**Lecture -7**
**Basic Tools of Economic Analysis and**
**Optimization Techniques (Contd…)**

So, we will continue our discussion on basic tool of economic analysis, what we were discussing in the previous class also. And particularly here we are trying to see how the regression analysis used for or as a tool for the economic analysis. So, typically we are talking about the regression techniques, how it is serving as a tool for the economic analysis in the previous class also and this session also. So, what is regression technique? Just a quick recap, what we discussed in the last class. And then we will continue our discussion on the different methods to find out the value of the slope intercept, and also what are the different methods that goodness to fit, and all this. But before that we quick recap what is a regression technique.

(Refer Slide Time: 01:07)
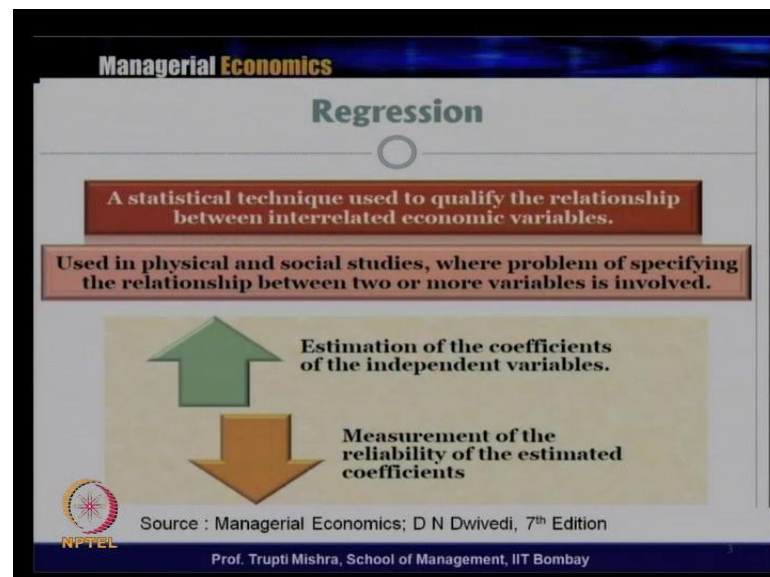


So, regression technique is a statistical technique used to qualify the relationship between the interrelated economic variable. Generally, it is used in physical and social studies, where problem of specifying the relationship between two or more variable is involved. And the first is to estimation the coefficient of the independent variable in the technique the first step is to estimate the coefficient of the independent variable, and then

measurement of the reliability of the estimated coefficient. So, first is to estimate the coefficient of the independent variable, because if you know regression, it is the relationship between the dependent and the independent variable. So, the first step of regression technique is to find out the value or the estimate the value of coefficient associated with the independent variable, and then to measure the reliability of this estimated coefficient. And for this generally we formulate a function, we just form formulate a hypothesis first, and on the basis of this, we formulate a function.

(Refer Slide Time: 02:10)



So, formulation of hypothesis is done on the basis of the observed relationship between two or more facts or event of real life. And after this, we generally translate the hypothesis into a function. So, suppose a hypothesis is sales growth is a function of advertisement expenditure, this hypothesis can be translated into the mathematical function, where it leads to Y is equal to A plus B x, where Y is sales, X is the advertisement expenditure and A and B are the constant.

(Refer Slide Time: 02:43)



Then, after a translating the hypothesis into the function; the basis for this is to find the relationship between the dependent and independent variable need to be specified, and stated in the form of equation.

(Refer Slide Time: 02:57)



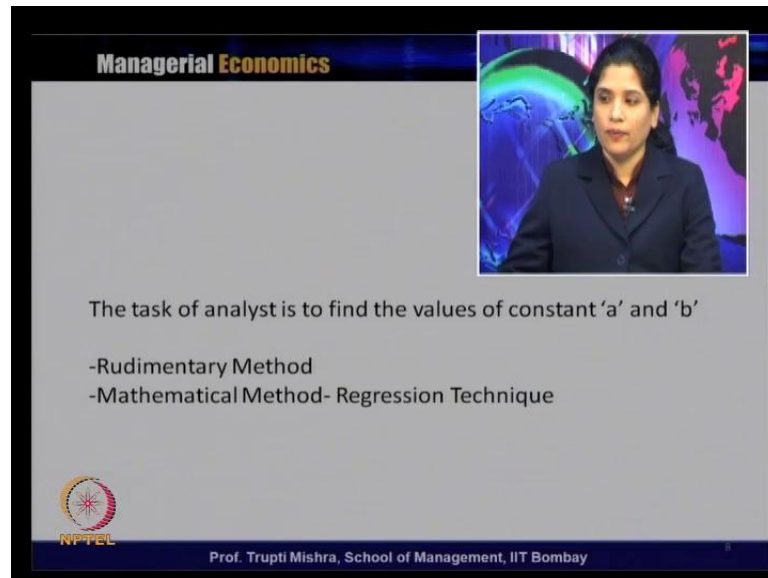So, here in the equation a is the intercept, it gives the quantity of sales without advertisement, when X is equal to 0. And b is the coefficient of Y in relation to X, gives the measure to increase the sales due to certain increase in the advertisement expenditure.

(Refer Slide Time: 03:14)



Now, here the task of analyst comes, and what is the task of analyst. The task of analyst is to find the value of the constant a and b, where a is the intercept and b is the slope. Because on the basis of value of a and b we will come to know, what is the relationship between the dependent variable and the independent variable. Generally two methods is being followed to find out the value of a and b; one is the rudimentary method and second is the mathematical method, generally typically known as the regression technique. So, we will start with the rudimentary method, that how this value of a and b is decided, and then we talk about the regression technique. So, first we will start the rudimentary method, and for suppose we take here the value of a constant a and b, we need to find out through rudimentary method.
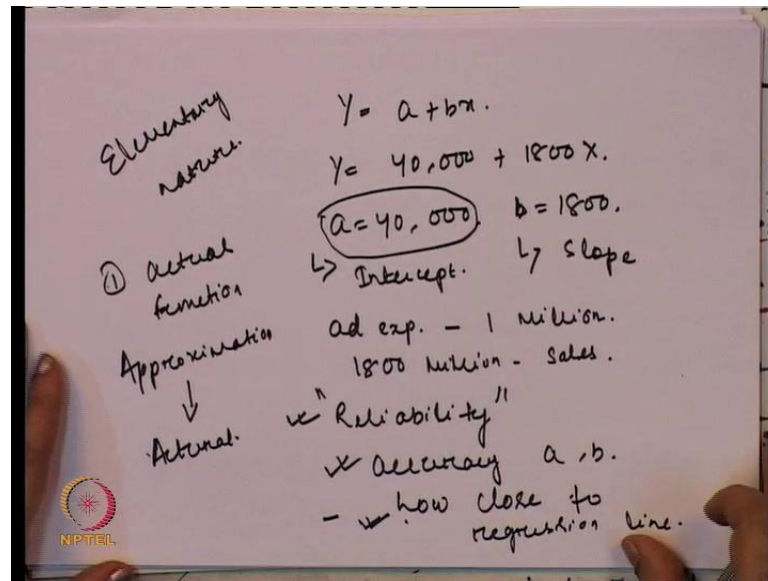
So, Y is equal to a plus b x is the functional form, or suppose we take a value Y is equal to value of a is 40,000, and this is 1800 x. So, here we can say a is 40,000 and b is 1800, and this is the slope, this is the intercept. So, if Y is the sales and x is the advertisement expenditure, when there is no advertisement expenditure, the value of Y will be equal to 40,000.

So, total sales without advertisement expenditure will 40,000. And if since value of b is 1800, we can say that if the measurement unit is in the million term advertisement expenditure of one million, will bring 1800 million increases in sales. Because this is the slope of the, value of the slope is 1800. So if the measurement unit is in million terms, advertisement expenditure of 1 million will be 1800 million increases in the sales. Now, the question is that, if this is the value of a and b, what is the reliability of this value. Whether how reliable this value, or what is the accuracy of this value a and b, or we can say that how close this value of a and b to the regression line. This cannot be solved through the rudimentary method, because this is generally a crud method being followed or the.

This is very elementary in nature, this rudimentary method is very elementary in nature, this cannot address this question that, how reliable is this or how accurate is this or how close to this, in the this typical regression line, because it indicates the visualization of function, not the formulation of the actual function. So, if you look at, it is not formulate

the actual function rather they visualize that, this is the value it has to be done. So, in this case, this is a approximation, because this is a crud method this is a approximation, not the actual relationship between two variables, and that is why rudimentary method is ruled out, or rudimentary method is not being followed much to find out the value of a and b, because they are not considering the actual function rather they are considering the approximate function. Now, the second method comes here to find out this value of a and b, typically the value of slope and intercept in order to understand the relationship between two variables. Two variables are economic variables in nature, and we generally use the second method as the regression technique to understand the relationship between two economic variables.

(Refer Slide Time: 07:43)
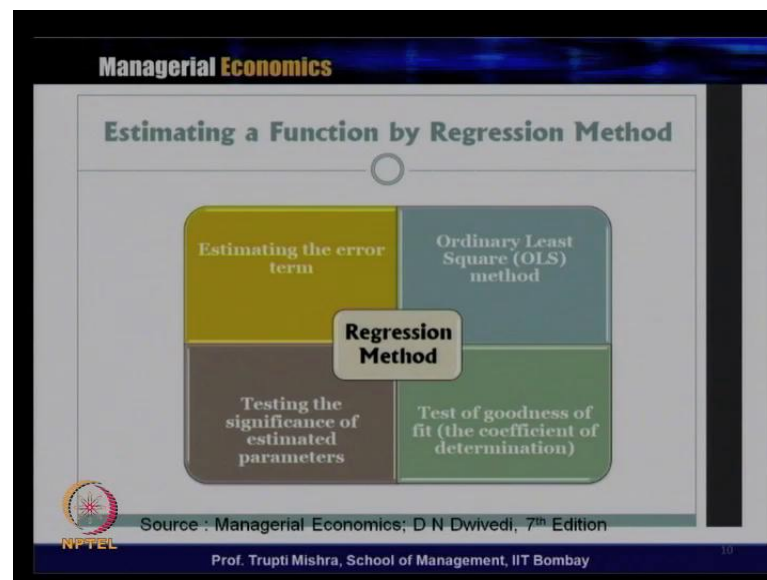


So, the second variable comes as the regression methods, here generally four components of these regression methods what we are going to discuss; first is to estimate the error term that through the ordinary least square method. Then we will test the significance of the estimated parameter. And then we will do the test of goodness of it, or to find out the coefficient of determination, because that will talk about the overall explanatory power of this model, how the model is or how the relationship between the dependent variable and independent variable is shown through the regression technique. So, typically this four stages being followed in case of a regression method, when we estimate a and b following this method; first we estimate the error term, then we do this through the ordinary least square method, then after getting the value of the parameter,

estimated parameter, we test their significance that how significant what is the level of significance of this estimated parameters.

And then we do the test of goodness of it to understand the overall explanatory power of this model, or overall explanatory power of the relationship between this two variables; that is dependent variable and the independent variable. So, first we will see how basically this regression line is being done, or how regression, the simple regression analysis. Then we will see how the error term comes, and then we will see how the level of significance can be done from the estimated parameter, what we have calculated using the ordinary least square method, and finally we will do the goodness of it.

(Refer Slide Time: 09:24)
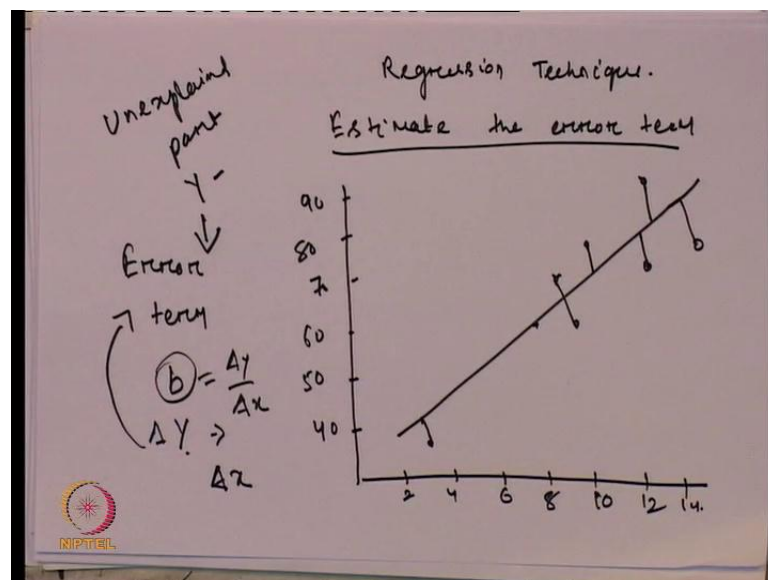


So, the first step is always to, in a regression technique the first term is always to, the first step is always to, estimate the error term. Let us take an example that what will be, since this is a relationship between the advertisement expenditure and sales, we will take different value of advertisement expenditure and how it is affecting the sale. So, suppose this is 40 50 60 70 80 90, or we can call it 2 4 6 8 10 12 14. So, we can just plot a regression line, and in the regression line some points we will find above the line and some points we will find below the line. So, all the points, all the actual data points is not on the regression line. There may be also some deviation in the regression line. So, all the advertisement expenditure do not fall in the regression line. So, since all the actual data point is not in the regression line; that leads to some error term in the model.

Why we call it error term, because there is a deviation in the regression line and the actual data point. And since there is a error term, the slope b does not explain the total change in Y with respect to change in x, because there is some error term. The slope b is what, slope b is basically del Y with respect to del x. So, we can say that, since there is a difference between the regression line and the actual data point; that leads to the error term, and since there is error term in the model, that implies that the slope b is not explaining the total change in Y with response to change in the x. So, there is some unexplained part of the Y, and this unexplained part of the Y is generally the error term. Now, we will see, what is the error term?

(Refer Slide Time: 12:10)



**Managerial Economics**

**Estimating the error term**
The error term refers to the deviation of the plotted points from the straight line drawn through the center of plotted points.

Graphical Explanation

Prof. Trupti Mishra, School of Management, IIT Bombay

So, error term refers to the deviation from the plotted point. So, what is the error term, error term is the generally the deviation from the plotted point, from the straight line drawn through the center of the plotted point. So, this is the center of the plotted point and that is why we call it the regression line. And what is error term, error term basically the deviation from the actual data point and this line. So, this is actual data point, this is the line, so the deviation is the error term. This is the actual data point, this is the line, this is the deviation is the error term. This is the actual data point, this is the line, this is the deviation from the error term, and similarly this is the error term, because that there is a deviation in the. Whenever there is a deviation in the actual data point, and the regression line or the line that is from the center of the data points; that gives us the center of the plotted point; that gives us the error term.

(Refer Slide Time: 13:12)



Now, why this error comes? This error comes, because there is an inaccurate recording or inaccurate specification of the sales data. So, sometimes there is inaccurate recording of this advertisement expenditure and the sales data, or also to specify the sales data there is inaccurate and that is leads to the error, and that is why we get a deviation between the actual point and the regression line.

(Refer Slide Time: 13:45)



There are two kind of errors; one is specification error, second one is the measurement error. What is specification error, specification error generally comes, because other

factor influencing sales could not be specified and included in the function. So, specification error comes from the fact, that whatever the other factor influencing the sales, could not be specified and included in the function, because here specifically we are saying that how advertisement expenditure is influencing the sales. But there may be many other factors which is influencing sales, that is not being consider, and that could not be specified, that could not be added in the function and that is why this specification error comes.

(Refer Slide Time: 14:18)
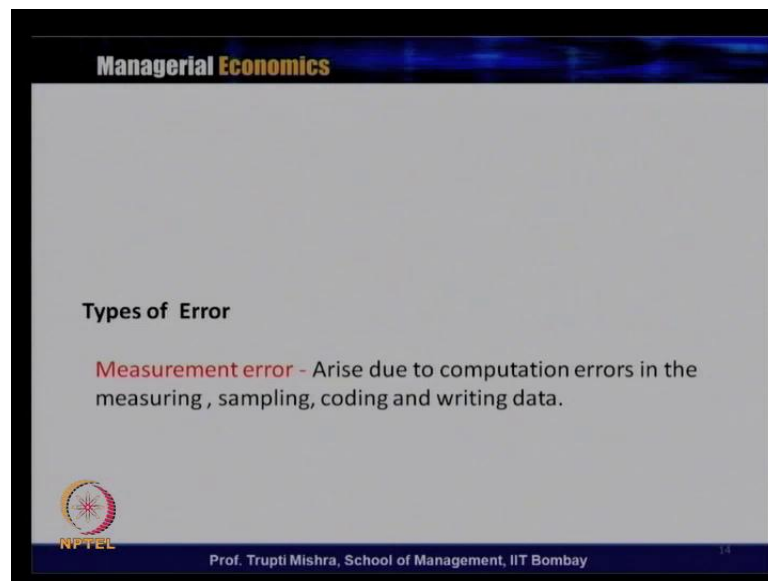


And second is the measurement error; this generally arise due to computational error in the measuring, sampling, coding and writing the data. So, specification error when we missed out some variables which influence the sales, and measurement error comes, typically its more mechanical kind of thing, this is arise due to this technical problem; that is computational error in measuring, sampling, coding or writing the data. So, we will see that, how we can understand this error term, error term like this, how this error comes, and how to minimize this error using the least square method.

(Refer Slide Time: 15:00)



So, we can say this Y t is the observe value, and Y c is the estimated value. What will be our error term, error term will be Y t minus Y t minus Y c. And if Y c is equal to a plus p x, this is to be the best fit. So, here again we will understand this, we will take this graph to understand this whether what is a best fit. So, here if you look at, if x is equal to eight million in a specific year, then in this case suppose this is the data point, this is the actual data point, here we can call it may be this is this is m, this is n, this is p. So, this p n is what, p n is basically the error term, because this is the difference between the actual data point and the original or the regression line. So, which one is best fit, best fit is one, where this Y t will be equal to Y c.

So, in this case, we can say that error term is equal to Y t minus Y c, and it will be if Y t is equal to Y c, then we will get error is equal to zero, and we will get it a best fit. So, but since we always assume that, there is some amount of error, because there is a deviation between the regression line, and the actual data point to we have some error term, and this error term is Y t minus Y c and Y is can be again called as the a plus b x, and this is the functional form for the error term. 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)
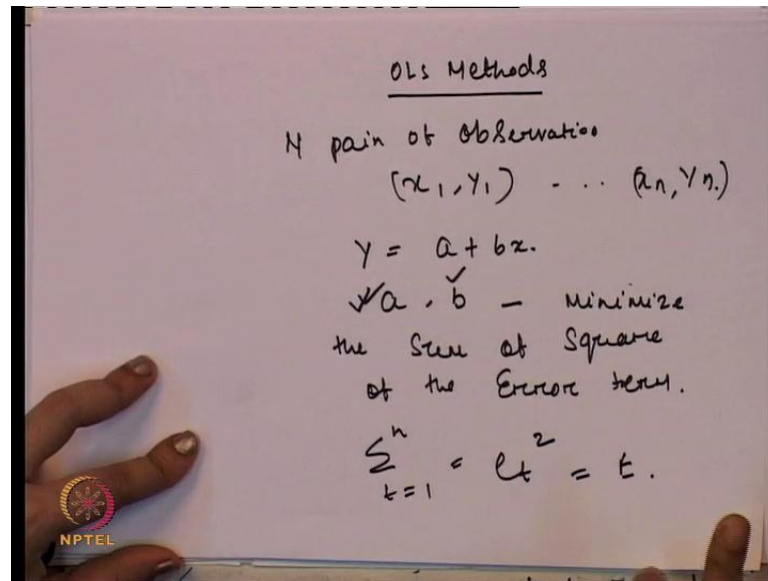


So, we have, suppose we have n pair of observation. We can call it from both the variable x 1 y 1 to x n y n. Here we basically want to feed the regression line, given by the equation Y is equal to a plus b x, and what is the motivation to fit the line. We need to find such value for a and b, that will 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 cases; 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.
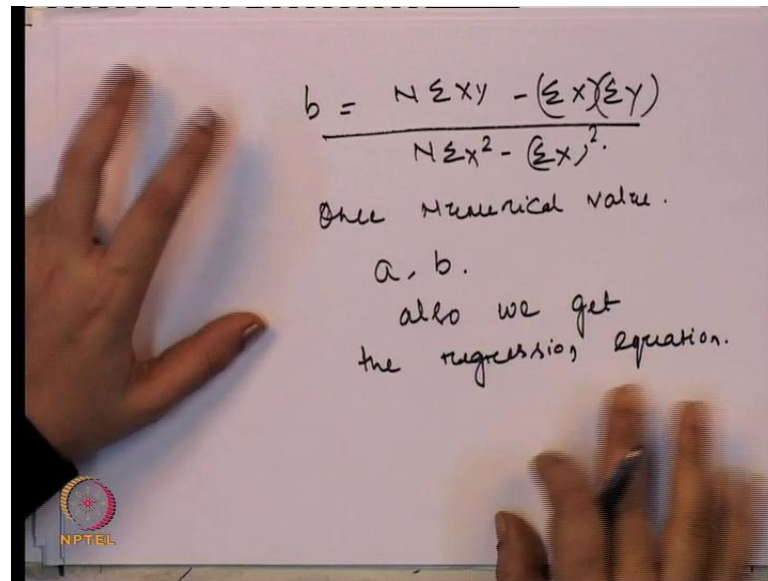
(Refer Slide Time; 22:47)



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 equations 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.
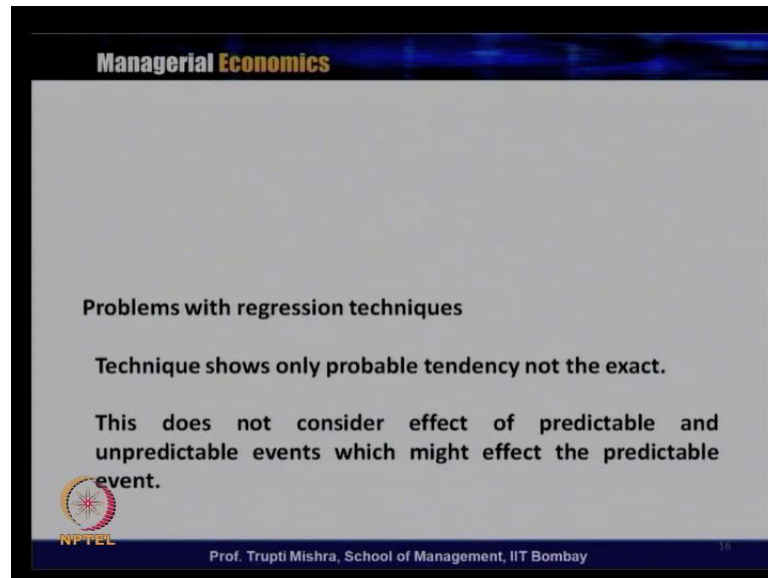
(Refer Slide Time: 24:03)



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.

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.

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)



**Managerial Economics**

**Test of significance of estimated parameters**

**Null hypotheses**
**Probability of rejecting the null hypotheses – finding level of significance**

**Level of significance determined by standard ratio and t statistics**

Prof. Trupti Mishra, School of Management, IIT Bombay

We take it in all 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)
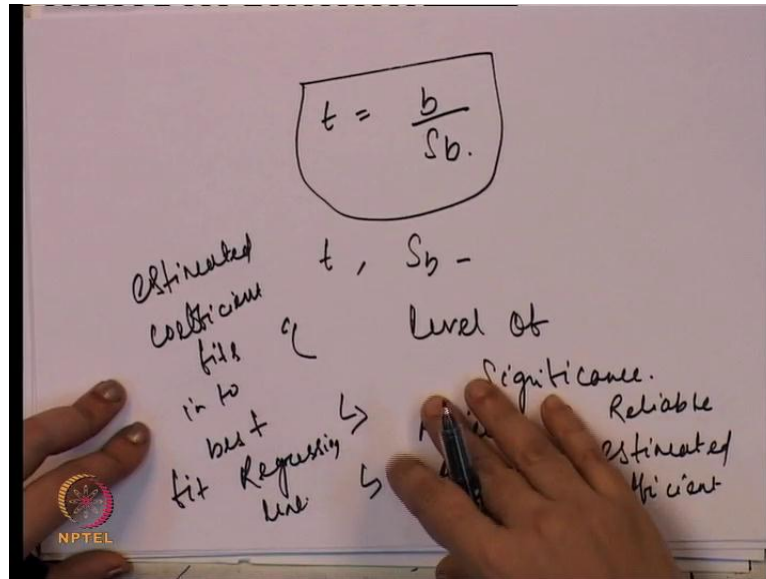


And once standard error is calculated; then we can find out the t ratio, and t ratio on the basis of the value of b and standard error of the b. So, this t ratio with the value of t and s b, we can find out the level of significance. And the level through the level of value of the level of significance, 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 fits into the regression line, best fit regression line.

(Refer Slide Time: 31:20)
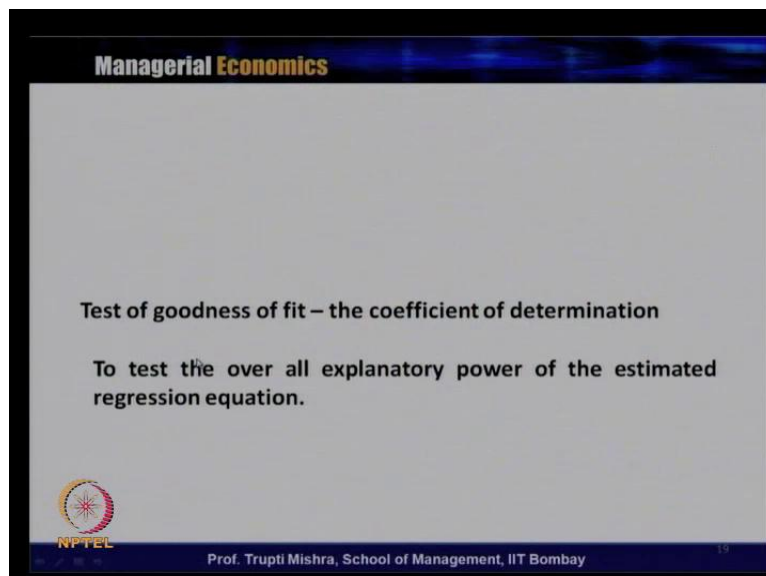
So, this is, this level of significance, or this test is only to find out how reliable is the estimated coefficient, but apart from this also, there is a test of goodness of it, or we call it is the coefficient of determination. It is to test the overall explanatory power of the estimated regression equation, or the estimated regression model or the estimated regression equation. So, through this coefficient of determination, or the test of goodness of it, 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)



$$R^2 = \frac{\mathcal{E}(Y_{te} - \bar{Y})^2}{\mathcal{E}(Y_t - \bar{Y})^2}$$

Total variation - Explained, Unexplained.

Unexplained $= \sum\limits_{t=1}^{n} (Y_t - Y_{te})^2$

Total variation.

$\mathcal{E}(Y_{te} - \bar{Y})^2 + \sum\limits_{t=1}^{n} (Y_t - Y_{te})^2$

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)



$$R^2 = \frac{\mathcal{E}(Y_{te} - \bar{Y})^2}{\mathcal{E}(Y_t - \bar{Y})^2}$$

Value of $R^2$ -
total variation in
dependent variation
explain through independent
variable.

$R^2 = 0.91$

$91\%$. -

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 w 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)
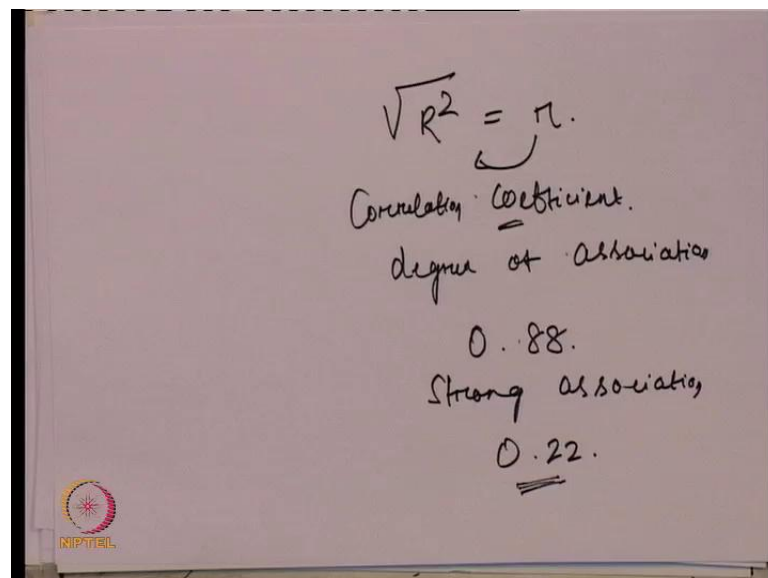


Next, we will find out one more relationship between the two; that is, from the R square. And this square root of R square gives us R, and R is nothing but the coefficient of correlation coefficient, and what is the role of this correlation coefficient in regression equation? This measures the degree 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 this two variables are not strongly associated, or this 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.