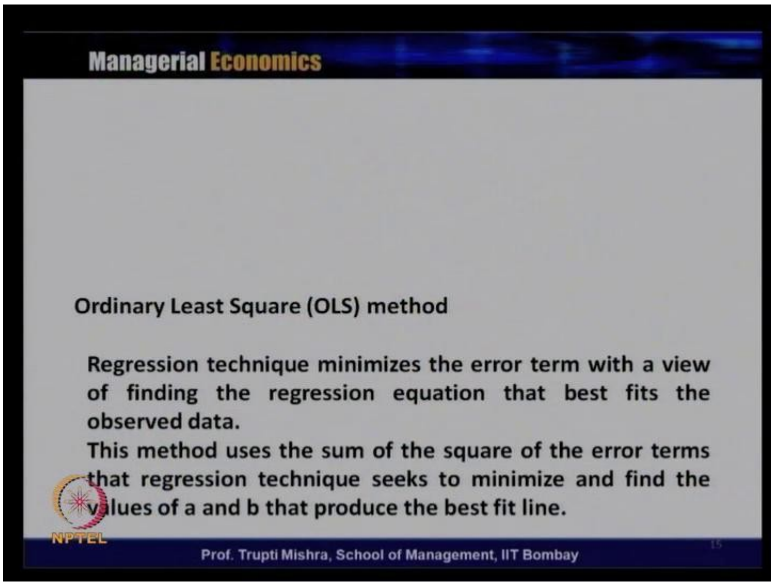
Managerial Economics
Prof. Trupti Mishra
S. J. M School of Management
Indian Institute of Technology, Bombay

Lecture -14

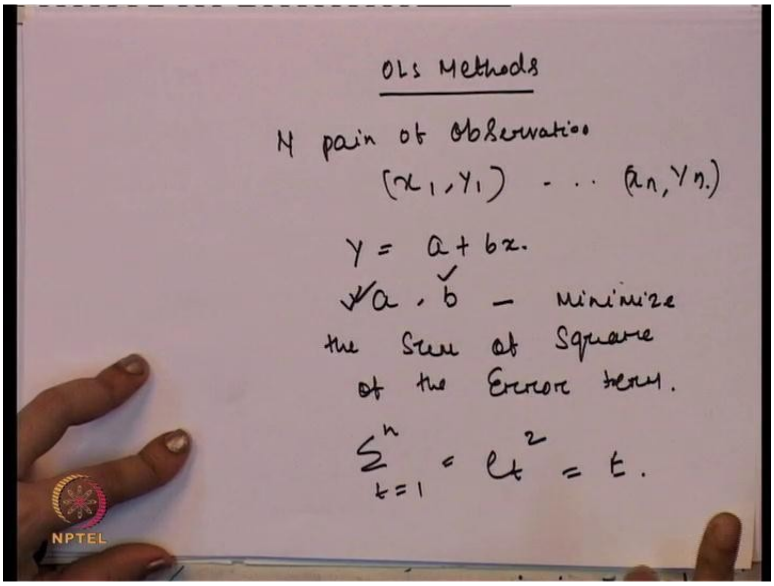
Now, we need to minimize this functional form to understand, or this functional form to know that, how to minimize this. This functional form error term we need to minimize this using the least square method. Now what is a least square method, because we will be using this method to minimize the error term; that comes between the regression line, and the actual data point.

(Refer Slide Time: 17:50)



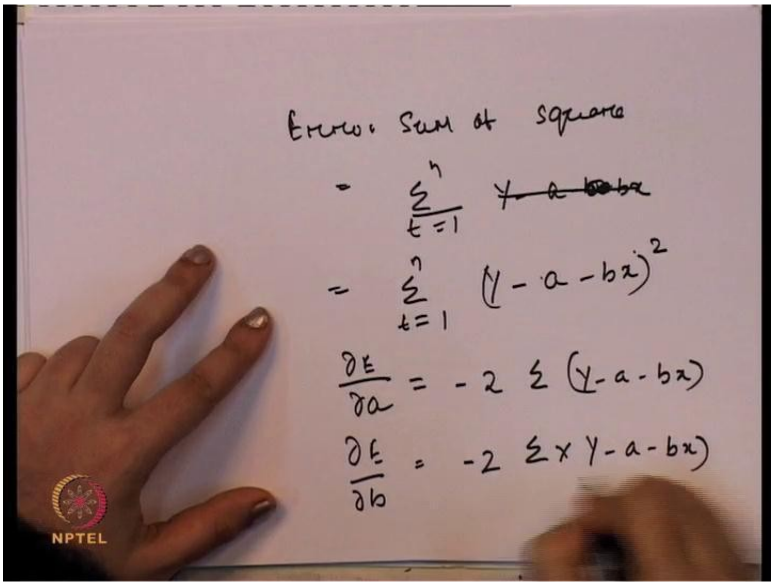
What is this least square method. Generally, regression technique minimize the error term with view of finding the regression equation that best fits the observed data. This method use the sum of the square of the error term, that regression technique seeks to minimize and find the value of a and b, that produce the best fit line. Now, how this OLS method then minimize this error term. They find the sum of the square of the error term, that regression techniques trying to minimize. They will take the sum of the square of the error term, the deviation between the actual data point and the regression line, what through the regression technique they are trying to minimize, and find the value of a and b, that will best fit the regression line, or the with the observed data point or the actual data point. So, we will see now this OLS method; that how this OLS method is being used to minimize the sum of the error term, by finding the value of a and b.

(Refer Slide Time: 18:56)



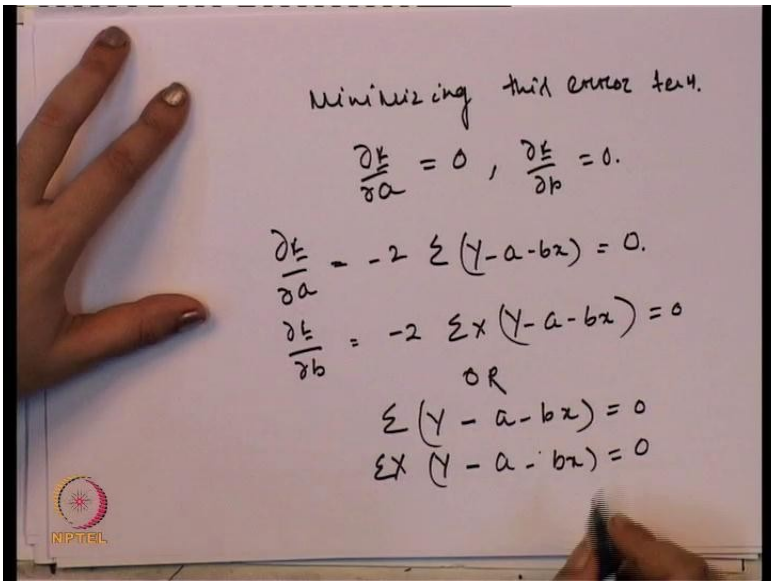
So, we have, suppose we have n pair of observation. We can call it from both the variable x 1, y 1 to x n, y n. Here we basically want to feed the regression line, given by the equation Y is equal to a plus b x, and what is the motivation to fit the line. We need to find such value for a and b, that will minimize the sum of square of error term. So, we need to calculate the value of a and b, which will minimize the sum of the square of the error term. So, this is e is equal to t from 1 to n; that is e t square, or we can call it as the e; that is the error term. So, in through OLS method, we are trying to find the value of a and b, which will minimize the sum of the square of the error term, and this is the sum of the error term square.

(Refer Slide Time: 20:26)



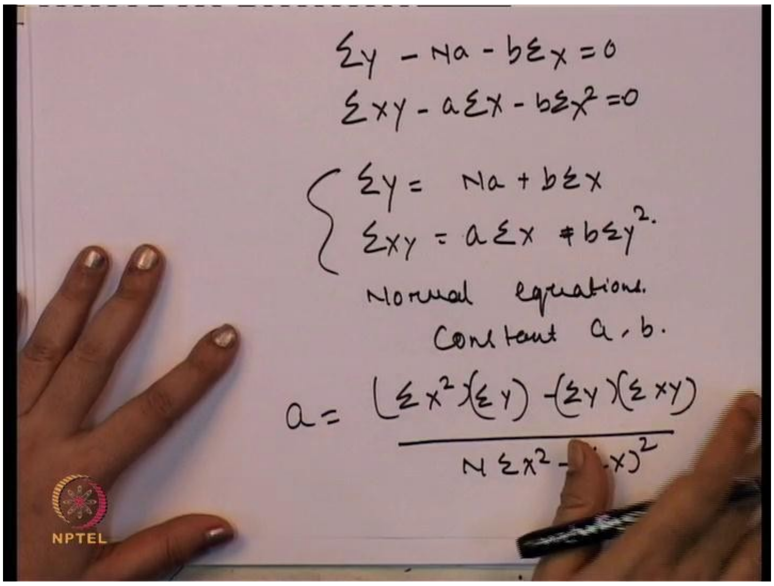
Now, we will see how we are going to minimize this error sum of square; so, error sum of square generally t 1 to n that is y minus a minus b x. So, e is equal to t is equal to 1 by n. So, Y minus a minus b x square, and we need to differentiate it with respect to a, and with respect to b. So, if it is differentiated with respect to a then we get 2 sigma y minus a minus b x, and in this case minus 2 sigma x y minus a minus b x. And we need to minimize this, and for minimization we need to equalize this with 0.

(Refer Slide Time: 21:33)



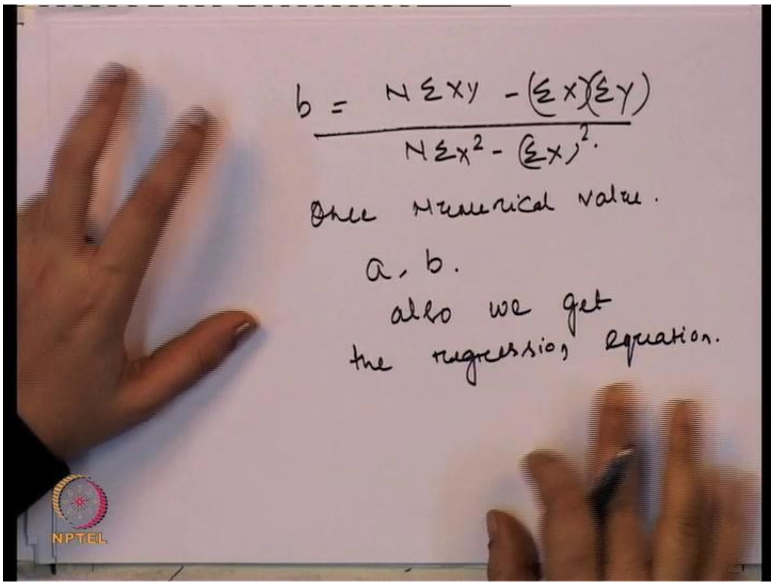
So, for minimizing this error term; so, minimizing this error term, we have to set d e by d a has to equal to 0, and d e by d b has to be 0. So, d e by d a, which is minus 2 sigma y minus a minus b x is equal to 0. And d e the d b is minus 2 e x, y minus a minus b x is equal to 0. Alternately, see minus 2 is a common factor in both this case; we can consider this, take out this minus 2. So, this will be a sigma y minus a minus b x has 2 equal to 0 and x y is equal to a minus b x has to be equal to 0.

(Refer Slide Time; 22:47)



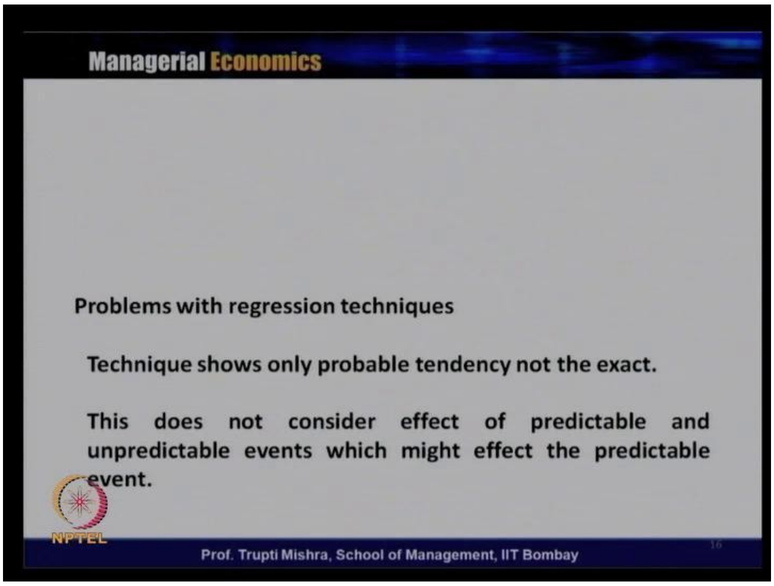
Now, we can rewrite this equation as, e y minus n a minus b e x is equal to 0, e x y minus a sigma x b sigma x square is equal to 0. So, arranging this, we can call it u y is equal to n a plus b e x, and e x y is equal to a sigma x plus b e y square. So, these are the normal equations, and this normal equation can be solved by determining the value of constant a and b. So, we need to find out the value of a on the basis of this normal equation, so this is e x square, e y minus e y, e x y, and by N e x square minus e x whole square.

(Refer Slide Time: 24:03)



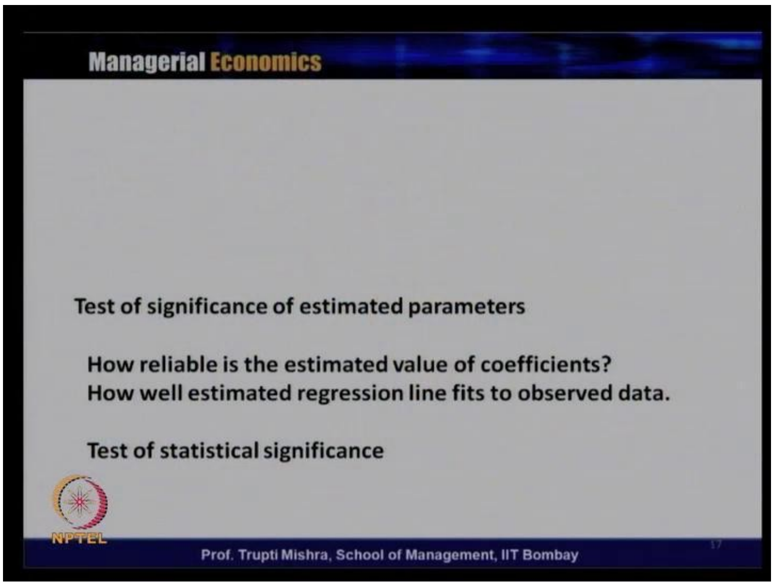
And for b, we can find out this N e x Y minus e x, e y, N e x square minus e x whole square. So, once we put the numerical value here, for x and y. Once we apply this numerical value of x and y, we get the value of a and b, and also we get the regression equation. So, this is the formula to find out the a and b. So, in that formula we can put the value of x and y, through that we can get the value of a and b, and from there again we can get the regression equation, which will best fit to the data. Now, So, we use OLS method to find out the, OLS method to find out the value of a and b, which will best fit to the regression line and also which will minimize the error term, what comes from the difference between the, deviation between the, original regression line and the, actual data point or the observed data point.

(Refer Slide Time: 25:24)



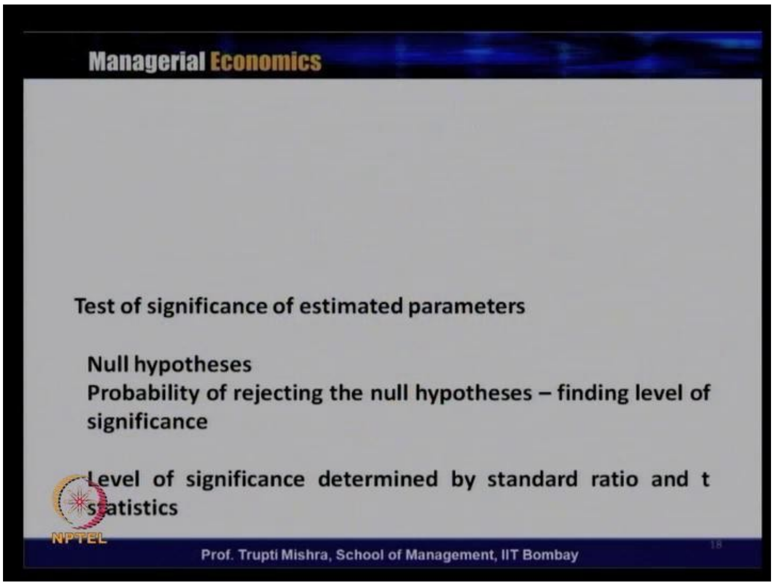
Now, what is the problem with this regression techniques? This technique shows only the probable tendency not the exact tendency, and that is how you cannot say that whatever the, it cannot be exact the value after getting the value of a and b, also it may not possible to get the best fit line. And this technique does not consider the effect of predictable and unpredictable event which might effect the predictable event. So, neither it, consider the effect of predictable, nor unpredictable event, but which might effect the result, or which might effect the value of a and b, and that is why there is a problem with this method.

(Refer Slide Time: 26:05)



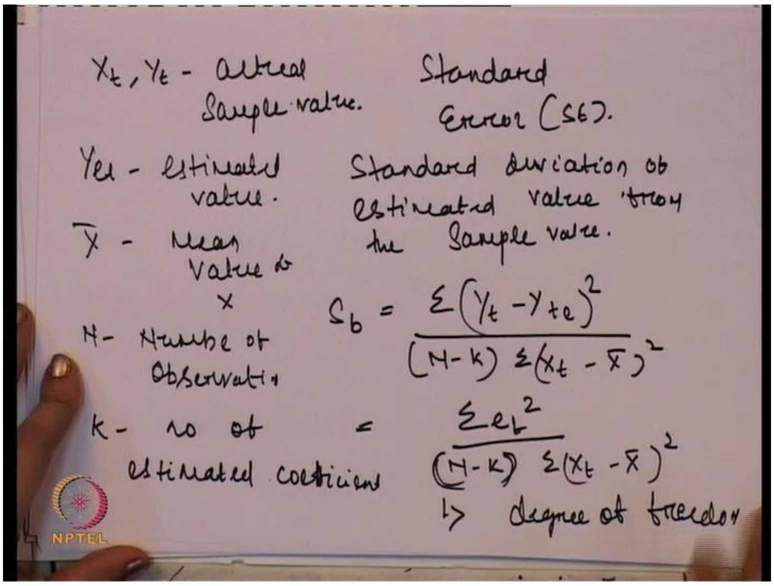
Now, how to overcome this, we can find out how reliable is the estimated value of coefficient, how well estimated regression line fits to the observed data; and that we can find out through the test of strategically significance. We can find out the estimated value of coefficient, and how estimated, how well estimated regression line will fit to the observed data. How we will do this test of significance. Generally, this is the test of significance of the estimated parameter, so how we generally do it.

(Refer Slide Time: 26:37)



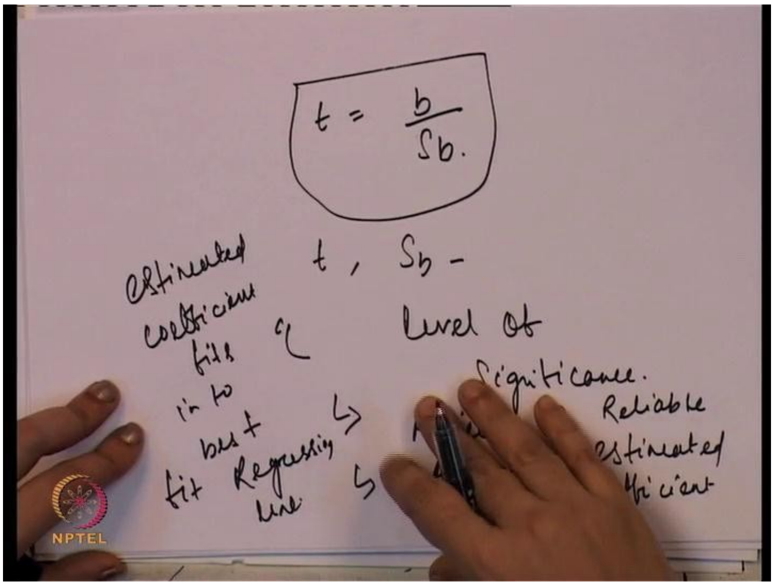
We take it in null hypothesis, and we try to see that whatever the null hypothesis, what is the probability of rejecting the null hypothesis. We find a level of significance, and we say that if the level of set of a limit and set of a percentage; that if the level of significance of this, then the null hypothesis has to accept. If the level of significance is this, then the null hypothesis has to reject; so for generally, hypothesis testing also with this the, level of significance. Then this level of significance is determined by the standard ratio and the t statistic. So, next we will find out the formula, to find out how to what is the standard error or what is the how to find out the standard error, and how to find out the t ratio, because that will through the standard error and t ratio, we can find out what is the level of significance or what is how reliable is the estimated coefficient, or how the estimate is estimated coefficient will fit the, regression line into the observed data.

(Refer Slide Time: 27:46)



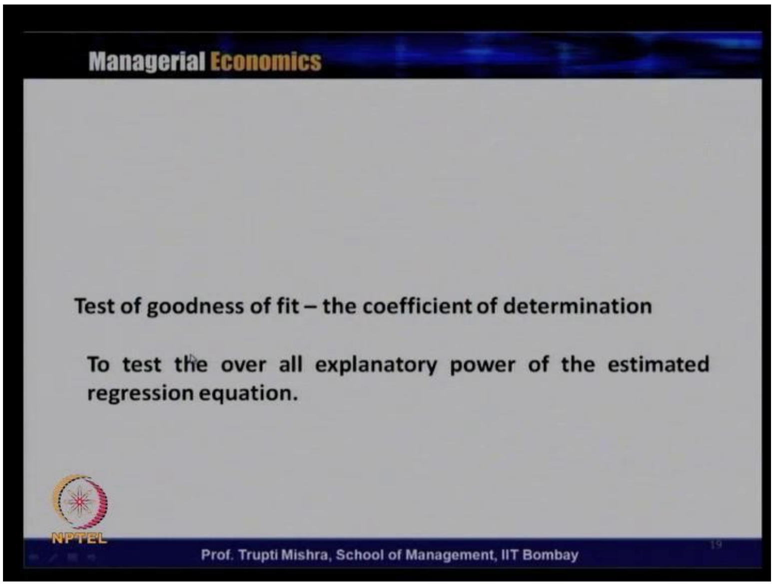
So, we will find out the standard error. So, standard error is it is nothing but the standard deviation of estimated value from the sample value. So, standard error of coefficient b can be. This is the formula to find out the standard error for coefficient b; that is sigma y t minus y t e square by n minus k e x t minus x bar square. Or rewriting this e t this is the error term square N minus k e x t minus x bar square. So, here x t and y t, this is the actual sample value for x and y for the time period t, y e t is the estimated value, x bar is the mean value of x, n is the number of observation, and k is the number of estimated coefficient. So, this n minus k is generally known as the degree of freedom. So, this is how we calculate the standard error, this is the formula to calculate standard error.

(Refer Slide Time: 30:06)



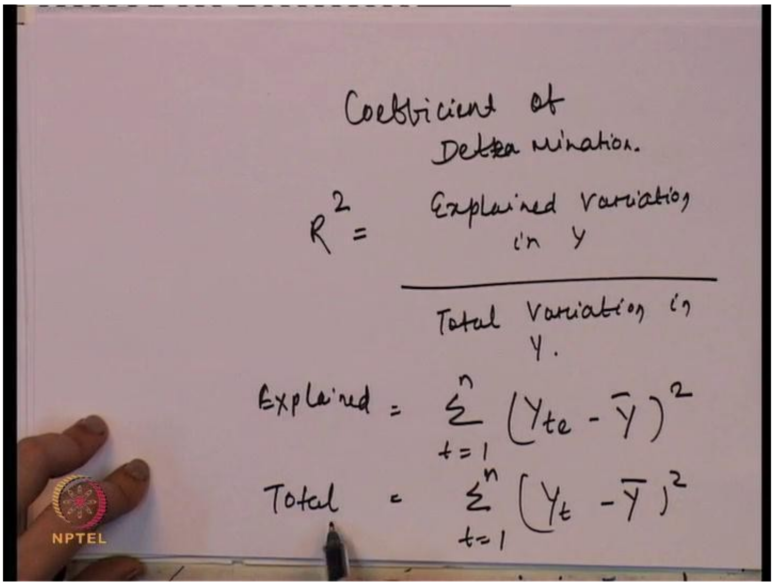
And once standard error is calculated; then we can find out the t ratio, and t ratio on the basis of the value of b and standard error of the b. So, this t ratio with the value of t and s b, we can find out the level of significance. And the level through the level of value of the level of significance, we can find out whether to reject null hypothesis or accept, and on that basis we can find out that, how reliable is the estimated coefficient. And how the estimated coefficient will fit into the regression line, best fit regression line.

(Refer Slide Time: 31:20)



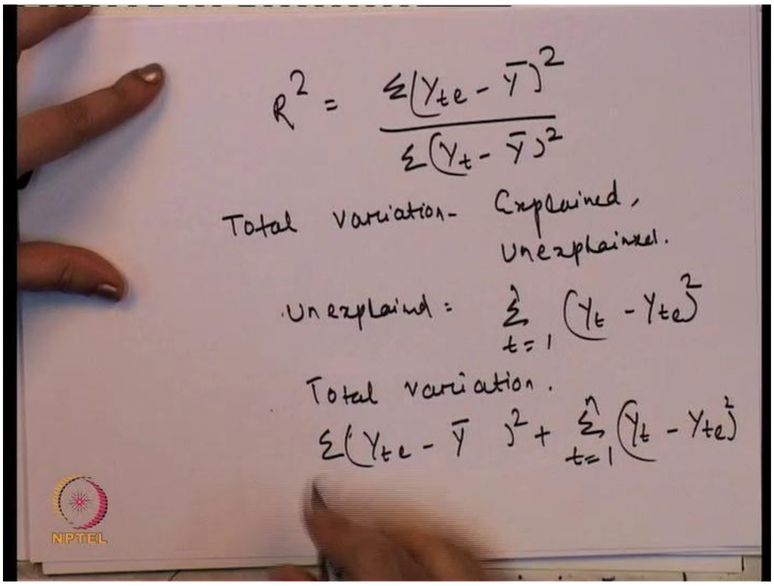
So, this is, this level of significance, or this test is only to find out how reliable is the estimated coefficient, but apart from this also, there is a test of goodness of it, or we call it is the coefficient of determination. It is to test the overall explanatory power of the estimated regression equation, or the estimated regression model or the estimated regression equation. So, through this coefficient of determination, or the test of goodness of fit, generally this test is conducted to test the overall explanatory power of the regression equation, because that gives the clarity, that gives the accuracy, that whether the regression equation is fitting into the best fit regression line or not.

(Refer Slide Time: 31:59)



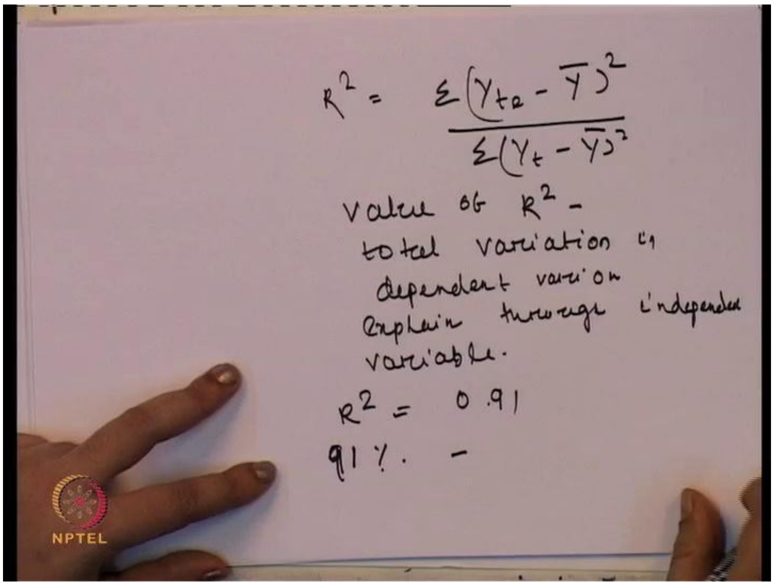
So, to test this goodness of it or to find out the coefficient of determination, we will see, what is the formula to do that? So, this coefficient of determination is otherwise known as R square, and R square is finding through explained variation of, explained variation in y and total variation in y. So, what is the explained variation in Y; that is, t is equal to 1 to n; that is, y t estimated mean value, this is the explained. And what is the total; total variation is e t is equal to 1 n then y t and y bar square. So, this is the explained variation in y, this is the total variation in y. Through this we will find out the coefficient of determination or the R square.

(Refer Slide Time: 33:19)



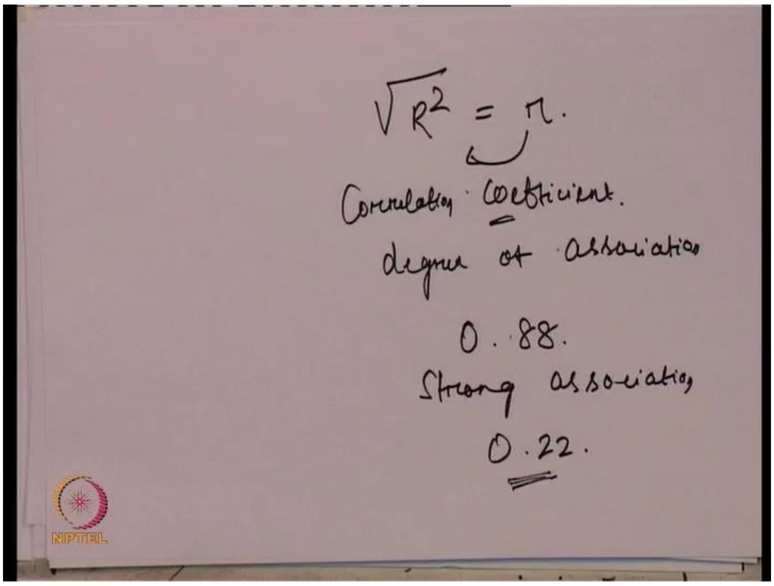
See, so through this R square is, since this is explained and total, so explained variation in y is y t e by y bar square, and total variation is y t minus y bar square. So, this total variation it has two parts; one is explained, another is unexplained. So, from here unexplained part is, e t is equal to 1 to n, so y t is y t e square, and total variation is; that is unexplained y t e minus y bar, this is explained plus unexplained; that is, y t minus y t e square.

(Refer Slide Time: 34:32)



So, now, this is the total variation, this how we got, that is through the explained and the unexplained part of the variation in y. So, R square is equal to, y t e minus y bar square, this is the explained variation, and y t minus y bar this is the total variation. So, this value of R square, once we get the value of R square. This talks about, what is the total variation in dependent variable explain through independent variable. So, suppose the value of R square comes as, R square is equal to, suppose 0.91. It means 91 percent of the total variation in the dependent variable is explained by the independent variable, and this is also we can say highly explanatory power, and the regression line is good fit, because of the value of R square is 91 percent.

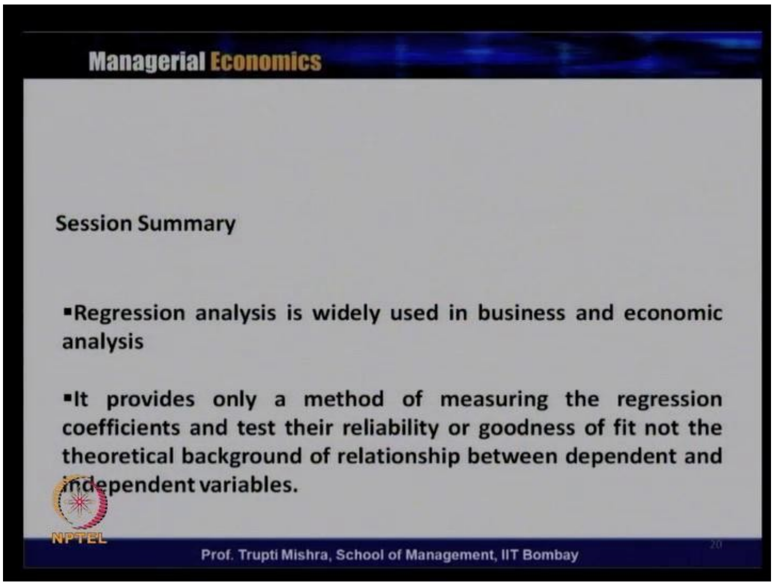
(Refer Slide Time: 36:01)



Next, we will find out one more relationship between the two; that is, from the R square. And this square root of R square gives us R, and R is nothing but the coefficient of correlation coefficient, and what is the role of this correlation coefficient in regression equation? This measures the degree of association between dependent and the independent variable. So, this correlation coefficient generally measures the degree of association between dependent and independent variable, and how we get this R, this is from the square root of R square. So, this suppose we get a value of this is 0.88. So, how we can explain this 0.88 or what is the implication of this 0.88.

There is a strong association between dependent and the independent variable, because the value of R is 0.88. Suppose the value of R is 0.22, so out of one, if it is the value of association is just 0.22, we can say that these two variables are not strongly associated, or these two variables are not may be associated with each other; means, if a change in one variable it is not going to effect the it is not going to change the other variable.

(Refer Slide Time: 37:36)



So, if you summarize whatever we discussed today about the use of regression technique or use of rudimentary method to understand this, or to find out the relationship between the economic variable. Generally, regression analysis is widely use in a business and economic analysis, but when it comes to that what is the contribution in term of regression technique. They only the provides the method of measuring the regression coefficient, how both of them they are related, that is through correlation coefficient, and may be what is the magnitude, what is the change in the dependent and independent variable, that is through the regression coefficient.

So, this regression analysis is providing only a method to measure the regression coefficient, and also to test the reliability or goodness of it, but it is not providing the theoretical base of the relationship between the dependent variable and independent variable. So, the null hypothesis always there that there is a dependent variable and the independent variable; they are related; how they are related, and how independent variable and dependent variable they are related.

So, regression technique is not contributing to that, that how this theoretical relationship is being build up; or what is the basis of this kind of relationship between the dependent variable independent variable. They are not contributing to the theoretical structure of this relationship between the dependent variable independent variable. Only they are providing, the methods which empirically talks about the relationship between the dependent variable and the independent variable. And it also talks that when we estimate the coefficient of the independent variable, how reliable it is, and also that whether it fits to a good regression line, or best fit regression line or not. So, with this, we conclude the discussion on the regression analysis, and then we will start the next module, that is on the theory of demand.