uth observation and corresponds to the response variable for u minus f Xu alpha beta, sum over u. This is the least square function and we are given this, this is the non linear function in alpha and beta. So, I can write this as Yu minus alpha minus 0.49 minus alpha e to the power minus beta Xu minus 8.

So, this is the non linear function we are given. We have to basically estimate alpha and beta in such a way that this least square function is minimum. That is what the least square method is. Now, since the given function f or the model is non linear, so all the normal equations are going to be normal. It is very difficult to solve a system of non linear equations.

So, what we do is that what we have learnt in the non linear estimation is that we approximate this non linear function by Taylor’s series and we approximate this non linear function by linear function. So, here is the Taylor series, it involves the derivative of this function. Let me write down. This is called the linearization.

(Refer Slide Time: 05:43)



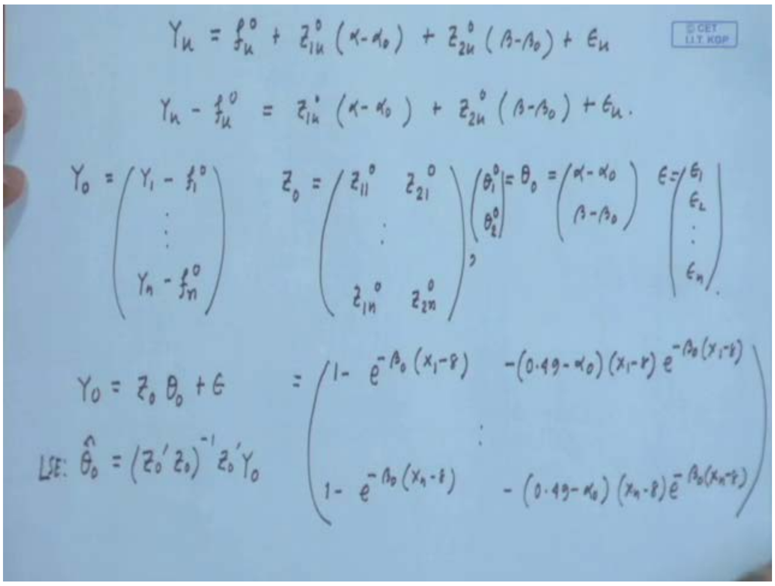
So, we linearize this function f Xu alpha beta which is equal to alpha plus 0.49 minus alpha e to the power of minus beta Xu minus 8. So, we linearize this non linear function by Taylor series. The derivative of this function with respect to alpha is 1 minus e to the power of minus beta Xu minus 8.

The derivative of this partial derivative with respect to beta is equal to minus 0.49 minus alpha e to the power of minus beta Xu minus 8. Now, the Taylor series expansion of this function expansion of f alpha beta about the point alpha naught beta naught is f, maybe I should write Xu also here, Xu alpha beta. Now, we write this non linear function. I mean of course, I mean approximate this non linear function by linear function using Taylor series and here is the approximation.

So, this one recalled to f Xu alpha naught beta naught plus df d alpha at the point alpha naught, so that is equal to 1 minus e to the power of minus beta, beta naught at the point alpha naught, so beta naught Xu minus 8 into alpha minus alpha naught plus derivative of this function with respect to beta at the point alpha naught beta naught. So, that is minus 0.49 minus alpha, alpha naught of course, into e to the power of minus beta naught xu minus 8 and then beta minus beta naught. Now, we can see that this one is linear in alpha beta. So, this is a constant term. This one is also constant because we have plugged the value alpha naught beta naught. So, this is linear in alpha beta.

So, we will write this in notation. This one is equal to fu naught plus this partial derivative at the point alpha naught beta naught that we will write at as Z1u naught and then alpha minus alpha naught plus Z2u naught beta minus beta naught. So, what we did is that we wrote this non linear function using; I mean we approximate this non linear function by linear function in alpha beta using Taylor series expansion. Now, we have to estimate the parameter alpha and beta for a linear function. We know how to do it using the multiple linear regression technique. Now, you can we are in the position to use ordinary least square technique.

(Refer Slide Time: 11:13)



So, Yu can be written as fu naught plus Z1u naught into alpha minus alpha naught plus Z2u naught beta minus beta naught plus epsilon u. So, I mean in the original model now this becomes a linear model here. This can be written as of course, as Yu minus fu naught which is equal to Z1u naught alpha minus alpha naught plus Z2u naught beta minus beta naught plus epsilon u. Now, here you can see that now this one is multiple linear regression model involving two parameters. Well, I would like to write this now in matrix form.

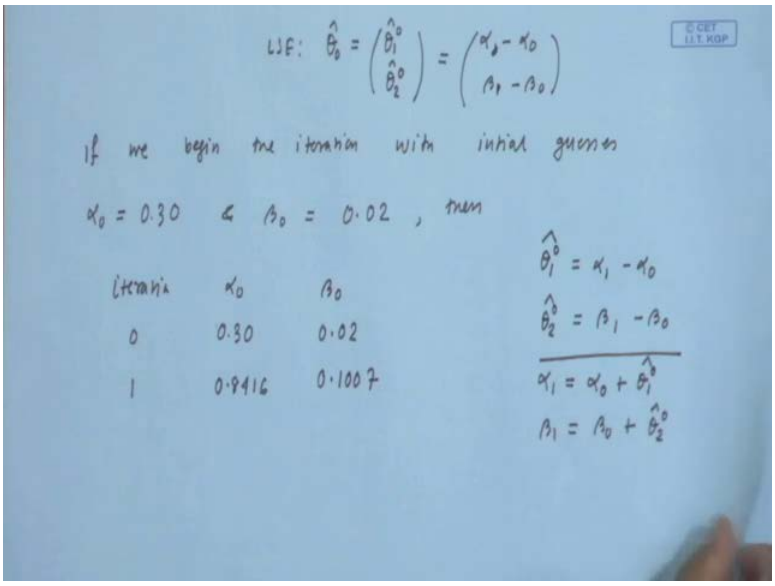
So, I will write use the notation Y naught for response vector, so Y1 minus f1 naught, and similarly, Yn minus fn naught. I will write my Z naught matrix is for Z11 naught, Z21 naught, this is for the first observation and similarly, Z1n naught, Z2n naught, so this is for the coefficient matrix. My parameter vector theta naught is equal to alpha minus alpha naught and beta minus beta naught. So, we want to estimate the parameter alpha naught, sorry alpha and beta. We approximated this function about alpha naught and beta naught. Well, we will see now how to estimate alpha and beta because that is what our aim is and epsilon is, of course, epsilon 1, and epsilon 2, up to epsilon n.

I am sure that you understand what the meaning of this one is. See this one is the partial derivative of the non linear function f with respect to alpha at the point alpha naught beta naught. So, this one is basically 1 minus e to the power of minus beta naught into X minus, so X1 minus 8 and Z21 is basically minus 0.49 minus alpha naught and X1 minus 8 e to the power of minus beta naught X1 minus 8. So, you can see that this is derivative of the function f with respect to beta at the point alpha naught beta naught.

Similarly, you know this one is 1 minus e to the power of minus beta naught Xn minus 8 and this one is minus 0.49 minus alpha naught Xn minus 8 e to the power of minus beta naught Xn minus 8. So, this is what the z naught matrix is. We know that in matrix form, this can be now written as Y naught is equal to Z naught X beta, so z naught theta naught plus epsilon and we know that then theta naught hat, which is equal to Z naught prime Z naught inverse Z naught prime Y naught. So, let me put some more notations also.

So, this one, I will call alpha minus alpha naught. I will call that theta one. In fact, you know too many notations for this non linear estimation. So, theta 1 naught is basically alpha minus alpha naught, theta 2 naught is basically alpha, sorry beta minus beta naught. Well, so we have found the least square estimate of theta naught. So, this is least square estimate.

(Refer Slide Time: 17:50)

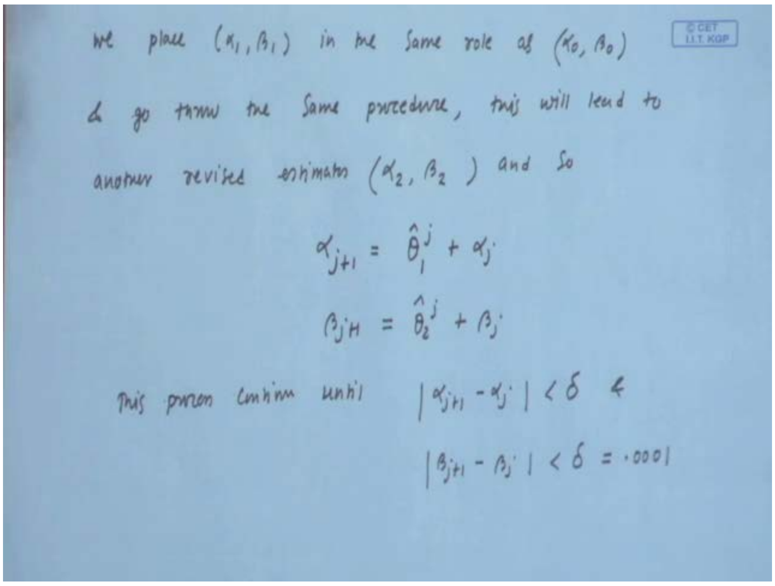


So, what we have observed is that we got theta naught hat which is equal to theta 1 naught hat, theta 2 naught hat that is alpha 1 minus alpha naught beta 1 minus beta naught. So, estimate of all these things, so what we do is that if we begin the iteration with initial guesses say alpha naught equal to 0.30 and beta naught which is equal to 0.02, then what we do is that actually, we iteratively we improve this alpha beta. So, the first iteration is, say 0, we have alpha 0. We took alpha 0 as 0.30 and beta 0 as 0.02. We approximated the function about this point alpha naught beta naught using Taylor expansion and we made it linear.

So, once we have the linear approximation of the function using the Taylor series, then we use the least square technique to estimate the parameters. So, this is how we got the estimation of the least square estimation of the parameters theta. Then, what we do is that, I should naught not write 1 here. What we have at this moment is that we have theta 1 naught hat and which is equal to alpha 1 minus alpha naught.

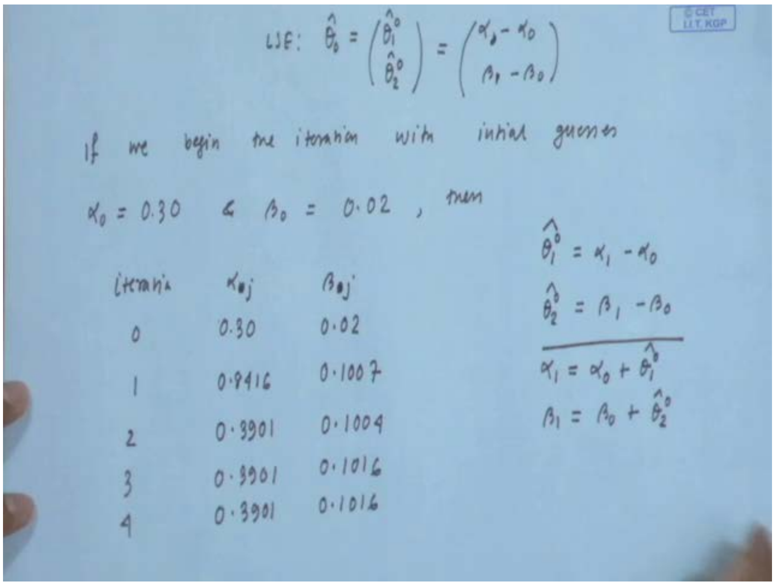
Also, we have theta 2 naught hat which is equal to beta 1 minus naught, in fact, beta alpha 1 in fact, it is alpha minus alpha naught, but what we do is that we considered this alpha 1 as improvement of alpha. Now, alpha 1 is alpha naught plus theta 1 naught hat and beta 1 is equal to beta naught plus theta 2 naught hat. So, we started with alpha naught beta naught. Then, now the improved value of alpha 1 and beta 1 are in the first iteration. They are 0.8416 and 0.1007. What we do is that again we place this alpha 1 and beta 1.

(Refer Slide Time: 21:44)



So, we place alpha 1 and beta 1 in the same role as alpha naught, beta naught. We go through the same process. So, this will lead to another revised estimate say alpha 2, beta 2. So, we started with alpha 1, beta 1. Now, we have, sorry, we started with alpha naught, beta naught. Now, we have alpha 1, beta 1. Again, next step, in the next iteration, we will have alpha 2, beta 2. So, at the jth step, we will have alpha j plus 1, which is equal to theta 1 j hat plus alpha j and beta j plus 1, which is equal to theta 2 j hat plus beta j. So, we continue this process. So, this process continues until alpha j plus 1 minus alpha j is less than delta and beta j plus 1 minus beta j is less than delta. So, delta is a pre specified small number.

(Refer Slide Time: 24:06)

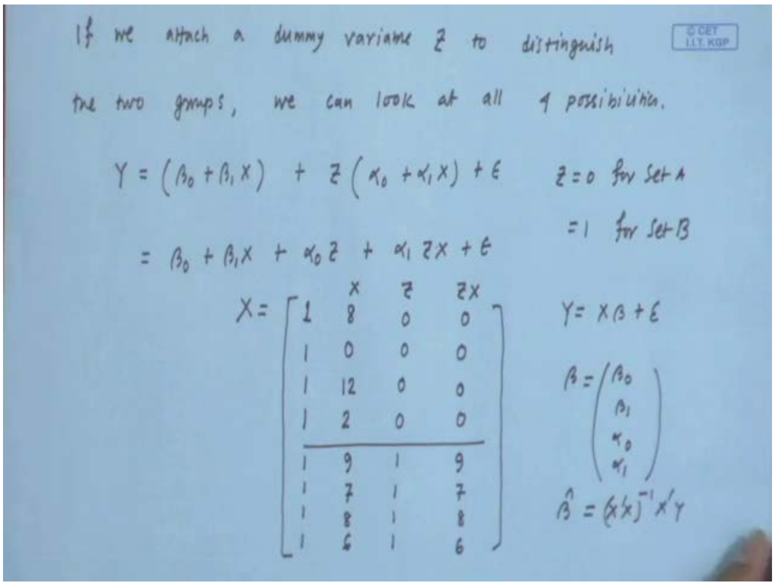


So, in our case, what we do is that at this moment, we have alpha 1, beta 1. Is this right? This is my alpha naught, beta naught. So, I should write here alpha j, beta j. So, in the first zeroth iteration, this is alpha naught, beta naught, alpha 1, beta 1. In the next iteration, I will get. You can check that the value of alpha 2 will be 0.3901 and beta 2 will be 0.1004, third iteration, it will be 0.3901 and here it is 0.1016.

In the fourth iteration, you will see now see alpha is naught changing. So, 0.3901 and 0.1016, so at the fourth stage, you see that there is no difference between the third and fourth step. So, alpha 4 minus alpha 3 is less than equal to this quantity. I mean, similarly, beta 4 beta 4 minus beta 3 is there is in fact, no difference, it is 0. So, we can stop here. Well, this is the first example from the non linear estimation. Next we will consider a problem from dummy variable.

So, dummy variables are utilized to separate blocks of data. So, here is an example. Look at this data. I do not know whether to fit two straight lines, one straight line or what. So, we have two sets of data, set A and set B. We do not know whether to fit one straight line to all the data together or two straight lines or what, we do not know. So, he has two sets of X, Y data given below, which cover the same X range. How do you resolve this dilemma? Describe and give model details and things he needs to do. So, we have learnt the use of dummy variables. So, will be fitting a general model involving two dummy variables including, say Z naught for this problem.

(Refer Slide Time: 27:08)



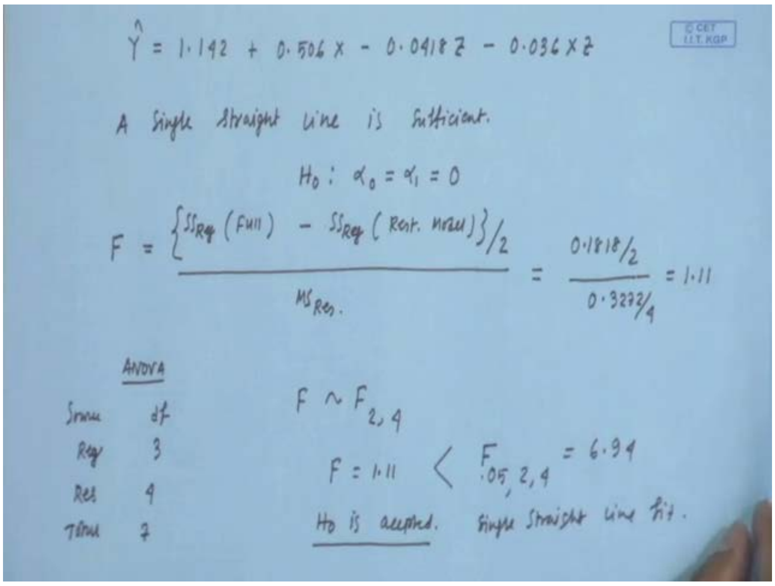
So, if we attach a dummy variable Z to distinguish the two groups, we can look at all four possibilities. You understood what I mean by four possibilities. Well, so the general model is Y equal to beta naught plus beta 1 X plus Z in o alpha naught plus alpha 1 X plus epsilon. So, Z is equal to 0 for set A and it is 1 for set B. This can be written of course as beta naught plus beta 1 X plus alpha naught Z plus alpha 1 Z X plus epsilon.

This is the model we are going to fit. You can see that here it is multiple linear regression model. What is the X matrix here? The X matrix has 1, 2, 3, 4 columns. So, the first column is all 1 of course. Let me put Z naught, I can put also here or X naught or let me put only 1 here. Then, I will have a column for X. I will have a column for Z. I will have a column for Z X.

So, first set has four observations. The observations are 8, 0, 12, and 2. So, I will put them, 8, 0, 12, and 2. For the first set, set A, my Z is equal to 0, so 0, 0, 0, 0. Then, ZX is of course, all 0. So, this is for first set and for the second set, 1, 2, 3, and 4. You can check that the X values are 9, 7, 8, and 6. For the second set or set B, Z is equal to 1. So, I will put Z is equal to 1, 1, 1, and 1. Then, of course, ZX is equal to 9, 7, 8, and 6.

So, this is what my X matrix is in matrix notation. Of course, you know. I am sure that you understand the difference between this X and this X. Well, so the model can be now written as Y equal to X beta plus epsilon. Of course, this beta is vector is beta naught beta 1 alpha naught. So, this beta is beta naught, beta 1, alpha naught, alpha 1. So, you know how to fit this model. This beta hat is equal to just X prime X prime inverse X prime Y. So, let me write down the fitted model now.

(Refer Slide Time: 31:24)



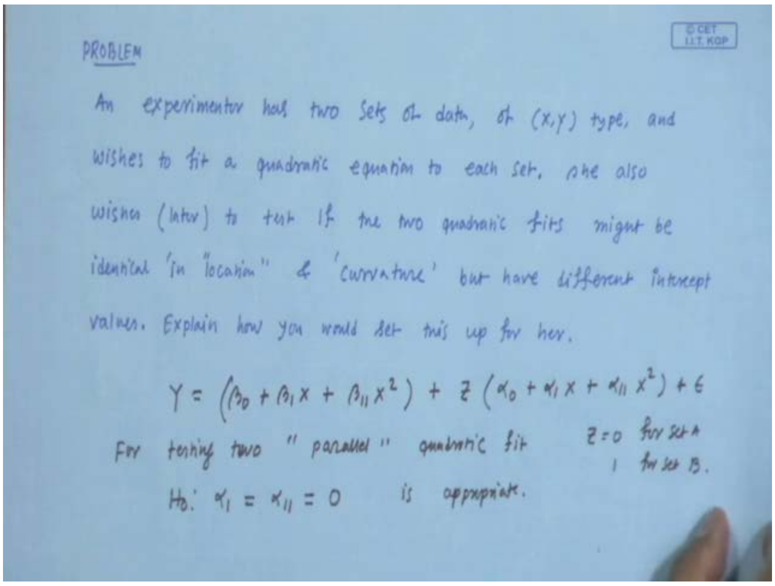
So, for the fitted model is Y hat equal to 1.142 plus 0.506 X minus 0.418 Z minus 0.036 XZ. So, this is my fitted model. Now, the question is whether a single straight line is sufficient. If there is not much difference in the response level, we can go for a single statement line fit, but we need to, see want, we have the general model. Now, we can test whether this is I mean one single line is sufficient. For that, what we have to test is that we have to test the hypothesis H naught that alpha naught equal to alpha 1 equal to 0 because we have considered the general model is this one.

Now, if I test, to test that whether a single straight line is enough, we have to test that alpha naught is equal to alpha 1 equal to 0. You know how to test this using the extra sum of square technique. So, F statistics is SS regression for the full model minus SS regression for the restricted model. So, the restricted model does not involve these two terms and this by degree of freedom 2 by MS residual, before doing all these things, you know just I will construct the ANOVA table, ANOVA table for the full model source. Total there are 8 observations.

The degree of freedom is 7 and the regression has 4 parameters. So, it will be 3 and the residual has degree of freedom 4. Now, the restricted model has only 2 parameters. So, this will have degree of freedom 1. So, 3 minus 1 is equal to 2. You can check that this is equal to 0.1818 by 2 by image residual is 0.3272 by four 4, is equal to 1.11.

Now, this F has degree of freedom 2, 4. Now, compute, sorry, you just check the tabulated value of F, F 05, 2, 4 is equal to 6.94. So, my observed value F which is equal to 1.11 is smaller than this one. That means H naught is accepted. H naught is accepted that means we can go for single straight line fit. So, this is in the problem I wanted to discuss from dummy variable topic.

(Refer Slide Time: 35:52)

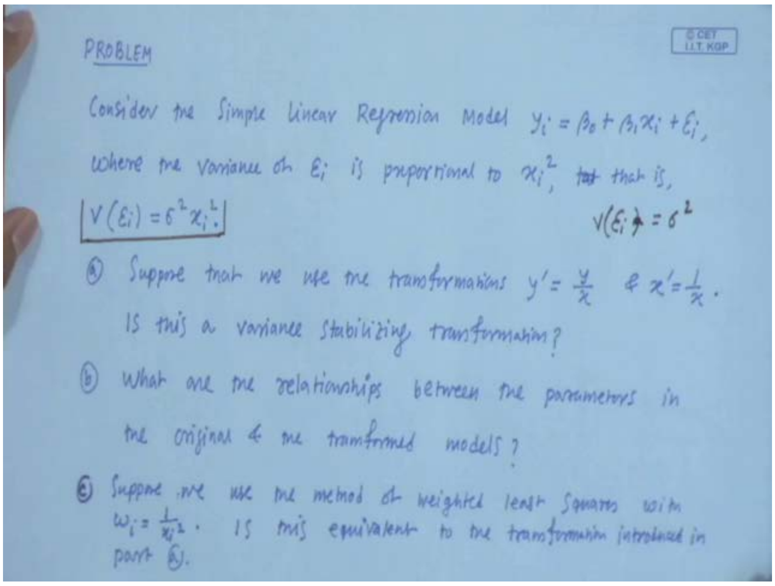


We have another problem involving dummy variable. So, it says that an experimenter has two sets of data of X, Y type and wishes to fit a quadratic equation to each set. She also wishes to test if the two quadratic fits might be identical in location and curvature, but have different intercept values. Explain how you would set this up for her.

So, what she has is that she has two sets of data on X and Y. She wants to fit quadratic equation. So, she should go for the general model like Y equal to beta naught plus beta 1 X plus beta 11 X square plus Z alpha naught plus alpha 1 X plus alpha 11 X square plus epsilon. So, of course, you know that Z equal to 0 for set A and 1 for set B. Now, she wants to check whether she can go for two parallel quadratic fit, for testing two parallel quadratic.

What you mean by parallel quadratic? They have the same location and curvature. Only they differ in the intercept. What yet to test is that you have to test whether alpha 1 is equal to alpha 11 is equal to 0. So, you test the hypothesis that alpha 1 equal to alpha 11 equal to 0. For testing two parallel quadratic fit, this one is appropriate. I am sure that you understand how to test this hypothesis using extra sum of square technique.

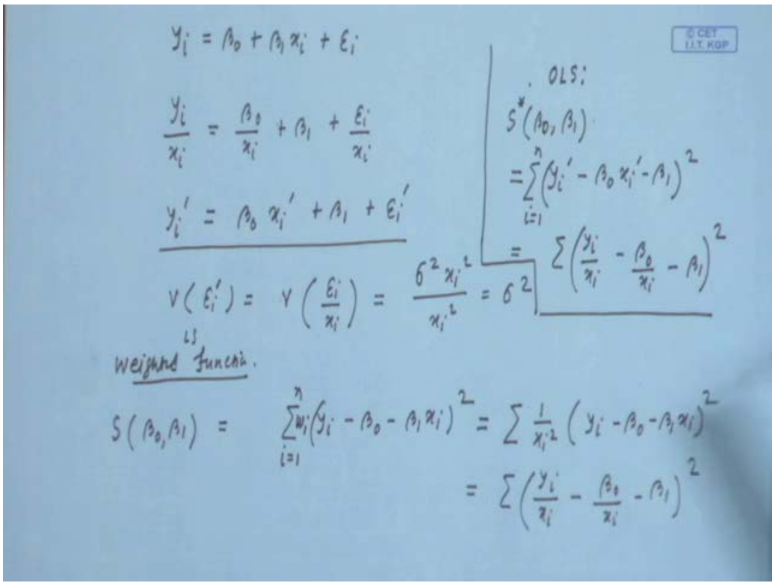
(Refer Slide Time: 38:36)



Now, I will be considering a problem from a topic called transformation and weighting to correct model in adequacy. There we talked about variance stabilizing transformation. So, I will be considering one problem from variance stabilizing transformation. Well, here is the problem. Consider the simple linear regression model yi equal to beta naught plus beta 1 xi plus epsilon i where the variance of epsilon i is proportional to xi square that is the variance of epsilon i is equal to sigma square xi square.

So, this means the assumption of constant variance is not satisfied. So, usually if epsilon i follow, epsilon i, the variance of epsilon i is equal to sigma square, then we go for the ordinary least square technique, but that is not true here. Variance is changing for different i and it is proportional to xi square. Suppose that we use the transformation y prime which is equal to y by x and x prime which is equal to 1 by x. Is this a variance stabilizing transformation? First solve this problem.

(Refer Slide Time: 40:26)



So, we start with the model yi equal to beta naught plus beta 1 xi plus epsilon i. Then, I am considering the transformation yi to yi by xi. Then, the model becomes beta naught by xi plus beta 1 plus epsilon i xi. So, if I call this i as yi prime, then yi prime is equal to beta naught and my xi prime is 1 by xi plus beta 1 plus epsilon i prime. So, this is the transformed model. Now, you can check in this transformed model, variance of epsilon i prime is equal to variance of epsilon i by xi. We know that variance of epsilon i is sigma square xi square and then by xi square. Now, the variance of transformed error Ei prime is constant, so yes.

So, the answer to the first problem is that yes, it is a variance stabilizing transformation. What are the relationships between the parameters in the original and the transformed model? Well, what is the relation? I hope that the relation is, so here you can see this is my transformed model. So, the slope in the original model becomes intercept in the transformed model. The intercept in the original model becomes slope in the transformed model. I mean that is what I felt.

So, next the next problem is that suppose we use the model of, sorry, suppose we use the method of weighted least square with wi is equal to 1 by xi square. Please recall what is weighted least square. Is this equivalent to the transformation introduced in part 1? I mean considering weighted least square with this weight, is it same as the transformation we consider in part 1? So, that is the question. So, you have to recall what is weighted least square. Let me may solve this problem. So, what is weighted least square?

Weighted least square is about finding the least square estimate of regression coefficients, but we consider the weighted least square function to estimate the parameters. So, the weighted least square function least square function is S, say beta naught, beta 1. So, what we do in the usual case is that, we are just minimizing this S, S residual that is yi minus beta naught minus beta 1 xi square i from 1 to n. So, this is the SS residual. So, we minimize this quantity to estimate beta naught and beta 1 in such a way that this is minimum. That is what the ordinary least square technique is.

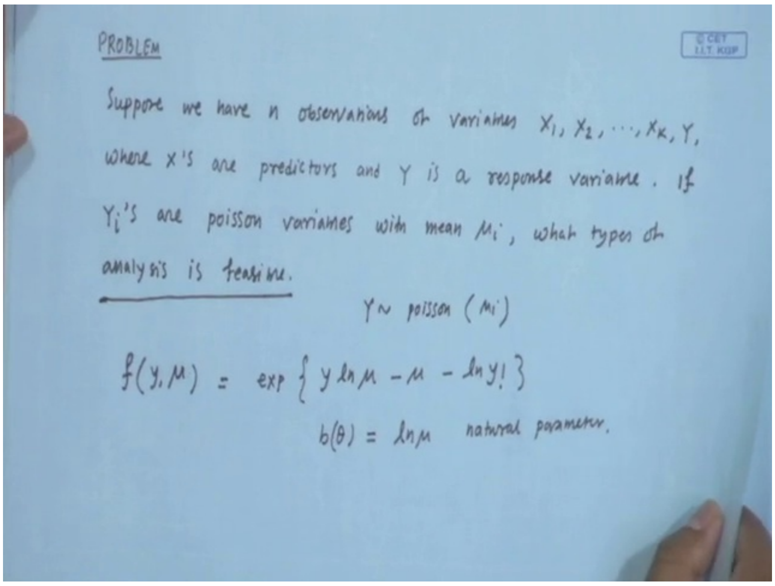
Now, in the weighted least square, we put a weight for the ith observation that is wi and here, but see part c or part 3 of the problem, it says that the weight is 1 by x square that is 1 by xi square. So, wi is 1 by xi square, which is equal to yi minus beta naught beta 1 xi square. So, this is the weighted least square function.

This can be written as yi by xi minus beta naught xi minus beta 1 whole square. So, using the weighted least square technique, we will estimate the parameter beta naught, beta 1 such that this is minimum. This is what the weighted least square technique suggests. Now, the question says whether this is equivalent to the transformation introduced in the part 1. Now, in part 1, so this is the transformation we considered. So, here to compute beta naught and beta 1, here you can see that the epsilon dash is the constant variance.

So, we can go for ordinary least square technique. The function will minimize is, say call it S star beta 1, sorry, beta naught beta 1. So, this is equal to, this is equal to yi prime minus beta naught xi prime minus beta 1. So, we will minimize this to estimate beta naught and beta 1. Now, you can see that this one is equal to this one. This is nothing but yi by xi minus beta naught by xi minus beta 1. See. So, we are minimizing. Now, we can see that these two here, this is the least square function for the ordinary least square and this is the weighted least square function for the weighted least square.

Now, we can see that both are same. This one is same as this one. So, the function we are considering in the transformed model to estimate beta naught and beta 1 is the same as the function we are considering to estimate beta naught and beta 1 using weighted least square technique. So, the answer to the problem part 3; is this equivalent? The answer is yes, this is equivalent we considered from the variance stabilizing transformations. So, this is one problem we considered from the variance stabilizing transformations.

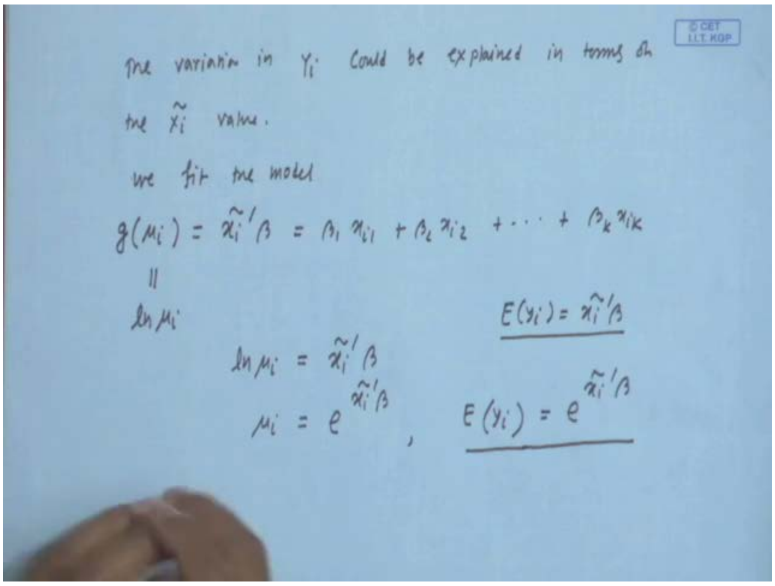
(Refer Slide Time: 48:29)



Next, I will be considering one more problem. This problem is from generalized linear model technique. Well, so the problem says that suppose we have n observations of variables X1, X2, Xk, Y where Xs are regression variables and Y is a response variable. The question is if Yi's are Poisson variables with mean mu i, what types of analysis is feasible? So, the objective of the generalized linear model topic was if the response variable is not following normal distribution variable, but if the response variable follows some distribution from the exponential family, then how to fit a model for that?

So, that was the objective of generalized linear model. Here you can see that this response variable Y follows Poisson distribution. So, Poisson distribution is a member of exponential family. So, we will see how to solve this problem. So, Y follows Poisson with mean mu i. Then, we know that the probability mass function of Y can be written as f y, mu which is equal to exponential y ln mu. You can check this minus mu minus ln y factorial and here b theta or b mu, I should write may be is ln mu, which is the natural parameter. What we have learned in this topic called generalized linear model is that how to fit a model when the response variable is not normal.

(Refer Slide Time: 51:41)



So, the variation in Yi could be explained in terms of the regression values. So, what model we fit that we fit the model fit the model some g mu i equal to xi prime beta, so which is equal to beta 1 xi1 plus beta 2 xi2 plus, say beta k xik. This g mu is the link function, which is nothing but the natural parameter that is ln mu i. So, the model we go for is that ln mu i is equal to xi prime beta. So, this can be written as mu i is equal to e to the power of xi prime beta, which is nothing but the expectation of yi. So, that is mu i, which is equal to e to the power of xi prime beta.

So, this is the model we need to fit if the response variable follows Poisson distribution. Usually what you fit is that if y follows normal distribution, then we fit the model E yi is equal to xi prime beta, but if it is following the Poisson, then we follow, we fit this model. If y follows say binomial, then depending on the natural parameter, we get the model to be fitted. So, I have tried to you know cover problem from different topics I considered in this course and that is also we need to stop now.

Thank you.