

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

Lecture -13

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)

The slide is titled "Managerial Economics Regression". It contains the following text:

- Regression**
- A statistical technique used to qualify the relationship between interrelated economic variables.
- Used in physical and social studies, where problem of specifying the relationship between two or more variables is involved.
- Estimation of the coefficients of the independent variables.
- Measurement of the reliability of the estimated coefficients

Source : Managerial Economics; D N Dwivedi, 7th Edition
Prof. Trupti Mishra, School of Management, IIT Bombay

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)

Managerial Economics

Method of Formulating a Function

Formulation of a hypothesis

- It is done on the basis of the observed relationship between two or more facts or events of real life.

Translating the hypothesis into a function

- Suppose a hypothesis, the sales growth is a function of advertisement, this hypothesis can be translated into a mathematical function
$$Y = A + Bx$$
where, Y = sales, X = advertisement and A and B are constants.

Source : Managerial Economics; D N Dwivedi, 7th Edition

NPTEL
Prof. Trupti Mishra, School of Management, IIT Bombay

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 Bx , where Y is sales, X is the advertisement expenditure and A and B are the constant.


(Refer Slide Time: 02:43)

Managerial Economics

Translating the Hypotheses into a Function

Translate verbal hypotheses into the form of estimable function.

To formulate, the relationship between the dependent and independent variables need to be specified and stated in the form of an equation.

 NPTEL

Prof. Trupti Mishra, School of Management, IIT Bombay

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)

Managerial Economics

Translating the Hypotheses into a Function

The constant 'a' is the intercept, it gives the quantity of sales without advertisement, when $X = 0$.

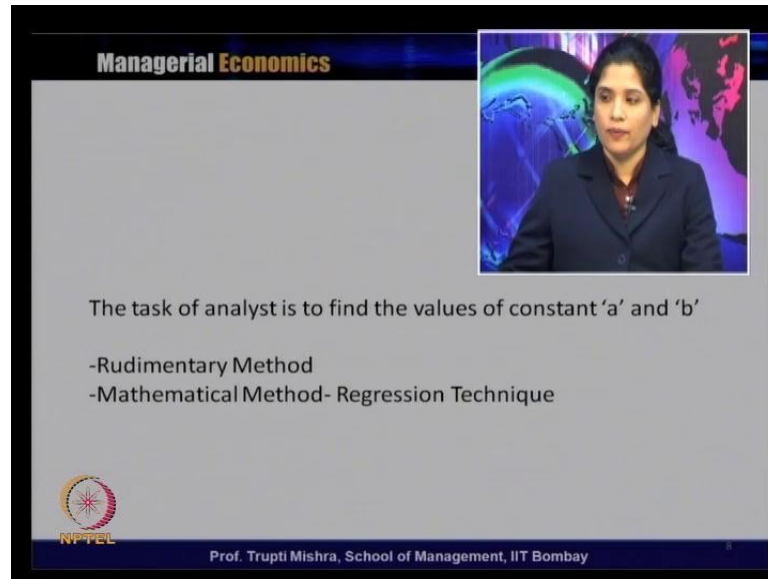
Constant 'b' is the coefficient of Y in relation to X- Gives the measure to increase in sales due to a certain increase in advertisement expenditure.

 NPTEL

Prof. Trupti Mishra, School of Management, IIT Bombay

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)



Managerial Economics

The task of analyst is to find the values of constant 'a' and 'b'

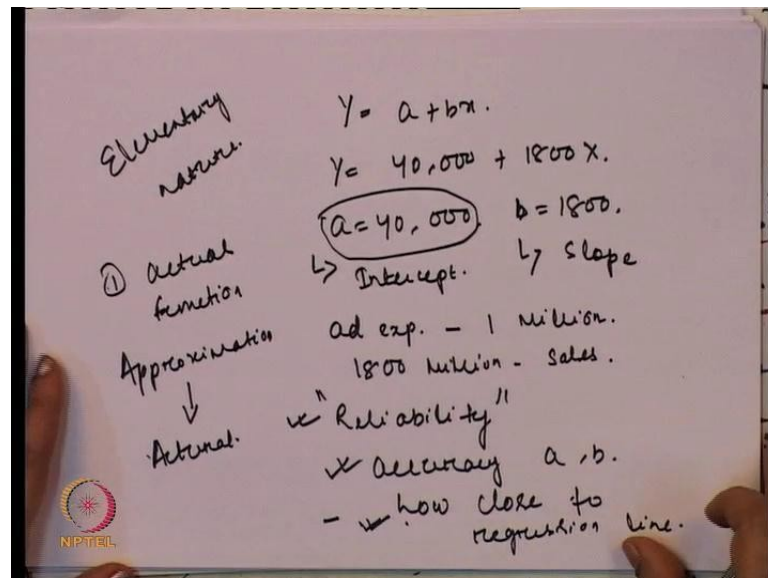
- Rudimentary Method
- Mathematical Method- Regression Technique

NPTEL

Prof. Trupti Mishra, School of Management, IIT Bombay

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.

(Refer Slide Time: 04:12)



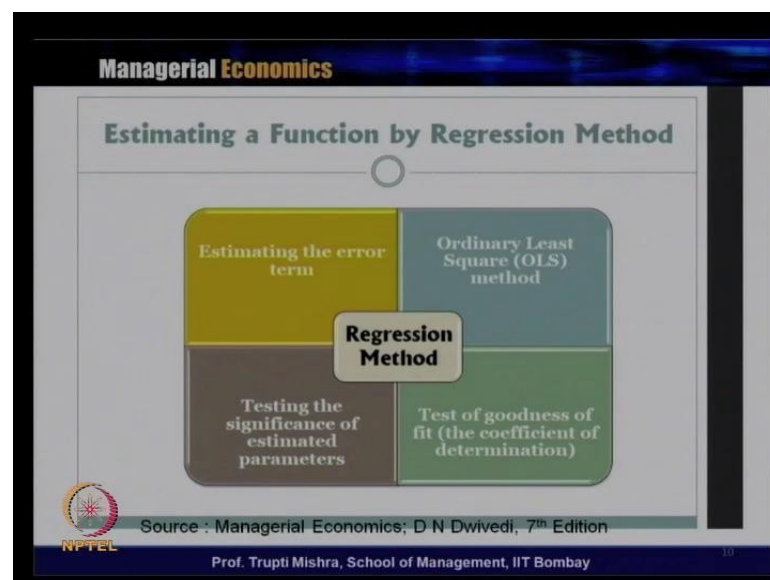
So, Y is equal to a plus b x is the functional form, or suppose we take a value Y is equal to value of a is 40,000, and this is 1800 x . So, here we can say a is 40,000 and b is 1800, and this is the slope, this is the intercept. So, if Y is the sales and x is the advertisement expenditure, when there 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 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 an approximation, because this is a crud method this is an 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.

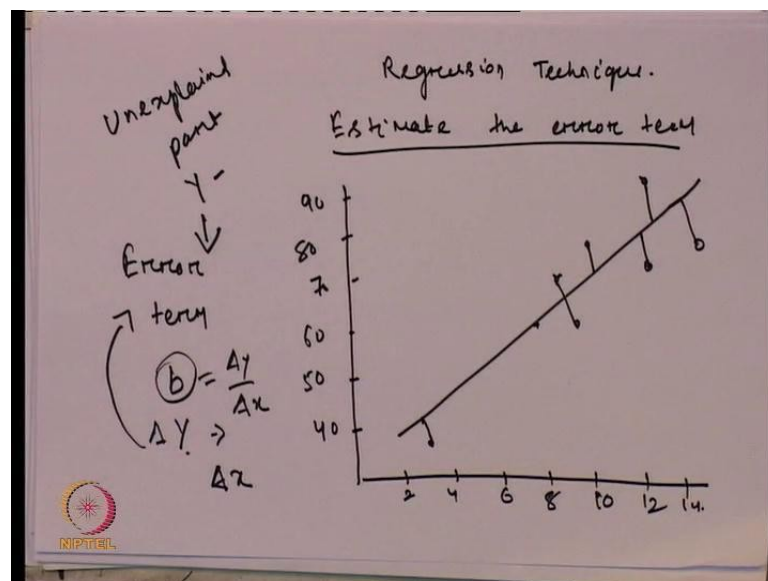
(Refer Slide Time: 07:43)



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 fit, 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 these 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 these estimated parameters.

And then we do the test of goodness of fit to understand the overall explanatory power of this model, or overall explanatory power of the relationship between these 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 fit.

(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 ΔY with respect to Δx . 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

NPTEL

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)

Managerial Economics

Types of Error

- Sampling error**- arising out of inaccurate recording or specifications of Sales data
- Specification error
- Measurement error

NPTTEL

Prof. Trupti Mishra, School of Management, IIT Bombay

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)

Managerial Economics

Types of Error

- Specification error** – arising because of other factors influencing sales could not be specified and included in the function.

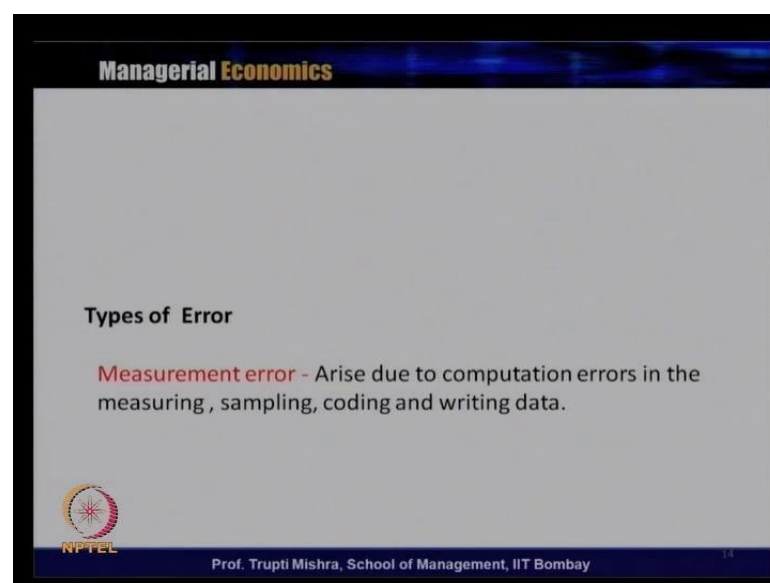
NPTTEL

Prof. Trupti Mishra, School of Management, IIT Bombay

13

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)

The image shows a whiteboard with handwritten mathematical definitions and equations. The text is as follows:

$$Y_t - \text{observe value.}$$
$$Y_c = \text{estimated value.}$$
$$e_t = Y_t - Y_c.$$
$$Y_c = a + bx - \text{best fit.}$$
$$Y_t = Y_c. \quad e_t = 0.$$
$$e_t = Y_t - Y_c.$$
$$[e_t = Y_t - (a + bx)]$$

In the bottom left corner of the whiteboard, there is a logo for NIPTEEL, which consists of a stylized sun or starburst icon above the text "NIPTEEL".

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 $a + bx$, 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 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_c is can be again called as the $a + bx$, and this is the functional form for the error term.