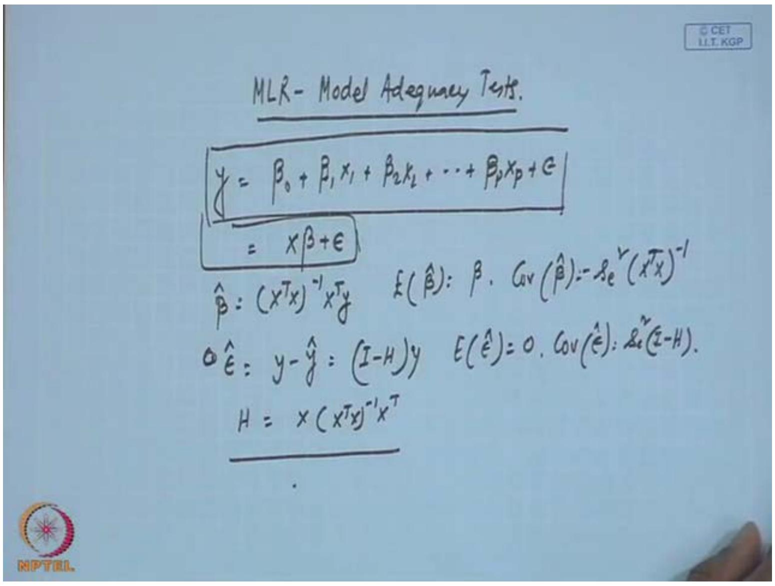
Applied Multivariate Statistical Modeling
Prof. J. Maiti
Department of Industrial Engineering and Management
Indian Institute of Technology, Kharagpur

Lecture - 23
MLR- Model Adequacy Tests

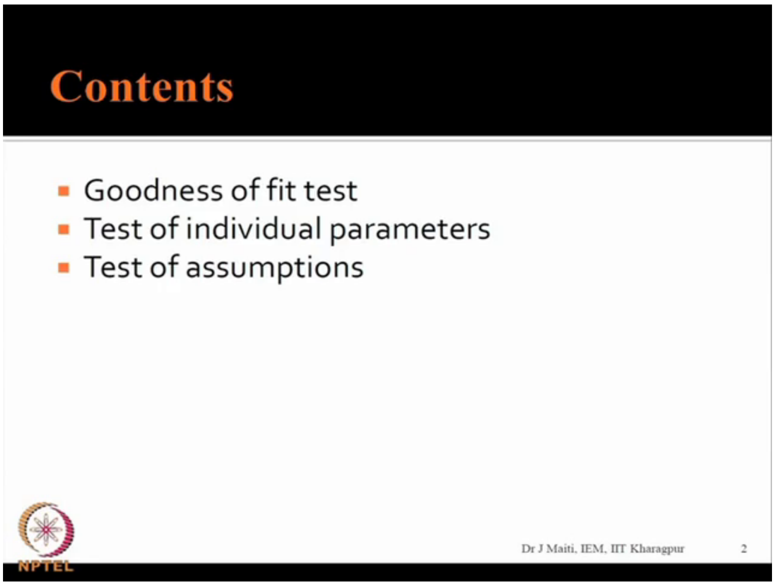
(Refer Slide Time: 00:22)



Good morning. Today we will discuss multiple linear regression Model Adequacy Tests. So, what we have learned so far about multiple linear regression, we have seen that the general equation is like this beta 0, beta 1 X 1, beta 2 X 2, beta p X p plus epsilon, and which we can write in terms of X beta plus epsilon. We have already estimated the value of beta, which is X transpose X inverse X transpose y. We also estimated the expected value of beta cap which is beta and covariance of beta cap which is actually S e square X transpose X inverse.

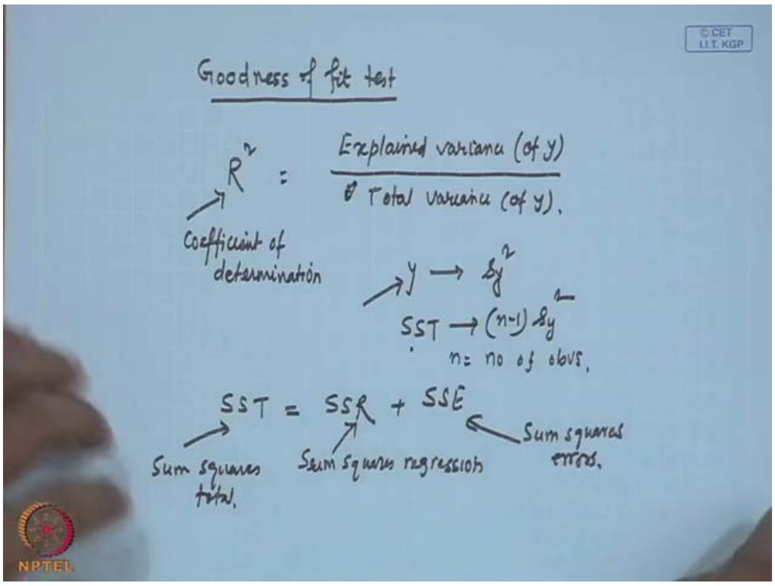
We also have seen epsilon cap, which is basically y minus y predicted and this one we have seen in terms of i minus H into y .We also have seen expected value of epsilon cap is zero and covariance of epsilon cap that is S E square into i minus h, where H is the hat matrix, which is X into X transpose X inverse X transpose up to this much we have covered. We want to test today that the model whatever we have we will develop under this equation or this equation. So, whether the model is adequate enough to explain the data given.

(Refer Slide Time: 02:59)



So, in this slide today's contents are goodness of fit test, test of individual regression parameters and test of assumptions, these three things we will consider.

(Refer Slide Time: 03:17)

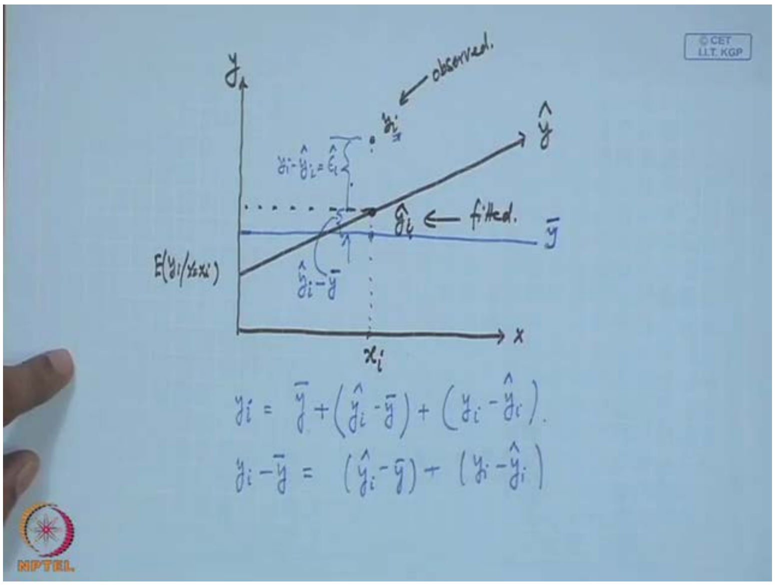


Let us see the first the goodness of fit test in goodness of fit test we will use one statistics called r square, which is known as coefficient of determination. This will give us a ratio of explained variance definitely variance of y by unexplained by total variance unexplained plus explained by total variance of y. So, that mean what you mean to say that y is having the y values they vary from one to another and there is variance of y, which is s y square s stands for the variance and standard deviation and related to y to the square that is the variance.

Then the total variability in y if we define like this S S T sum square total that will be your n minus 1 s y square, where n is number of observations. So, this total variability will be decomposed into two parts model explained variability plus unexplained variability. So, what is your will be doing now? That S S T can be decomposed into S S R that R stands for regression. That is sum square regression plus, what is unexplained that is sum square, sum squares error, this one is the sum square, total sum squares total.

So, all related to the dependant variable we are talking about we are talking about S S T sum square total for y. Then this variability is explained by regression equation the explained portion is coming under S S R and unexplained portion is coming under S S E. It is similar to in ANOVA, what we have seen the same thing.

(Refer Slide Time: 06:21)



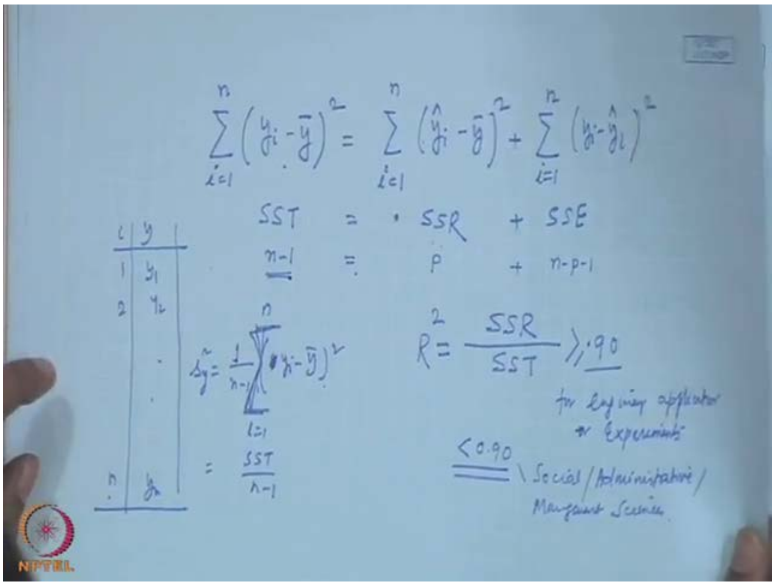
So, in order to establish this relationship let us explain in with one variable, let it be that y versus X regression line let it be like this. So, this one is your y cap and let there is one observation of X i that is ith observation on x. Then if you want to predict what will be the expected value of y i expected value of y i given X equal to X i. This is nothing but this value, so this one is the predicted value of y. Now, this is y i predicted let the original value of that y i is somewhere here this is the observed value. So, observed this is fitted. Let us also define the average, let the average of y is here actually it can be little up.

Anyhow, we will just assume that it is here then this value is y bat, this value is y i cap this value is y i. So, what is this value then? y i minus y i cap this value is nothing but error value correct. Then what is this value this one, so if you have no information about X that mean you have only information about y, then the best fit value is y bar. Suppose, you want to say what is the next value you will consider y bar as X is available as a result what is happening you are able to find out some fitted or predicted value y i bar i cap. Definitely, it is what can i say that closer to y i, which is the observed one.

So, that means this is the amount for y i, which is explained by the regression line if there is no regression line that mean no X information is with you. You will go by y bar whether this is the if you have only the y information this is the best fit. So, then this amount what is this y i? This y i cap minus y i bar? This is what is explained by the regression line then y i minus y i cap, which is error cannot be explained by the regression line. When you are talking about one variable case X equal to one y equal to is one always one in multiple regression.

So, then if I want to write this, why I cannot write like this y bar plus? When this one plus this one plus this one. So, y i, y i cap minus plus y i minus y i cap you can write like this y bar y bar cancelled out y i cap and y i cap will be cancelled out y i will be there. So, if you recall the ANOVA part, then you can very well understand that what we are going to do y i cap minus y bar plus y i minus y i cap you require to square the terms first. Then take summation over all observations you will be getting this square plus this square definitely sum of this square plus sum of this square plus some c o. That c o cross product that portion will become 0. So, ultimately it is similar to your this one, that ANOVA part.

(Refer Slide Time: 11:20)



So, what will happen? Ultimately, you will get something like this. This i equal to one to n equal to i equal to one to n y i cap minus, this square plus i equal to one to n y i minus y i cap whole square. So, what is this one? This is our S S T, because if i 1, I have y with y one to like y n i equal to 1 to n. I am asking you, what will be the variance component? here you will say that 1 by n minus 1 into y i minus y bar square. That is sum i equal to 1 to n this one is nothing but this quantity y i minus y bar square .This is that is S S T by n minus 1, so this is S S T.

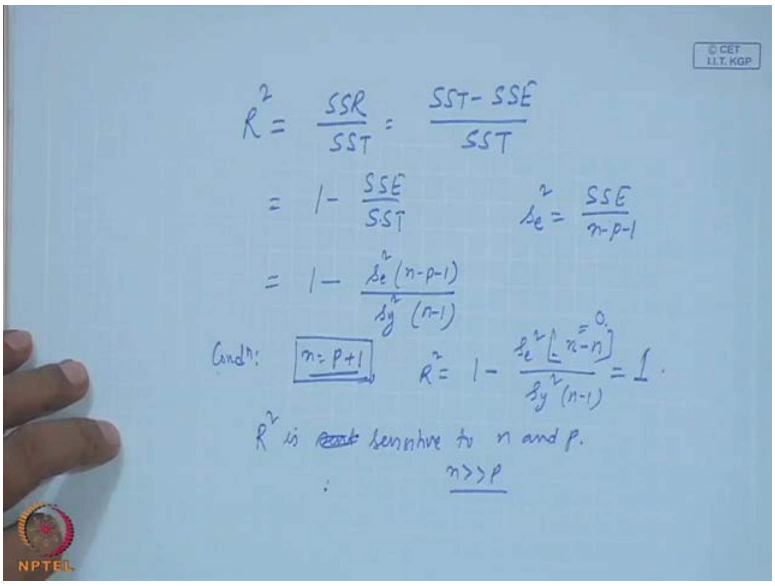
So, similarly this one will be S S R, because this is the amount, which is more, which is basically explained by the regression. Then this is another amount, which is quantitative, which is not explained by the regression. So, similarly, what will be your degrees of freedom here n minus 1? What will be the degrees of freedom here? How many parameters you are estimating p plus 1?

Student: p plus 1.

So, p plus 1 minus 1 that will be p, then what will be the degrees of freedom here n minus p minus one getting? So, for total y the S S T part sum square total as you have n observations n minus one is the degrees of freedom, you are estimating p plus one parameter including inter shift. So, that is why p plus one minus one this is the degrees of freedom and S S E this n minus p minus one this is the degrees of freedom. So then what is r square r square? Is explained variance what is explained variance S S R divided by unexplained total variance, that is S S T.

So, this r square value should be greater than equal to 0.90 for engineering application or if you do experiments laboratory experiments, but if you collect data from field maybe less than 0.90 is also acceptable. Because, you are collecting data from field, which is more what I can more variable in nature more dynamic more volatile in nature. So, there what will happen ultimately it will be very difficult to get r square equal to 0.9. You may get less, but you can then you can go for suppose for social sciences, social administrative management sciences. It can be less than 0.90, but it all depends on the accuracy required.

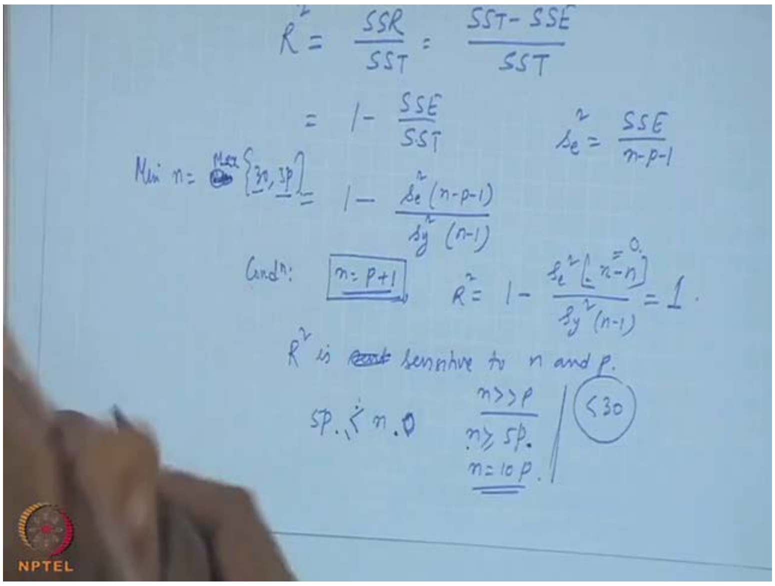
(Refer Slide Time: 15:40)



Now, in r square there is one problem the problem is if I write down R square little bit different way. That this is R by S S T, which is S S T minus S S E by S S T, because this is quite obvious. So, one minus S S E by S S T is correct. What is your S S E? In last class I think, yes last class only I told you that S E square equal to S S E by n minus p minus one. We have shown this one, so instead of S S E i will write S E square n minus p minus one divided by S S T. I can write s y square into n minus one you see s y square equal to S S T by n minus one. So, S S T equal to s s y square into n minus one, so you can write in this manner.

Now, condition is like this that n equal to plus one what will be the value of r square one minus S E square into n minus n. Because, n equal to p plus one p plus one divided by s y square n minus one this is zero. So, it is one this is the limiting condition what will happen if you number of observation is equal to number of parameter estimated your r square will become one saturated. So, that means r square is sensitive to n and p and you require n much much greater than p to rely on our square. So, in multiple regression what it is said, that it is expected that.

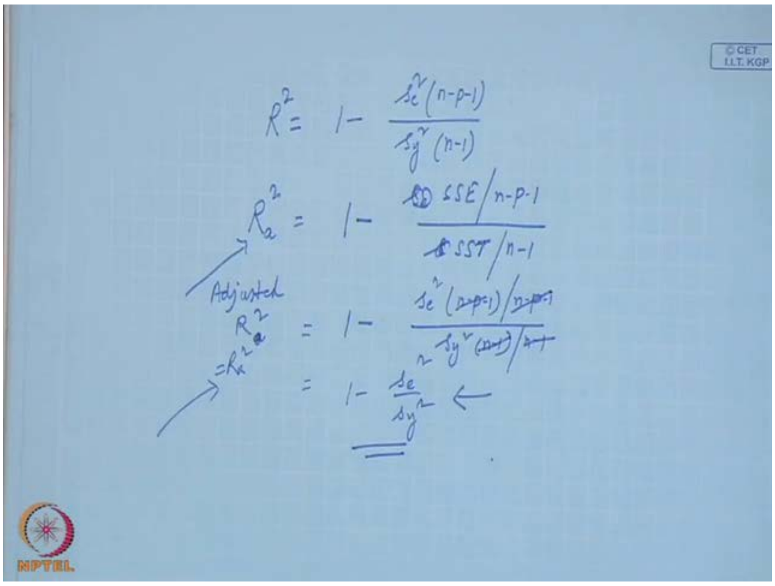
(Refer Slide Time: 18:23)



That n is will be like this less than equal to 5 p not less than it should be what mean to say it should atleast be greater than equal to 5 p desirable is 10 p. If it is more than 10 p no problem, but again very large n also not desirable in many cases suppose if you go for ki square test sometimes in later part of the lectures, we will see that ki square will be using many some places. Then there if ki square is also sensitive to sample size more sample size also not good.

So, for regression purpose what I mean to say here that if your sample size is atleast five times of p that is the starting point you have to have this amount of data and 10 p is the desirable one and in no case. Ultimately, it should be less than 30, suppose there is only one p, so that may you may think that it will be 5 into one 5, that is not the case. So, I can say that the minimum value is 30 minimum n equal to minimum of 30 and 5 p whichever is minimum maximum of this you write should not be minimum maximum of this 30 and 5 p.

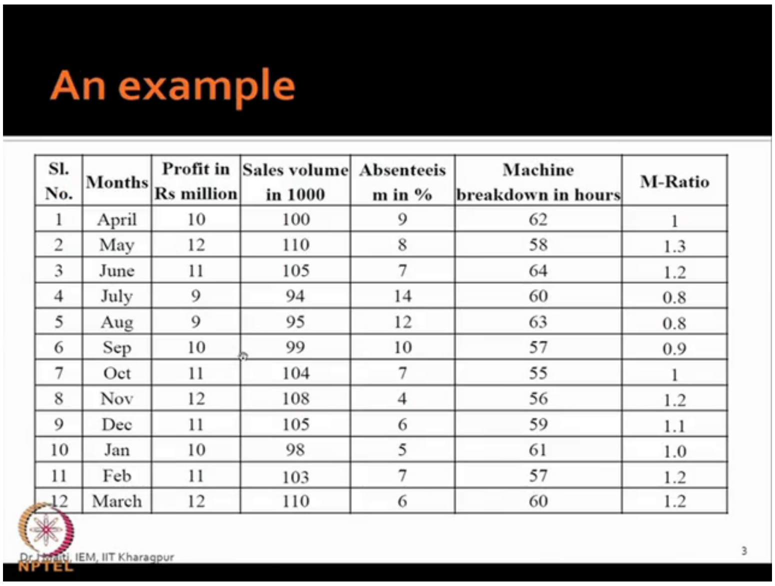
(Refer Slide Time: 20:35)



So, in order to avoid this that dependency on n and p, if we modify r square in this manner, what will happen? We have written this r square equal to one minus S E square into n minus p minus 1 by s y square n minus one. Let us create one more statistics R a square, which is known as adjusted R square in this manner this divided by S S E by n minus p minus one and S S T by n minus 1. That means the sum square error divided by its degrees of freedom sum square total divided by its degrees of freedom.

Then what will happen? You can write S E square n minus p minus 1 divided by n minus p minus 1 by s y square n minus 1 divided by n minus 1. So, n minus p minus 1 n minus p minus 1 will cancel this also will be cancelled. So, it will be one minus S E square by s y square it is irrespective of sample size it will give you a value and this is R a square. So, adjusted r square, which is known as R a square, this is parsimonious. This is known as parsimonious fit measure in case of multiple regression.

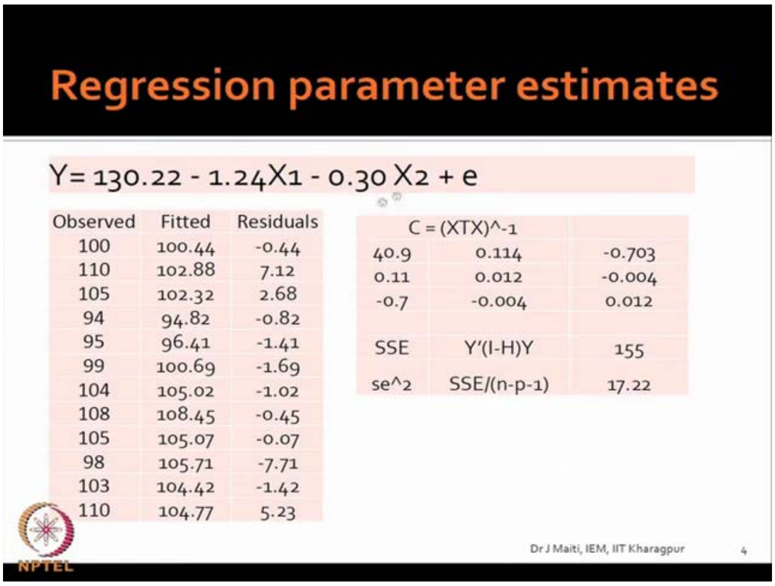
(Refer Slide Time: 22:31)



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

So, with this we want to see that the example we are discussing the same example and what are the values? This is the data, we will be using and here sales volume is our y absenteeism and machine breakdown. These two is our independent variables for explanation of multiple regression.

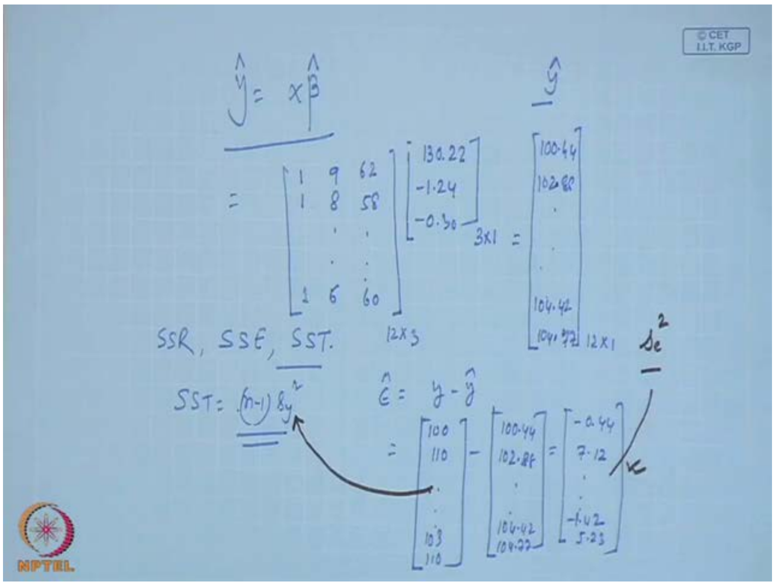
(Refer Slide Time: 22:50)



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

Then last class we have fitted this one, we found that y equal to 130.22 minus 1.24 X 1 minus 0.30 X 2 plus error. This is my regression line using this regression line you have found out the fitted values getting. Basically, how we got this fitted values? We are saying y equal to X beta cap, what are your X values X values one, like 12 values then X 1 values are there X 1 values are X 1 values are absenteeism and machine breakdown show absenteeism like 9, 8 like this up to 6 and breakdown hours 62, 58 like this 60.

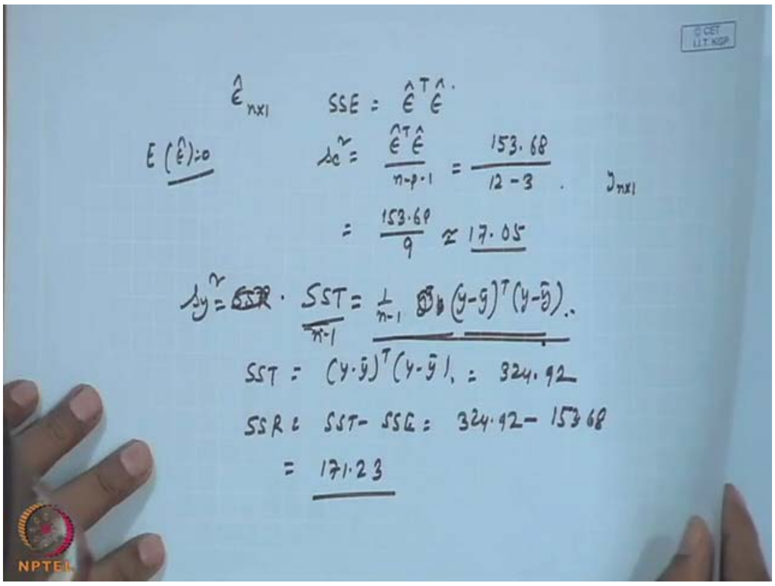
(Refer Slide Time: 23:16)



This is your X you have computed beta what is the beta cap values one 30.22 then minus 1.24 minus 0.30. So, this one is 12 cross 1, 2, 3 then this is three cross one. So, what you will get you will get a 12 cross one vector of y cap. This is what is fitted values you say in the fitted values are 100.44, 102.88. So, like this you will be getting upto at the end 104.42, 107.77.

So, what we require now? That fitted values are there y is also available you require to calculate S S R, S S E, S S T that is our work. Now, S S T you will be you will be calculating very well like this using this n minus one into s y square that is the formula. Now, if I want to calculate S S E you are required to find out what is the residuals. This residuals will be this is y minus y cap so if you you will be getting already y values are there y values are 100, 110. So, in the same manner 12 values 103 and 110, then fitted values you got 100.44, 102.88 in the same manner you will be getting here 100.4 and 204.77. That means finally, the error values will be minus 0.447 point one two same manner you come minus1.42, 5.23.

(Refer Slide Time: 27:09)

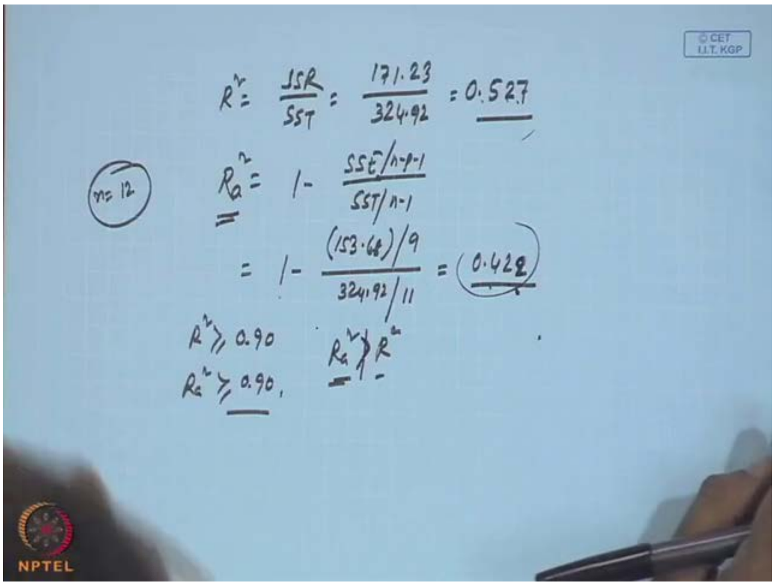


So, you have your error part also from here y value you calculated this one from this you can calculate S E square getting me, how to calculate? So, our epsilon cap is n cross one if I make a one to calculate S S E, which is epsilon cap transpose epsilon cap S E square is epsilon cap transpose epsilon cap by degree of freedom n minus p minus 1. So, this value this value will be epsilon t this value will be around 155 I think 153.68 by n is 12 p minus 1 is 3. So, 153.68 by 9, so it will be 17.05, let it be similar this value will be there, so then S S R [FL].

Now, we have to find out S S T, S S T value we have to check S S E I calculated S S T value will be 324.92, how do I calculate same way your y is y is n cross 1. So, 1 by n minus 1 if you do y transpose y you will be getting it I think y y minus y bar transpose y minus y bar this one you have to do, but here we have not done this manner, but we assume that this will be 0, but actually when you calculate that expected value of epsilon that may not be 0. Because, your data is this one data is this, so what will be the expected value that is mean value that sum of this epsilon i cap by n it is theoretically it should be 0, but it may not be 0. If you calculate like this I think this value will be 155.

So, 0 that mean value is 0 then this will become 155 by calculated, but there is definitely that slight difference will be there. Because, of small data set other way it should follow even though that expected value of epsilon cap is not 0, but it should not be high value. So, using this S S T you will be you are able to calculate S S T. S S T is this one I was talking about s y square equal to this by n minus 1 equal to this correct so then S S T is known. So, S S T is y minus y bar transpose y minus y bar this value we have found out that 324.92. So, my S S R is S S T minus S S E. Now, so 324.92 minus 153.68, which is 171.23, so S S R is known, S S T is known.

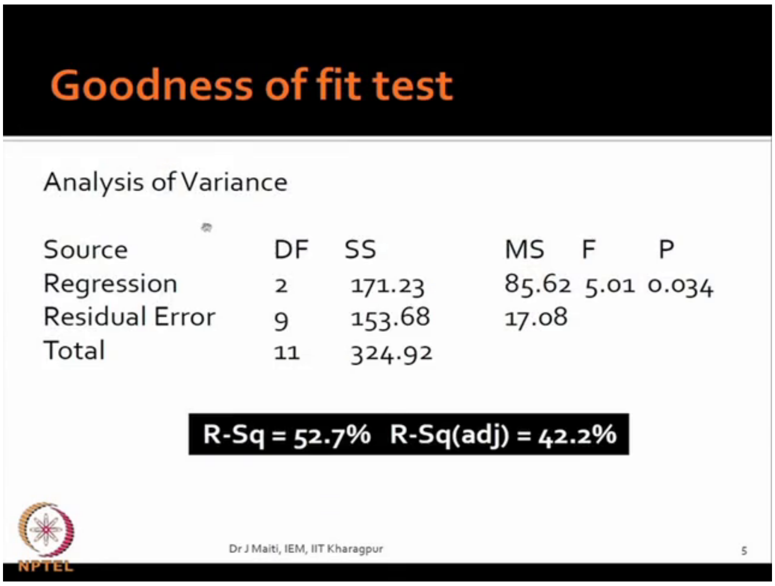
(Refer Slide Time: 31:28)



So, R square value will be S S R by S S T, which is 171.23 by 324.92. It is 52... 0.527 less very less. Then what is your R a square value? This is 1 minus S S E by n minus p minus 1 by S S T by n minus 1. So, that is 1 minus S S E is how much 153.68 divided by 9 that is the degree of freedom by S S T is 324.92 divided by 11 n minus 1 and this quantity is 0.422. So, from 0.52 from 0.527 to 0.422, it is a I think reasonable change this is, because your n is only 12. This is because of this value change is there see R a square drastically change to 0.42 from 0.527.

Now, what can you conclude form this straightaway? You can say that as R square is 0.527. This model is not fit what I said R square should be greater than equal to 90. Similarly, R a square should be greater than equal to 0.90. So, both the things definitely whenever you will get R square greater than R a square greater than R square. That should not that you get, or not you will not get it may be r square equal to less than equal to R a square less than equal to R square and that is what is happening here.

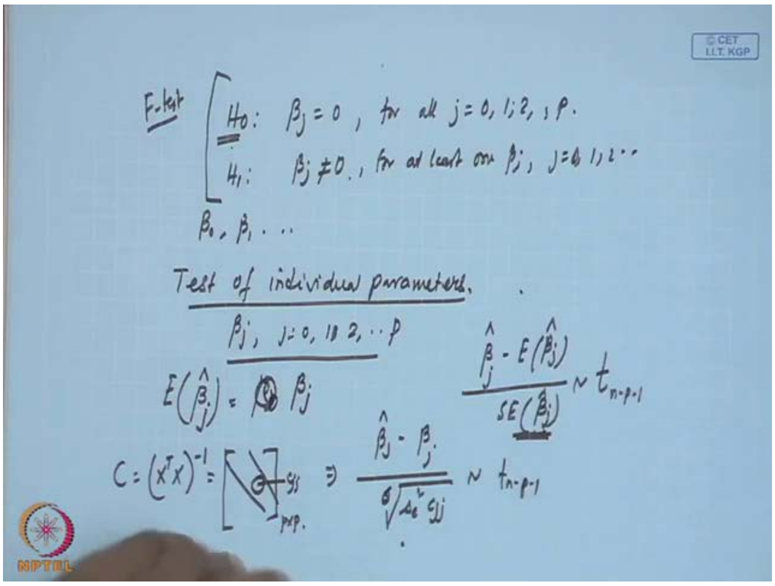
(Refer Slide Time: 34:29)



However, in this particular example it is not fit, but some example it will you will fit that you get a good fit. So, here using R square you will basically or this S S T, S S R you are what you are creating? Here you are creating a ANOVA table. You see the ANOVA table analysis of variance source regression residual oblique error total degrees of freedom. There are three parameters to be estimated, so 3 minus 12 residual errors is 9, because total degrees of freedom is 11, 11 minus 2 that will be 9 and the sum squares 71, 150.

So, we can create M S, what is M S? S S R by degree of freedom and your M S E will be S S E by degrees of freedom. The quantity M S R by M S E will follow f distribution with the respective numerator and denominator degree of freedom it is similar to ANOVA case. Here what you are finding out your R square is saying that your model is not good, but your f statistics is saying that the p value is how much here. 0.034. It is it is less than 0.5, Then what is the difference between this vis, a vis this what we are testing here. I think you can understand this thing also in reference to ANOVA.

(Refer Slide Time: 36:11)



So, when you make this type of ANOVA table, our null hypothesis is that beta j equal to 0 for all j. That mean there is no regression coefficient, which is significant and our alternative hypothesis is that there is at least 1 beta j for at least 1 beta j 0. Like this that mean if beta 0 beta 1 or anyone of the beta parameter are significantly contributing in explaining the variability of y.

Then this H 0 will be rejected R square is a consolidated measure that how much variability of y you are able to explain. If it is 52 percent, you are not happy you want 90 percent or more, but 52 even if 50 percent is explained, but that means some parameters some of the independent variables, they have importance. They are explaining that 52 percent, it is not 0 percent.

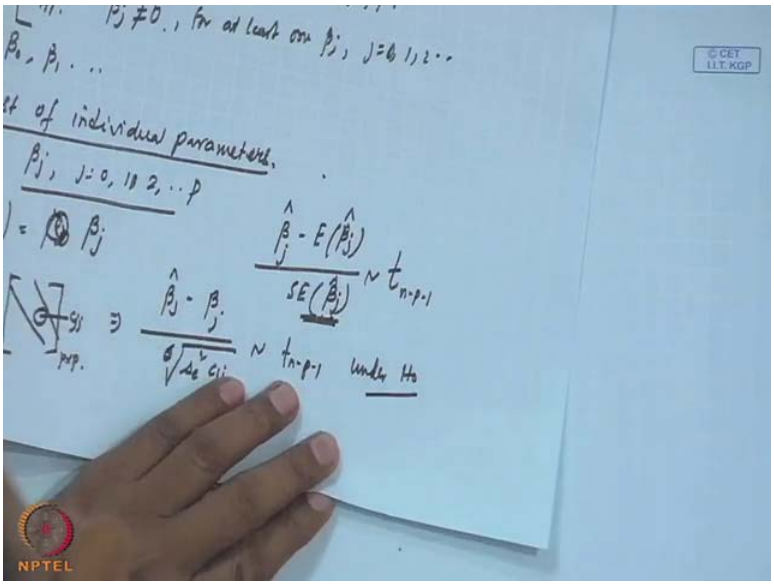
So, as a result you are doing another test collectively you are basically checking that there is that influence of all the regression parameters are 0 and alternative hypothesis is atleast one of the regression parameters has effect. Now, this is the case and under this situation our test says that my model is fit. So, it is not contradictory it is it is r square it is the overall measure it is saying that, yes overall I you may not be happy with the explanation, because you are not able to explain the total variation of the y variable, but you are able to explain 52 percent. So, that means some variables are contributing and that is being tested here.

If this is the case then definitely we want to test the individual parameters. Now, collectively this f test is saying that there are some parameters, which are contributing f test is saying r square is saying that overall variability explanation is poor. Now, of R square you may disconnect the model absolutely no problem we may not go further, but as 52 percent variability is explained 52.7 percent variability is explained. So, then you may be interested to know also that it is something, which you know now. So, i want to know, which variable is contributing towards this direction? So, then next topic is test of individual parameters.

So, individual parameter mean beta j 0, 1, 2 p you have seen earlier that the beta cap is the estimate of beta. So, that means expected value of beta j will be beta j. Ok? Now we can also create one statistics like beta cap j minus expected value of beta j cap by standard error of beta j cap. This will follow t distribution n minus p minus 1 will be the degrees of freedom. So, that means this quantity this quantity becomes beta j cap minus beta j by now this value S E beta j can you remember, what will be this value?

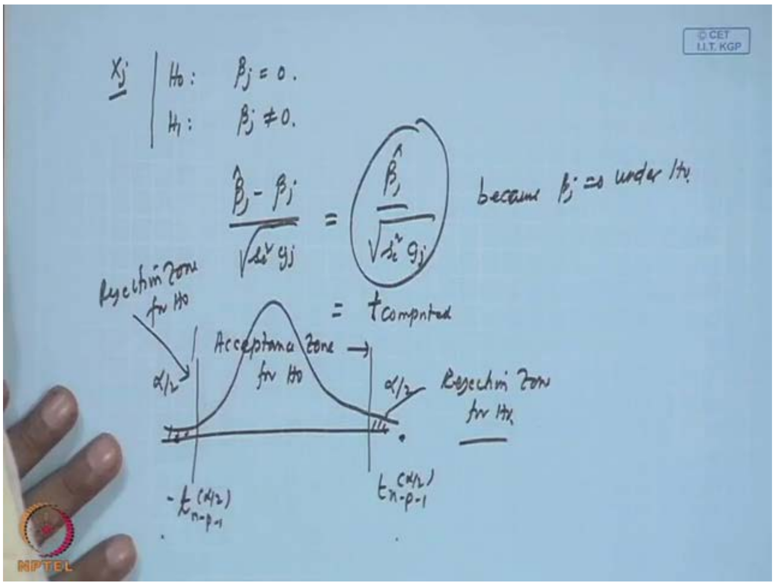
Last class S E square into C j j in last class we have seen this one, C j j where c is nothing but X transpose X inverse this will be a matrix of p cross p. Somewhere, in the diagonal line that jth variable is there this you are taking as C j j fine. Now, this we are saying it will follow this distribution.

(Refer Slide Time: 41:57)



Now, this will follow t distribution we are we are saying this follow under H 0, so we will test some hypothesis here.

(Refer Slide Time: 42:04)

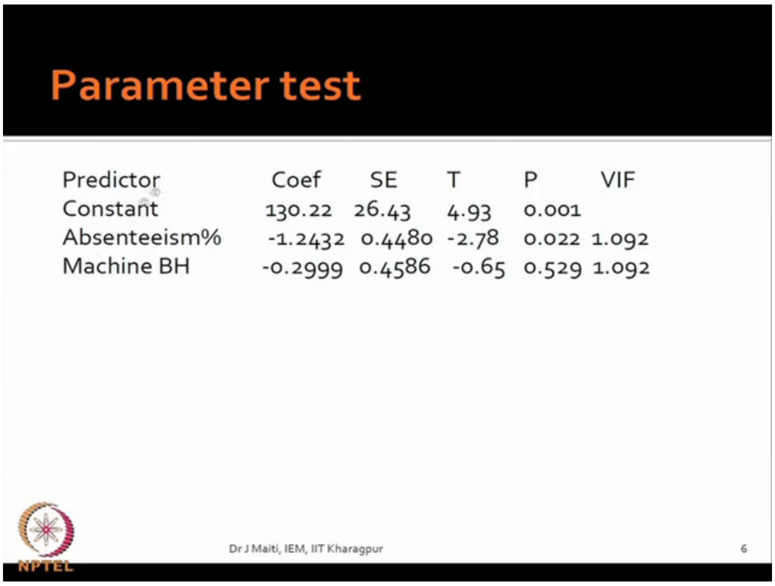


What is this H 0? H 0 is beta j equal to zero then H 1 is beta j not equal to zero what is the difference this versus f test. Here we are saying beta j equal to zero for all and alternatively atleast 1 is not 0 here you are not testing collectively you are testing for a particular variable X j, its contribution the contribution is beta j. So, you are saying H 0 beta j equal to 0 and beta j not equal to 0, we are not considering the other variables getting me.

So, under this case what will happen to this that beta j cap minus beta j by S E square C j j will become beta j by S E square C j j, because beta j equal to 0 under H 0. So, we are putting beta value 0, so this is what is t value you find out in any software if you run you get, what I can say that output in that output you will find out one table individual parameters. Then the standard error then t values this t value is this beta, this one. Now, this quantity follows t distribution, what does it mean it is a two tailed distribution? So, this side let it be alpha by 2, this side also alpha by 2.

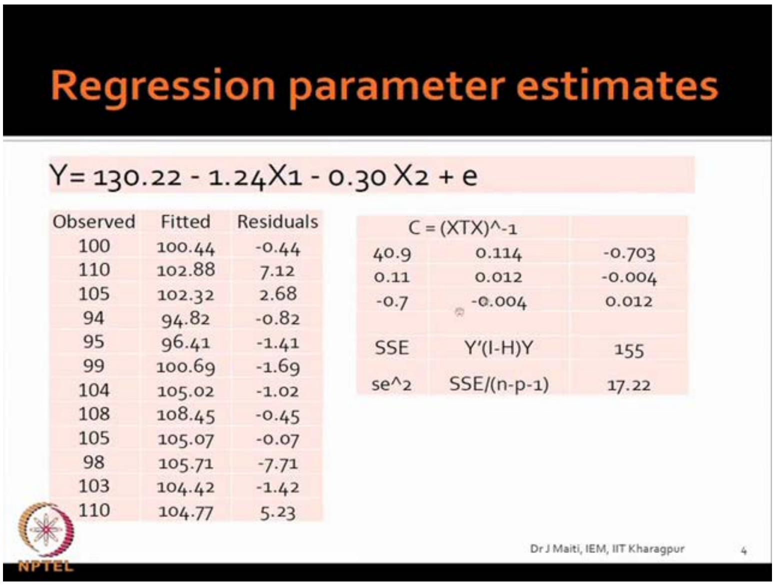
Then this is t n minus p minus 1 alpha by 2 t n minus p minus 1 alpha by 2. This is minus, this is plus, so if your computed t this is the t computed beta j cap by its standard deviation. If this one falls here or here either the left extreme or right extreme, that the rejection zone then H 0 will be rejected. So, this is our acceptance zone for H 0 then this side is rejection zone for H 0 and this is also rejection zone. Correct?

(Refer Slide Time: 45:26)



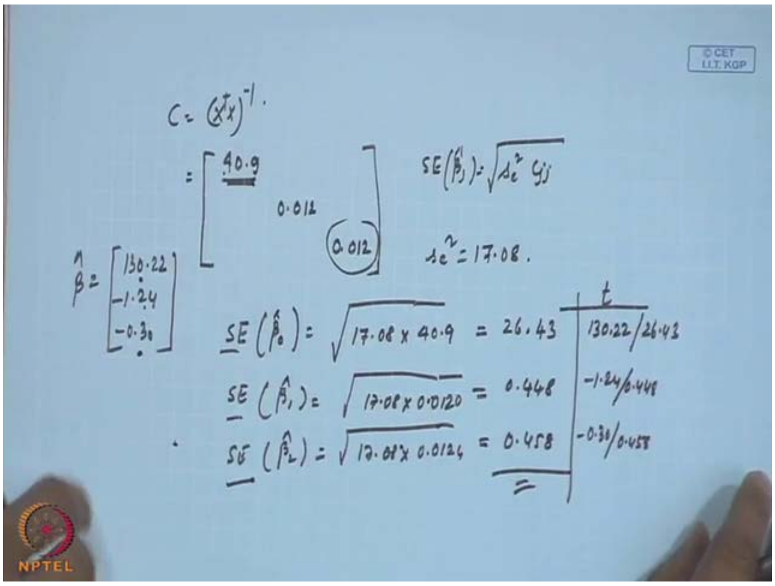
So, let us see what is I think you will be able to do this one with all the information are with you all information. I we have already seen beta j C j j also given to you, S E square is known to you, everything is known to you. So, you will be able to compute this. Now, this is the parameter test table in this table you see this is the predictor or explanatory variables or independent variable one is beta 0, which is the constant term Absenteeism machine breakdown hours. Now, we have seen that beta 0 is one 30.22 Absenteeism minus 1.24 and machine breakdown hours minus 0.30. Since, that we have given, now these are the standard error the standard error is this.

(Refer Slide Time: 46:07)



Let me see this you see this is c matrix. In this C matrix, how these things are coming?

(Refer Slide Time: 46:16)



You can write that C is X transpose X inverse. This is inverse, this one I have written in X l this is X transpose X inverse. Now, this value is 49, I am writing only the diagonal 1.2 0.12 off diagonal elements are also there. So, then this one is C and your standard error of beta j cap this is S E square C j j its square root [FL]. What is S E square value? We got S E square I think I told you that 17.08.

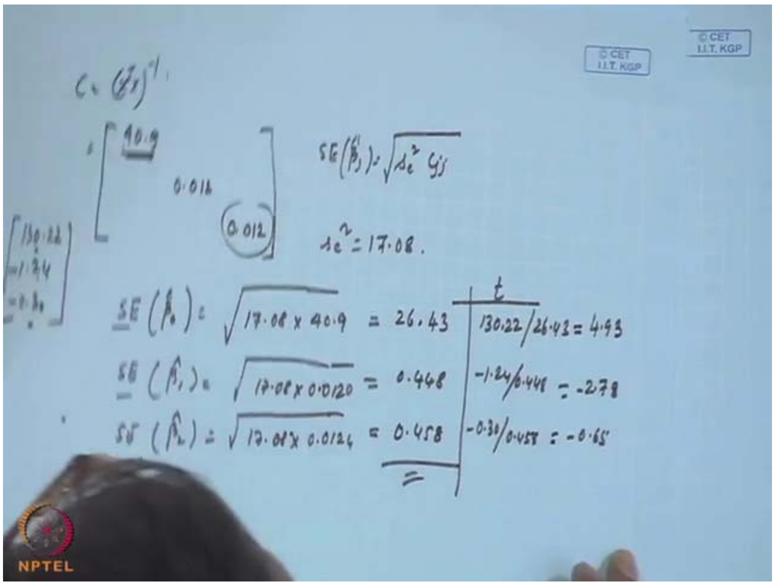
So, S E square is 17.08 correct, so if I want to find out standard error of beta 0 cap then you will be writing 17.08 into what you multiplied 49. So, this value you will get some value this value will be basically, 26.43. So, similarly standard error of beta 1 cap it will be 17.08 into 0.012. This value will be your 0.448 and standard error of beta 2 cap this is also same 0 8 into this one is also 0.02012 and some values are there 0.125 or 4 something there, this one is 0.058 here it will be 0 like this. So, rounding errors effects are there.

Once you know this values that S E are known, what will be your t values corresponding t values can you not find out t values are estimate by standard error first one estimate is 130.2 beta. I think you can still remember this one 30.22 minus, 1.24, minus 0.30, that I have given you earlier. So, this is the beta j beta 0 beta 1 beta 2 all cap estimate value. So, this divided by standard error standard error is 26.43 second one is how much minus 1.24 divided by standard error 0.448. Then third one is your minus 0.30 divided by 0.5458.

Student: ((Refer Time: 49:40))

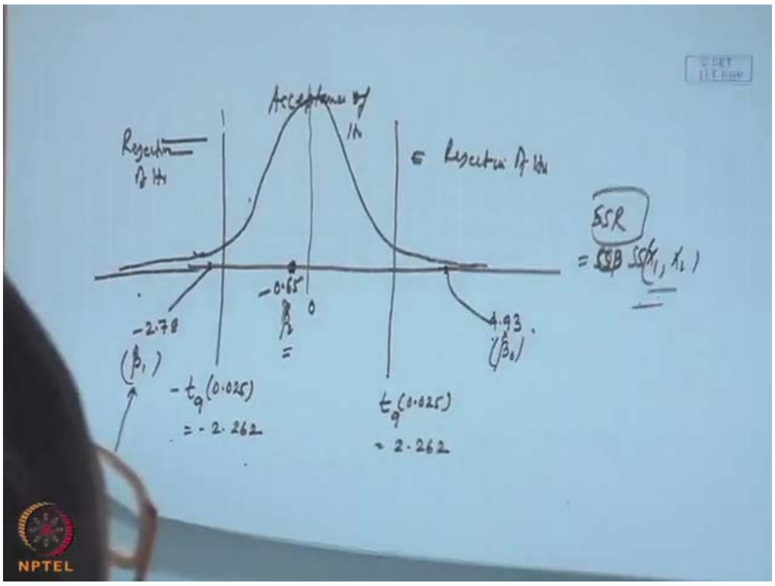
It is zero H 0 is saying that beta j equal to 0.

(Refer Slide Time: 49:58)



Then this values, all these values, this value will become 4.93. This is minus 2.78, this is minus 0.65. Correct?

(Refer Slide Time: 50:11)

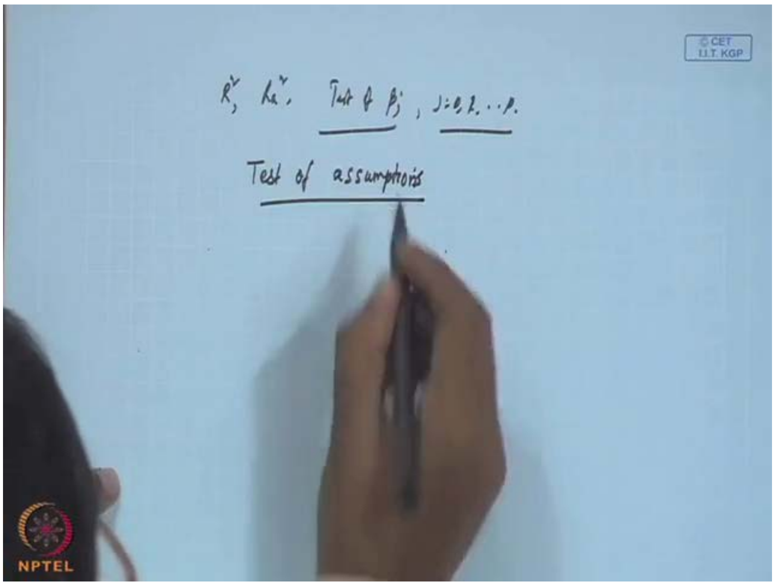


If this is the case, then you see one thing that this is my t distribution. So, I am creating my acceptance zone, this is t n minus p minus 1 mean 9 alpha value. Let it be 0.025 that is 0.0205 by 2 this is minus t 9, 0.025, this value will be I think 2.62. So, what is your beta 0 value? 4.93, so 4.93 mean somewhere here. So, 4.93second one is minus 2.78, somewhere here third one is minus this is my 0 minus 0.65. So, you have seen this is your rejection zone rejection of H 0 rejection of H 0 acceptance. Now, this one is related to beta zero cap, this value this is beta one cap related to this is beta two cap.

Beta 0 rejected beta 1 also rejected only beta 2 is accepted. So, that is why, what happen atleast one of the variables. There are two variables X 1 and X 2 constant is the intercept term we are interested in this two concept is also there it is also significant. So, beta 1 is significantly contributing, if it is contributing, what is its contribution that also you maybe interested to know?

So, you also be interested to know, what will be the contribution of X 2? This is the parameter that if one unit change in X 1 is there beta 1 unit change in y will be there, but from regression S S R point of view S S R can be your s s beta. That is beta one also or X 1 I can say S s X 1 and X 2 all those variables they are also contributing. So, you think if you can find out actually in regression, we only require this thing. We require this beta value, so this is what is our parameter test?

(Refer Slide Time: 53:37)



So, essentially I told you that R square R a square and test of parameters test of beta beta j j equal to 0 to p this is the test after that you have to do the test of assumptions. Your model is fit and you are considering the model, then only the individual parameters and test of assumptions are required. Otherwise, if you think that the model fit is not good you are disconnecting the model no need of fitting the test of assumptions further, but when you accept a model you must do the test of assumptions. I think test of assumptions we will consider in the next class.