

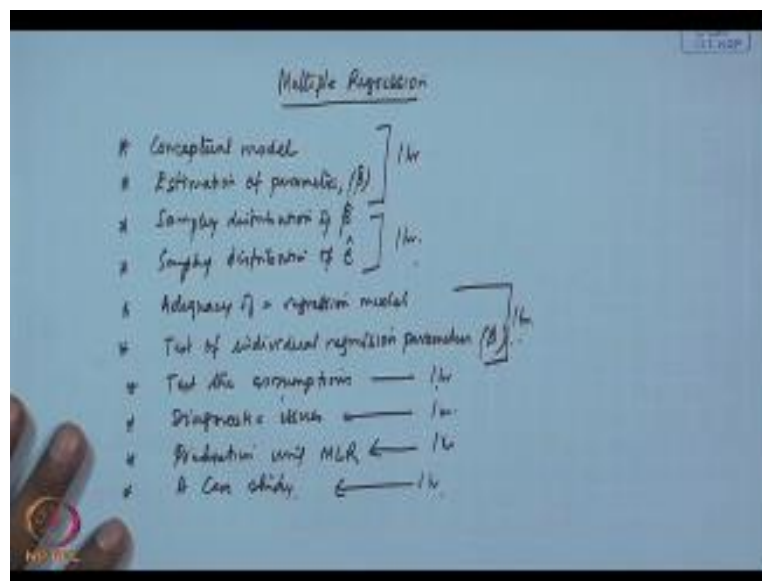
**INDIAN INSTITUTE  
OF  
TECHNOLOGY  
KARAGPUR  
NPTEL  
National Programme  
On  
Technology Enhanced Learning  
Applied multivariate Statistical Modelling  
Prof .J.Maiti  
Department of industrial engineering and management  
IIT karagpur**

**Lecture-21**

**TOPIC  
Multiple Regression-Introductions**

Good morning today we will discuss multiple regression.

(Refer Slide Time: 00:25)



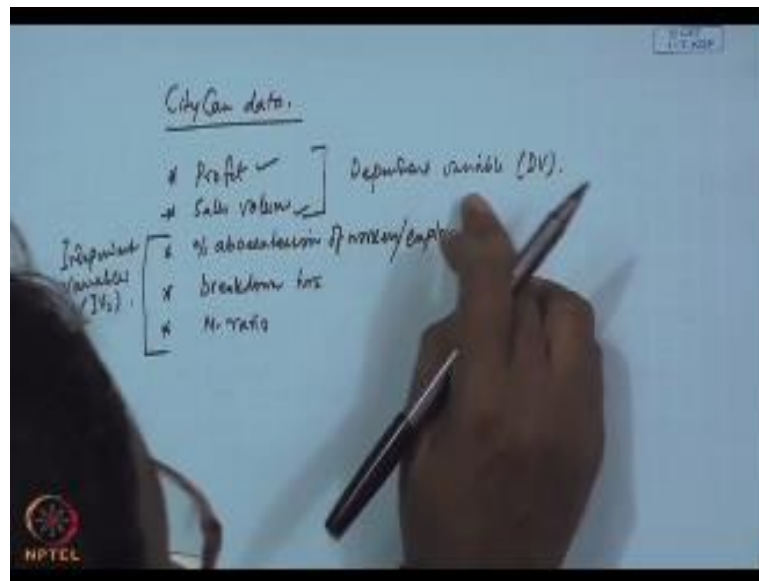
The contents of multiple regressions are first we will start with a conceptual model, then we will describe the estimation of parameters. Then we will go for sampling estimation of parameters in terms of  $\beta$  vector then sampling distribution of sampling distribution of  $\beta$  estimated. Then we will go for sampling distribution of error sampling distribution of error

estimated, and then we will go for that model fitting lesson that is adequacy of a model regression model.

Then we go for test of individual regression parameters, which is all  $\beta$  s. After that, we will test the assumptions related to regression, then there are certain diagnostic issues, then prediction using multiple linear regression. Finally, a case study will be showed, so this is in totality will be covered under multiple regressions. So, we will go by the first the model, then estimation of parameters sampling distribution. I think, today we will be able to cover this one, conceptual model and estimation of parameters.

I think this is for one hour then sampling distribution of that  $\beta$  estimates and sampling distribution of error estimate, I think one more hour is required there. Adequacy of regression and this test of individual parameter maybe one more hour, test of assumption one hour, then diagnostic issues one hour and prediction one hour, and case study one hour. So, 1,2, 3, 4, 5, 6 hours, 6 hours we will be discussing on multiple regression. Now, let us start with an example, I think in first class and also in subsequent lectures. I have given one example for the City can data.

(Refer Slide Time: 04:13)

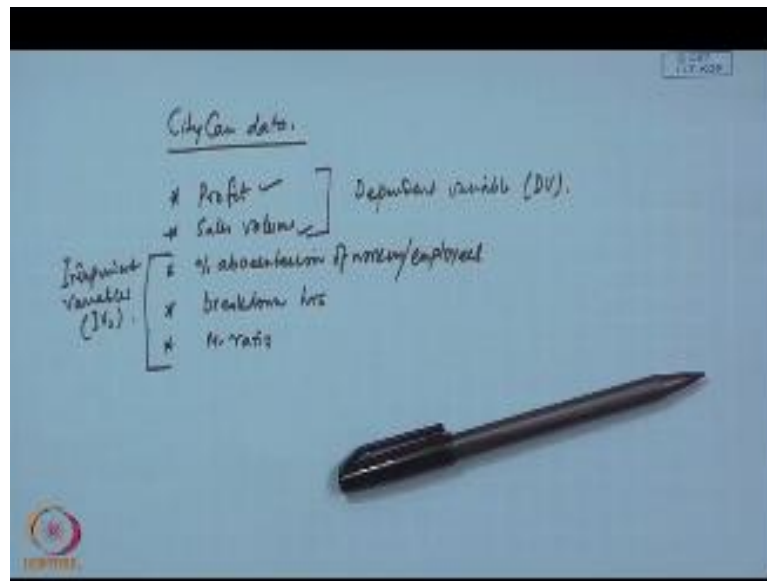


City can is a small company working in the local market and we have seen that the data structure was like this. Other way, we can start with the variables the important variables for this particular company profit, then sales volume, and then they have found out the absenteeism, percentage of absenteeism of workers or employees. Then, related to machine, that machine breakdown hours monthly, then they also have a marketing department whose performance is measured through M ratio.

The company's primary interest is how to improve the profit that definitely through maximizing the sales and apart from many other things. The company is interested to know how these three variables that absenteeism, breakdown hours; M ratio is affecting sales volume as well as affecting profit. Now this is the case then they under this situation, there are two types of variables one called dependent variables or dependent variable DV, another set is known as independent variable which is IV.

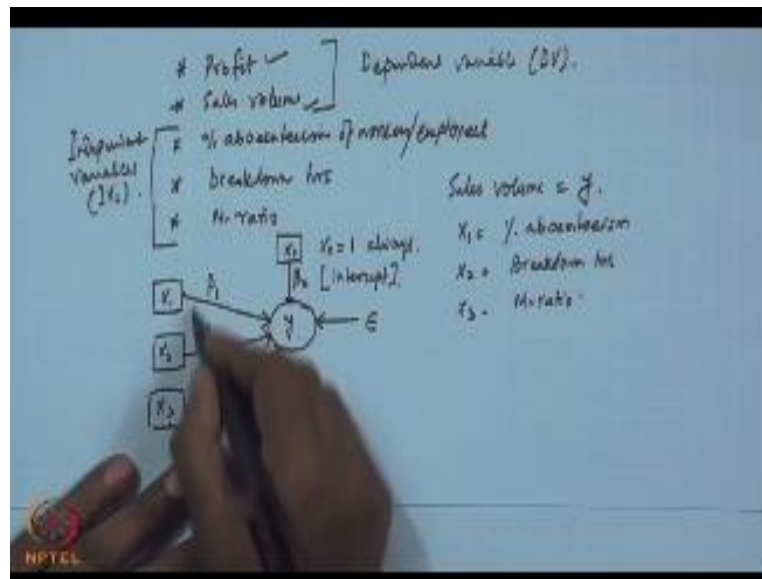
So, in this case profit is one dependent variable, sales volume is another dependent variable and there are three independent variables like absenteeism, breakdown hours and M ratio. Now, what is of interest in this particular case here, we want to test that whether the independent variables are contributing.

(Refer Slide Time: 06:34)



In explaining the dependent variable one at a time it is happening or not, if we can find out that.

(Refer Slide Time: 06:50)



Yes there is influence of the independent variables on the dependent variable. Then, depending on the influence measure some action can be taken to control the dependent variable with the help of controlling the independent variables. So, pictorially if we now consider that one dependent variable, let it be sales volume which is  $y$ , we are denoting like this. Other variables we are denoting like this  $X_1$  = to percentage absenteeism,  $X_2$  = to breakdown hours and  $X_3$  is M ratio.

So, we are not considering profit at present, later on we will see that how profit also can be included. So, under this situation, if I define like this that  $y$  within a circle or ellipse and  $X_1$  within a rectangle,  $X_2$  with another rectangle  $X_3$  with another rectangle. I can create some straight line curved with the arrow head to  $y$ , and then this will determine that the respected IVs are influencing the dependent variable that is the measure.

So, that means  $X_1$  influences  $y$  that is why the arrow head even it is from  $X_1$  and terminate at  $y$  that is the case. So, I can say that if  $X_1$  has influence of  $\beta_1$ ,  $X_2$  has influence of  $\beta_2$  and  $X_3$  has influence of  $\beta_3$ , what does it mean? It means that if I change one unit of  $X_1$ , there will be  $\beta_1$  unit of change in  $y$  if you change one unit of  $X_2$  there will be  $\beta_2$  unit changes in  $y$ . If you change one unit of  $X_3$  there will be  $\beta_3$  change in  $y$ . If you change one unit of  $X_1$ , there will be  $\beta_1$  unit change in  $y$ , similarly if you change one unit in  $X_2$  there

will be  $\beta_2$  units change in  $y$ , one unit change in  $X_3$  causes  $\beta_3$  units changes in  $y$  this is the meaning of this influencing parameters.

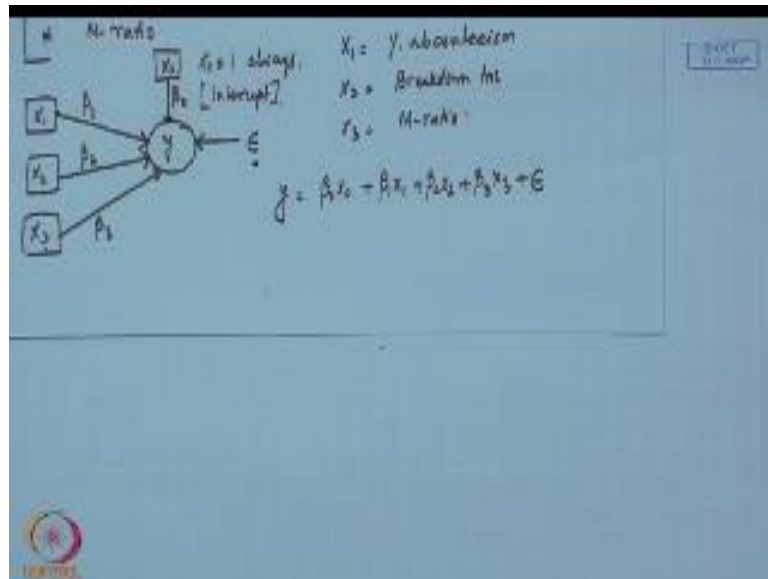
Now, what will happen, suppose none of them for example, absenteeism, break down and M ratio, none of them contributes to  $y$  will not be your sales. There will be sales also, it is not necessarily that they have to contribute always, so in that case we require some other parameter which is known as  $\beta_0$  that is coming through  $X_0$  where  $X_0$  takes value of 1 that is always  $X_0 = 1$ .

Always by saying this, what we are trying to say that even if there is no influence of  $X_1, X_2, X_3$  still there will be some amount of  $y$  values, this is the constant term  $\beta_0$  is added there. Now, if you take one observation of  $y$  followed by second, third like this what will happen? You will find out that you are not able to explain using  $\beta_1, \beta_2, \beta_3$  and  $\beta_0$ , and all these values not able to explain the total variability in  $y$ .

This means there are some other variables or some other, either they are controllable or uncontrollable variables are there which are also contributing towards the variability of  $y$ . So, as a result we require another measure which is known as error that mean, in multiple regression case you will get three types of I can say parameters one is  $\beta_0$  constant parameters which is known as intercept. Then, the influence of the independent variables, these are known as that influencing parameters or regression coefficients, also  $\beta_0$  is also regression coefficient, intercept coefficient.

Then, another one is a random one which is error which also contributes, so essentially then what will happen? If I want to put the entirety in terms of an equation, I can write like this  $y = \beta_0 X_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$ .

(Refer Slide Time: 12:32)



You can very well write this one, so that means this figure is the pictorial representation of multiple regression model which in mathematical term you can write using an equation. Now, as we have said that that  $X_0$  will take value of 1 always, so this can be written like this  $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$ .

(Refer Slide Time: 13:26)

$$X = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}$$

(p+1) x 1      (p+1) x 1

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \epsilon \leftarrow$$

Now, if we generalize this one that we are not interested only in three variables, we will go for a variable vector which is having what I say three variables let it be. So, I am writing like this  $X_1 X_2 \dots X_p$ , so there are  $p$  variables which are contributing towards  $y$  and as we have already seen, there is one intercept parameter and with related to one constant  $X_0$  that one we are making it 1. So, we can write if we want to create this  $X$  in terms of also you want to add  $X_0$ , then you can write this one  $p + 1$  into 1.

See this  $X_0$  is need not required to be written here, later on it will be taken care of essentially this  $X_1$  to  $X_p$  is the variables IVS  $X_0$  is a constant that intercept term will come for that. So, then if we go by  $X_1$  to  $X_p$ , this is  $p \times 1$  variable vector under this situation what will be the  $\beta$  because there will be influence for each of the variables. So, I can write  $\beta_1 \beta_2 \dots \beta_p$  which is  $p \times 1$  but see there is one intercept parameter intercept parameter related to  $X_0$  so that is  $\beta_0$ , so you require to estimate  $p + 1$  parameters regression coefficients here.

Now, I can write like this  $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon$ . So, this is in general the general regression equation this is the general regression equation you can write later on we will write it in terms of matrix equation. Now, let us assume that you are going to you are collecting data,



(Refer Slide Time: 16:16)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \epsilon$$

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \epsilon_i$$

variables

$$y_i = E(y_i | x_{i1}, x_{i2}, \dots, x_{ip}) + \epsilon_i$$

$$= \hat{y}_i + \epsilon_i \Rightarrow \epsilon_i = y_i - \hat{y}_i$$

So let  $n$  data points are collected or you can collect if you say to be collected this will be the all observations will be random. So, in this case our data matrix will be first one is the dependent case, so  $y$  it will be  $y_1, y_2, y_3$  like your  $y_n$  data points. Then, you are collecting  $X$  that will be  $n \times p$  is the number of observation,  $p$  is the number of variables, so we will use here  $X_{11}, X_{12}, X_{1p}, X_{21}, X_{22}, X_{2p}$ . So, like this  $X_{n1}, X_{n2}, \dots, X_{np}$ , so this is related to IVS, these two sets of data you will collect.

Now, if you use this data set into the equation given here, so for every data set or every data points this equation will become true. So, in order to incorporate that every data points into consideration if I write one  $y_i$  here, and then here it will be  $X_{i1}, X_{i2}$  like  $X_{ip}$ , this is what you are saying the general observation related to this. Now, if you want to write the equation here, the equation then what you will write, we will write  $y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_p X_{ip} + \epsilon_i$ .

Then, what is this quantity, this quantity you see it is a weighted linear combination of several variables here basically multiple random observations. So, can you not write-down, this is variate, this is weighted linear combination, and this is variate. Then, what does this variate represent? In this case, this one represent the expected value of  $y_i$  given  $X_i$  what is this  $X_i =$  to? Basically,  $X_{i1}, X_{i2}$  like this  $X_{ip}$  getting me without error term, this one is the expected value of  $y_i$ . If I know the  $i^{\text{th}}$  observation for the individual variable like this,

then what will be the expected, what will be the value of  $y_i$  that is the predicted value. As we do not know that what are this IVS for the  $i^{\text{th}}$  observations, so accordingly we also do not know what will be the  $y_i$  value, but by this variate we are saying that what will be the expected value of  $y_i$ . If you collect  $i^{\text{th}}$  observation  $X_{i1}$  to  $X_{ip}$ . And then that mean what is happening then  $y_i = \hat{y}_i + \epsilon_i$ . This expected value is nothing but the predicted value; suppose you want to use this later on this will be the better. So, I can write from here that  $\epsilon_i = y_i - \hat{y}_i$ .

So, essentially what are the things we have found out here, we have found out all the coefficients regression coefficients also the error term? We also described that that this linear combination of the variate and this variate will basically talk about what will be the value of expected value of  $y$  given these conditions.

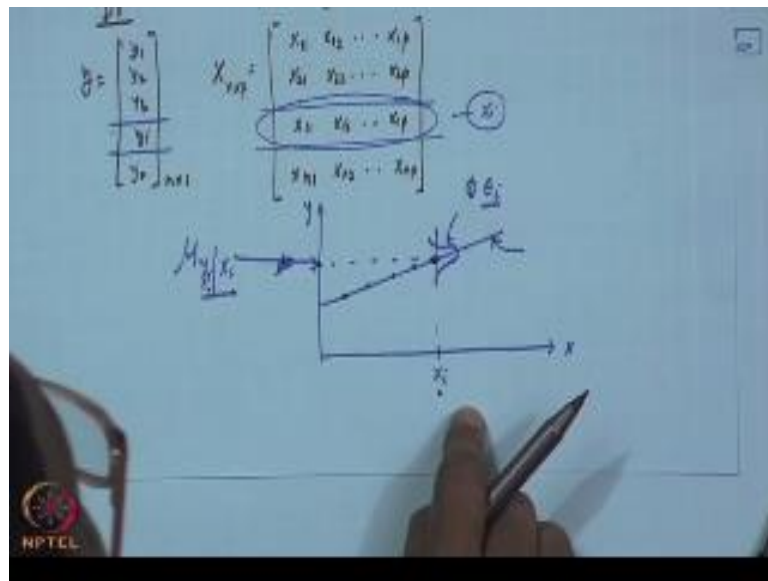
(Refer Slide Time: 21:28)

n data points are collected

$$\begin{array}{c}
 \text{DV} \\
 y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{bmatrix}
 \end{array}
 \quad
 \begin{array}{c}
 \text{IVs} \\
 X_{\text{exp}} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \dots & \vdots \\ x_{i1} & x_{i2} & \dots & x_{ip} \\ \vdots & \vdots & \dots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}
 \end{array}$$

So, these are conditional  $y$  we say we also write this one as  $\mu$  of  $y$  given  $X$  in general term, we will write  $\mu$  of  $y$  given  $X$ . If I say  $y_i$ , then given  $X_i$  where  $X_i$  is this is the observation.

(Refer Slide Time: 22:12)



So, if we further assume little bit this is because you just assume that this is one X and one y and you are fitting a linear line, suppose this is your  $X_i$ . So, what will be the regression line, the linear line then you go like this you will be meeting you are getting this point here. What is this value this value here? This is nothing but this value  $\mu$  mean value of y for the i the observation given i th observation.

Now, why we are talking about mean value, the reason is suppose you collect for the i th observation, if you collect one, now you will get some value here. If you collect two, the value may change, so ultimately there will be one error part. So, this one is taken care of by the concept error, so any point on this regression line which will be fitting later on. These points are the mean value of y given x, so we will see later on when we fit all those things I will tell you that how to make all those things. Now, we have collected n data points and we have also a regression equation that

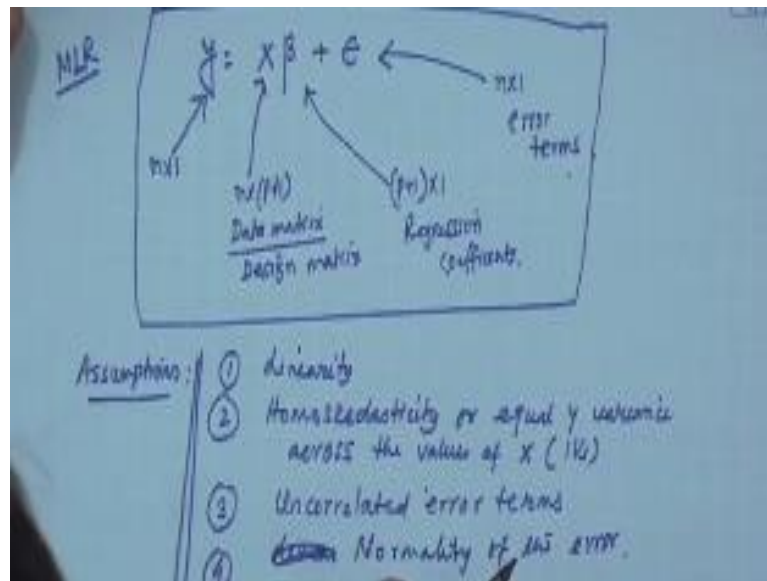
(Refer Slide Time: 23:48)

The image shows a whiteboard with handwritten mathematical equations. At the top, the general form of a linear regression equation is written:  $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \epsilon_i$ . Below this, a system of equations for multiple observations is shown, grouped by a large curly brace:  $y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \dots + \beta_p x_{1p} + \epsilon_1$ ,  $y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \dots + \beta_p x_{2p} + \epsilon_2$ , and  $y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \dots + \beta_p x_{np} + \epsilon_n$ . At the bottom, the matrix representation of these equations is shown:  $\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$ . Dimensions are indicated below the matrices:  $n \times 1$  for the response vector,  $n \times (p+1)$  for the design matrix, and  $(p+1) \times 1$  for the parameter vector.

$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \epsilon_i$ . If I write it for all the observations, then  $y_1$  will be  $\beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \dots + \beta_p x_{1p} + \epsilon_1$ . For  $y_2$ , you will be writing like this  $\beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \dots + \beta_p x_{2p} + \epsilon_2$ . If you go on writing, ultimately up to the  $n$ th term, you will be writing like this  $\beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \dots + \beta_p x_{np} + \epsilon_n$ . So, if I write in matrix form, now can you write  $y_1, y_2, \dots, y_n$ , for the left-hand side you can write this, so this  $n \times 1$  vector I can write this to. If here for  $\beta_0$  there is only one that is 1, so 1, 1, 1 for  $\beta_1 x$ , what is this 1, 1,  $x_{21}, x_{11}, x_{n1}$  for  $\beta_2 x_1, x_2, x_1, x_2, \dots, x_n$ .

Similarly, for  $\beta_p x_1, x_2, \dots, x_p$  then  $x_{i1}, x_{i2}, \dots, x_{ip}$ , this one is now what will happen, here there are  $n$  into  $p+1$  variables, one for the intercept. So, if I multiply it by  $\beta_0, \beta_1, \dots, \beta_p$  this will be  $p+1$  into 1, then the multiplication of this will give you  $n \times 1$ . It is giving you  $n \times 1$ , so  $n \times 1 = n \times 1$  which is basically this portion is taken care of by this + there is error terms  $\epsilon_1, \epsilon_2, \dots, \epsilon_n$  this is  $n \times 1$ .

(Refer Slide Time: 27:17)



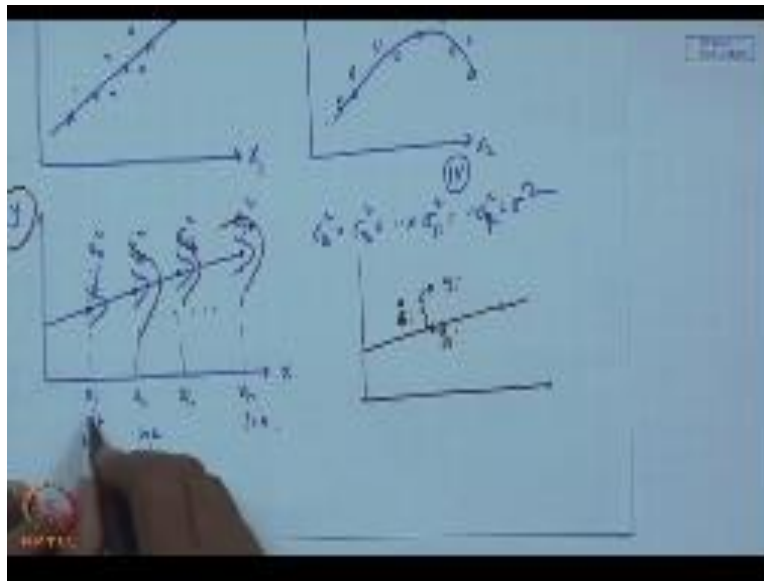
So, the resultant can be written like this  $y = X\beta + \epsilon$ , so where  $y$  is  $n \times 1$  vector of observations of the dependant variables,  $X$  this is  $n$  into  $p + 1$ . The data matrix of IVS including intercepts, this is also known as data matrix. Actually,  $X$  is this one including intercepts, so we say this one as design matrix,  $\beta$  is the  $p + 1 \times 1$  regression coefficients and  $\epsilon$  this is  $n \times 1$  error terms  $n \times 1$  error terms.

So, this is a nutshell equation for multiple linear regressions in matrix form. Now, we will see some of the assumptions of multiple regressions, first assumption is linearity, so that is linear relationships, second assumption is homoscedasticity or we say  $y$  variance across the values of  $x$ . Other way, you can say IVS given observations of IVS that homoscedasticity

I will show you how it will be done. Variance  $\propto y \propto X$   $y$  variability remain constant, then third one is we have seen that the error term is there. So, the error term will be uncorrelated error terms, which mean there are  $n \times 1$  errors. These vectors this will not be correlated, uncorrelated error terms and definitely linearity, sorry normality of the error terms normality of the error.

So, these four assumptions needs to be tested later on after fitting the model we will test this, but at present at this moment is should say something about what is this. So, first understand what linearity is,

(Refer Slide Time: 30:47)



You consider your relationships  $y$  versus  $x$ , for example it is  $X_1$ , and then when you plot the data if you plot the data. So, suppose you are getting this type of relationship when you are plotting with  $X_2$  you may be getting this type of relationship, so in the first case there is linearity in the second case it is not a linear one. So, if your data that  $y$  versus  $X$  that relationship is non-linear then this model is not applicable, what you require to do?

Then, you require converting this to linearity; you have to transform the data. Primarily, we want the transform IV first case from linearity point of view transform it. So, the data becomes linear the relationships become linear, now your relationships can be negative, positive no problem, but linearity is an issue. Second one is your homoscedasticity, I told you just few minutes back that suppose this is my  $X$  this is  $y$ , let it be because as we two dimensional 1, 1, 1,  $X$  your regression equation let it be like this.

We are assuming that one fitting is possible linear fitting is possible, so you have collected data point's  $n$  data points. Suppose, this is  $X_1$ , this one is  $X_2$  and then you will go for this one is  $X_i$ . So, similarly last one let it be  $X_n$ , so what I said that this observation, this is the first observation  $i = 1$  this is  $i = 2$ , so like this  $i = n$ . We are saying that you will collect one sample, but you may collect several samples also. Under such condition what we will have done for a particular value of  $x$ , you may get several values of  $y$ , so if you plot this several values for a particular value of  $x$ , then whatever variability of  $y$  you observe that must

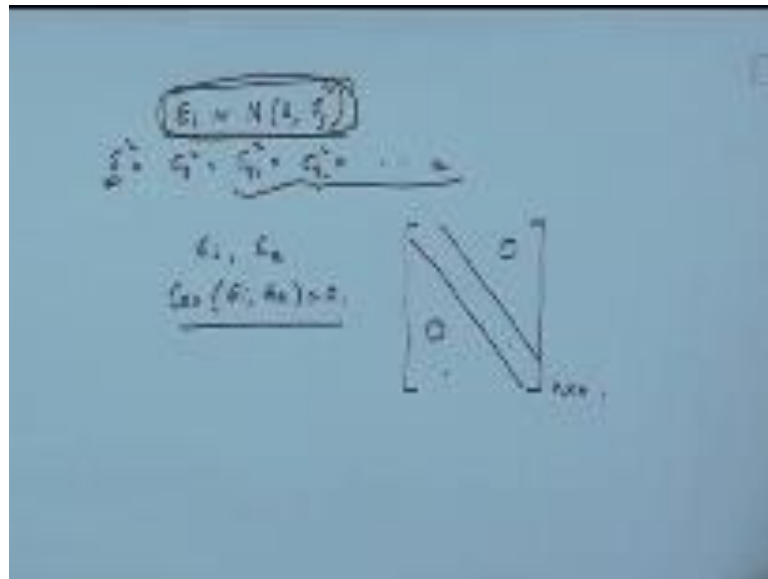
be  $\bar{y}$ . So, that mean we are saying this one if it is  $\sigma^2$ , suppose  $\bar{y}$  to 1, so  $y_1 \sigma$ ,  $y_2^2 \sigma$ ,  $y_3^2 \sigma$ ,  $y_n^2$ . I think this is  $i^2$ , so what is the condition  $\sigma y_1$  means at  $X = 1$ , this  $y_2 y_i y_n^2$ , this should be  $\sigma^2$ .

This is what is homogeneity from variance variability of  $y$  point of view, if there is violation of this, for example your plot may not be like this your plot something this here it is, but here may be it is like this here may be again small here may be big errors. So, across  $X$  values  $y$  is not homogeneous from variability point of view that assumption if violated you require to transform variables, but in that case you have to transform they variable.

If there is linearity problem, you transform the variables, you can go for both, but it is preferable that you transform the  $X$  variable. If heterogeneous nature, that heteroscedasticity is there, that mean not equal variance is satisfied, so in that case you have to transform the  $y$  variable. Then, third is uncorrelated error terms, actually you see this equation this figure, so when you use regression equation and you predict you will basically predict this points for  $X = 1$ , this value  $X = 2$  this value 3 i this value.

So, this value is the remaining portion is the error. Suppose, your case is like this your original value is here  $y_i$ , then based on this equation you are saying this is  $\hat{y}_i$ , so this portion is my error portion, this is the error. Similarly, everywhere you will be getting error, what will happen here it is  $\sigma \epsilon_1, \epsilon_2, \epsilon_i, \epsilon_n$  error uncorrelated error terms because each error is random, now

(Refer Slide Time: 36:44)



So the error will be will follow this type of distribution  $\epsilon$ . I will follow normal distribution with mean 0 and what will be the variance, variance will be  $\sigma^2$   $\sigma y^2$ , that means what is this one  $\sigma y^2$  we are writing it here  $\sigma^2 = \sigma y 1^2 = \sigma y 2^2$ . For all observation for equal variance, this condition is very important because we are considering these variability part will be taken care by the error.

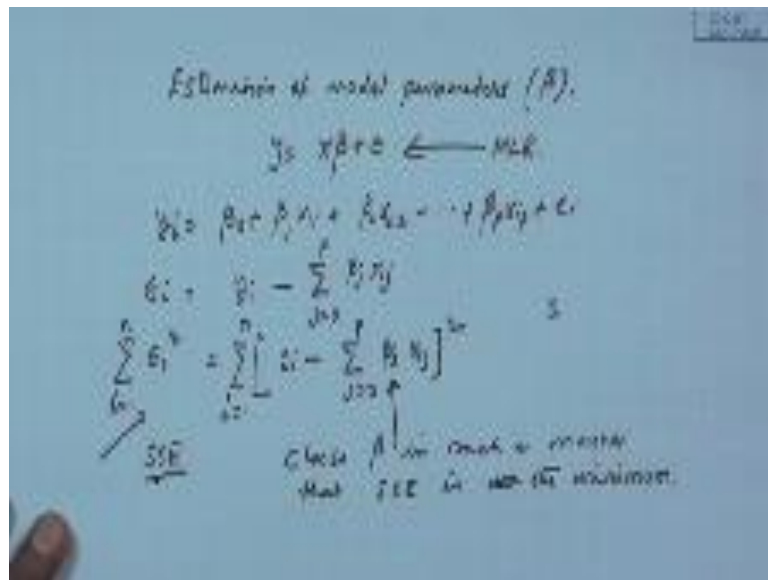
So, error will follow normality with mean 0 and variance  $\sigma^2$  where this  $\sigma^2$  is the basically the y variability for each of the observation of x. Now, what we are trying to say further that suppose that there are two error terms  $y_i$  and  $y_k$ , then the covariance between these two will be 0, which is uncorrelated error terms. So, as there are n errors, so if I see the covariance between the errors terms what will be the  $n \times n$  matrix we will be getting diagonal elements will be the variance off diagonal elements should be 0.

That is what uncorrelated error terms and normality of error is this error will follow normal distribution with  $\sigma^2$  variability mean 0. So, these are the assumptions we will test all those assumptions later on when we fit regression equation and it is required to be tested,



So once you are satisfied with the data that examining the data you find out that that assumptions are reasonably valid. Then, you will go for fitting that is estimation modelling or estimation of parameters.

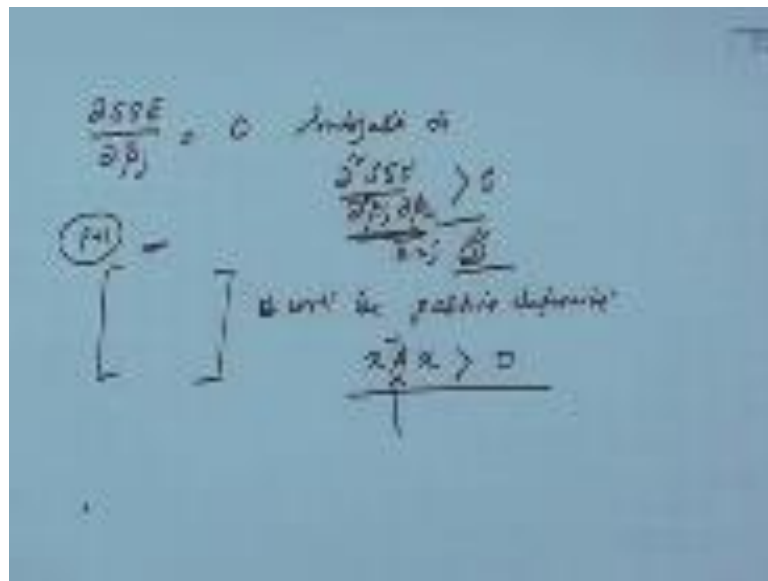
(Refer Slide Time: 39:03)



Estimation of parameters model parameters, that is  $\beta$  we are talking about how to estimate the  $\beta$  where  $y$  is  $X\beta + \epsilon$ , this is my MLR, this is my MLR, so if you recall that we say  $y_i$  is  $\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ij} + \epsilon_i$  and if you do little modification, here then this will be  $y_i - \sum_{j=0}^p \beta_j X_{ij}$  when  $j=0$ , this  $X_{ij}$  will takes the value of 1. So, if I make a square of it, then you will be getting  $(y_i - \sum_{j=0}^p \beta_j X_{ij})^2$ . So, this is for a particular observation, but we have  $n$  observation,

so if I take summation over  $n$  what is this quantity, total error sum of squared errors. So, this one is SSE, now we will choose  $\beta_j$  values in such a manner that SSE will be minimum, so your ultimate optimization is used here choose  $\beta$  in such a manner that SSE is the minimum. So, you can do very easily this one

(Refer Slide Time: 41:33)



So we can write SSE also in some other way also. What will happen you have to go for derivation SSE by  $\beta_j$ ? This you put equal to 0 subject to what subject to that your  $\frac{\partial^2 SSE}{\partial \beta_j \partial \beta_k}$ , this is greater than 0. This is greater because it is a minimum case, so we basically know what happen, were here I am saying that only for one variable case,  $\beta_j$ , here it is  $\beta_k$ . Also, we have taken that means the covariance part is two variables, this is coming into actually what is happening here there are so many variables p variables.

Sir here should we means  $\frac{\partial^2 SSE}{\partial \beta_j \partial \beta_k}$  or only  $\frac{\partial^2 SSE}{\partial \beta_j^2}$ , you are talking about when there is only one variable we say this by now if  $k = j$  it will take care of that issue. It is general form we have written here basically. It is basically hessian matrix. So, actually that I am trying to say as there is p +1 estimates, you are making, so you will begetting a matrix, so hessian matrix.

That matrix will positive definite will be positive definite, that is what I mean to say, that is why I have written like this. So, every component you will be calculated and a matrix will be found and that matrix will be positive definite by positive definite we mean that suppose A is a  $n \times n$  matrix. Now, you get any vector X, if you find out that  $X^T A X > 0$ , that is that means A is positive definite, so what the hessian matrix will come that must be positive definite then it is the minimum condition.

(Refer Slide Time: 44:02)

$$SSE = e^T e = (y - X\beta)^T (y - X\beta)$$

$$\frac{dSSE}{d\beta} = -2X^T(y - X\beta) = 0 \quad \text{SSCP}$$

$$-X^T y + X^T X \beta = 0$$

$$X^T X \beta = X^T y$$

$$\beta = (X^T X)^{-1} X^T y$$

Now, let us write in terms of matrix then we found that  $\epsilon$  is  $\epsilon_1 \ \epsilon_2 \ \dots \ \epsilon_n$  that is  $n \times 1$ . We are making SSE mean <sup>2</sup> sum <sup>2</sup> error can you not write like this, this is a vector, so  $1 \times n$ ,  $n \times 1$  this will give you  $1 \times 1$  a scalar quantity. If this is true, you can write like this  $y - X\beta$  transpose  $y - X\beta$ , you can write because our regression equation is this. So,  $\epsilon$  is  $y - X\beta$ , you can write like this, now if I take derivatives against  $\beta$ , then you will derive this one. It is nothing but the <sup>2</sup> term.

So ultimately what will happen, it will come like this  $2X$  transpose  $y - X\beta$ ,  $X$  transpose will come from here then this will remain this because we are deriving the derivatives for  $\beta$  any one will come from here. This is with matrix compatibility; we will get this equation, so this equation we will put to 0. I can write now that  $-X$  transpose  $y + X$  transpose  $X\beta = 0$ , so I can write  $X$  transpose  $X\beta = X$  transpose  $y$ .

If I multiplied both side by inverse of  $X$  transpose  $X$ , then what will happen, what I have done,  $X$  transpose  $X$  inverse I multiplied both side. Now,  $X$  transpose  $X$  inverse  $X$  transpose  $X$  in the covariance lecture I have given you that SSCP matrixes  $X$  transpose  $X$ . Can you remember that data of covariance type and I say that this is basically a square matrix symmetry matrix,

So the inverse times this will be identity matrix. I think as generally it is programmed that is why. That is why case, so this one a identity matrix, so I can say that  $\beta$  we are writing, now  $\beta$  cap because we have taken the X transpose X all this fixed values which are collected from the sample. So, this will become X transpose X inverse X transpose y, so this is your formula for estimating regression coefficients, I will show you one example for the estimation part. Let us solve this one problem,

(Refer Slide Time: 47:21)

$$y = \begin{bmatrix} 10 \\ 20 \\ 30 \\ 40 \\ 50 \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 5 & 7 & 10 & 12 & 20 \end{bmatrix}$$

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

$$X^T X = \begin{bmatrix} 5 & 54 \\ 54 & 149 \end{bmatrix}$$

$$(X^T X)^{-1} = \begin{bmatrix} 1/5 & -1/5 \\ -1/5 & 1/5 \end{bmatrix}$$

$$\hat{\beta} = \begin{bmatrix} 5 \\ 54 \end{bmatrix}$$

Suppose y is 10, 20, 30, 40, 50, this is my  $5 \times 1$  vector and my design matrix is 1, 1, 1, 1, 1 5, 7, 10, 12, 20. We are taking it is basically  $5 \times 2$ , so that mean y 1 variable independent variable, we have taken that is X and this one is your X 0 1 1 we have taken. So, what you require to calculate, you require to calculate  $\beta$  cap which is X transpose X inverse X transpose y.

So you first find out your step 1 is find out X transpose X, so that will be your 1, 1, 1, 1, 1 5, 7, 10, 12, 20. This is X transpose, then your again X transpose X only, so one 1, 1, 1, 1, then 5, 7, 10, 12, 20, now this matrix become  $2 \times 5$ . This is  $5 \times 2$ , the resultant will be  $2 \times 2$  matrix.

So, you will get this X transpose X, if you 1 into 1 like this, so there are data points, so you will be getting 5 first this into this, this into this like this. Now, this into this into this like this sum of this 5 + 7, 12, 22, 34, 54, other side also this  $\times$  this that will become 54, now 5 into

this  $25 + 49 + 100 + 144 + 400$ . So,  $9 + 14 + 418$  then  $4 + 48 + 2, 10 + 1, 1, 4, 4, 10$  and  $4 + 1 + 7$ , so 718. So, you second what you require you require computing inverse, so is say this is my step one.

(Refer Slide Time: 50:32)

$$\text{Step 2: } (X^T)^{-1} = \frac{1}{|X^T|} \text{adj}(X^T)$$

$$|X^T| = \begin{vmatrix} 5 & 54 \\ 20 & 718 \end{vmatrix} = 5 \cdot 718 - 54 \cdot 20 = 3590 - 1080 = 2510$$

$$\text{adj of } \begin{bmatrix} 5 & 54 \\ 20 & 718 \end{bmatrix} = \begin{bmatrix} 718 & -54 \\ -20 & 5 \end{bmatrix}$$

$$(X^T)^{-1} = \frac{1}{2510} \begin{bmatrix} 718 & -54 \\ -20 & 5 \end{bmatrix} = \begin{bmatrix} 0.286 & -0.0215 \\ -0.00797 & 0.00199 \end{bmatrix}$$

Now, step 2 step two I want X inverse X transpose X inverse, so this is my determinant of X transpose X that one by this one by this adjoin of X transpose X, so determinant of X transpose X equal to determinant of 554, 54, 718 which is  $5 \text{ into } 718 - 54^2$ . This will become 5 into 8, how much it is  $40 + 935$ - this quantity will become 16, 96, 26,16 and the resultant quantity will be 674. Now, ad joint of this matrix 554, 54, 718, this will become, so 718 will come - 54- 54, 5, this will be like this. So, your X transpose, X inverse then will become 1by 674 into 718 - 54 - 54, 5.

This will become the quantity will become something like this, if we go by this by this will be 1.07, - 0.08, - 0.08 then 0.007. If you go little, if you want to make it then 0, 1 something will come for digit case okay. Then, what you require to know, now you require knowing X transpose y that is another thing you have to compute.

(Refer Slide Time: 53:04)

Step 3: Compute  $X^T y$

$$X^T y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 5 & 7 & 10 & 12 & 20 \end{bmatrix} \begin{bmatrix} 10 \\ 20 \\ 30 \\ 40 \\ 50 \end{bmatrix} = \begin{bmatrix} 50 \\ 1970 \end{bmatrix}$$

Step 4:  $\hat{\beta} = (X^T X)^{-1} X^T y$

Then, your step 3 is compute X transpose y, so this one can be written like this X transpose y is 1, 1, 1, 1, 1 5, 7, 10, 12, 20, this is my  $2 \times 5$  into  $5 \times 1$  the 10, 20, 30, 40 and 50, so this is  $5 \times 1$ . So, it will be  $2 \times 2 \times 1$ , resultant will be  $2 \times 1$ , so that n all will be summed here 10, 30, 60, 150. Finally, then this is  $50 + 140 + 300 + 480 + 1,000$ .

I think this quantity will become 0, this side, this  $5 + 4, 9 + 8, 7$  then  $4 + 3 + 1, 8 + 19$ , this is 1970. So, your step 4 is  $\hat{\beta} = (X^T X)^{-1} X^T y$  what is our X transpose X inverse, we found out we found out X transpose X inverse. I think X transpose X inverse is  $1.07 - 0.0801, 0.0801, 0.0007$  into X transpose y is 150, 1970.

(Refer Slide Time: 55:08)

Handwritten mathematical derivation on a blue background:

Step 4:  $\hat{\beta} = (X^T X)^{-1} X^T y$

$$= \begin{bmatrix} 107 & -0.0021 \\ -0.0021 & 0.107 \end{bmatrix} \begin{bmatrix} 150 \\ 198 \end{bmatrix}$$
$$= \begin{bmatrix} 2.70 \\ 1.775 \end{bmatrix}_{2 \times 1}$$

Calculations shown to the right of the matrix multiplication:

$$160.50 - 157.80 = 2.70$$
$$-12.015 - 13.79 = -25.805$$

So, this quantity will be this is  $2 \times 2$ ,  $2 \times 1$ , it will be  $2 \times 1$ , this into this minus this into this the quantity will become like this  $160.50 - 157.80$  and it will become  $2.70$ . Similarly, other one will become  $-12.015 - 13.79$ , resultant quantity will be  $1.775$ , and this is the case.

(Refer Slide Time: 55:51)

The image shows handwritten mathematical work on a blue background. At the top, the regression equation is written as  $y = \beta_0 + \beta_1 X_1 + \epsilon$ , which is then simplified to  $y = 2.70 + 1.775 X_1 + \epsilon$ . To the right, a vector  $X$  is shown as  $X = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 5 \\ 7 \\ 10 \\ 12 \\ 20 \end{bmatrix}$ . Below this, the predicted values  $\hat{y}$  are calculated as  $\hat{y} = 2.70 + 1.775 X_1$ . The error term  $\epsilon$  is then calculated as  $\epsilon = y - \hat{y}$ , with a list of values:  $\begin{bmatrix} 10 \\ 20 \\ 30 \\ 40 \\ 50 \end{bmatrix}$ . To the right of this, a column of calculations shows the predicted values for each  $X_1$  value:  $2.70 + 1.775 \times 1$ ,  $2.70 + 1.775 \times 7$ ,  $2.70 + 1.775 \times 10$ ,  $2.70 + 1.775 \times 12$ , and  $2.70 + 1.775 \times 20$ . A small logo is visible in the bottom left corner of the slide.

So, my regression equation is now  $y = \beta_0 + \beta_1 X_1 + \epsilon$  to  $2.70 + 1.775 X_1 + \epsilon$ , this is your regression equation your  $y$  cap will be  $1.70 + 1.775 X_1$  this is  $y$  cap. If you want to find the  $\epsilon$  cap the error term that is  $y - y$  cap which is  $y - y$  cap, we have to find out. Now, this one you have  $y$  values, you have  $y$  values like 10, 20, 30, 40, 50 and you require to find out  $y$  cap.

How do you find out you have  $X$  values, also  $X$  values are 1, 1, 1, 1, 1, 5, 7, 10, 12, 20, so this  $2.70 + 1.775$  what is  $X_1$  value  $5 \times 2.70 + 1.775$ .  $X_2$  value  $7 \times 2.70 + 1.775$ ,  $X_3$  value  $10 \times 2.70 + 1.775$ ,  $X_4$  value  $12 \times 2.70 + 1.775$  into 20. This will give you your error values it will be  $5 \times 1$  okay, I think this is what estimation we have calculated very simple problem only one variable at a time.



(Refer Slide Time: 58:03)

$y = f(x) = \beta_0 + \beta_1 x_1 + \epsilon$   
 $y = 2.70 + 1.775x_1 + \epsilon$   
 $A = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{bmatrix}$   
 Simple regression:  $\hat{y} = 2.70 + 1.775x_1$   
 $\epsilon = y - \hat{y} = \begin{bmatrix} 10 \\ 26 \\ 30 \\ 40 \\ 45 \end{bmatrix} - \begin{bmatrix} 2.70 + 1.775 \times 1 \\ 2.70 + 1.775 \times 2 \\ 2.70 + 1.775 \times 3 \\ 2.70 + 1.775 \times 4 \\ 2.70 + 1.775 \times 5 \end{bmatrix}$   
 $p+1 = 3$   
 Multiple regression  
 $y \rightarrow$  single DV

So, when you take  $y$  is function off  $x$  in the linear mode like  $\beta_0 + \beta_1 X + \epsilon$ ,  $\epsilon$  only one variable, this equation is known as simple regression, simple linear regression when that mean  $p = 1$  case  $p = 1$ . Now, if it is  $p$  is number of parameter to be estimated, then it is two intercept like this if I say  $p$  is the number of variable case, this is only 1.

So, when it is  $p \geq 3$  or more number of parameters to be estimated, other way I can write  $p + 1$  should be 2, then it is simple regression when  $p + 1 \geq 3$  that will be multiple regression. All those case,  $y$  is a single DV  $y$  is our single DV only one DV at a time we are considering. So, next class we will see the sampling distribution of  $\beta$  the  $\beta$  you have estimated this  $\hat{\beta}$  I am saying sampling distribution of  $\hat{\beta}$ ,

(Refer Slide Time: 59:26)

The image shows a handwritten derivation of a linear regression model. At the top, the general equation is given as  $y = \beta_0 + \beta_1 x + \epsilon$ . Below it, a specific example is shown:  $y = 2.70 + 1.775x + \epsilon$ . The matrix form is also shown:  $y = X\beta + \epsilon$ . The design matrix  $X$  is a column vector with values 1, 2, 10, 12, 20. The response vector  $y$  is a column vector with values 10, 20, 30, 40, 25. The regression coefficients  $\beta$  are shown as a column vector with values 2.70 and 1.775. The error term  $\epsilon$  is shown as a column vector with values 10, 20, 30, 40, 25. The expected value of the beta cap is calculated as  $E(\beta \text{ cap}) = \beta$ . A diagram shows the sampling distribution of  $\beta \text{ cap}$  as a normal distribution centered at  $\beta$ . The diagram also labels  $\beta$  as a constant and  $\beta \text{ cap}$  as a random variable.

That estimate basically, so this  $\beta$  cap you have estimated using a sample. If you go for several samples, the  $\beta$  cap value will change and it will become random. It is a random variable what about  $\beta$  is it a random variable the  $\beta$  the regression coefficient. From population point of view, you have  $y = X\beta + \epsilon$  which we are talking about it is our regression line for the population  $\beta$  is the population parameter.

So, that constant and they are unknown SI that is why you are obtaining this  $\beta$  cap which is estimate of  $\beta$  and expected value of  $\beta$  cap will be  $\beta$ . This is the unbiased estimation, so  $\beta$  is constant and random constant and unknown constant and unknown and  $\beta$  cap is random variable, but it is known. Now, when you are collecting data you are getting this value, so we will go for sampling distribution of  $\beta$  cap next class.

**NPTEL Video Recording Team**

**NPTEL Web Editing Team**

**Technical Superintendents**

**Computer Technicians**

**A IIT Karagpur Production**

[www.nptel.iitm.ac.in](http://www.nptel.iitm.ac.in)

**Copyrights reserved**