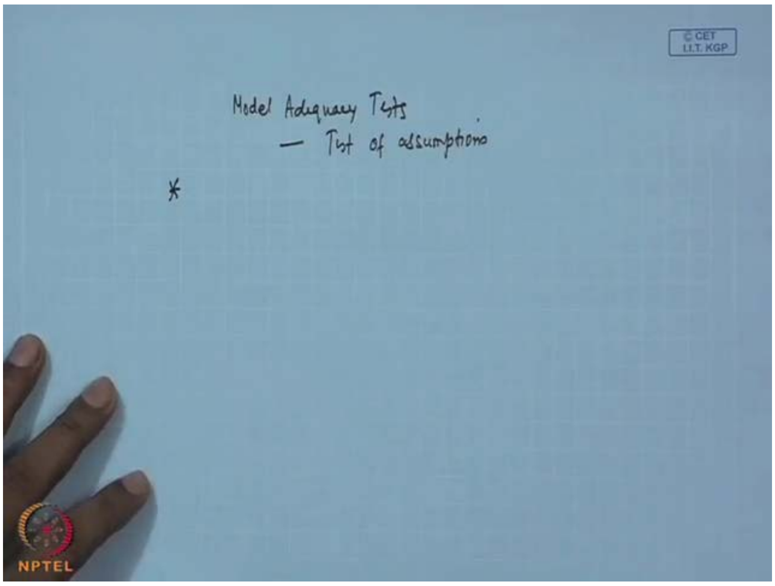
Applied Multivariate Statistical Modelling
Prof. J. Maiti
Department of Industrial Engineering and Management
Indian Institute of Technology, Kharagpur

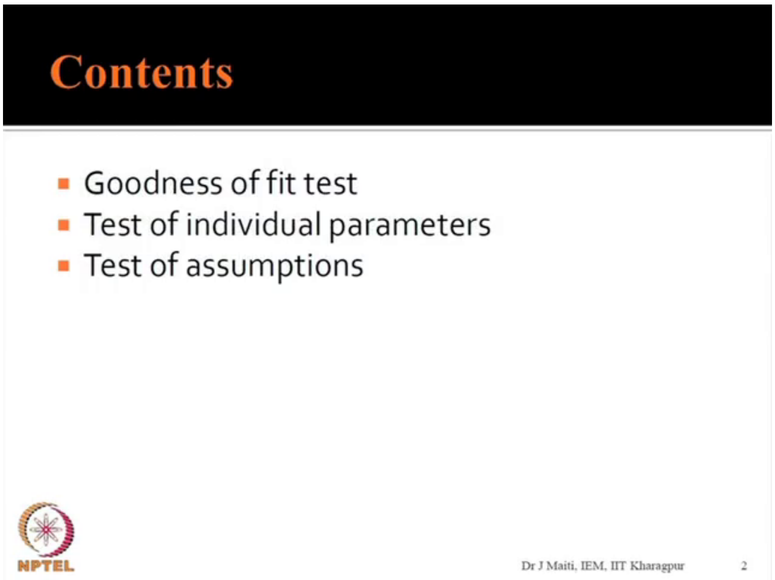
Lecture - 24
MLR - Test of Assumptions

(Refer Slide Time: 00:23)



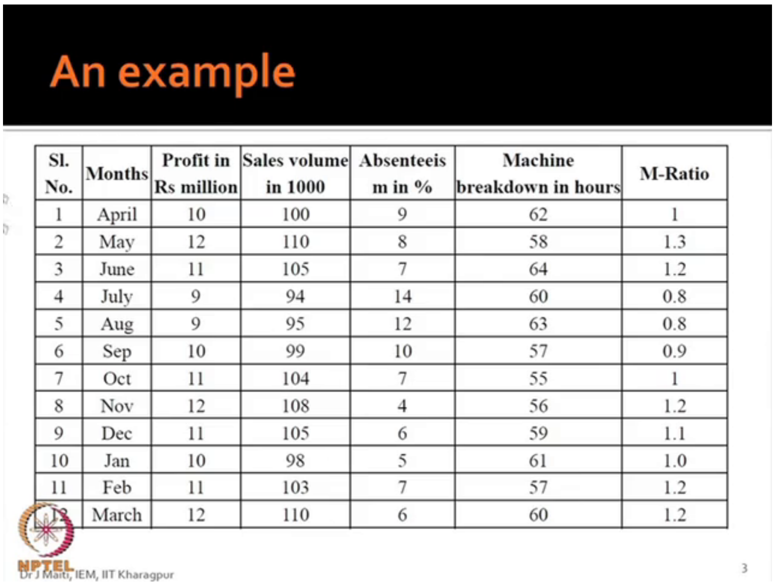
Good afternoon. Today, we will continue model diagnostics. Today we will continue model adequacy tests, adequacy tests. Under this, today's topic is tests of assumptions.

(Refer Slide Time: 00:53)



Now, let us look back what we have completed so far under multiple linear regression.

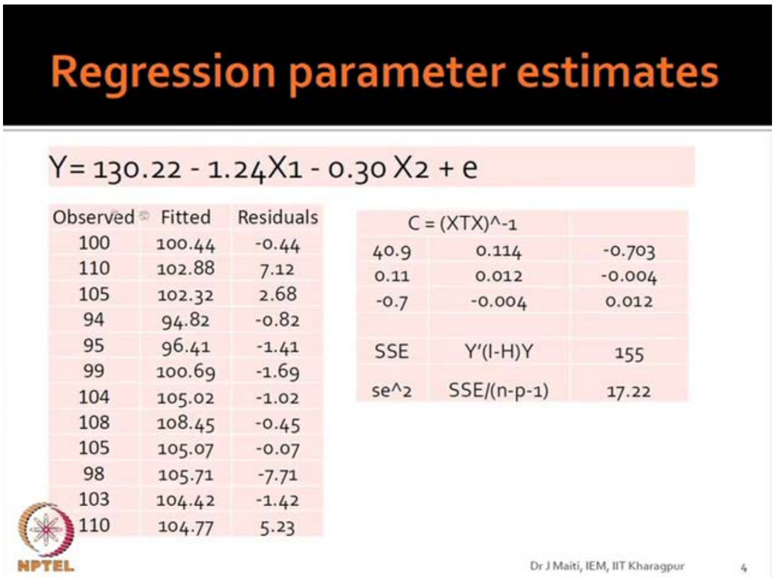
(Refer Slide Time: 00:56)



Sl. No.	Months	Profit in Rs million	Sales volume in 1000	Absenteeism in %	Machine breakdown in hours	M-Ratio
1	April	10	100	9	62	1
2	May	12	110	8	58	1.3
3	June	11	105	7	64	1.2
4	July	9	94	14	60	0.8
5	Aug	9	95	12	63	0.8
6	Sep	10	99	10	57	0.9
7	Oct	11	104	7	55	1
8	Nov	12	108	4	56	1.2
9	Dec	11	105	6	59	1.1
10	Jan	10	98	5	61	1.0
11	Feb	11	103	7	57	1.2
12	March	12	110	6	60	1.2

We have taken this data set.

(Refer Slide Time: 01:01)

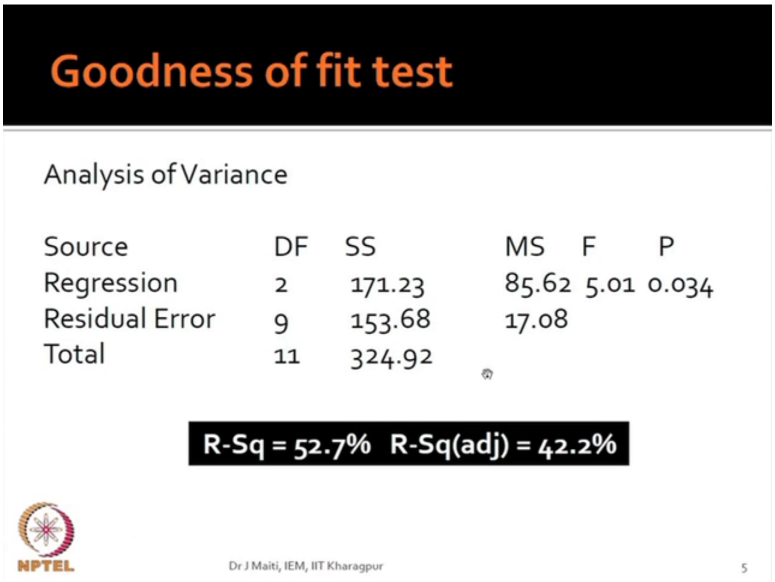


Observed	Fitted	Residuals
100	100.44	-0.44
110	102.88	7.12
105	102.32	2.68
94	94.82	-0.82
95	96.41	-1.41
99	100.69	-1.69
104	105.02	-1.02
108	108.45	-0.45
105	105.07	-0.07
98	105.71	-7.71
103	104.42	-1.42
110	104.77	5.23

	C = (XTX)^-1	
40.9	0.114	-0.703
0.11	0.012	-0.004
-0.7	-0.004	0.012
SSE	Y'(I-H)Y	155
se^2	SSE/(n-p-1)	17.22

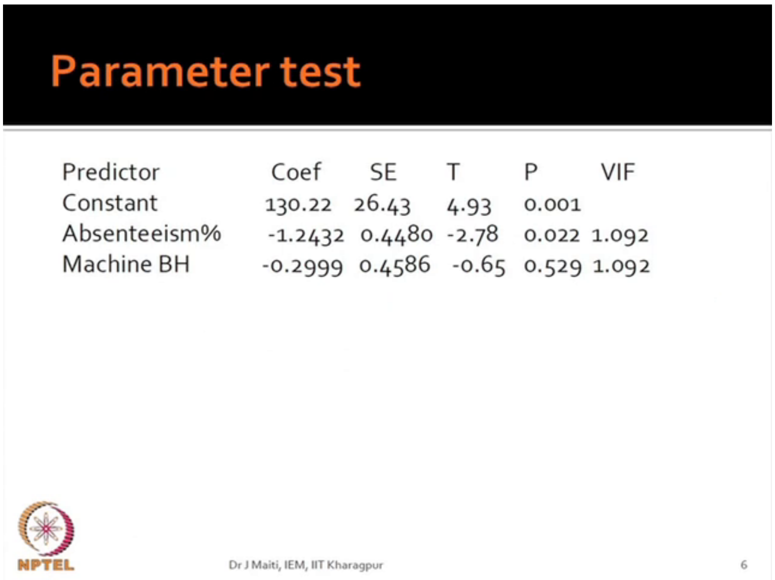
Then, we have fitted the regression equation with sales volume Y and your machine breakdown hours X2 and absenteeism X1 and the observed value. This is the fitted value and residuals. This is your X transpose X inverse on that matrix and which we have used to calculate some of the regression statistics.

(Refer Slide Time: 01:30)



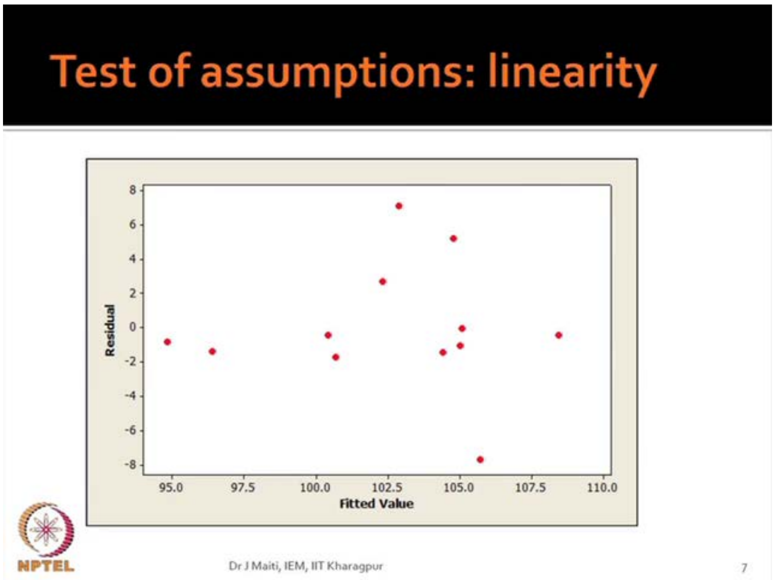
Analysis of variance, we have completed.

(Refer Slide Time: 01:36)



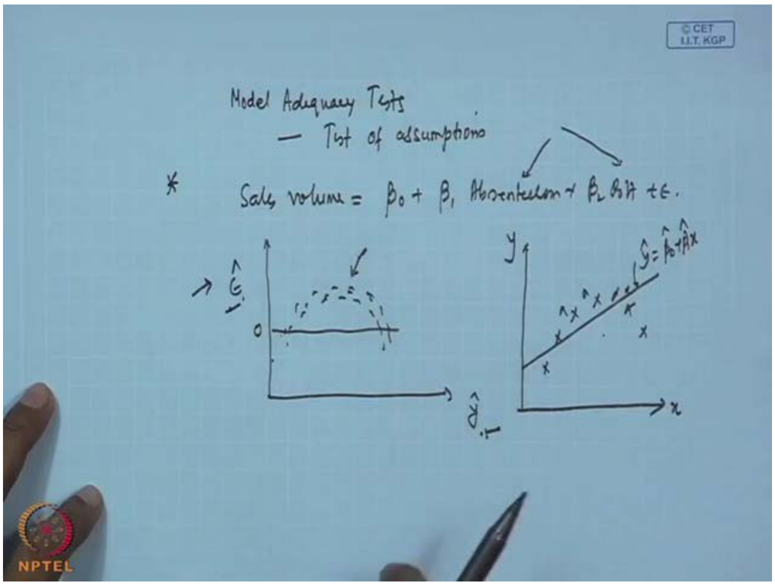
We have also completed the parameter tests.

(Refer Slide Time: 01:38)



Then, today's topic is test of assumptions. One of the assumptions is linearity.

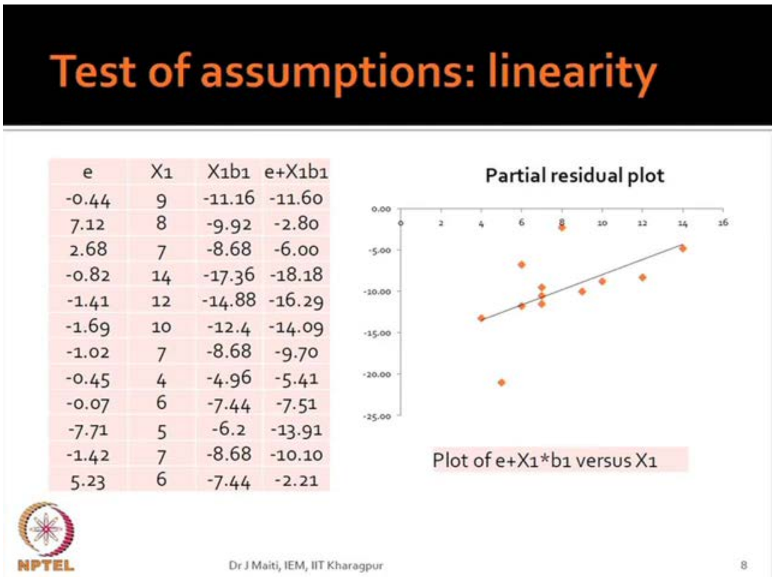
(Refer Slide Time: 02:17)



Now, if you look into the results of the given problem that is the small company is doing business in local city where sales volume is a function of your absenteeism and breakdown hours plus error, so this diagram is related to the that fitted value that is y cap beta is the residual. If there is any anything that is departure from linearity, then what will happen? The scenario will be like this. Suppose you took y versus x, then your data points are like this. Now, if you fit linear model, this is y cap beta 0 plus beta 1 x beta 1 cap plus beta 1 x, this is the regression line.

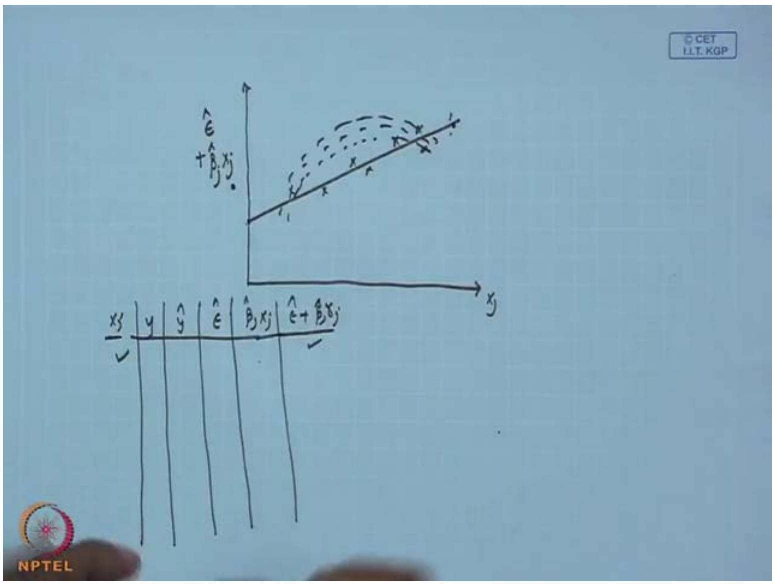
Then, what will happen? When you plot the error versus that is the fitted residuals versus predicted value, you will get the line, you will not get straight. Your non linear part will come here it will come here like this. This will be the zero value error part. Now, in this particular figure, when you see the residual versus fitted values, I do not find that any departure from the linearity. Now, if there is departure from linearity, first of all you will be when you feed the residual with fitted values, you will be getting this type of plot. Next, when there is more than one variable like this case, then if there is any departure, we are interested to know that for which variable, this non linearity is there? So, that is also possible.

(Refer Slide Time: 04:42)



That can be done through partial residual plot. Now, what is partial residual plot? Partial residual plot looks like this.

(Refer Slide Time: 05:02)

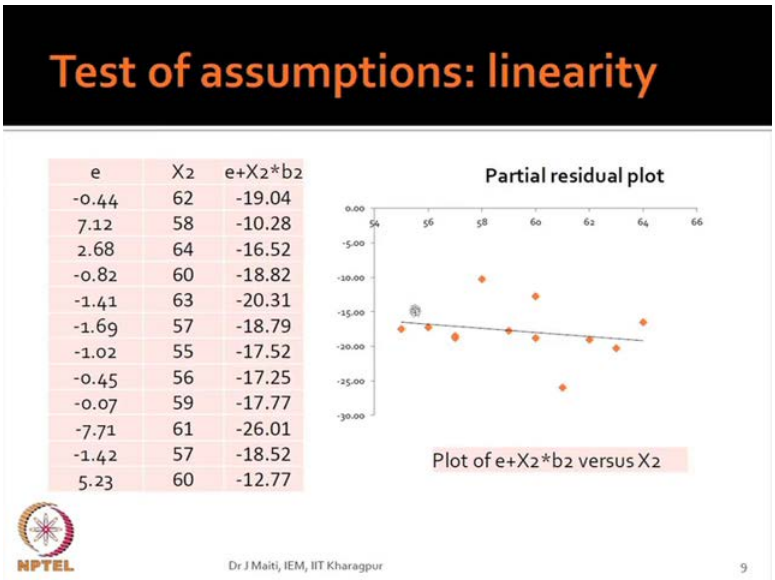


Here, instead of plotting your residuals versus fitted values, we will plot residual plus suppose the jth variable beta j, xj and xj. So, from the fitted your y sales volume, then y cap, then you are getting the epsilon that is these are the residuals. In addition, you have to have beta j xj. Then this also you have xj somewhere here, so all those values. Now, you require this plus beta j cap xj and this versus this, when you plot, you will be getting a plot. When there is no departure from linearity, you will be getting a straight line, for example like this and your data will be hovering around this straight line.

If there is departure from normality linearity, if there is departure from linearity, then what will happen? In this case, the non linear portion will come and it will look like this. Then what is the difference between these plots? Beta is the residual plot, this is known as residual plot and here you see your non linear portion is coming under straight line, which is parallel to y cap. Here, this value is zero, error value is zero. Now, in this case, these residuals, these are basically hovering around this line. Here, these points, it is not zero. It is basically epsilon plus beta j cap xj. So, that is the difference.

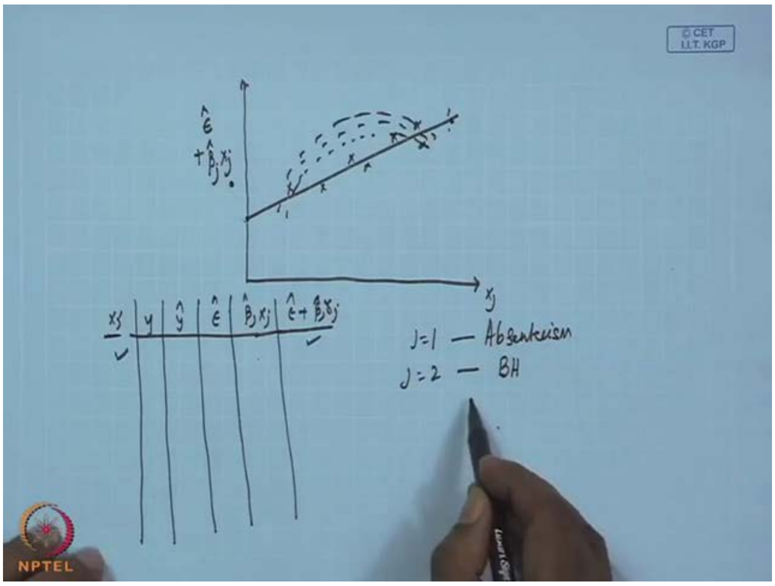
Now, if we want to see what has happened to that problem given, if I go by e plus X1 beta 1 versus X1, this side it is X1 and in this side, it is e plus X1 beta, beta1, then you are getting a plot like this. So, from this plot, it cannot be said that there is departure from linearity.

(Refer Slide Time: 08:05)



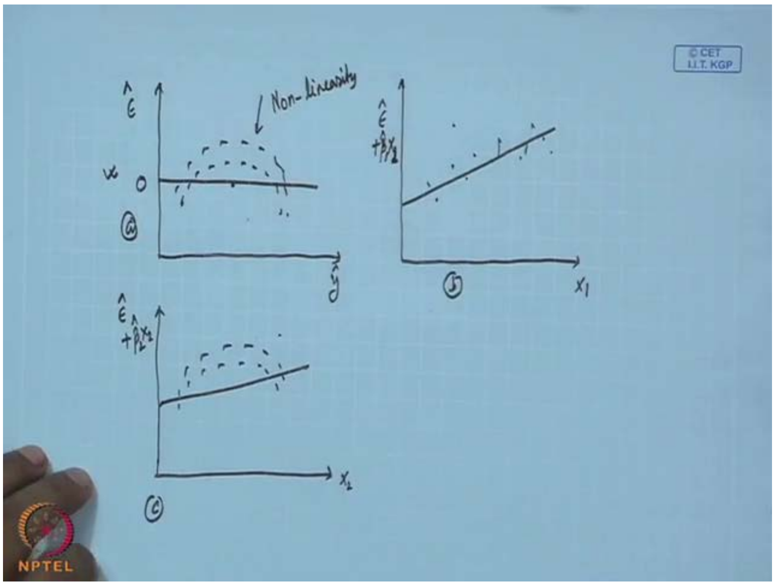
In a similar manner, you can see the other one. So, we have taken two independent variables here.

(Refer Slide Time: 08:11)



When j equal to 1, it is absenteeism. When j equals to 2, this one is our breakdown hours. So, for breakdown hours plot, partial residual plot, if you see here also, you see there is no departure from linearity. Then what will happen if we find figure like this?

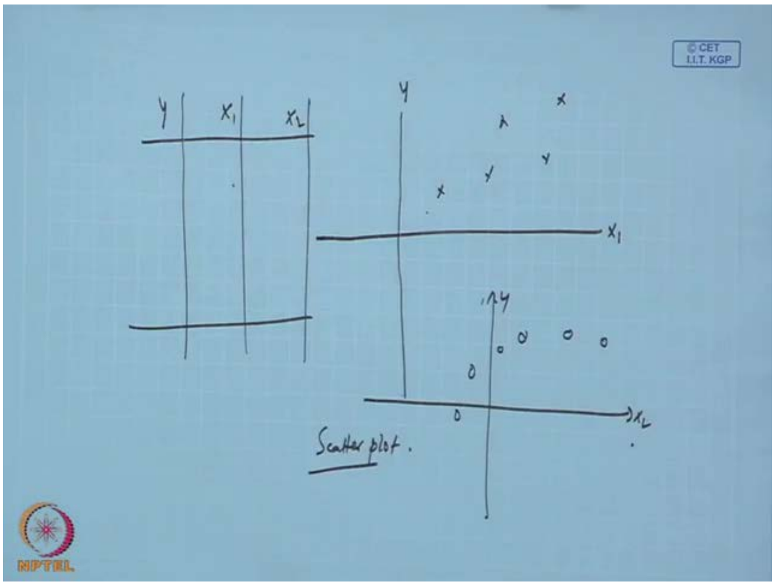
(Refer Slide Time: 08:50)



Suppose, if you plot first residual plot, which is epsilon cap, bishop is y cap, this is zero value and your plot residual plot looks like this, so it indicates non linearity. Now, suppose there are two variables, x1 and x2, representing the independent variables. So, you are doing like this beta 1 cap x1 and this side x1 and you found a curve like this.

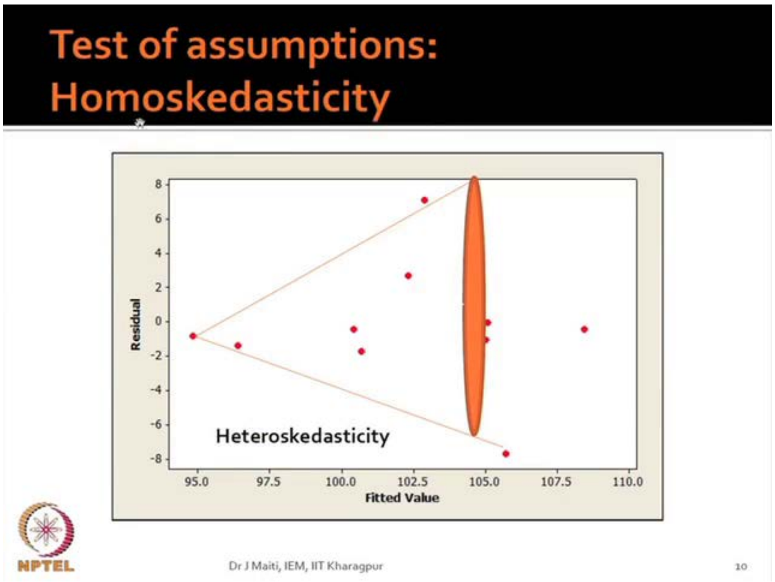
When you plot for the second one x2, then epsilon plus beta 2 cap x2, suppose this one is like this, but residuals like this. So, first plot is showing you there is no linearity. Second plot if I say this is a, then b and c, second plot is showing that this non linearity is not due to x1. The third plot is showing this non linearity due to x2. Should I repeat? So, linearity assumption is vital.

(Refer Slide Time: 10:54)



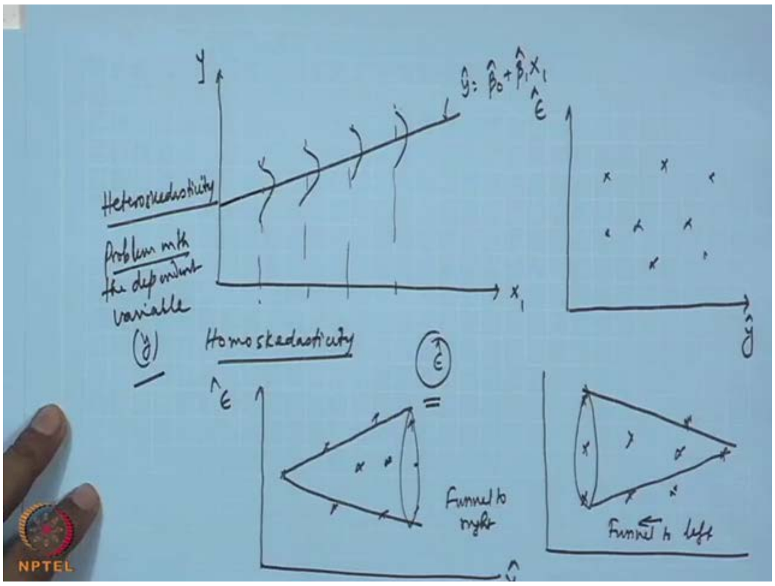
When you collect data that time, suppose you have data y, x1 and x2 like this. So, for linearity, you are just developing your scatter plot. If you develop scatter plot, suppose this is y, this one is x1. Then by seeing the scatter plot, you will be able to find out whether there is linearity or non linearity. Suppose y versus x2, so plot may be something like this. Even though when there is, you are not able to see any departure from linearity from the scatter plot, the scatter plot, what you require to do, even though you require testing the residuals. So, residual test is just confirmation that whether there is linearity or not because residual will give you little better picture; even if there is a small amount of non linearity that will be captured through residuals.

(Refer Slide Time: 12:17)



Our second assumption is homoskedasticity. By homoskedasticity, we mean that equal error variance.

(Refer Slide Time: 12:29)



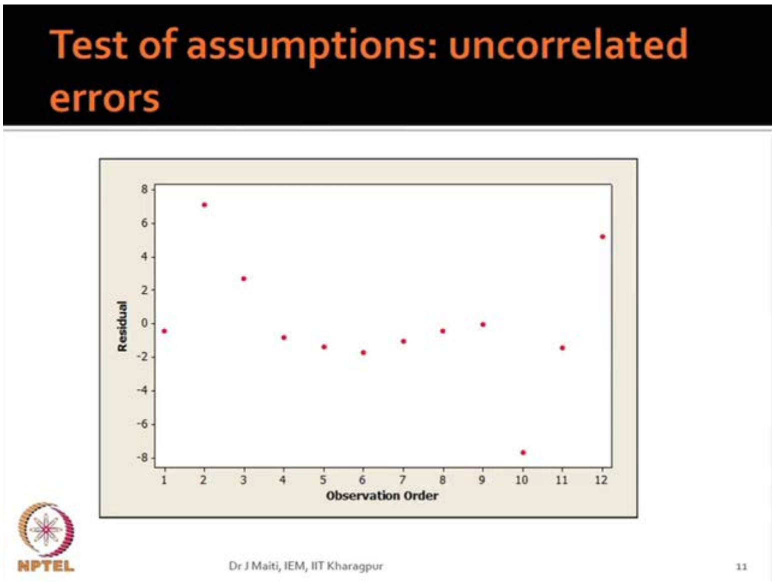
Basically, our assumption was like this suppose you are going through x and y plot. Let x is only one variable and you are doing regression fitting. Your this one is y cap equal to beta 0 cap plus beta 1 cap, let it be x1, this also is x1. So, there are many values x1. We assume that that here across all values of x1, y's variability remain constant that is equal variance homoskedasticity. This is this is the condition called homoskedasticity.

Now, when you are fitting regression line, you are getting this line. Then where this variability part is going? This variability part will be accounted by your error term. So, if you do the residual plot, you will get you will get the picture. If there is heteroskedastictiy, it will be rebuilt. If your thing is like this, suppose your error versus residual plot is like this, there is no trail nothing is there. Then this is homoskedasicity condition but what will happen? If your plot is like this, suppose this is the case, your plot is like this or this side is your epsilon cap and sigma cap or your plot may be like this like this.

So, if we do little work, more work here, what I you can find out that this, this one is like this, this one is like this. So, a funnel is created here, so funnel to right and this one is funnel to left. So, if your residual plot resembles this type of picture, it is the heteroskedasticity condition that means equal error variances are not satisfied. So, the problem we have discussed, the same problem when you plot the residual versus fitted value, you see this is like this. So, I have created funnel shape because it resembles a funnel shape.

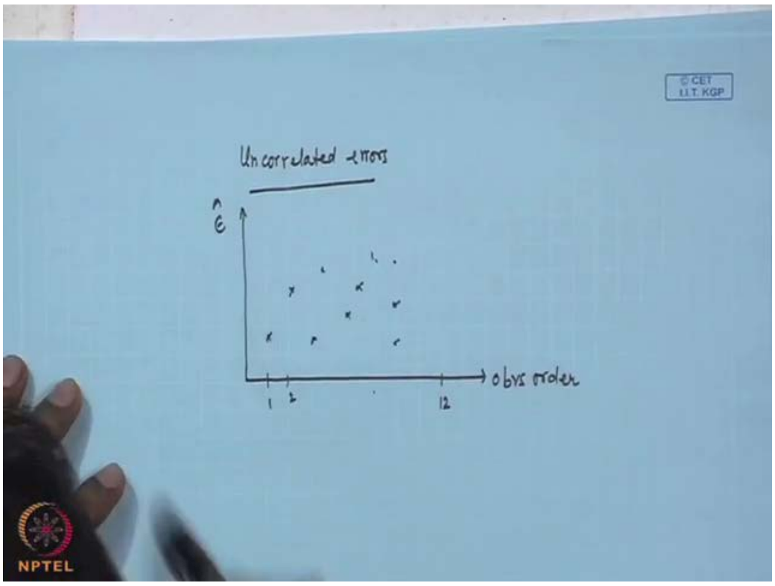
So, it simply indicates that the small problem what we have discussed problem there is heteroskedasticity. Heteroskedasticity is a serious issue and it is related to the dependent variable. It is a problem with the dependent variables. So, when we talk about heteroskedasticity, it is a problem of dependent variable y. So, later on, we will see the remedy. Under remedy, we have to transform y.

(Refer Slide Time: 16:31)



Another assumption was uncorrelated errors.

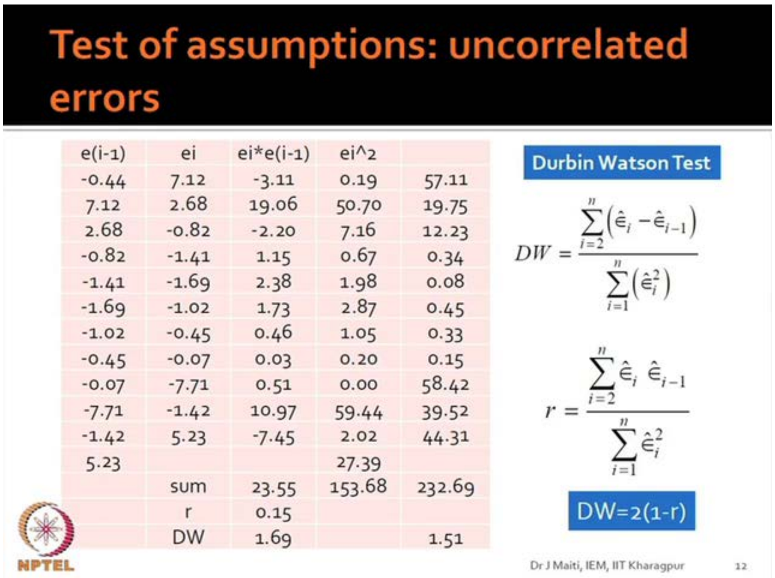
(Refer Slide Time: 16:37)



So, here the plot is residual versus observation, order observation, order versus residual, so 1, 2, like this suppose 12 observations. Then there should, again this should be random. What we mean to say is that there should not be any pattern. So, if there is randomness, then it is fine, otherwise there is correlation between the observations, observed data, which is not desirable.

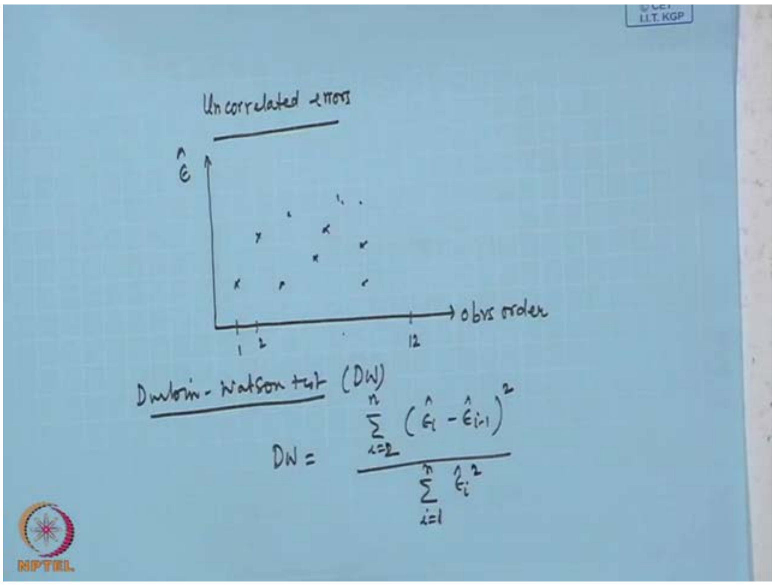
If you see the problem what we have discussed, it is difficult for me to say whether there is a particular trained or not because data points are limited, only 12 data points, but if we see like this, this side, there something some non linear type of things are coming. But, we have seen for this problem, the data is linear. So, here I think it is very difficult. So, whenever you find difficulty, I think there is some quantitative test also, which can be done to see the uncorrelated error condition.

(Refer Slide Time: 18:06)



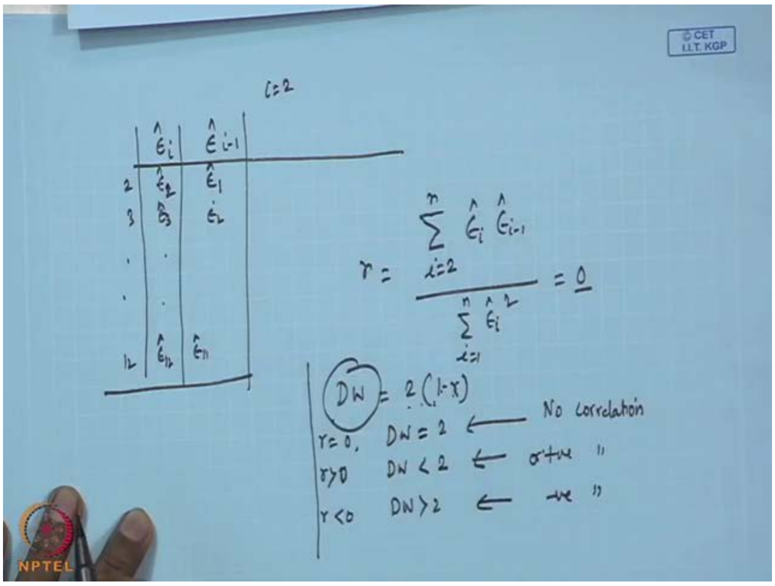
The test commonly used in is Durbin Watson test.

(Refer Slide Time: 18:14)



Durbin Watson test for uncorrelated errors, this test if I say that DW, it is basically that i equal to 2 to n epsilon i cap minus epsilon i minus 1 cap divided by sum total, i equal to i equal to 1 to n case and here there will be square and this i square. So, all of us know the correlation part also.

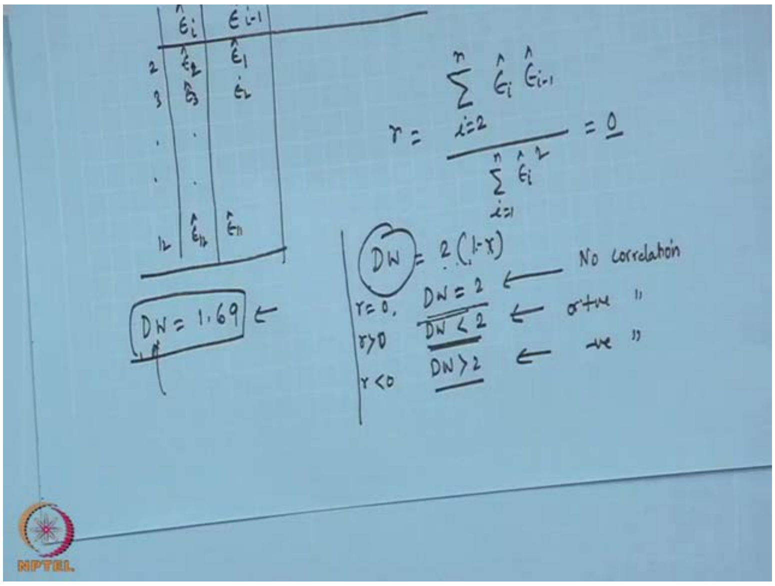
(Refer Slide Time: 19:15)



Suppose, what is happening here, if there is correlation, suppose this is epsilon i and this one if I write epsilon i minus 1, so when your i equal to two, so if I write 2, then it will be 2, this will be 1. If you write like this, then this will be 3, and this will be 2. So, ultimately this will be twelve this will be 11. So, that means if there is correlation, for now for different values of i, you can think of that i minus 1, you have given here. So, you can go for different lags also. Then r value will be sum total i equal to 2 to n epsilon i cap epsilon i minus 1 cap by sum of i equal to 1 to n epsilon i cap square. So, this is correlation.

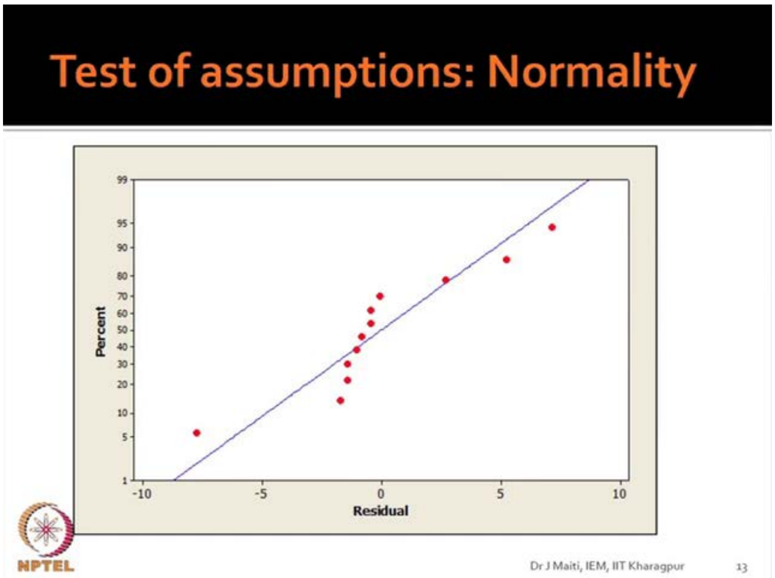
Now, this value should be 0. Now, there is relation with DW and this correlation coefficient. So, it is 2 minus r. So, when r equal to 0, DW equal to 2. So, the Durbin Watson test, this value, if it is two, it indicates no correlation. If r is greater than 1, then what will happen? If r greater than 1, if I say r greater than 0, let it be r greater than zero, then this will be less than 2. So, that means this less than 2 means positive correlation and when r less than 0, DW greater than 2, this is negative correlation. So, most of the software give these DW values. So, you find out what value it is coming. If it is two, then it is fine and no auto correlation. If this value is less than or greater than 1, then there is certain amount of auto correlation.

(Refer Slide Time: 21:55)



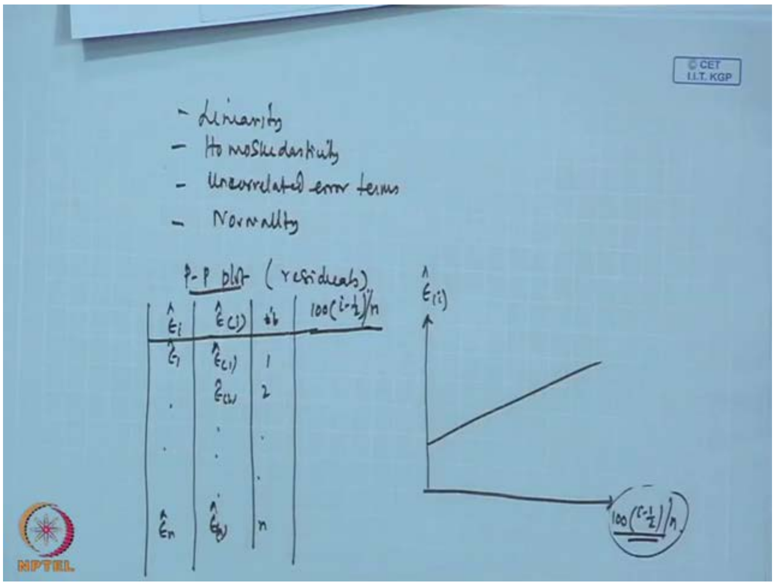
So, the problem what we have taken that our statistic problem, so there DW value is 1.69. So, it is little less than 2. It is positive correlation, positive correlation, little positive correlation is there, but it is the point estimate only. You may be interested to interval estimate also and you have to go for that statistic distribution. Accordingly, you will be able to find out.

(Refer Slide Time: 22:31)



So, we have seen the first one is linearity.

(Refer Slide Time: 22:32)

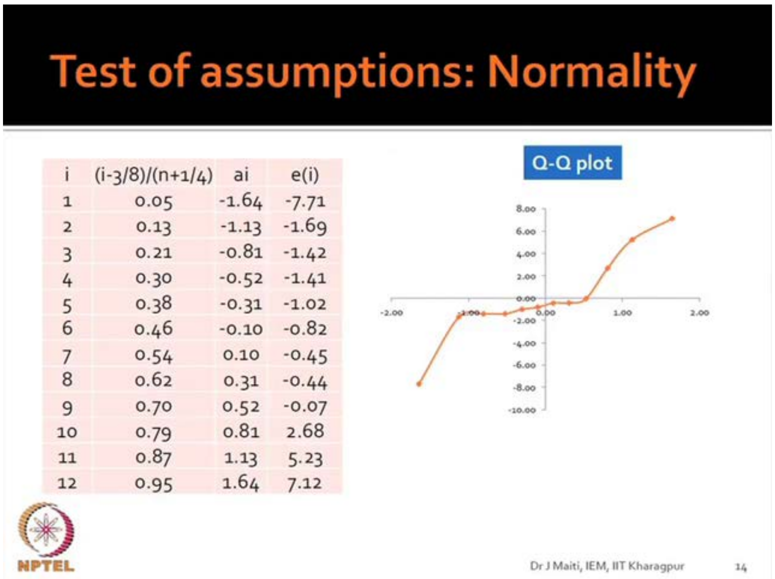


Second one we tested that I think homoskedasticity. Third one is uncorrelated error terms, uncorrelated error terms. Now, we will discuss normality. I think I have discussed normality earlier with respect to ANOVA. There what I have discussed, there I have discussed the PP plot, probability probability plot. Definitely, this is related to residuals. What we have done there? We have seen that suppose the residual is coming up to epsilon cap 1 to n. Then we found out the ordered one that is order residual like this. Then you all already have that i equal to basically, this is suppose ordered, i ordered means order residual i equal to 1, 2, like this to n.

Then, what we have created? We have created 100 into i minus half by n. Then we have plotted the two. You have plotted these two that means epsilon that ordered one and your 100 into i minus half divided by n and you got a straight line. So, just one question is why is the 100? It is basically making your percentage figure only, nothing else i minus half by 1 is all residual. Yes, that is residual in probability limit. So, in a software, this axis will be giving this one, but if you change the axis, there will be straight line, it will remain this one.

Now, you see that for our data, the problem we have discussed, this one is showing like this. So, there is departure. It is not normal. Actually, the interesting thing is that when there is heteroskedasticity, there is also usually normality problem. So, that is why, if you transform for heteroskedasticity to solve the heteroskedasticity problem, then you will find out that normality problem is solved. Many a times, normality if you solve, heteroskedasticity problem will go also.

(Refer Slide Time: 25:37)



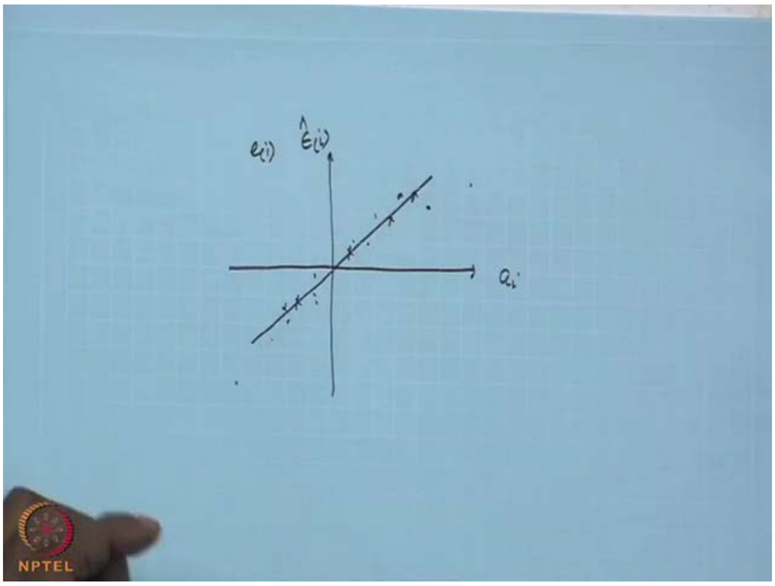
So, then another plot I will show you here that is QQ plot, quantile quantile plot. So, the error is normally distributed. Each of the error term that should be normally distributed with mean 0 and your variance component will be a c square, I think i minus h, we have discussed in last class. So, with this considering that property that normally distributed with mean 0 and variance some variance component, then here you are transforming, you can do these type of things also n minus half by n. Then also you can use this formula, but there is another formulation, which is used i minus 3 by 2, 3 by 8, i minus 3 by 8 divided by n plus 1 by 4.

It is found that this formulation will give you better result instead of i minus half by n, Here, for all i, you are getting these probability values. Then you are calculating the corresponding z values, so z values. Then you are plotting this z value, what we are saying here ai values, these ai values with ei values. So, earlier you have plotted probability versus the residuals.

Student: Sir, here ei, is it ordered ei?

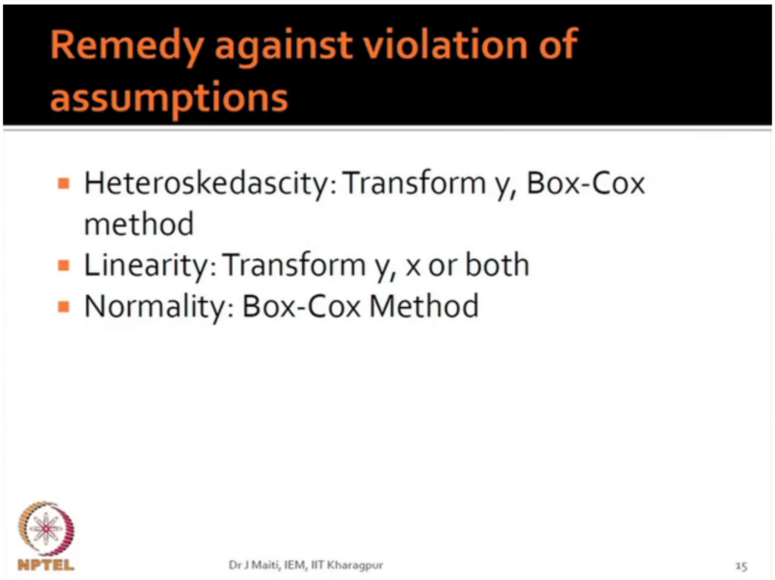
Yes, ordered ei within it is keeping ordered ei. So, then you must get a straight line, which will pass through the origin. Here, things will be like this.

(Refer Slide Time: 27:34)



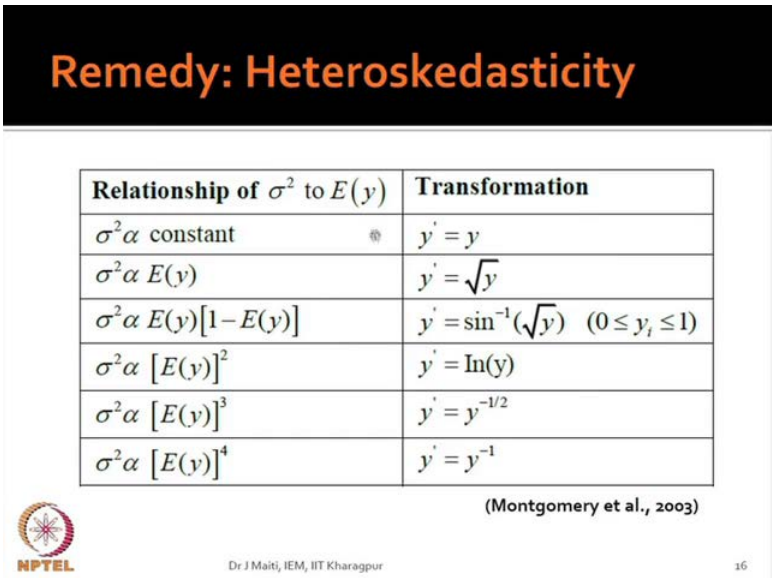
Suppose this side if I write that epsilon i or ei, we have written in this, ei also you can write and this side is the z value, what we are saying that ai here. So, this point should lie on this or go around it. If you get this type of plot, then it is normal. The normality assumption is satisfied, otherwise not. In this particular figure, we found out that really it is not normal. Our y values are non normal values. So, the errors are also showing this non normality. So, if there is violation to any of the assumptions, how do you go about it? What you will do so that this violation can be avoided and your multiple regression equation can be fit or multiple modelling regression modelling can be done.

(Refer Slide Time: 28:51)



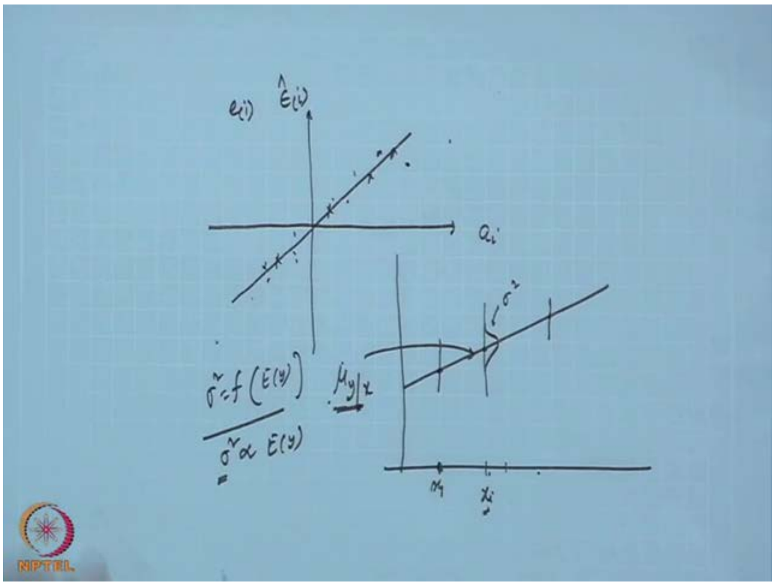
Now, if there is heteroskedasticity, then transform y and Box-Cox method is one method, which is used for transforming y. If there is linearity, you can transform y, you can transform x or you can transform x, y both. So long, if there is a linearity problem, If there is linearity problem, so if there is heteroskedasticity problem, you will be transforming only y. Box-Cox method is also power transformation of y. If there is linearity, you can go for x, y or both. For normality, we will use Box-Cox method.

(Refer Slide Time: 29:39)



Now, what type of transformation is applicable under what condition? Actually, if you see the regression equation, what you are doing?

(Refer Slide Time: 29:53)



In regression equation, this fitted line is nothing but this is the conditional mean values that mu y given x, this line, all values are like this. So, if for every value of x, it is constant, then the first one is coming into consideration. Sigma square is constant, sigma square, this sigma is related to the variability, this variability of y for a particular value of x. I told you in last class suppose this one is xi, they are doing something.

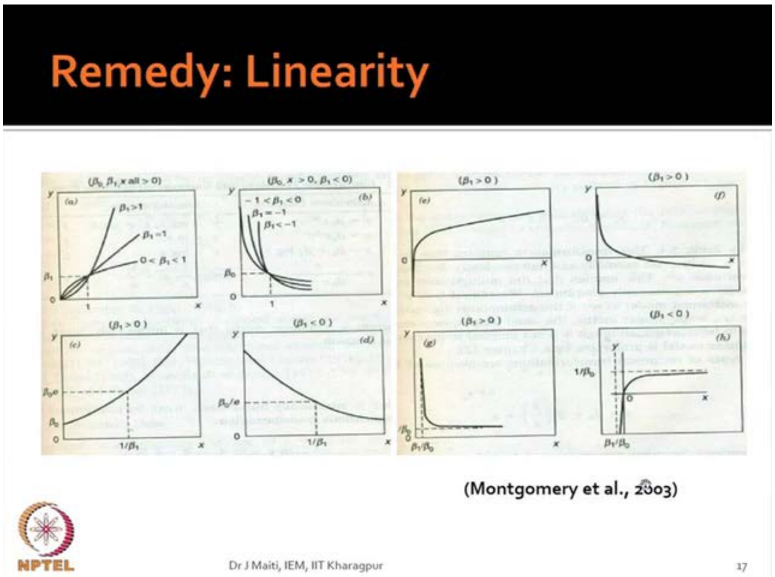
So, what I mean to say here you see that for xi, this is the sigma value, this sigma, sigma square, you will write. Now, if it is constant mean across all values of i, it is constant, then you are going for no transformation. We are writing our f dash equal to y, but it may so happen that suppose the sigma value is a function of your expected y.

That means what we are saying that when expected y means here this value is greater than this value, this, this value is conditional mu y x mean value of y given x equal to x1 or x2 or something; whatever it is, let it be x1. This one is for xi. Now, this is the regression line. So, it is increasing. When x is increasing, y is increasing, so what is happening? This mean value is higher than this. This conditional mean is higher than this like this.

Now, if sigma square is proportional to expected value of y, so then means sigma square is also increasing, sigma square is also increasing. So, under this condition, you will go for square root transformation. Then third one is sigma square is proportional to their mean value two values. It is similar to that when we talk about binomial case p into 1 minus p type of things, so this is similar to this. Suppose the expected value of binomial variable is this is the p this one will be the 1 minus p for the other value. So, under such condition, your transformation will be sine inverse square root of y.

If your this one that variability, y variability is proportional to square of the expected value of y, it is said that log y. If it is in terms of square or cube of expected value of y, then y to the power minus half that mean 1 by root over and if it is 4, then it is 1 by y. So, these are the different types of transformation what is given in Montgomery 2003, Montgomery. Then Peck and Vining, one book is there, introduction to linear regression analysis. It is only for regression analysis. That is a very good book.

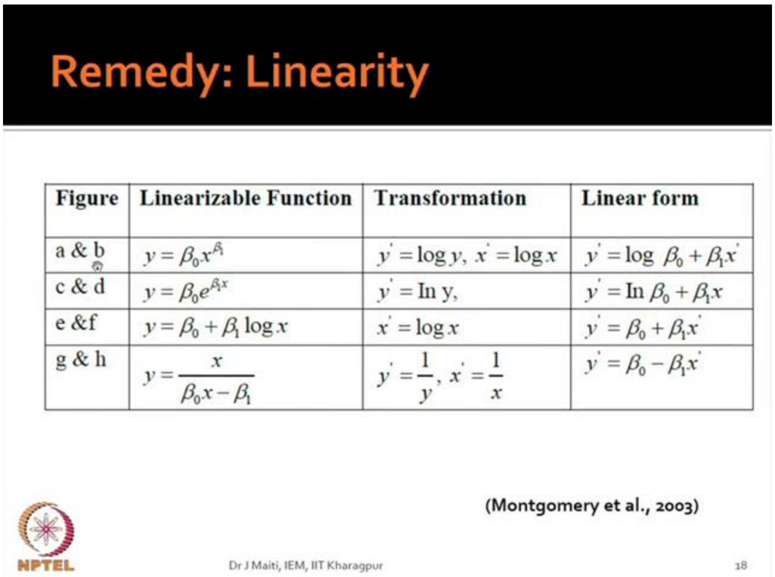
(Refer Slide Time: 34:06)



Now, if there is a linearity problem, see there are several types of linearity problems. For example, y versus x when you are seeing, your things will be like this, like this, like this depending on the beta values, also rehression coefficient values. So, under these different types of non linearity present in y versus x relationship, Montgomery et al in the same book, they have given this transformation for difference.

Suppose this one is figure a, figure b, figure c, figure d, figure e, figure f, figure g, figure h. So, first two cases, because from the bivariate scatter plot, you will be able to find out what type of picture you are getting, the relationship. Then suppose the relationship is like this, the first one, then what you will do?

(Refer Slide Time: 35:04)



Your transformation for a and b, you see this and this, the transformation is y. It is equal to log y x log x. This is very useful because you can develop the scatter plot always. Once you know this, now you see the relationship. Then for figure a and b that linearizable function is y beta 0 x to the power beta 1.

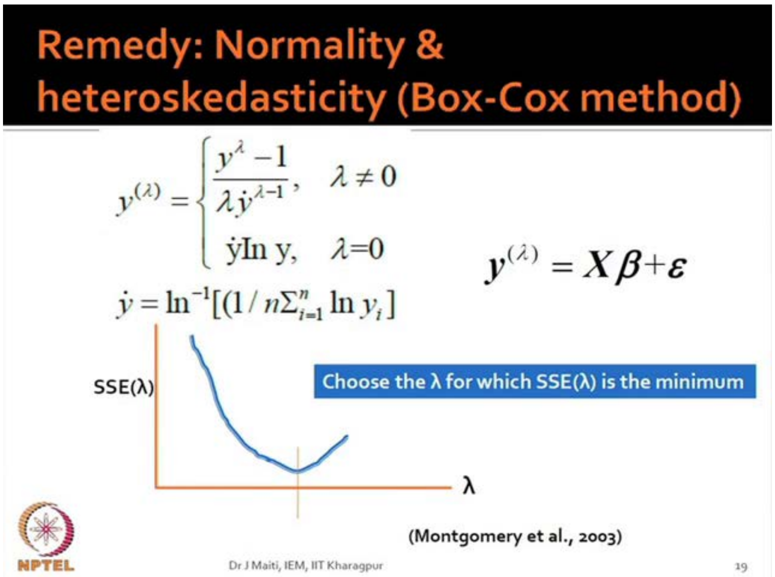
Now, if you take log x, sorry log of this equation, then log y equal to beta 1 log beta 0 beta 1 x that is same. You will be getting log beta 0 plus beta 1 x, you will be getting. If your figure is like this c and d, this as well as this, either it is increasing, linear increasing or decreasing. Under this condition, your linearable function is y beta 0 e to the power beta 1 x and your transformation only y that y dash is log y. Then this linear form will be like this.

If your data is e and f like this type of function, then you go for y equal to beta 0 plus beta 1 log x. So, that mean you are making, transforming x into log x. You will be able to get linear relationship for the last part that means that it is extremely coming, very low, very low value. After this one, after some values of y, y is very low and here also initially low. Then there are again constant values here, here also, some constant values.

Then what is happening here? Your transformation function is y equal to x by beta 0 x minus beta 1 and 1 by y and 1 by x. So, this will be the case. It may so happen that you may go get some other form, then what you will do? But, I think this list is exhaustive, almost the form is given, and almost all the types of relationships are given here.

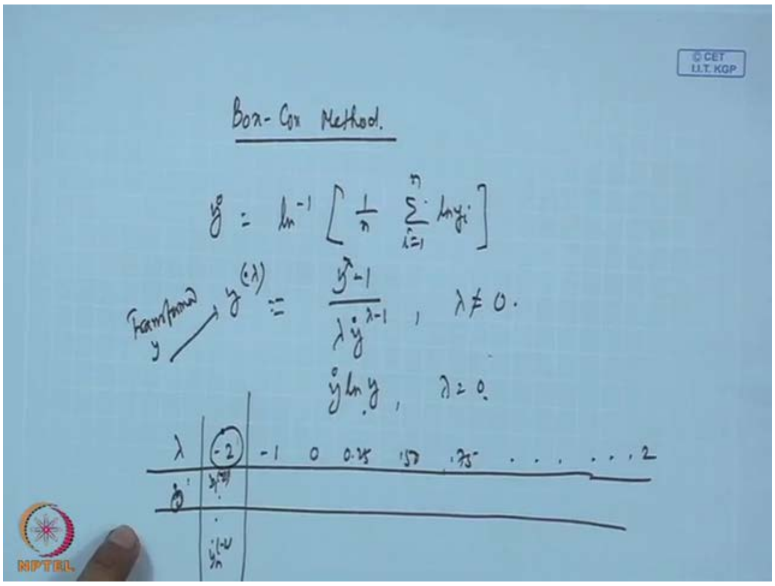
Student: Basically, sir first one or two cases we have to see basically what is the process of finding this. If we know the process of finding this from our scatter plot, we will be able to find the other one. Yes. In fact, fitting if you do, then the non linear part automatically will come. There you see that what type of equation is possible and whether it can be linearized or not.

(Refer Slide Time: 38:19)



Now, I will show you another important consideration, which is known as Box-Cox method.

(Refer Slide Time: 38:29)



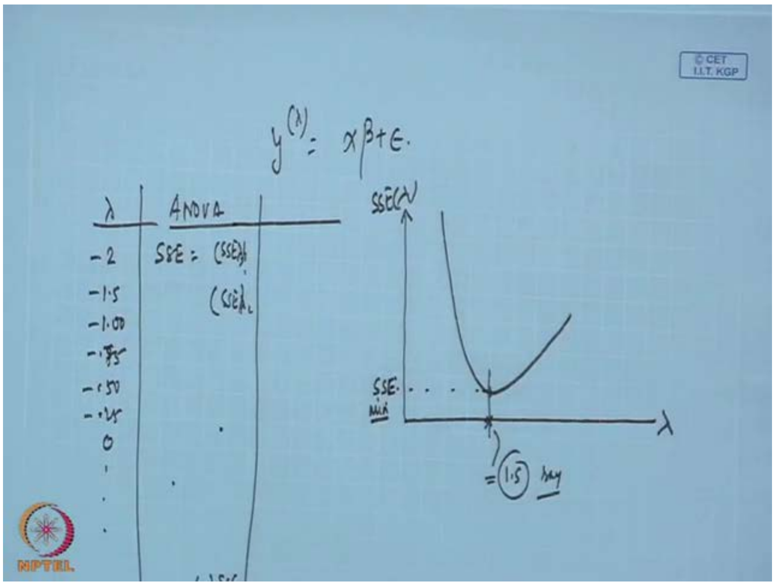
Box-Cox method transformation, the transformation is like this. There are several versions of this power transformation. Finally, Box-Cox given this transformation, you have to choose a lambda value in such a manner that it will give you the minimum sum square error. Then lambda will be put here and your transformation will be like this. Depending on the value of lambda, this one will show like this. So, y to the power lambda is within bracket. This lambda is y to the power lambda minus 1 by lambda, then y does lambda minus 1, when lambda not equal to 0. When lambda equal to 0, this will be y dot log y, where y dot is nothing but it is basically geometric mean.

So, y dot is log inverse 1 by n, this sum of the logs. So, y dot is log inverse 1 by n, i equal to 1 to n log of y i, this one. Then you are doing that you are creating this is the transformed variable, transformed y which will be y to the power lambda minus 1 by lambda y dot lambda minus 1, if your lambda not equal to 0 or y dot log of y if lambda equal to 0. So, what you will do? Then you will choose several values of lambda. For example, let it be minus 2, minus 1, 0, suppose 0.25, 0.50, and 0.75. It can go up to 2. This side also, you can do like this. Then you find out what is this y lambda values. For this, for lambda equal to minus 2, I think you find out or other way, this side you do.

y equal to lambda 2, find out that y 1 minus 2 minus 2. Similarly, y n minus 2, but here to the power, it is not to the power minus 2, it is to the power minus lambda. Then when you are writing minus 2 here, you are writing, you are changing all the values here.

Every value will be changed. Each one will be changed. So, then which value you will choose? Then you regress. Your regression equation, you fit like this.

(Refer Slide Time: 41:52)



Y is equal to lambda equal to x beta plus epsilon. You will be getting ANOVA table for every case. Suppose I will write lambda value like this minus 2, minus 1.5, minus 1, minus say 0.75, minus 0.5, suppose minus 2.5, 0, like positive up to may be plus 1. Then in your ANOVA table, one component is there, SSE value. Find out SSE value here, SSE value like this, all SSE value will be there. I can say it is SSE lambda 1, SSE lambda 2, like this SSE, let lambda k some values are there. Then you are plotting for every lambda, you are plotting SSE lambda, you may get like this. So, this is the minimum SSE. So, our lambda is this value. Let it may be your suppose 1.5 say.

Then, you consider this 1.5 and transform each of the y values with this formulation and go for regression analysis complete.

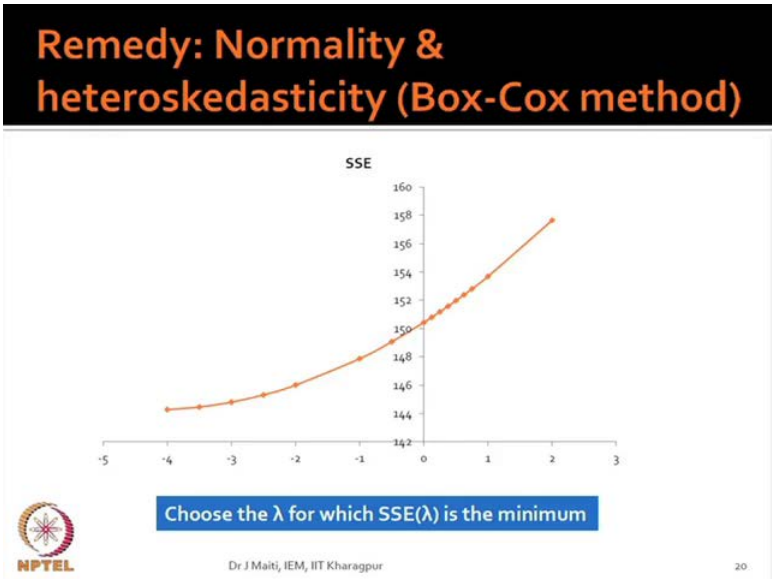
Student: Sir, for us to find out this, first for various lambda value, we have to find out the y.

Yes, yes.

Student: So, in that way means every time y is minimum...

That one that lambda you will consider for that regression equation that is your regression equation. This is Box-Cox method what is given in Montgomery et al 2003. This is the way Box-Cox transformation is done.

(Refer Slide Time: 44:03)



For our case that example data, we have seen that this is non linear, no, not non linear, non normal and heterogeneous that is heteroskedasticity problem was there. Then, we have, I have seen that ultimately what happened? When we have gone for different type of hat, different values of lambda, the power transformation, and then see that it is basically we are going that is at minus 4, this value is coming almost 144 point something, then 146, 148, like this increasing.

But in this zone, there the increase is not much. Then I think it is may be in minus 2 level also. It is 146 point something is there, but we have taken only small amount of data. So, if you take large that may be different, but it is saying that transformation is required. So, for this test of assumptions, I think these are the methods. If you have any question regarding this, please ask me. So, next class, we will discuss model diagnostics. So, after 10 minutes, we will start.