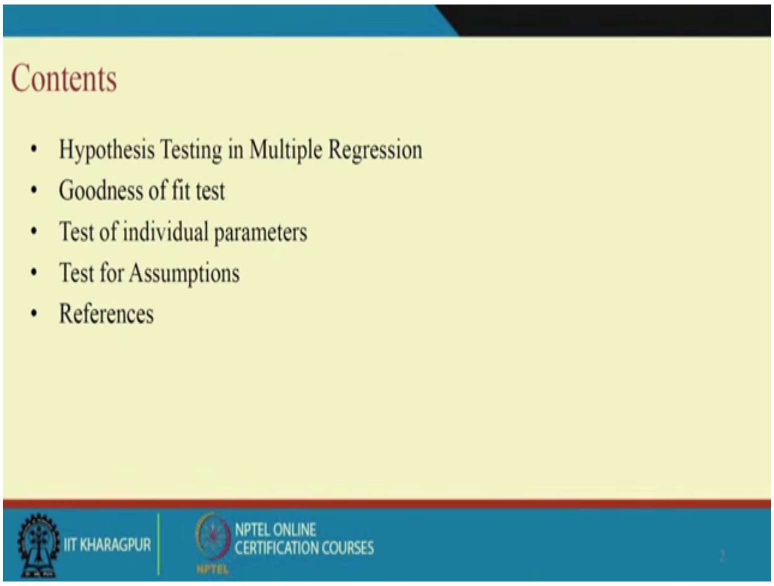
Design and Analysis of Experiments
Prof. Jhareswar Maiti
Department of Industrial and Systems Engineering
Indian Institute of Technology, Kharagpur

Lecture – 22
Multiple Linear Regression: Hypothesis Testing and Model Adequacy Test

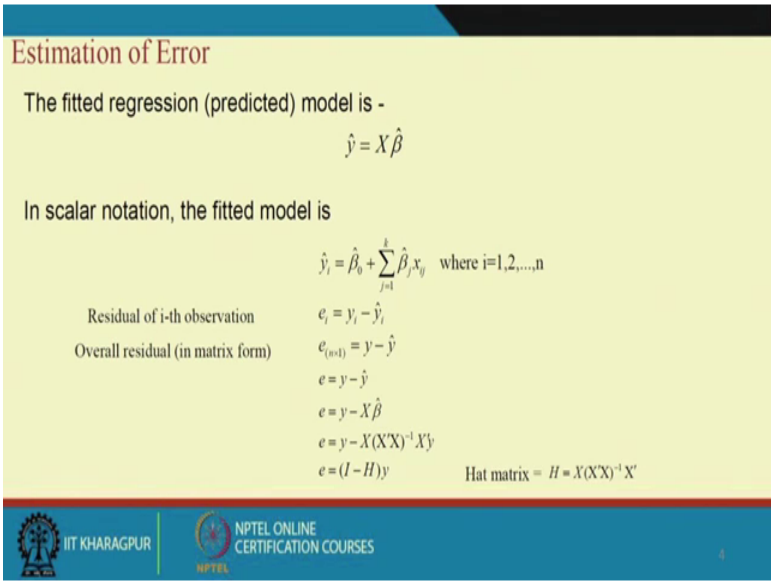
Welcome. We will continue Multiple Linear Regression. Today I will talk about model adequacy test primarily whether regression equation is adequate or not from the data or the experimental data you obtain that basically I have given you a proper response surface or not.

(Refer Slide Time: 00:32)



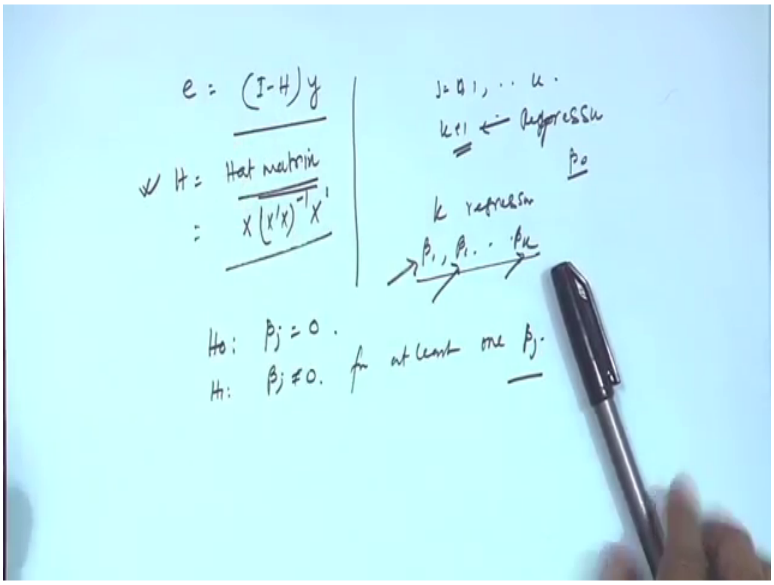
So, first I we will start with the overall hypothesis testing using f test then I will give you some goodness of fit test measure like R square, adjusted R square. Test of individual parameters we will revisit because earlier I have shown you how to do and another important one is the test of assumptions. So, when I talk about test of assumptions I will give you what are the assumptions and how those assumptions must be must be verified that they are really true.

(Refer Slide Time: 01:09)



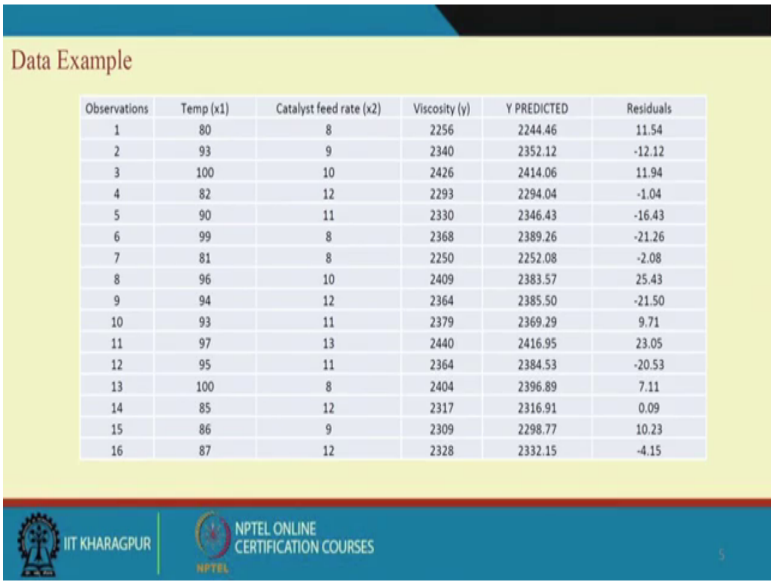
So, straightway let us go to that a regression equation here that y equal to X beta plus epsilon now if you when you estimate the y it will be y equal to X beta cap. And all of you know then in case scalar notation we can write like this and ultimately the difference between the actual observation minus the predicted one or otherwise other way I can say fitted one is the residual. So, e i y i minus y i cap. So, as there are n number of y values so that means, there will be n number of residuals. So, in matrix form then that e equal to y minus y cap it is n cross one vector and if we just do certain level of manipulation for e we will find out that e will be i minus H into y where H is X, X transpose, X inverse X transpose.

(Refer Slide Time: 02:19)



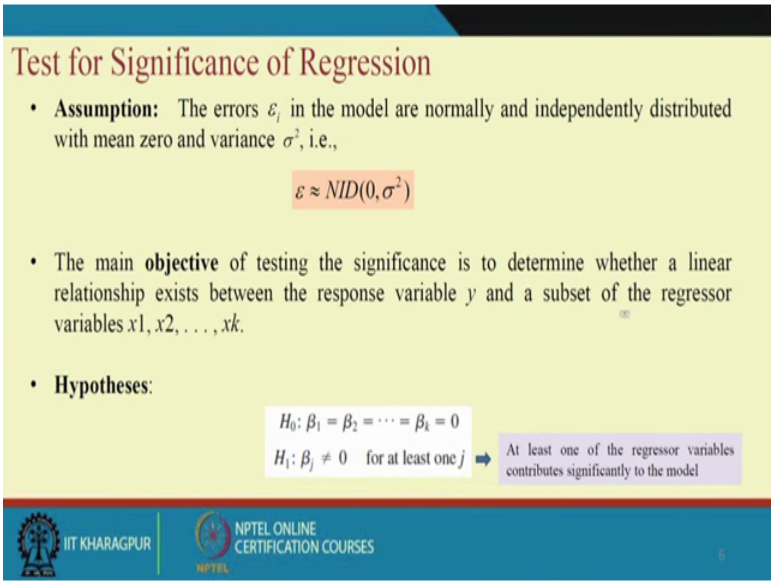
So, we can see that the residual can be written like this i minus H into y, where i is the identity matrix and H is known as hat matrix H is known as hat matrix which is X, X transpose X inverse X transpose this is the part. It is a very interesting matrix because it has lot of implications it can be used for doing lot of tests. Particularly if suppose we are interested to know what are the contribution of individual observations that time you will find out that the diagonal element of hat matrix will talk about the contribution of individual observations suppose you have n number of observations that diagonally will talk all these things.

(Refer Slide Time: 03:14)



So, anyhow let us see one a one data example. We have 16 observations we have 3 2 factors X temperature and catalyst feed rate and our dependent variable is viscosity that is to be predicted let it be and we have fitted the model and which we have already shown you earlier and then from there this model we found out the predicted values or the fitted values here exactly these are fitted values. Then this, this viscosity observed values minus fitted value is giving you the residuals these residuals e values this is the last column.

(Refer Slide Time: 03:55)

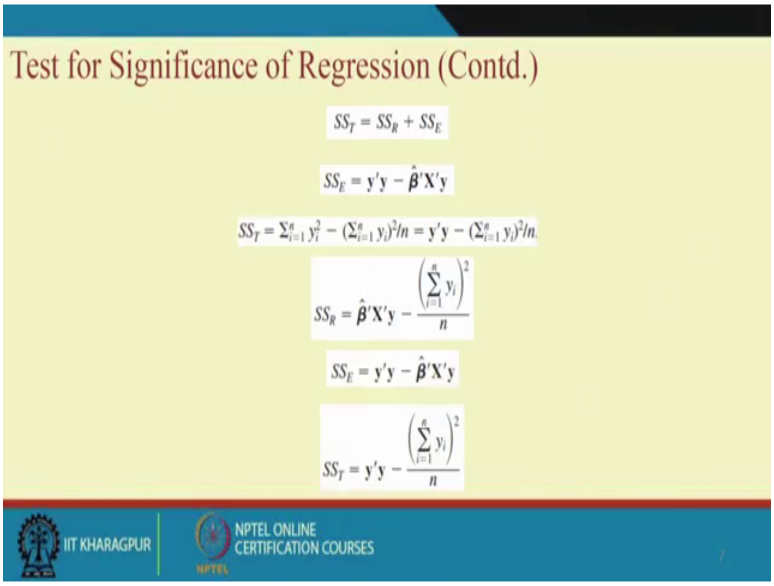


Now, what we will do here we will we will actually the one of the important assumptions here is that the errors are normally distributed identical and normally distribution iid, independent and identically distributed and that is also normally distributed that is one.

So, anyhow that test we will see later on. But what is the hypothesis overall fit test here overall fit test is that we are we have j equal to 0, 1 to k; that means, k plus 1 regression, regression. If we do not consider this one that the con intercept that beta 0 then we have k regressions, k regressors and stab; that means, beta 1 beta 2 like this beta k and that is important because beta 1 related to variable X 1, X 2 and X k like this important which of the variables are contributing or not. So, here what we will do we will over for overall test we say that beta j not equal to 0; that means, none of the factors are contributing towards the response what we observed during the experiment and H 1 is beta j not equal to 0. For at least one beta j let me at least beta 1 or beta 2 or beta 3 sum is contributing and no not that all beta j are 0.

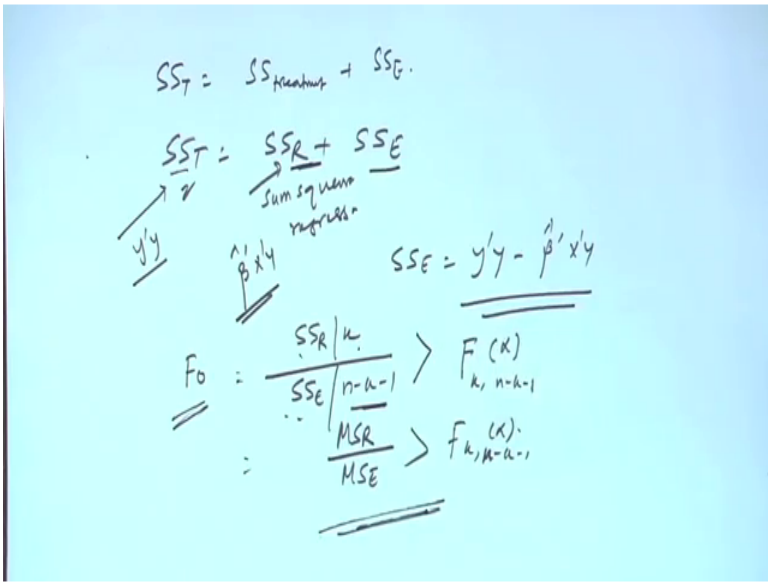
This is overall test, this overall test is done through f statistics.

(Refer Slide Time: 05:39)



You have seen in ANOVA that sum square total equal to sum square I think that the treatment plus sum square error in one way ANOVA you seen this one.

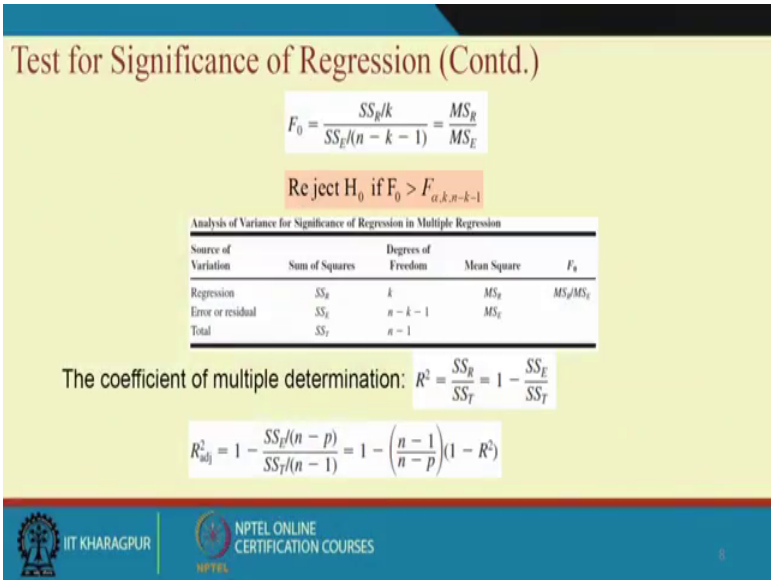
(Refer Slide Time: 05:45)



Here also in regression we can write sum square total equal to sum square regression plus sum square error this is sum square regression or we can say that it sum square model. And all of you know that y transpose y will give you this and sum square regression will be that beta cap transpose X transpose y from this and essentially then S E will be SS T minus SS R which is will be y transpose SS E will be y transpose y minus beta cap transpose X transpose y this will give you the SS E.

So, fine then SS R also beta s and then finally, what happened finally, what happened I have given probably little difficult equation differently. Let us see that SS T all of we know will be this following this equations. So, here y transpose y when is mean subtracted is there that time if you subtract by y is means after that formula, but otherwise what happen is SS T y i square minus this; that means, this quantity is coming here. Similarly SS R also that this will be subtracted and SS E then this SS E y transpose y minus this and SS T equal to y transpose y minus this by n y transpose y minus y i square by n that is what we have seen. It is a basically some kind of duplications we have made, but whatever may be the thing you please remember you will be able to compute SS R you will be able to compare SS T and SS E also you will be able to compute from errors residuals, but other way SS T minus SS R will give you SSE.

(Refer Slide Time: 08:11)



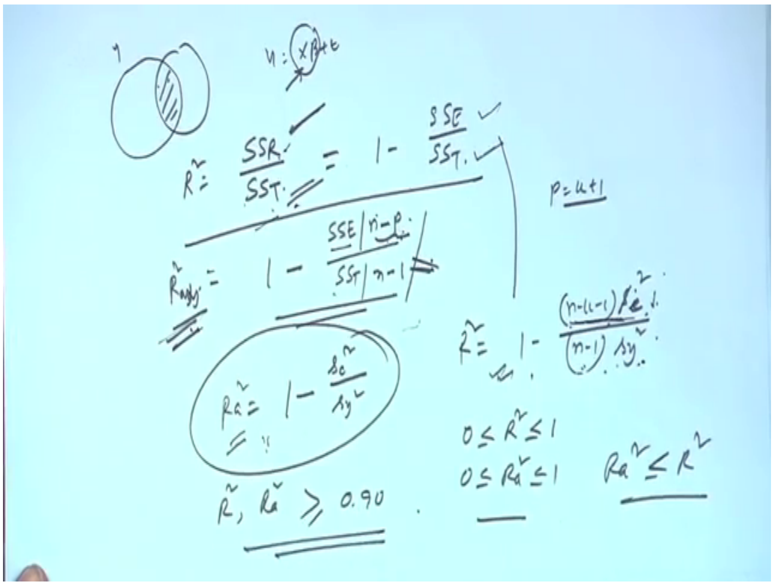
So, then what happened you will create a statistics called f 0 this is nothing, but SS R divided by k by SS E divided by its degrees of freedom n minus k minus 1. So, if this,

this one is greater than F k n minus k minus 1 may be alpha. So, then what we will say H 0 rejected, H 0 is rejected.

Now, what is this SS R by k? This is a MS R what is SS E by degree of freedom MS E so; that means, this MS R by MS E if it is greater than k n minus k minus 1 alpha will this F 0 it is and then we will say that the null hypothesis that none of the factors are contributing is not correct. This is what is in terms of ANOVA table if you see the ANOVA table here it is in terms of a ANOVA table we have shown the same thing.

Now, this is overall f test overall f test will say that whether at least one of the regression coefficient contributing or not if none of the regression coefficient, contribute regression coefficient is significant or none of the factors X variables are contributing then and then is that is what H 0 will be accepted and none will be nothing in no influence. But what happened here you may be interested to know that suppose if test it says that H 0 is rejected then we want also we want to know that what is the variability of y is explained by X that was this is if I say the y variability may be the regression y is X beta plus epsilon then this X beta this portion may be able to explain this much. So, what is this portion?

(Refer Slide Time: 10:11)



So, this portion it is some kind of absolute major that the variability of y explained by the regression model that is if divided by the variability not explained by the regression model or total variability will give you a measure which is known as R square. So, R square is a major which is SS R by SS T that mean variability explained by the model and total variability of y variability of y explained by the model and total variability this is R square this one can be written like this 1 minus SS E by SS T also.

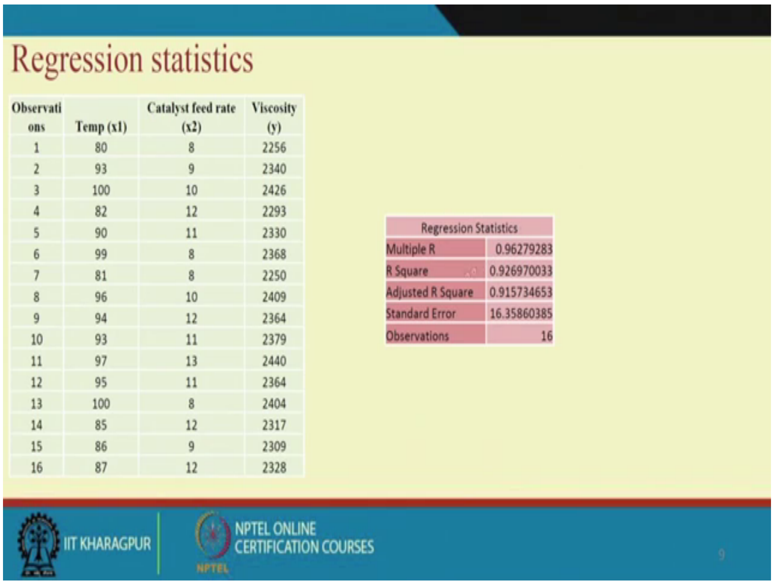
So, what is the problem here? Problem is if I expand this I can write what SS E we can write that SS R by SS T and SS E by this one. So, we can write like this SS E by n minus p by SS T by n minus 1 where p equal to k plus 1 if you write like this then this quantity will give you a measure which is known as suppose this quantity gives you a measure which is known as R square adjusted.

So, I will explain little further this one. So, R square I can write one minus SS E is nothing, but n minus k minus 1 into se square you have seen earlier se square similarly SS T you have seen n minus 1 into sy square. So, now, if you divide the thing this, this by degrees of freedom n minus k minus 1 or m minus p and this by its degrees of freedom you are getting this one. So, we are creating another coefficient measure are adjusted R square which will become then, then what happened Ra square if I write then this will be SS E by n minus p minus n minus k minus 1 that will this se square SS T means n minus se square by n minus 1 minus sy square. So, this portion, this value is this R square value is unaffected by n and p that is the sample size as well as the parameters number of parameters to regression coefficient to be estimated.

So, it is a speedo sample size this, this value will get and it is a better measure than R square, R square. R square will be impleted suppose n equal to p then this quantity what happen ultimately this quantity will become for example, here if I read n equal to p this will become 0. So, R square become 1 minus abnormally high R square value you will get. So, and whatever may be the case R square will lie in between 0 to 1 and Ra square also lie in between 0 to 1 and Ra square is greater than equal to R square. This R square, is that is coefficient of multiple determination ok. So, this is the goodness of test.

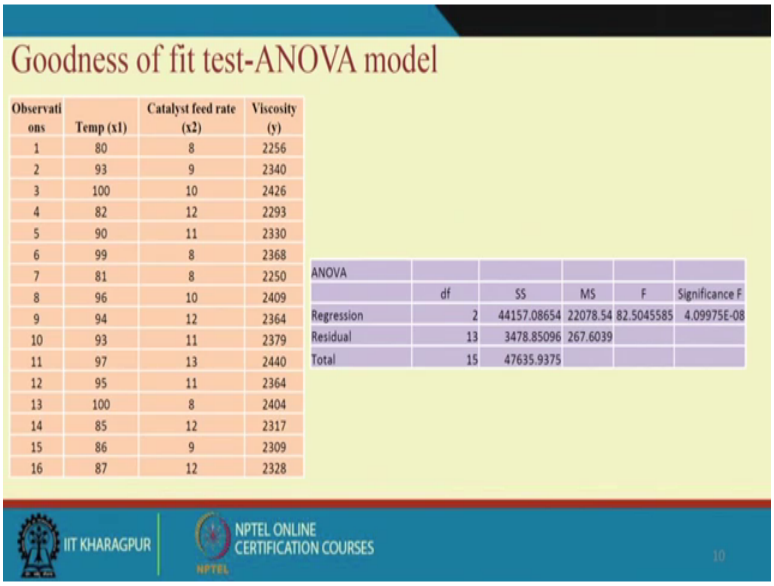
If R square value is greater than in case of a laboratory experiment whether R square or R a square if it is greater than equal to 0.90 then it is a fit data, is a good fit to the model regression model.

(Refer Slide Time: 14:28)



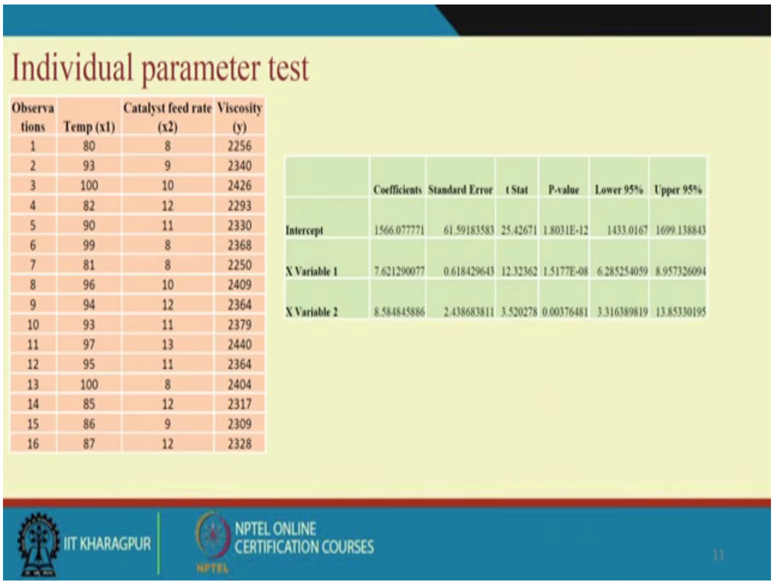
You just see that multiple R square value that 0.96, I think R square is SS by SS R by SS T, fine. So, R square value 0.92 adjusted R square 0.91 and standard error is 16.3 by observation 16. So, R square and adjusted R square both are more than 0.9. So, we can say that it is good the data is fit to the regression model.

(Refer Slide Time: 15:07)



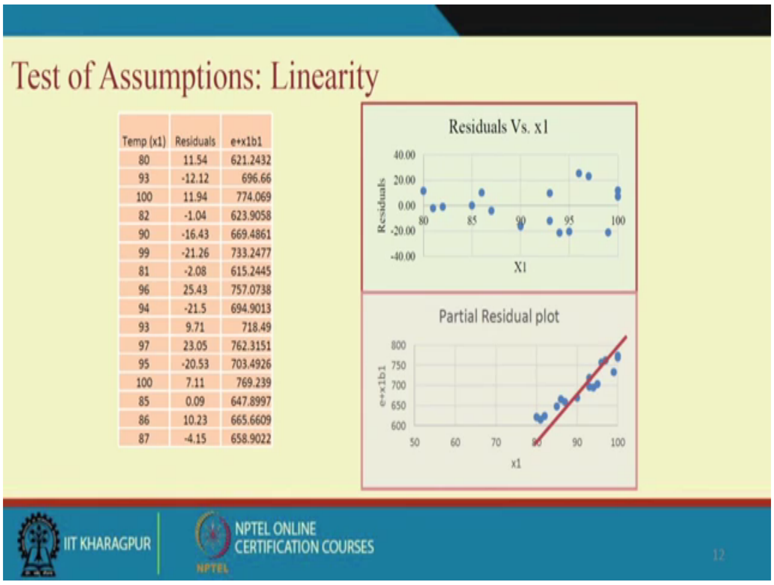
This is the f test what I explained to you. Now f value computed 82 and it is highly significant; that means, what happened model weather predictors of the factors are contributing.

(Refer Slide Time: 15:24)



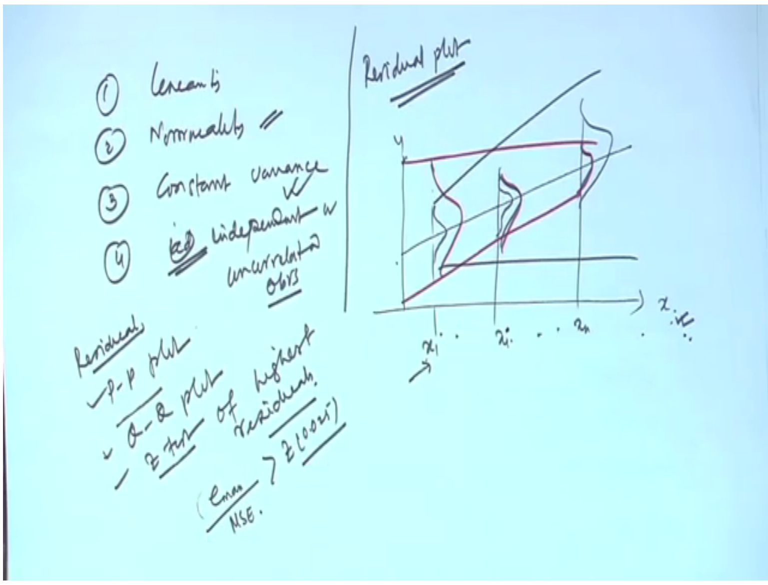
Individual parameter test earlier I have shown you and here this example is repeated.

(Refer Slide Time: 15:30)



Now another, now the test of assumptions; test of assumption is very very important in regression.

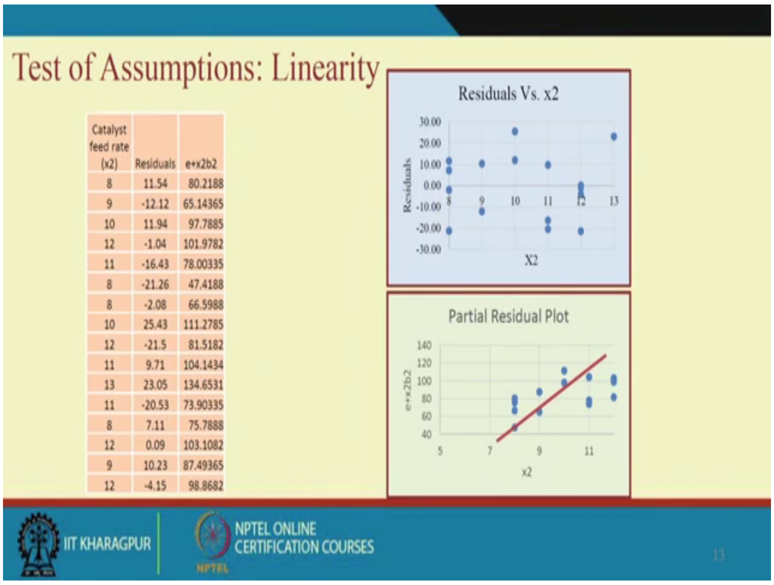
(Refer Slide Time: 15:41)



What are the assumption, multiple linear regression, one is the linearity normality 3 is your constant variance 4 is iid independent, independent in nature I can say independent or uncorrelated observations independent or uncorrelated observations.

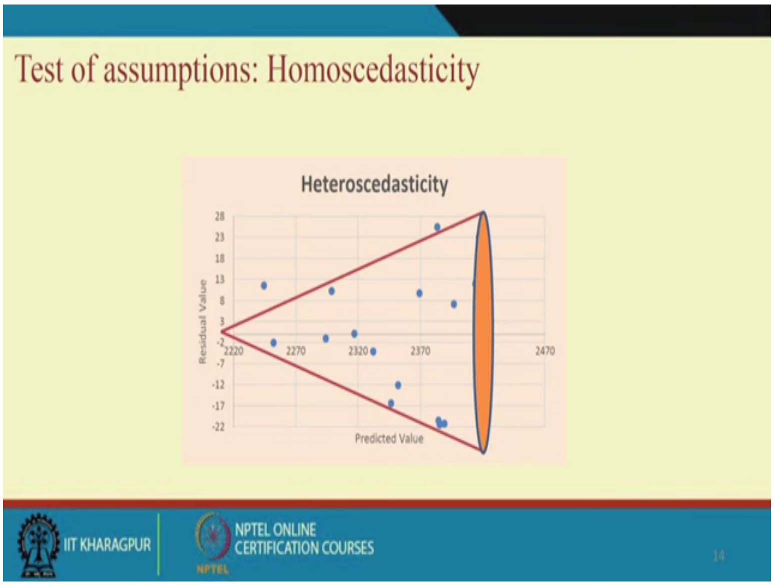
So, from residuals plot we will be able to find all those things residual plot, plot will help you in knowing that whether they are valid or not anyhow. Let us see that for this data set we have found out the residuals this is temperature versus residual. You see the residual plot here now in case of linearity we have shown here partial residual plot this is residual versus X 1; that means, these versus this and we do not find any pattern here. But when you go for partial residual plot; that means, the residual plus the X 1 contribution and versus X 1 then you see the linear plot is obtained so that means, linearly related.

(Refer Slide Time: 17:22)



Now, similarly for X 2, but this linearity is, this linearity is your little weak because there is more spread.

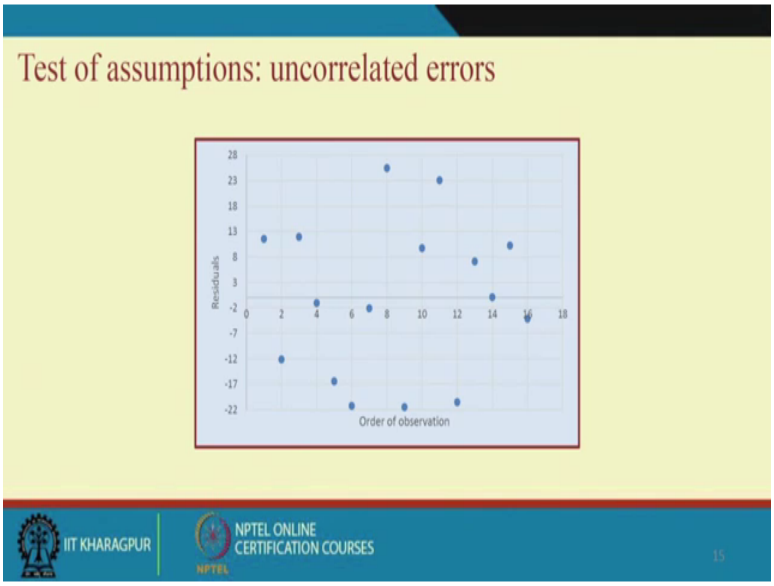
(Refer Slide Time: 17:38)



Now second one is that normal here I am showing that heteroscedasticity, it is nothing but that the error variance are not constant over x. If you recall my first I say if it is x and this is y suppose we are fitting a regression line when we are having only one predictor variables or factors then we say that for every fixed observations suppose this is x 1 this is x i and this may be x n every fixed observation, if I keep x at x 1 and do experiment several times I may get y like this y distribution like this here y distribution like this here. The assumption is that irrespective of your x value the y variability across x, will variability will be same or constant variance across the x this is known as homoscedasticity.

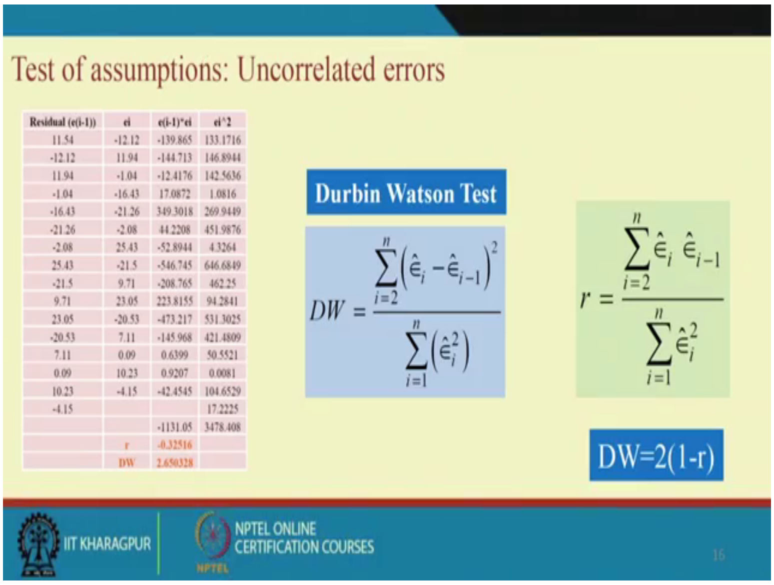
If there is violation like this the when X is low like the value is like this if high like this high like this or other way around. So, this one is big one or other way around, so this is the biggest one this is what is this and this is what is this. So, then you will have either funnel to left or a funnel to funnel to right this kind of situation then this is a called this is a situation for heteroscedasticity for not un constant error variance across y and this is a violation of constant variance and aggression estimates will be inflated or un or it will be wrong that is the issue. So, if you how do you know that it is happening. So, if you plot the residual value versus predicted value and found some kind of funneling effect then it is a heteroscedasticity problem, but if it is random then it is not.

(Refer Slide Time: 19:46)



So, for uncorrelated errors are independent, both or the order in which you have done the experimentation put this side and residual in the vertical y side. So, then you see that whether these observations are showing any, it is a random observations there is no systematic pattern in the plot then it is uncorrelated.

(Refer Slide Time: 20:12)

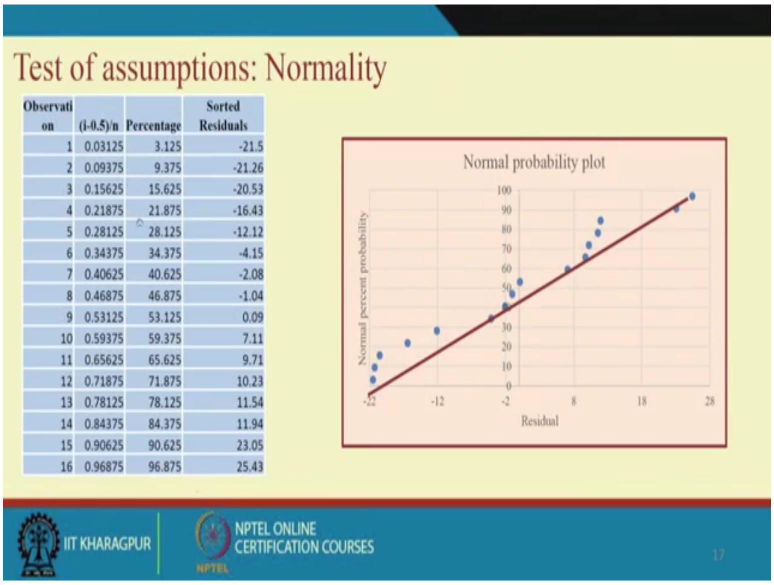


So, what about the linearity a normality part I think I have told you p-p, p-p plot are told earlier. So, for all residuals you do this, this is used for p-p plot, q-q plot, q-q plot or also you can say individual z test, suppose z test of highest error highest residual that test of highest residual maximum residual. So, you just find out the maximum residual e max from this data and then you divide it by your error variance that is MS E and if this one this is greater than suppose z 0.025 then we can say that there is the violation of normality.

So, for uncorrelated errors there is another test called Durbin Watson test and here what happened just to see that whether the errors are correlated or not, you create the you take the error in or in this order. And you keep create some lag maybe for 1 lag or some lag that 1 lag, 2 lag, k lag maybe and then you find out the correlation between the 2 and here what happened this is basically the residuals and then we created the i minus 1 and e i this then we have found used this formula and found out the R value and it is shown that the R value is minus 3.33 and Durbin Watson test value is 2.65 it is correlated. But 0.3 is significant I think because it is not less.

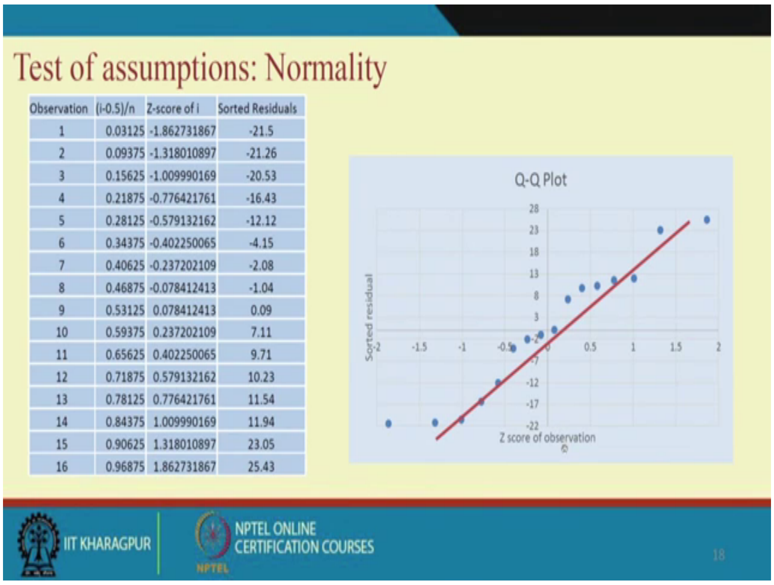
Again R square Durbin Watson value is also more than 2, but it can be considered other way also whether it is also R square and other things giving you better, but this is one assumption which is supposed to be violated here because the constant error sorry independent part.

(Refer Slide Time: 22:39)



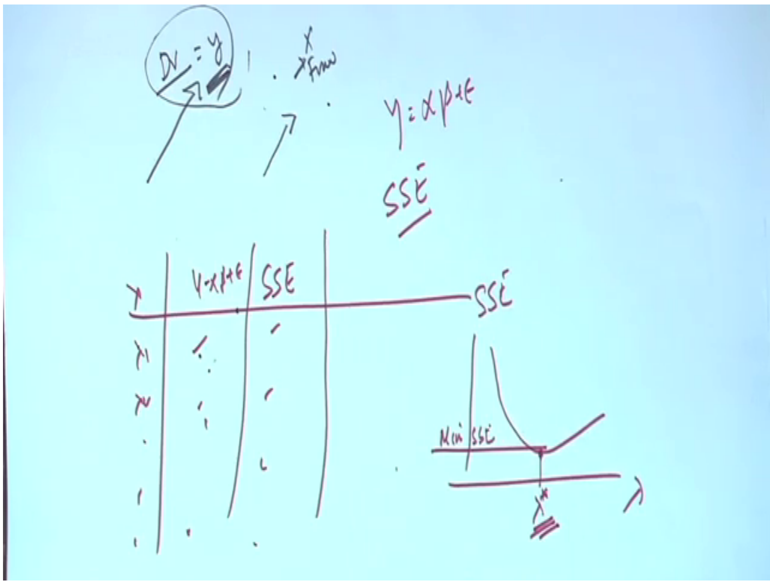
Now, I told, I talk to you about normality and this you know this normality earlier I shown that the residual that you use the normal probability plot and if there is a straight line kind of things and this i minus 0.5 by n. If you use this formula you will get a straight line that is normal there is another one is that residual versus z score means quantile quantile plot. This quantile quantile plot is also showing some kind of normality that is not deviating from normality a much.

(Refer Slide Time: 23:02)



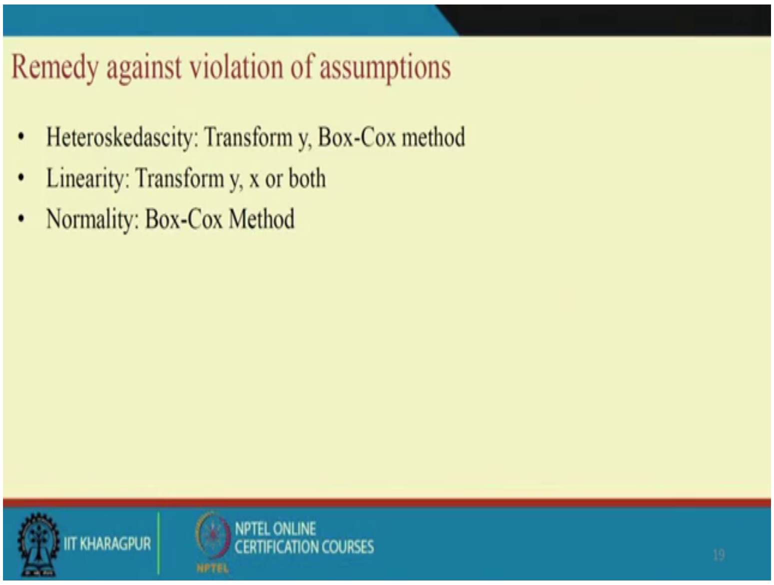
Last, but very important one also that means, what happened when the constant error variance as well as normality assumption is violated you have to transform the data. So, there are many methods of data transformation, and please remember this homoscedasticity it is a feature for the dependent variable y, dependent variable y.

(Refer Slide Time: 23:45)



We are talking about fixed effect model and X are fixed ok. So, when you talk about transformation so it is primarily related to y, but sometimes many times we also do X transformation so that the relationship can be change some of the assumptions can be satisfied for example, linearity; for example, may be normality also.

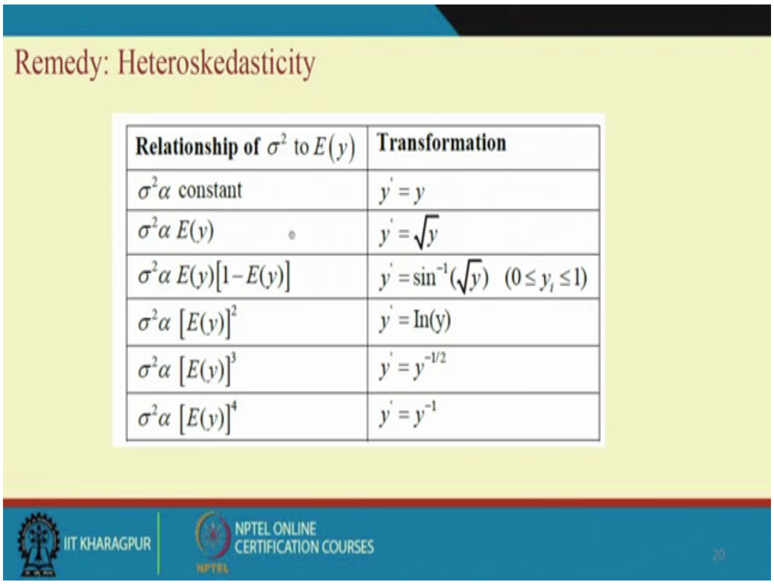
(Refer Slide Time: 24:21)



So, here what I will show you I will show you very quickly some of the methods for transformation one is Box-Cox method and another one is I think some more methods are there but primarily we will be considering on Box-Cox.

If there is heteroscedasticity then you transform y using Box-Cox method if there is linearity you can transform y, x or both it will help you. If it is normality again go for Box-Cox method.

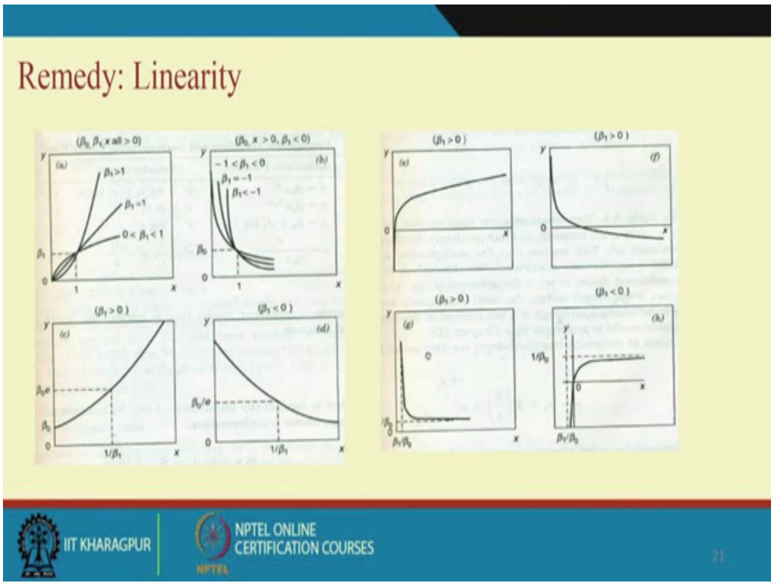
(Refer Slide Time: 24:58)



Relationship of σ^2 to $E(y)$	Transformation
$\sigma^2 \alpha$ constant	$y' = y$
$\sigma^2 \alpha\, E(y)$	$y' = \sqrt{y}$
$\sigma^2 \alpha\, E(y)[1-E(y)]$	$y' = \sin^{-1}(\sqrt{y})\quad (0 \le y_i \le 1)$
$\sigma^2 \alpha\, [E(y)]^2$	$y' = \text{In}(y)$
$\sigma^2 \alpha\, [E(y)]^3$	$y' = y^{-1/2}$
$\sigma^2 \alpha\, [E(y)]^4$	$y' = y^{-1}$

And this is what is the transformation kind of things. Suppose you may if constant variance is satisfied there is no transformation required, but if the error variance is proportional to the mean value then find out that take the square root. If error variance is proportional to mean and this quantity then make sin inverse square root transformation. If it is square of mean log transformation, if it is cuba mean then inverse square root transformation, if it is 4th power mean then you do 1 by y that is inverse of this. These are some guidelines you can use while you do some kind of analytics that is experimental data analysis.

(Refer Slide Time: 25:53)



So, you see that although this kind of transformation will give you the desired result and this is available in montgomery linear statistical models.

(Refer Slide Time: 26:05)

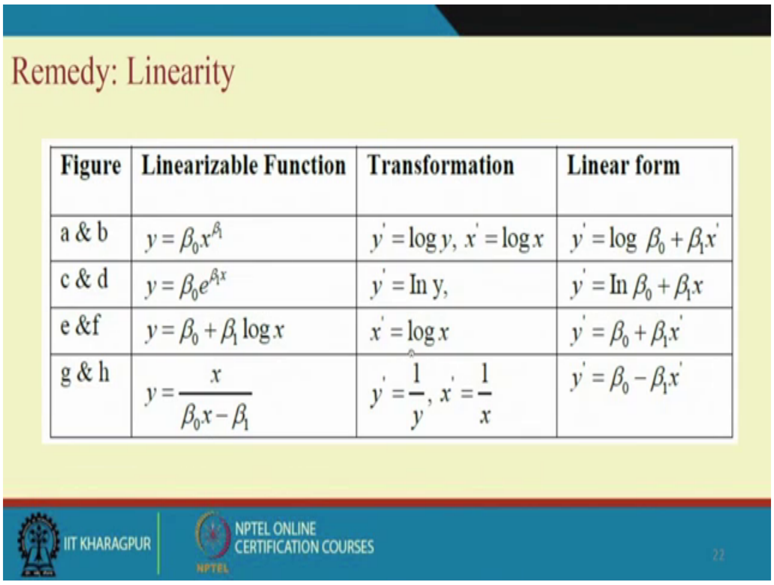
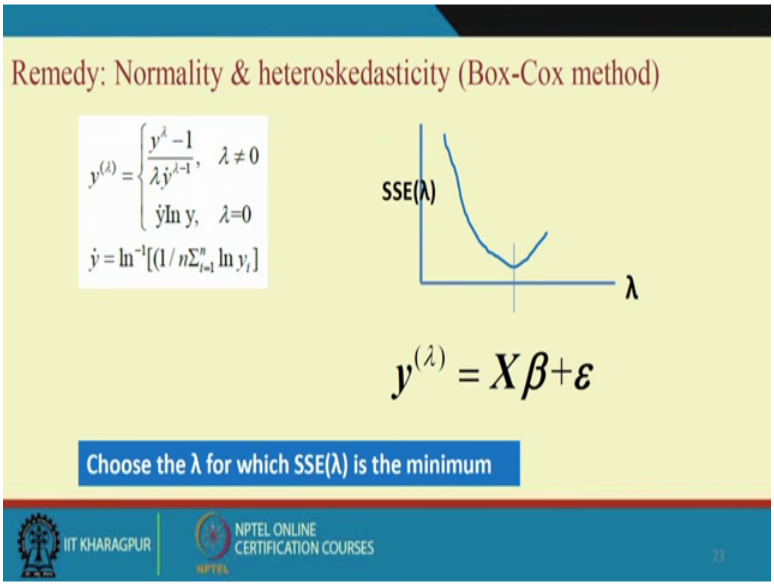


Figure	Linearizable Function	Transformation	Linear form
a & b	$y = \beta_0 x^{\beta_1}$	$y' = \log y,\ x' = \log x$	$y' = \log \beta_0 + \beta_1 x'$
c & d	$y = \beta_0 e^{\beta_1 x}$	$y' = \ln y,$	$y' = \ln \beta_0 + \beta_1 x$
e &f	$y = \beta_0 + \beta_1 \log x$	$x' = \log x$	$y' = \beta_0 + \beta_1 x'$
g & h	$y = \dfrac{x}{\beta_0 x - \beta_1}$	$y' = \dfrac{1}{y},\ x' = \dfrac{1}{x}$	$y' = \beta_0 - \beta_1 x'$

So, you see that if the function is something like this, then you go for log transformation function like this go for this, function like this go for distribute transformation and you do not know which transformation will help you, but that is why what happened you may do all kind of transformation and then find out that whether the assumptions are satisfied particularly in the linearity point of view here.

(Refer Slide Time: 26:34)

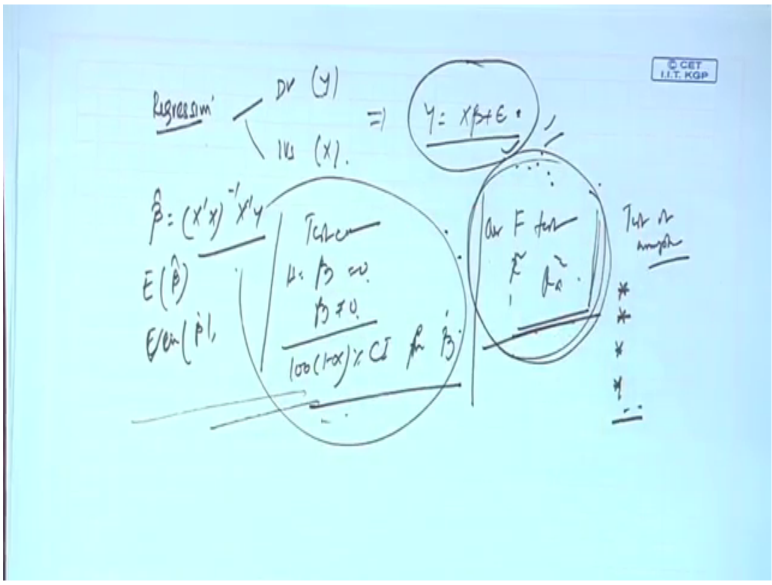


And you do Box-Cox transformation for normality and heteroscedasticity. Here what happened? You basically create a power lambda then this you are creating in the norm variable y to the power within bracket lambda which is y to the power lambda minus 1 by lambda into y dot lambda to the minus 1. Now, the lambda obviously, not equal to 0 in the first case if lambda equal to 0, you write y dot log y where y dot is inverse log one by n i equal to 1 to n log y i that mean you take the log of the original observations then take their sum divide it by the take the average, average of the log transport information and then its inverse log inverse.

So, you choose and this is basically y dot and now you choose different lambda values feed the regression equation y equal to X beta plus epsilon calculate SS E and then plot. So, what happened? You will choose lambda different lambda value different lambda value like this then your y equal to X beta plus epsilon some model will be there and then you will calculate SS E some value you will get SS E and then you plot lambda versus SS E; what will happen, you will find out a curve like this where this lambda is giving you the minimum SS E minimum SS E value this lambda this is a lambda step. This is the best transformation for you.

So, choose lambda for you is lambda SS E lambda is the minimum and now again fine that equation is also known you know which lambda is minimum an equation is known and that linear equation in huge. So, let me just do one let me summarize the thing.

(Refer Slide Time: 28:57)



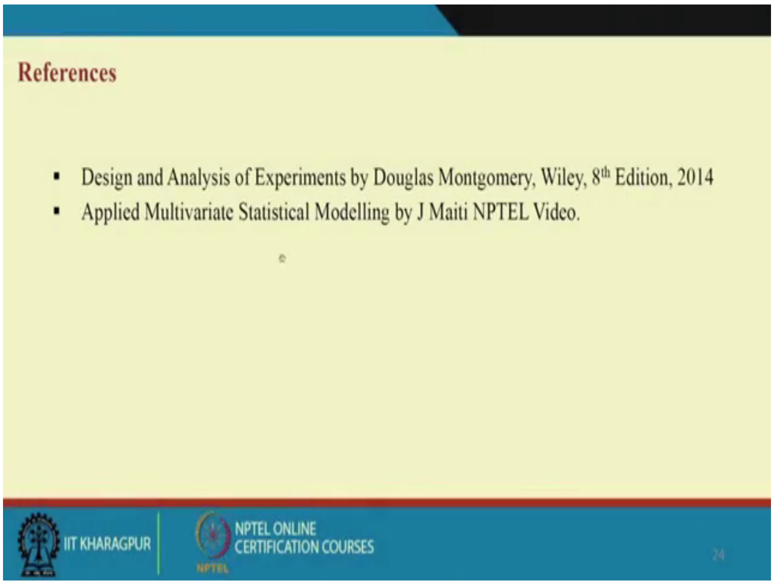
We talked about regression and primarily it is linear regression and in linear regression obviously, we have 2 sets of variable dependent variable 1, independent variable several.

We have are interested to find an X equation like this where epsilon is the error term. We have estimated beta using certain equation and then we have estimated the mean value and the variance of beta. Then what we have tested beta tested beta, beta j equal to 0, H 0 beta is equal to 0 beta j not equal to 0. We also found out the con hundred into 1 minus alpha percent confidence interval for beta j, beta j that also you found out.

And then also we found out that whether the model is fit or not using f test overall f test and your R square Ra square and then also we have that a say H 0 a set of variables contributing or not contributing both partial marginal test and subset hypothesis testing we have done. Then we have gone for test of assumptions test of assumptions both linearity, normality, homoscedasticity independence all those things and we have also seen some of the examples particularly here we are using 2 independent variables and one dependent variable and how this model is working or not.

Nutshell what I mean to say you must know that the data fit to this model. You must know that every model has certain assumptions those assumption must be tested. You must be aware that when you are saying a model that model must be fit to the data that is the adequacy test, overall fit test. When the model overall model is fit then you go for individual parameters whether individual they are contributing or not and accordingly you accept or discard those variables which are not contributing, again you rebuild the model with the significant parameters. You may find out some of the estimates are different than the earlier, if that is the case you please be careful if some of the variables become insignificant then there may be the partial correlations and there may be problem in from the beginning data collection to variable selection to maybe the test and test of assumptions and all these things.

(Refer Slide Time: 31:54)



So, thank you very much. Again I must tell you that the Montgomery book and my earlier NPTEL video lecture on multivariate statistical modeling. Thanks a lot.

Thank you very much.