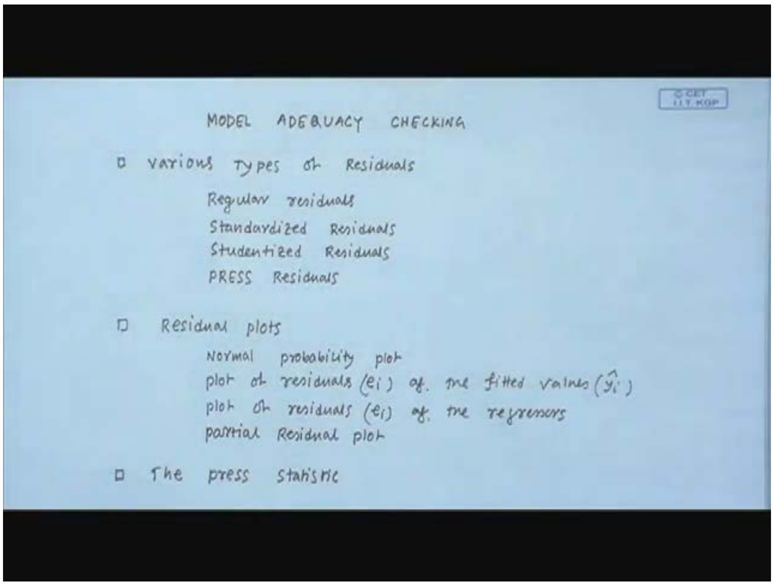
Regression Analysis
Prof. Soumen Maity
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture - 17
Model Adequacy Checking

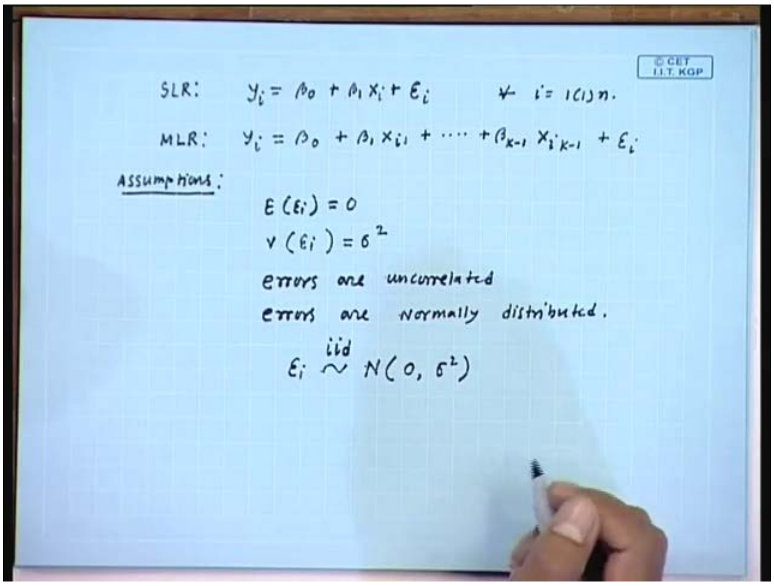
(Refer Slide Time: 00:21)



Hi, this is my first lecture on Model Adequacy Checking. And here is a content of this model, we will be talking about various types of residual. Like regular residuals, standardized residuals and then studentized residuals, press residuals and after that we will be talking about several residual plots, like a normal probability plot. A plot of residuals against the fitted response values y i hat, and plot of residuals against the regressors and partial residual plot. And finally, we will be talking about the press statistic, let me explain the objective of this module, given a set of observations say x I, y i while fitting a simple linear regression model or say you know multiple linear regression model. We make a several assumptions on erratum's like epsilon i is random variable with mean 0 and variance sigma square. You also assume that you know epsilon i is are uncorrelated that is covariance between epsilon i and epsilon j is equal to 0.

And we also assume that epsilon i is normally distributed random variable, with mean 0 and variance sigma square. The objective of this module is to present several methods for diagnosing (Refer Time: 02:38) of the basic regression assumptions, so given a set of observations, the question is you know how do you know the your observations satisfy the basic assumptions. So, that we learn here we will talk about several methods for diagnosing (Refer Time: 03:02) of basic regression assumptions here. What we will do in simple linear regression or in multiple linear regression, we make some basic assumptions on error. For example, you know we assume that the error has a 0 mean, and also we assume that error the term has a constant variance, and the errors are uncorrelated. And also we assume that the errors are normally distributed.

(Refer Slide Time: 03:57)

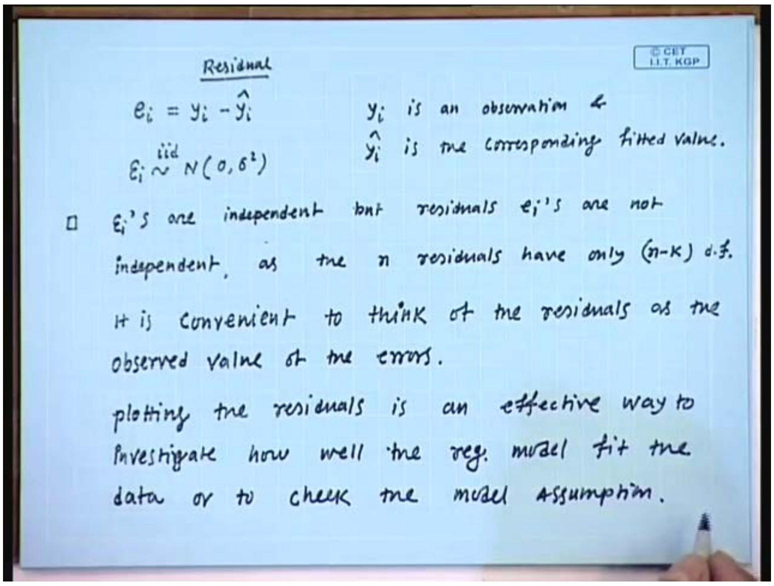


And let me just write the things formally recall the simple linear regression, the model here is y equal to beta naught plus beta 1 a x plus epsilon and for the i'th observation just put i, i, i here for i equal to 1 to n. And for the multiple linear regression y i equal to beta naught plus beta 1 x i 1 plus beta k minus 1 x i k minus 1 plus epsilon i, so this is a multiple linear regression model with k minus 1 regressor variables.

So, the basic assumptions here, the assumptions are expectation of E i is equal to 0, variance of E i that is the erratum's epsilon has constant variance, variance of epsilon i is equal to sigma square, errors are uncorrelated and also we assume that the errors are normally distributed. So, all together I can write that epsilon i follows normal with mean 0 and variance sigma square, and they are independent and identical distributed.

So, today what we will do is that we will I mean in this module basically, what we will do is that we will present a several techniques to check these basic assumptions on error, whether they are correct or not. So, you know gross variations of this assumptions may yield model which is very unstable right, so in this module we will learn how to given a set of data, whether the data set satisfy these basic assumptions or not. So, we will talk about you know several plotting of residuals, so based on that we will check a whether the these assumptions are correct or not.

(Refer Slide Time: 08:13)



The residual first, you know what is the residual in simple linear regression model or multiple linear regression model. So, e i the i'th residual is equal to y i minus y i hat, so y i is the i'th observation, so y i is an observation and y i hat is the corresponding fitted value. So, this is called a you know regular residuals and this is a e i the i'th regular residual, it measures the part of variability in the response variable, which is not explain by the model. Because, e i is the difference between the original data i mean response value y i and the fitted value.

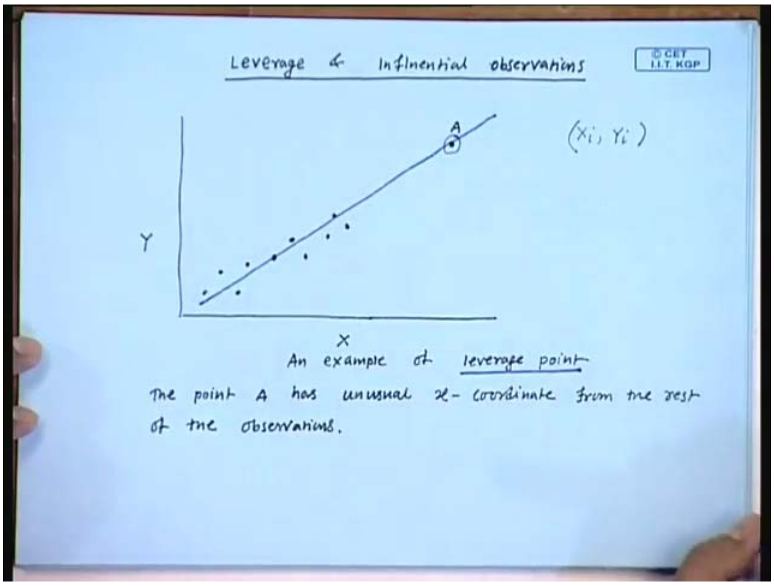
So, the part which has not been explained by the regression model is e I, so it is you know very convenient to treat this e i has the observed value of the epsilon i. Because, we want to test a the assumption on e i sorry assumption on epsilon i that is the error we assume that epsilon i follows normal 0 sigma square, and their independent and identical distributed. So, the observed value I mean this residuals e i's are treated as you know the observed value of the errors epsilon i's.

So, what we know about e i is that first of all epsilon i's are this is some observation you know you know that, epsilon i we assume that epsilon i's are independent, But, the residuals e i's are not independent, as the n residuals have only n minus k degree of freedom. Because, you know about this degree of freedom all this e i's you know the residuals we cannot choose a independently, I am talking about multiple linear regression model with k minus 1 regressors.

So, there are k constraint on involving e I, so we cannot choose all the e i's the residual independently. So, we can choose n minus k of them independently, and the remaining k e i's have to be chosen in such way that this satisfy those k constraint well. So, what I said is that you know it is since we are trying to check whether these assumption E i follows normal 0 sigma square, this i i d whether this is true or not, this is i i d or not it is convenient to think of the residuals as the observed value of the errors.

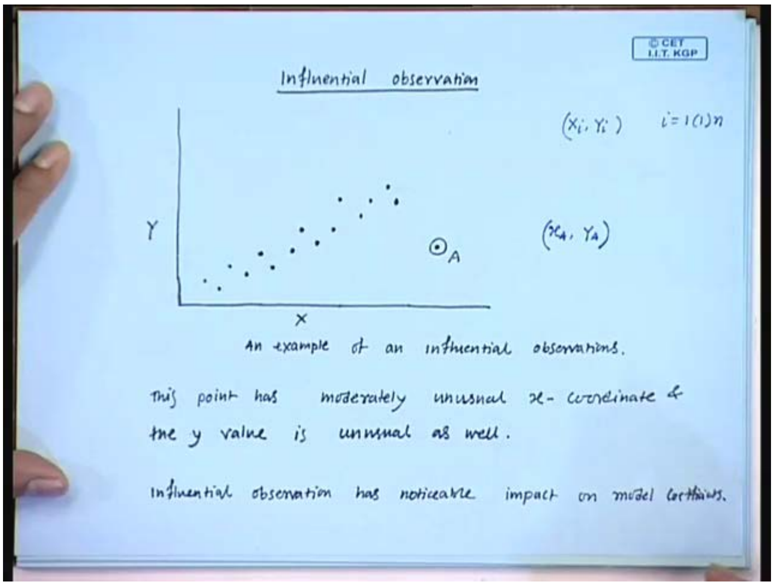
And you know plotting the residuals is an effective way to investigate how well the regression model fit the data or to check the model assumptions. So, we learnt about several residuals plots in this module, and before that you know I just want to introduce two definitions, one is called you know the I mean the leverage and the influential observation. Because, the things are connected let me just know first introduce what is mean by leverage point and influential observation.

(Refer Slide Time: 16:28)



So, see here is a scatter plot of the observations X i Y i, so suppose you know I have some observations X i Y i, and this is the a scatter plot of the given data. Now, you see the point A, this has unusual x coordinate from the rest of the observations. So, the x coordinate for this point is much larger than the x coordinate of the remaining observations. So, this is an example of a leverage point I will give some numerical example for leverage point and the influential observation also. So, if a data point has unusual x coordinate, but here you note that you know the, this point is lying on the general trend of the observations. So, if you fit the given data here the fitted model will be something like this, so this one you know lies on the fitted model, fitted line.

(Refer Slide Time: 18:30)

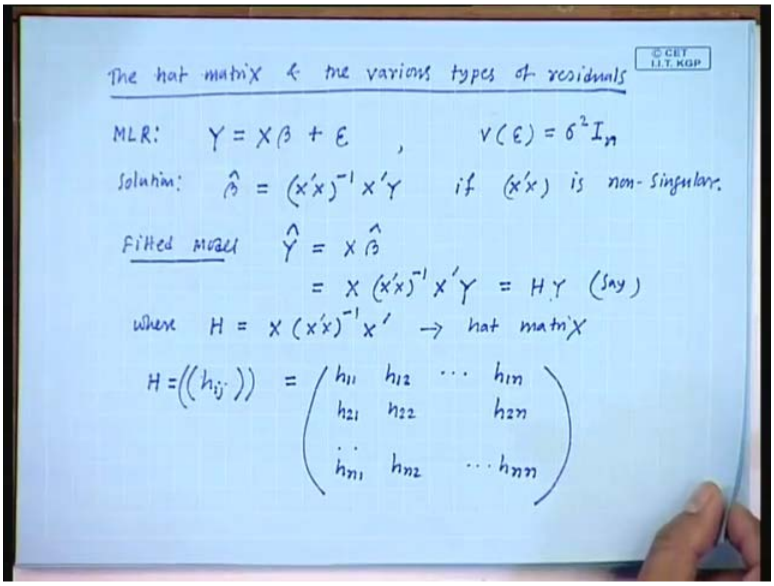


Now, I will talk about a influential observation, so again you know this is a scatter plot for the observation X i Y i i equal to 1 to n, and the point A here is called an influential observation. You check that you know this point has, moderately unusual x coordinate and the y value is also unusual, so a both the x coordinate and y coordinate. So, if I say this point is say X A, Y A then both X A and Y A are larger, you know compare not larger I mean they are for I mean what I want to say is that X A is different from the centre of x coordinate.

And similarly Y A is different from the centre of y coordinate well, so this is you know here A, this point A is not lying on the general trend of the data set. So, this is a leverage point as well is, it is not on the general trend of the data set, so this type of observations is called influential observation. And the influential observation has a noticeable impact on model coefficient, so just let me just you know again, let me tell that what is the you know leverage point and a influential observation.

So, a point set to be leverage point if it has unusual x coordinate, but the point may lie on the general trend of the data, but in case of influential observation point or an observation, you know is set to be influential observation. If it has unusual x coordinate, as well as it has unusual moderately unusual y coordinate, now again you know today what we will do is that we will talk about several scale residuals. So, first I will start with the hat matrix and then I will talk about several scale residuals well.

(Refer Slide Time: 22:20)

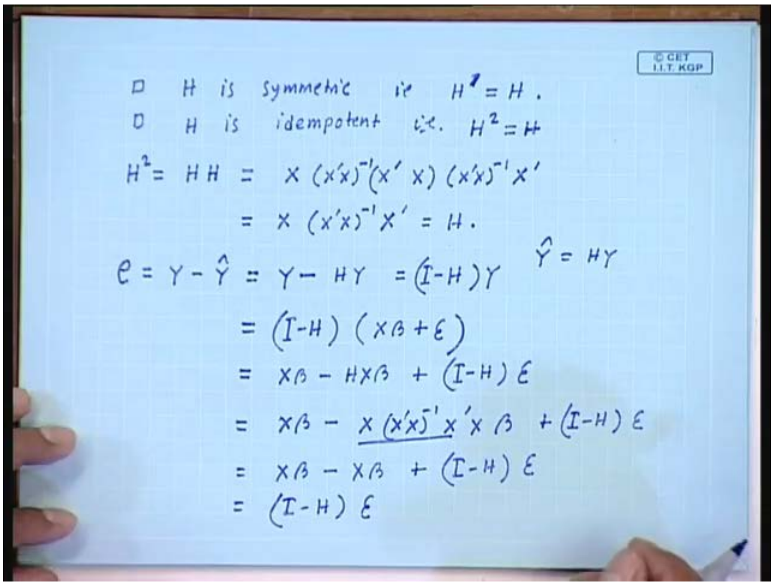


The hat matrix and the various types of residuals let me just to recall you know, the multiple linear regression model, in matrix form we write this as Y equal to X beta plus epsilon. So, Y is a vector of Y 1 I mean n responses, and beta is also vector of you know beta naught, beta 1 up to beta k minus 1 and epsilon is a vector, epsilon 1, epsilon 2 up to epsilon n. And what you assume here is that, the variance of epsilon is equal to sigma square I, this is a best assumption we make this is I n.

Now, solution of this multiple linear regression model is we know that beta hat is equal to X prime X inverse X prime Y, if X prime X is nonsingular. So, we know an how to this is the least covariance estimated of the regression coefficient X prime X inverse X prime Y. So, the fitted model is Y hat is equal to X beta hat, which is equal to X X prime X inverse X prime Y. So, you just know plug beta hat here and this is equal to equal this is equal to H Y, say where of course, H is equal to X X prime X inverse X prime.

So, this matrix is called the hat matrix because you know this is called hat matrix because it maps Y to Y hat that is why it is called a you know hat matrix anyway. So, the elements of H is equal to say h i j, which is equal to h 1 1, h 1 2, h 1 n, h 2 1, h 2 2, h 2 n, h n 1 h n 2, h n n this is you know what the hat matrix, and you know how to calculate the elements of the hat matrix because. X is known then you can compute the elements of the hat matrix H well.

(Refer Slide Time: 27:48)



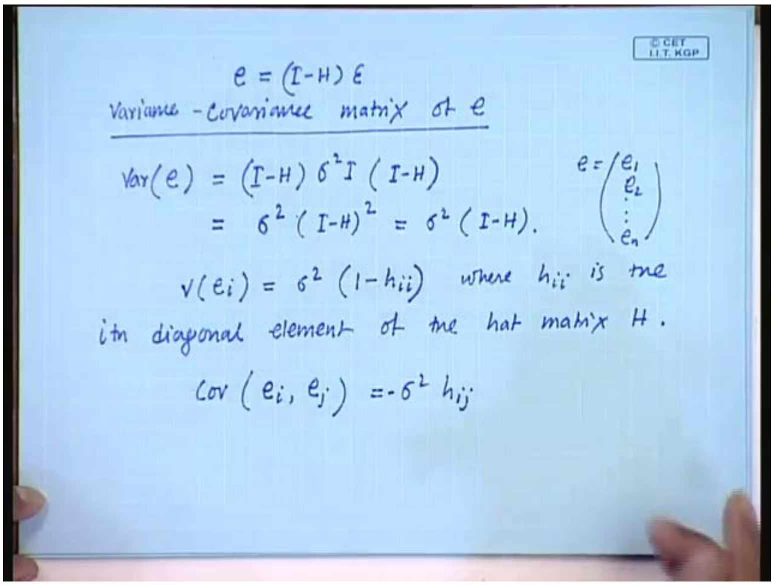
Now, we will talk about you know several properties of hat matrix first of all H is it can be verified that H is symmetric that is H transpose or you know H transpose is equal to H. And the second property is that H the hat matrix H is idempotent that is H square is equal to H well. Let me prove this one, what is H square, H square is equal to H into H my hat matrix H is equal to X X prime X inverse X prime and this is H into H. So, X X prime X inverse X.

So, this is equal to now this will cancel out, so this is X X prime X inverse X, which is equal to H again. So, the two properties of the hat matrix is the hat matrix is symmetric and it is also, and idempotent matrix and now the residual you know in matrix notation, the residual is equal to Y minus Y hat right. So, Y minus Y hat is equal to Y minus, now Y hat you know Y hat is equal to H Y, so H Y which can be written as I minus H into Y.

So, this is equal to I minus H and Y is equal to X beta plus epsilon, this is equal to X beta minus H X beta plus I minus H epsilon. So, this is equal to X beta sorry, so X H is equal to X X prime X inverse X prime X beta, this is H and then X beta plus I minus H epsilon. So, this one is nothing, but so this is X beta minus again X beta plus I minus H epsilon, this is equal to I minus H epsilon.

So, what I am trying do is that you know in this lecture, my aim is to introduce several scalar residuals, what we know till now is that, we know just regular residuals that is e i. Now, will we talk about several scalar residual, which are useful to for some purpose, so we will talk about you know standardize residuals, we will talk about student residual, and also we will talk about the press residual. So, for those purpose you know I need to find the variance covariance matrix of the residual. So, variance of epsilon i is equal to sigma square, but variance of e i with treat as you know, the observe value of the epsilon i, but variance of e i is not sigma square well. So, what I am trying do is that I am trying to find the variance of the i'th residual e i, so what I have at this moment, I have I know that e i is equal to 1 minus H into epsilon i.

(Refer Slide Time: 33:39)

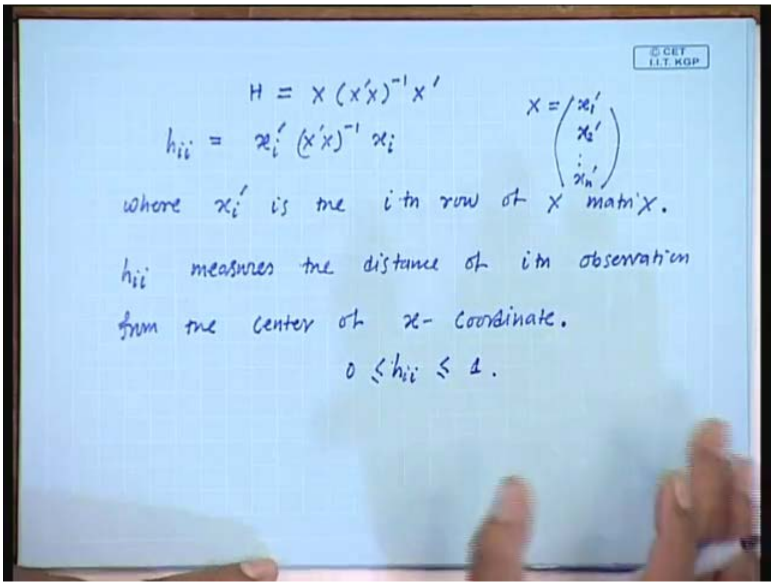


So, what I know is that e i e is equal to I minus H epsilon, now I can find you know the variance covariance matrix of e, so variance covariance matrix of e, e is a vector right. So, variance covariance matrix of e equal to I minus H sigma square I I minus H, it is a very standard one and this is equal to sigma square I minus H square, now as H we know that the hat matrix H is an idempotent matrix. Then you can check that if H is idempotent then I minus H is also idempotent, so I minus H square is equal to I minus H.

So, this is you know we can write as this is equal to sigma square I minus H, so this is variance covariance matrix of e, e is you know it is a vector you know I hope you understand that e is e 1, e 2, e n and then this is variance covariance matrix of e. So, from here I can write you know variance of e i is then equal to sigma square 1 minus the i'th diagonal element this 1 minus I 1 minus h i i. So, h i i is the i'th diagonal element of H, where h i i is the i'th diagonal element of the hat matrix H.

And similarly you know you can find the covariance between not necessary, but you know just covariance of between the i'th residual and j'th residual, see e i, e j is a you can find it from here that is equal to sigma square h i j of course, a minus here. Well I have something to say more about this the i'th diagonal element well.

(Refer Slide Time: 37:45)

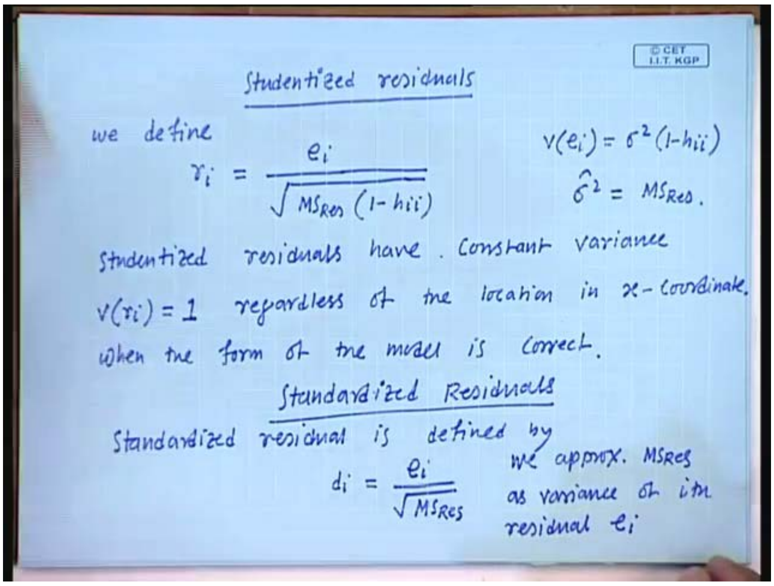


So, what is h i i h i i i is the i'th diagonal element of the hat matrix H and the hat matrix H is equal to X X prime X inverse X prime. So, what is this X, X is you know the coefficient matrix sort of, so X has the rows x 1, x 2 prime x n prime, so x i prime is associated with the i'th observation right. Now, I am interested in the i'th diagonal element of this hat matrix, so I hope you understand that h i i is then just x i prime x prime x inverse x i. So, where x i prime is the i'th row of X matrix.

So, it you know you can check that what h i i does is that, h i i measures the distance of i'th observation from the center of x coordinate, and you have to understand. So, this is enough to explain you know this is what the h i i is that, is this quantity and x i is the x i prime the i'th row of the X matrix. And then h i i measure the distance of i'th observation from the center of x coordinate, and it is not difficult to observe you know realize that h i i they are in between 0 to 1.

So, what massage I want to give from this h i i that you know you must have understood that h i i is the i'th diagonal element of the hat matrix H, and it measures the distance of i'th observation from the centre of x coordinate. And now you recall the definition of a leverage point, a leverage point is a point which has unusual x coordinate. So, it is quite obvious is that you know the h i i is going to be large, if the i‘th observation is a leverage point. So, you know somehow you know from h i i we can get information about the leverage point. So, this is what I wanted to mention here, next we move to you know some there are various type of residuals, till now we know only one residual that is e i that is called regular residual, and I will introduce you know three more scale residuals in this lecture.

(Refer Slide Time: 42:24)



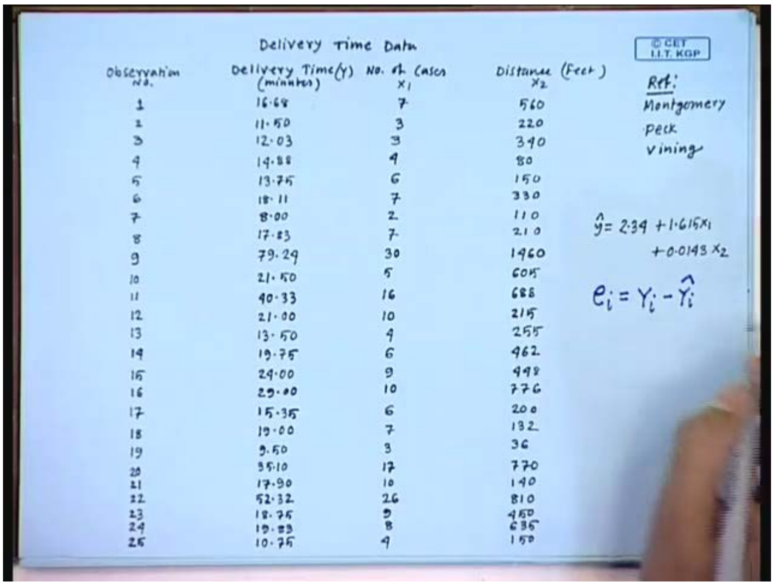
First I will talk about studentized residuals, so what is studentized residual we define r i has e i by standard deviation, what you know what we did just now is that, we have computed the variance of e i, variance of e i is not sigma square, we know that variance of e i is equal to sigma square into 1 minus h i i right. Now, the standard deviation of e i is a just square root of this quantity, and since sigma square is not known we generally know estimate sigma square by MS residual.

So, the studentized residuals r i is nothing but this thing MS residual into 1 minus h i i, so e i by it is the i‘th residual by it is divided by it is standard deviation. So, it is very easy to observe that studentized residuals have constant variance that is variance of r i is always going to be equal to 1, regardless of the location in x coordinate of course, when the form of the model is correct. So, this is a one scale residual, next we will be talking about of a standardized residuals.

So, what is a standardized residuals, standardized residual is defined by d i equal to i‘th residual divided by just MS residual. So, what we did here is that, we have just replaced you know we are just approximating the standard deviation of e i the actual standard deviation is this quantity or actual variance is this quantity. And what we do in the standardized residual is that, we approximate this quantity by sigma square, so here we approximate MS residual as variance of i‘th residual e i.

So, till now we know about a two scale residuals, one is standardized residual and the other one is studentized residual, and both the scale residuals they give almost a similarly information, but in some cases they are different. So, I will just give one example you know this example is from a book by Montgomery, there you know we have the value of studentized residual and standardized residual, we will compare you know we will see when the values both those two scale residuals are almost similar and when they are different.

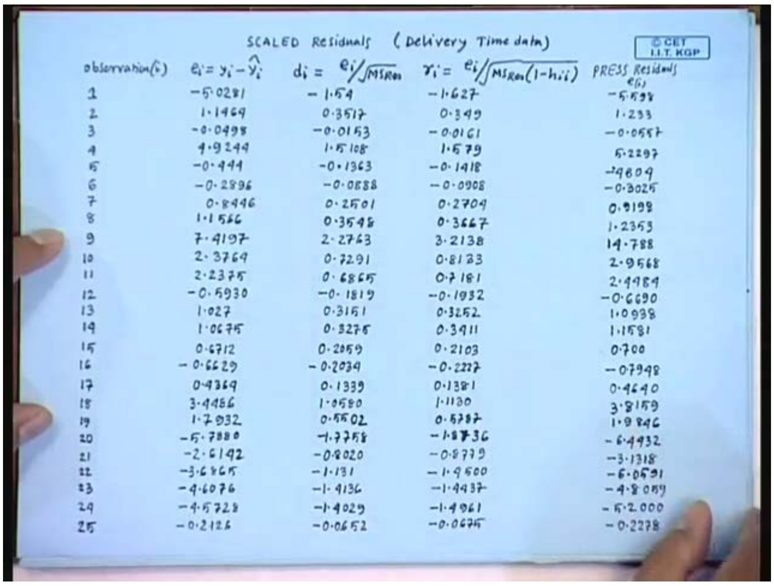
(Refer Slide Time: 48:50)



So, this is an example from a Montgomery book, here the first column is the observation number there are 25 observations, and second column is the delivery time, which is the response variable delivery time y, in minutes. And these are the values, and there are two regressors here one is the number of cases this denoted by X 1, and other one is X 2 which is the distance in fit.

So, this is an example of multiple linear regression model with two regressors, and a one response variable and you know how to fit multiple linear regression model to this data that we have discussed in the previous module. And here is the fitted model, once you have a fitted model, you know the actual value to the response variable, you know the fitted value of the observation of the response variable. Then you can compute you know you can compute e i which is equal to y i minus y i hat, so I have a table for this regular residuals here is a table well.

(Refer Slide Time: 50:54)

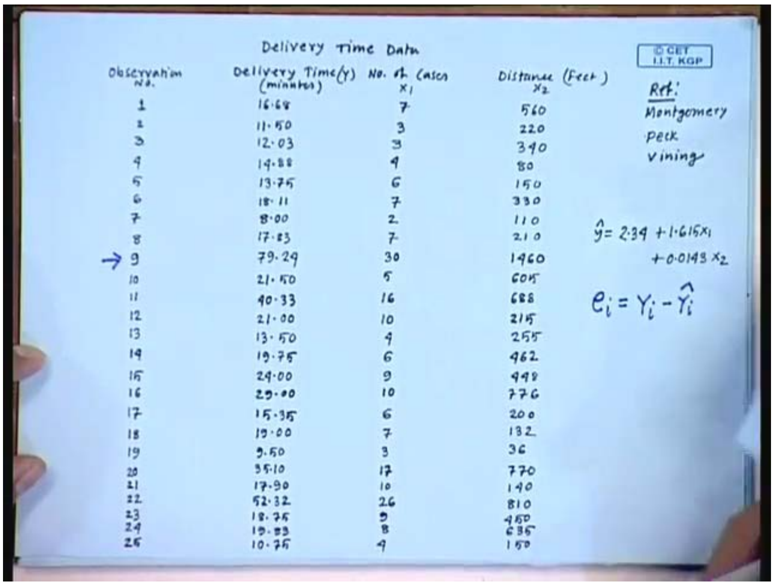


SCALED Residuals (Delivery Time data)

observation(i)	$e_i = y_i - \hat{y}_i$	$d_i = e_i/\sqrt{MS_{Res}}$	$r_i = e_i/\sqrt{MS_{Res}(1-h_{ii})}$	PRESS Residuals $e_{(i)}$
1	-5.0281	-1.54	-1.627	-5.598
2	1.1469	0.3517	0.349	1.233
3	-0.0498	-0.0153	-0.0161	-0.0557
4	4.9244	1.5108	1.579	5.2297
5	-0.444	-0.1363	-0.1418	-.4809
6	-0.2896	-0.0888	-0.0908	-0.3025
7	0.8446	0.2501	0.2704	0.9198
8	1.1566	0.3548	0.3667	1.2353
9	7.4197	2.2763	3.2138	14.788
10	2.3764	0.7291	0.8133	2.9568
11	2.2375	0.6865	0.7181	2.4484
12	-0.5930	-0.1819	-0.1932	-0.6690
13	1.027	0.3151	0.3252	1.0938
14	1.0675	0.3275	0.3411	1.1581
15	0.6712	0.2059	0.2103	0.700
16	-0.6629	-0.2034	-0.2227	-0.7948
17	0.4369	0.1339	0.1381	0.4640
18	3.4486	1.0580	1.1130	3.8159
19	1.7932	0.5502	0.5787	1.9846
20	-5.7880	-1.7758	-1.8736	-6.4932
21	-2.6142	-0.8020	-0.8779	-3.1318
22	-3.6865	-1.131	-1.4500	-6.0591
23	-4.6076	-1.4136	-1.4437	-4.8059
24	-4.5728	-1.4029	-1.4961	-5.2000
25	-0.2126	-0.0652	-0.0675	-0.2278

So, what it says is that you know the first column is the number of observations, the second column is e i minus y i minus y hat. So, the second column gives the residuals I mean regular residual e i, the third column gives the standardized residual e i by a root of MS residual, and the fourth column gives studentized residual.

(Refer Slide Time: 51:54)



Delivery Time Data

Observation No.	Delivery Time (y) (minutes)	No. of Cases x_1	Distance (Feet) x_2
1	16.68	7	560
2	11.50	3	220
3	12.03	3	340
4	14.88	4	80
5	13.75	6	150
6	18.11	7	330
7	8.00	2	110
8	17.83	7	210
→ 9	79.24	30	1460
10	21.50	5	605
11	40.33	16	688
12	21.00	10	215
13	13.50	4	255
14	19.75	6	462
15	24.00	9	448
16	29.00	10	776
17	15.35	6	200
18	19.00	7	132
19	9.50	3	36
20	35.10	17	770
21	17.90	10	140
22	52.32	26	810
23	18.75	9	450
24	19.83	8	635
25	10.75	4	150

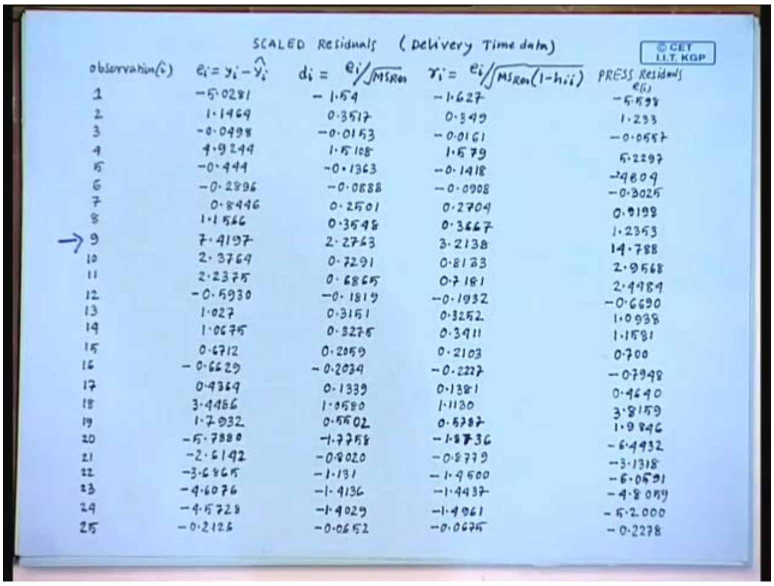
Ref: Montgomery Peck Vining

$\hat{y} = 2.34 + 1.615x_1 + 0.0143 x_2$

$e_i = Y_i - \hat{Y}_i$

Now, if we look at these observations carefully first let me you know refer the previous thing, here you can note that the 9'th observation which has unusual X coordinate. So, here the x 1 value is 30 where is the centre of X 1 is you know quite less compared to 30, and the value of the x regressor X 2 is 1460, which is also quite large compared to the centre of X 2 coordinate. So, it appears that you know this seems to be a 9'th observation seems to be a leverage point or an influential observation.

(Refer Slide Time: 53:14)



SCALED Residuals (Delivery Time data)

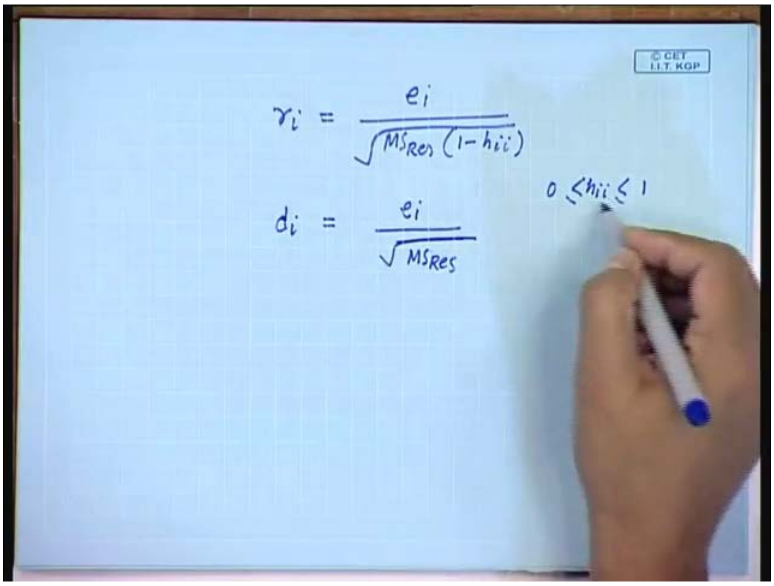
observation (i)	$e_i = y_i - \hat{y}_i$	$d_i = e_i/\sqrt{MS_{Res}}$	$r_i = e_i/\sqrt{MS_{Res}(1-h_{ii})}$	PRESS Residuals $e_{(i)}$
1	−5.0281	−1.54	−1.627	−5.598
2	1.1469	0.3517	0.349	1.233
3	−0.0498	−0.0153	−0.0161	−0.0557
4	4.9244	1.5108	1.579	5.2297
5	−0.444	−0.1363	−0.1418	−.4809
6	−0.2896	−0.0888	−0.0908	−0.3025
7	0.8446	0.2501	0.2704	0.9198
8	1.1566	0.3548	0.3667	1.2353
→ 9	7.4197	2.2763	3.2138	14.788
10	2.3764	0.7291	0.8133	2.9568
11	2.2375	0.6865	0.7181	2.4484
12	−0.5930	−0.1819	−0.1932	−0.6690
13	1.027	0.3151	0.3252	1.0938
14	1.0675	0.3275	0.3411	1.1581
15	0.6712	0.2059	0.2103	0.700
16	−0.6629	−0.2034	−0.2227	−0.7948
17	0.4369	0.1339	0.1381	0.4640
18	3.4486	1.0580	1.1130	3.8159
19	1.7932	0.5502	0.5787	1.9846
20	−5.7880	−1.7758	−1.8736	−6.4932
21	−2.6142	−0.8020	−0.8779	−3.1318
22	−3.6865	−1.131	−1.4500	−6.0591
23	−4.6076	−1.4136	−1.4437	−4.8059
24	−4.5728	−1.4029	−1.4961	−5.2000
25	−0.2126	−0.0652	−0.0675	−0.2278

And let me check the residual for the 9'th observation, note that the residual e 9 that is the residual for the 9'th observation is 7.41. And which is you know this is suspiciously a large, this residual compare to the other residuals and also let me check the value of standardized residual for the 9'th observation. That is 2.27 and the value of the standardized residual for the 9'th observation is 3.21.

Instead of now what I want to I want to make observation here, what I want to comment here you know my statement is that this r 9 is substantially larger than d 9. Whereas, you know if you compare you know r 8 and d 8 there is no much difference, similarly say r 7 and d 7 there is not much difference between the standardized residual and the studentized residual.

So, my final conclusion here you know what I have observed or a what you need to know is that the standardized residual and the a studentized residual, they give almost the same information similar information. But, there will be a substantial difference between the standardized residual and the studentized residual, if the associated observation is and influential observation or leverage point. So, if the given point is you know the given observation is leverage point or influential point, then there will there will be a substantial difference between the studentized residual and the standardized residual. Otherwise, they are almost a similar and you know you have to understand why it is so just give outline of the of this fact you know.

(Refer Slide Time: 56:26)



Let me just recall what is a what is studentized residual, studentized residual is r i which is equal to e i by MS residual into 1 minus h i i. This is the studentized residual, and standardized residual I think it is d i, d i is equal to e i by root over of MS residual, now if the i‘th if the observation is an influential observation, then h i i is going to be large. So, h i the limit for h i i is it is between 0 to 1, if this is large; that means, this is close to 1 and then this is small, the denominator is small means the whole thing is large.

So, that is why and in case of standardized residual h i i is stated as 0 always, so if the i‘th observation is an influential observation. Then h i i is going to be large, h i i is large means 1 minus h i i is small; that means, the denominator is small and then the whole thing is going to be large. So, that is why you know in case you know you can this is one way, you know the to identify whether the observation influential or not looking at the difference between the studentized residual and a standardized residual. I have to stop now.

Thank you.