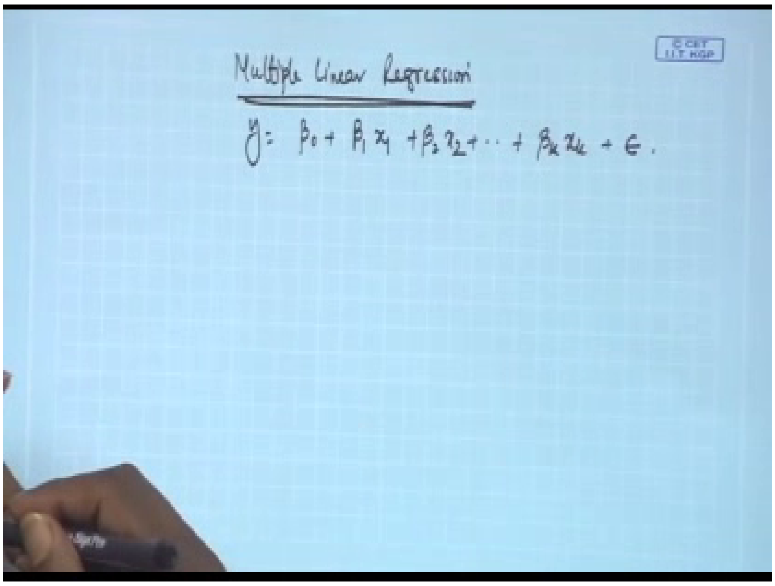
Design and Analysis of Experiments
Prof. Jhareswar Maiti
Department of Industrial and Systems Engineering
Indian Institute of Technology, Kharagpur

Lecture – 20
Introduction to Multiple Linear Regression (MLR)

(Refer Slide Time: 00:25)

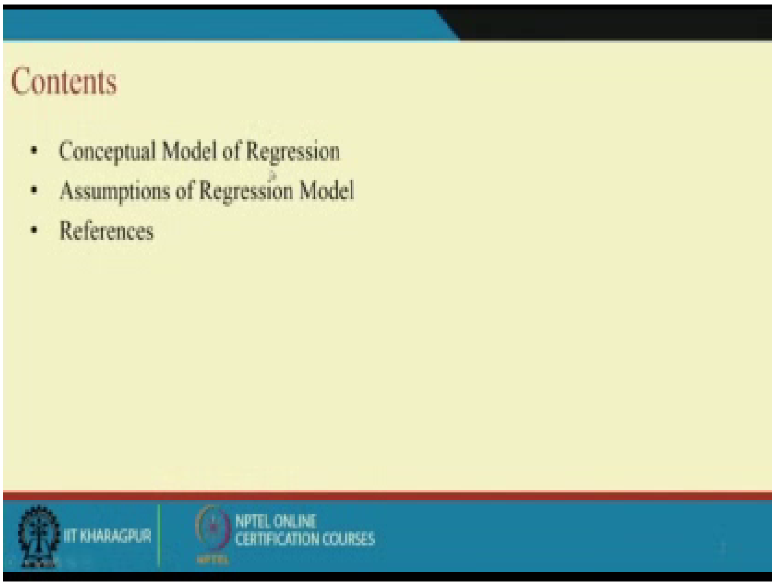


Multiple Linear Regression

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon.$$

Hello, we will discuss now multiple linear regressions. So, the presentation will be like this, this is a very basic lecture on multiple linear regressions.

(Refer Slide Time: 00:35)

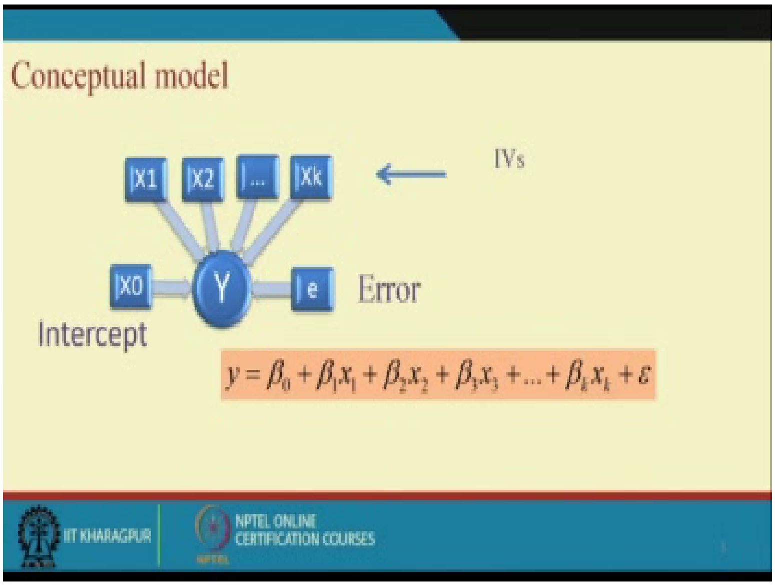


Contents

- Conceptual Model of Regression
- Assumptions of Regression Model
- References

So, as a result we will give you a conceptual model, when, what is a regression model and what are the assumptions of regression model. And if possible I will give you the equation for parameter estimation which will be discussed in next lecture also.

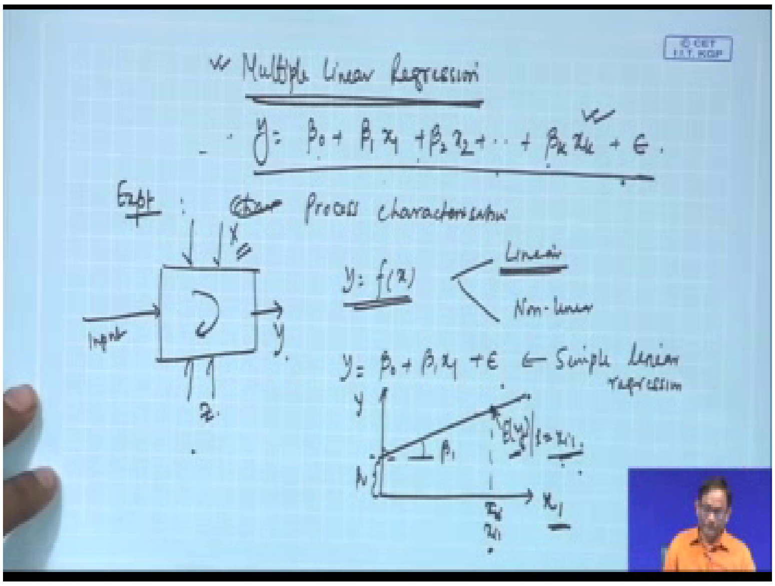
(Refer Slide Time: 00:57)



Conceptual model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_k x_k + \varepsilon$$

(Refer Slide Time: 01:00)



Multiple Linear Regression

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon.$$

Expt : Process characterisation

$$y = f(x)$$

$$y = \beta_0 + \beta_1 x_1 + \epsilon$$ ← Simple linear regression

$$E(y)|_{x = x_1}$$

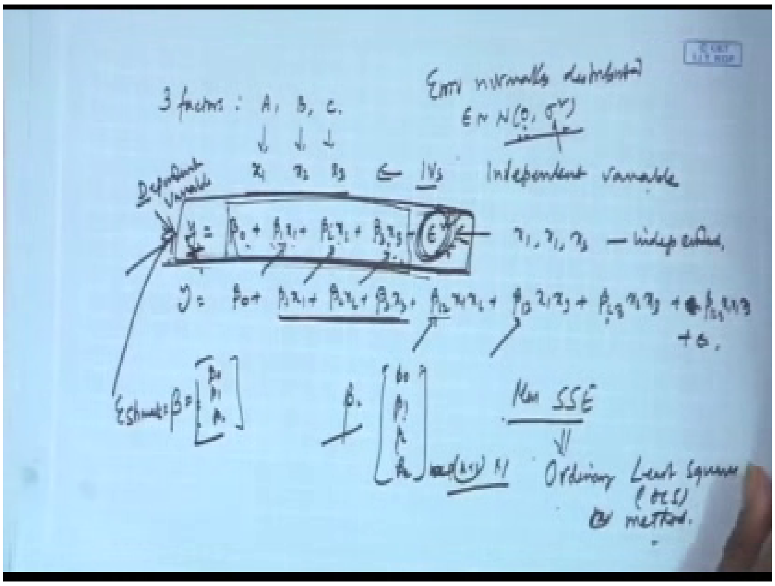
So, see this one. Suppose you just think of a experiment, where you are interested to characterize the process, or I can say process characterization is the issue. If you recall my that earlier lectures, you know that I have given a process model with controllable factors and controllable factors, and with y the output input controllable x may be uncontrollable level z and something like this

If you find out, suppose we want to characterize the process with reference to control level factor, then y is a function of x. So, that is what you want to find out. Now that is then y linear equation will be y beta 0 plus beta 1 x 1 plus epsilon. So, this is known as simple linear regression. So, it will be a regression line like this, this side your x 1, this side is y and this will be your beta 0, and the slope represent beta 1.

And this is your, this any point on this line is y that expected value of y given x equal to. Let it be x i x 1 i this one. Let it be or i 1 x i 1, then you did x i 1 y 1 1 for the x 1 variable i is the i th term generation. So, what is the value of point here or any point on this regression line, what it will depict? It will depict that y that the mean value of y with respect to x is equal to x 1.

Now, if there are more than one factor, then your equation will be like this. Suppose there are k numbers of factors. So, factors mean k number of. I am talking about controllable factors here. We are not considering z at this moment, let it be. We can improve z here also, but for simplicity I am not doing, because our aim is to see multiple linear regressions. So, then this one, this equation represent multiple linear regression. In multiple linear regression this one, suppose we have three factors A B and C.

(Refer Slide Time: 03:51)


3 factors : A, B, C.
x_1 x_2 x_3 ← IVs
Independent variable
Error normally distributed
$\epsilon \sim N(0, \sigma^2)$
Dependent variable
$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$
x_1, x_2, x_3 — independent
$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \beta_{123} x_1 x_2 x_3 + \epsilon$
Estimate $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_k \end{bmatrix}$
$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_k \end{bmatrix}_{(k+1) \times 1}$
Min SSE
Ordinary Least Squares (OLS) method.

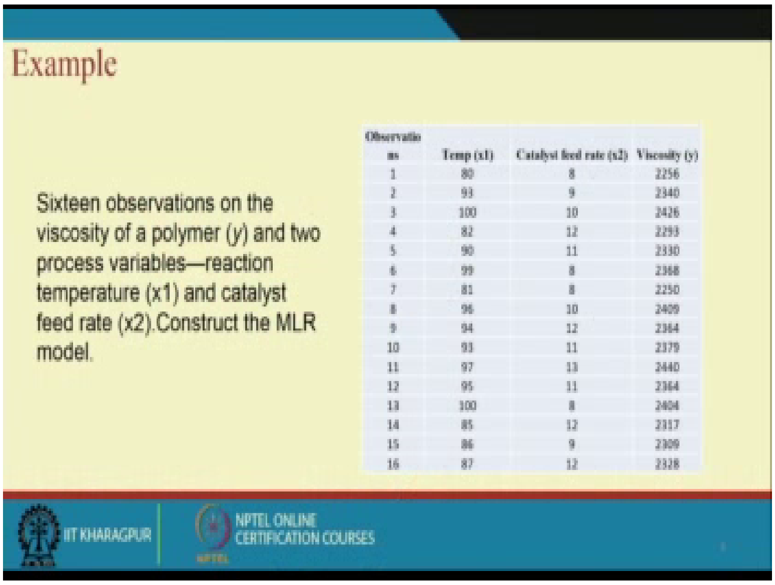
Let they are measured with continuous scale and we are saying x 1 x 2 x 3, and our quality variable is y.

So, I can write linear regression like this; beta 0 plus beta 1 x 1 plus beta 2 x 2 plus beta 3 x 3 plus epsilon. So, this is also a miller, but here what happened, this x 1 x 2 x 3 independent, if there is some amount of dependency between these x. So, we can also write something like this beta 0 plus beta 1 x 1 plus beta 2 x 2 plus beta 3 x 3 plus beta 1 2 x 1 x 2 plus beta 1 3 x 1 x 3 plus beta 2 3 x 2 x 3 plus epsilon. Here you see there are parameters related to the factors or the variables, and this, basically this variable x which are independent variables I Vs and y is known as dependent variables, dependent variable and independent and all x are independent variables.

So, if we assume that the independent variable are correlated. So, then we are creating some other coefficient, which takes care of their correlation part like this. So, you are getting additional variable, variable with the existing variable. So, this is the, this is some kind of. In first class I say that this is some kind of mean effects kind of things, and later on we will see these are beta 1 beta 2 are the main effects kind of things. Beta 1 these are interaction effects kind of thing. So, it may. So, happen that there are a three way interaction then beta 1 2 3, x 1, x 2, x 3 plus epsilon.

So, and then if I think that there is the non-linear part quadratic effect is there. So, slowly the non-linear part also can be added with this kind of regression equation for the time being. We will be discussing with this kind of relation or with this ok.

(Refer Slide Time: 06:55)



Example

Sixteen observations on the viscosity of a polymer (*y*) and two process variables—reaction temperature (x1) and catalyst feed rate (x2).Construct the MLR model.

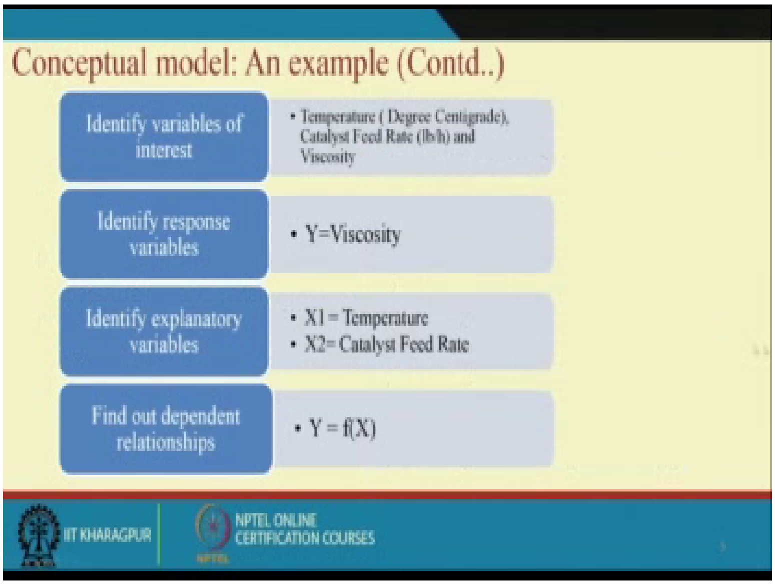
Observations	Temp (x1)	Catalyst feed rate (x2)	Viscosity (y)
1	80	8	2256
2	93	9	2340
3	100	10	2426
4	82	12	2293
5	90	11	2330
6	99	8	2368
7	81	8	2250
8	96	10	2409
9	94	12	2364
10	93	11	2379
11	97	13	2440
12	95	11	2364
13	100	8	2404
14	85	12	2317
15	86	9	2309
16	87	12	2328

So, now let us see one example here. Suppose the viscosity of polymer a job importance, whose behavior we wanted to assess and we also want to characterize these with reference to two process variables reaction; temperature x 1 and catalyst feed rate x 2.

And suppose we assume that they are linearly related. It may be show that they are not linearly related, but we are assuming that linearly related, and another one is that. Let we are assuming that they are independent x 1 x 2 are independent in nature, and there those things can be tested once the experiment is done over and the data is collected, that can be tested. Also suppose x 3 experiment is conducted, and you have 16 observations and with different temperature, and different catalyst rate your viscosity values are like this.

So, you see that in y there are variability of y, y values are changing and this may be, because of change in temperature and change in catalyst rate, and if they are linearly related then we can develop a multiple linear regression. So, the, what is the usual procedure.

(Refer Slide Time: 08:20)



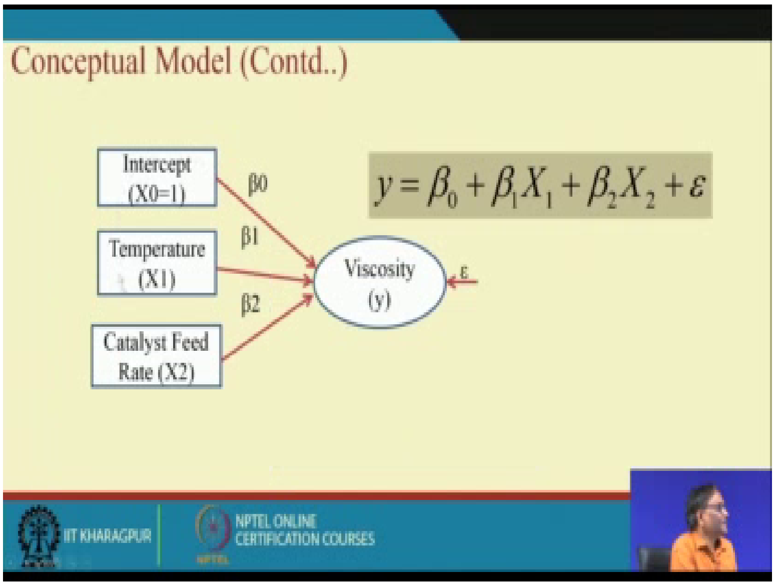
Conceptual model: An example (Contd..)

Identify variables of interest	• Temperature (Degree Centigrade), Catalyst Feed Rate (lb/h) and Viscosity
Identify response variables	• Y=Viscosity
Identify explanatory variables	• X1 = Temperature • X2= Catalyst Feed Rate
Find out dependent relationships	• Y = f(X)

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

Procedure is like this. Find out the interesting variables or the idle variable of interest. In this particular example, temperature catalyst feed rate and viscosity. Then finally, identify the response variable. It is the viscosity explanatory variable or independent variable some time. We said this is temperature in this, and our characterization relationship Y equal to function of X.

(Refer Slide Time: 08:48)



Conceptual Model (Contd..)

Intercept (X0=1)

Temperature (X1)

Catalyst Feed Rate (X2)

β0

β1

β2

Viscosity (y)

ε

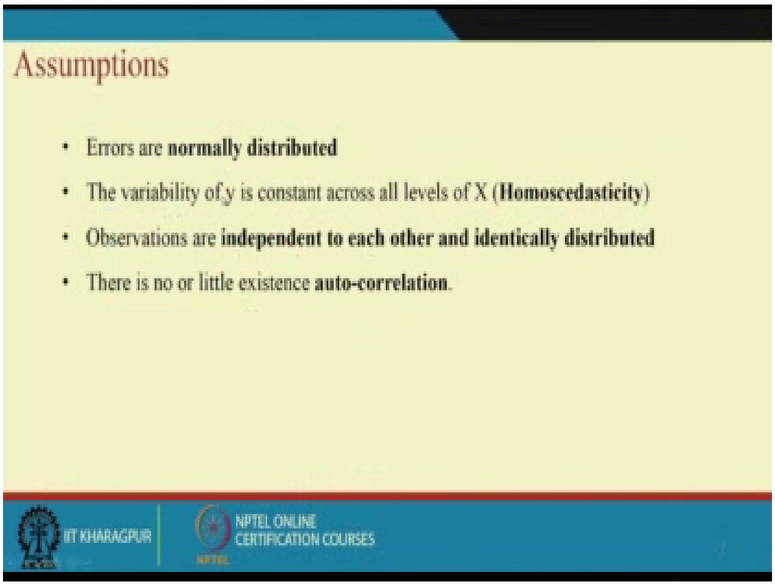
$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

So, then this model, linear model, multiple linear model will be like this beta 0, beta 1 x 1 plus, beta 2 x 2 plus epsilon. So, what is the job here? Our job is to estimate the beta values. What is our job? Suppose we consider this equation, our job is estimate beta which is basically a vector beta 0, beta 1 and beta 2 with respect to this equation, with respect to general k variable beta will be beta 0 beta 1 beta 2 like beta k k cross k plus 1 cross 1 vector.

So, these will be estimated. We will see later on that. We basically minimize the error, a major this error will be minimized with respect to all the observations considered. So, that is why we will consider sum square error and minimize sum square error. This will give you this value, using ordinary least square ordinary least square ols regression ordinary least square method. So, there can be mle, mle maximum likelihood estimation, but in this case multiple linear both will give you the same final formula.

(Refer Slide Time: 10:24)



Assumptions

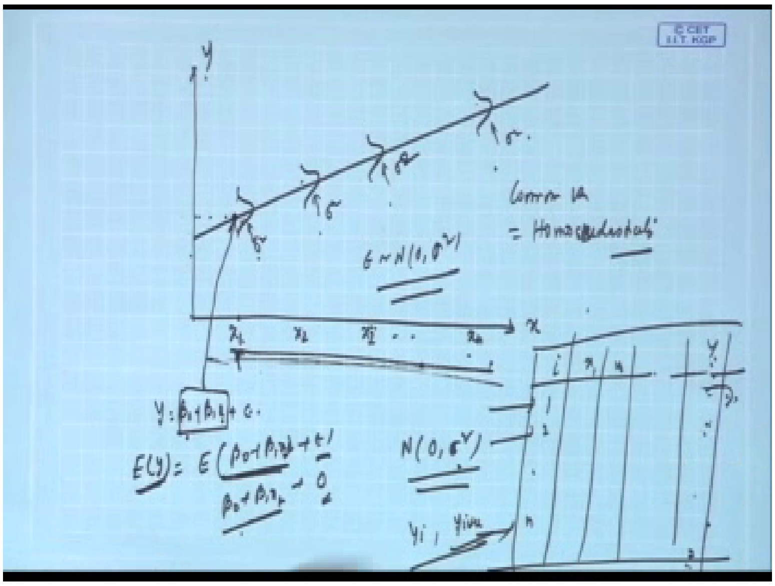
- Errors are **normally distributed**
- The variability of y is constant across all levels of X (**Homoscedasticity**)
- Observations are **independent to each other and identically distributed**
- There is no or little existence **auto-correlation**.

The assumptions are errors are normally distributed the variability of y is constant across all level of X, which is had homoscedasticity observations are independent to each other, and identically distributed. There is no or little existence of autocorrelation if you see the linear equation here. So, you see y is function of this, and here importantly these X values.

They are fixed values beta 0 beta 1 beta 2 they are constant values. So, this portion, this portion is a fixed portion pattern portion so, but y is variable. So, that this variability part the randomness part will be captured by error. So, other way in the first one, when you are saying that the samples drawn from normally distributed population means y is normal.

So, it indicates that the error is also normal. So, error is normally distributed, errors normally distributed. So, if I say epsilon, this is normal distribution with mu and sigma square. Sigma square will be variability mu is 0. So, it is 0 sigma square, whatever variability of y is there; that is going to error.

(Refer Slide Time: 11:58)


© CET I.I.T. KGP
y
x
x_1 x_2 x_i . . x_n
σ
Constant = Homoscedastic
$\epsilon \sim N(0, \sigma^2)$
$y = \beta_0 + \beta_1 x + \epsilon$
$E(y) = E(\beta_0 + \beta_1 x + \epsilon)$
$\beta_0 + \beta_1 x + 0$
$N(0, \sigma^2)$
y_i, y_{in}
i
1
2
n
y

So, with reference to simple linear regression, when one variable pictorially can be represented like, suppose this is x, this one is y, and this is my regression line. Suppose this is the, this is my x 1 and this value is x 2, suppose this value is x i. Similarly x n values are there

Then what happened this, as I told you that y is beta 0 plus beta 1 x plus epsilon. Here this is the fixed part, this for any x value. When you find the value on this line, this is basically this force, this part, this value is this. So, now, if I want to find out the expected value of y, then this is the expected value of this equation beta 0 plus beta 1 x plus plus epsilon. So, ultimately this portion for a fixed x value if it is a fixed value, this is beta 0 plus beta 1 x plus. What is the expected value of epsilon? It is 0.

So, that is why I say we said that, but this one gives you the expected value of y given x equal to x 1 in this case, or x is equal to x i. So, now, the 0 that is why this 0 yg, this the first part. Second part is that variability where it will go, it will be variability will be represented like this. So, everywhere for a fixed value of x you can get different value of y, and that variability is like this. So, x i this variability is like this x in

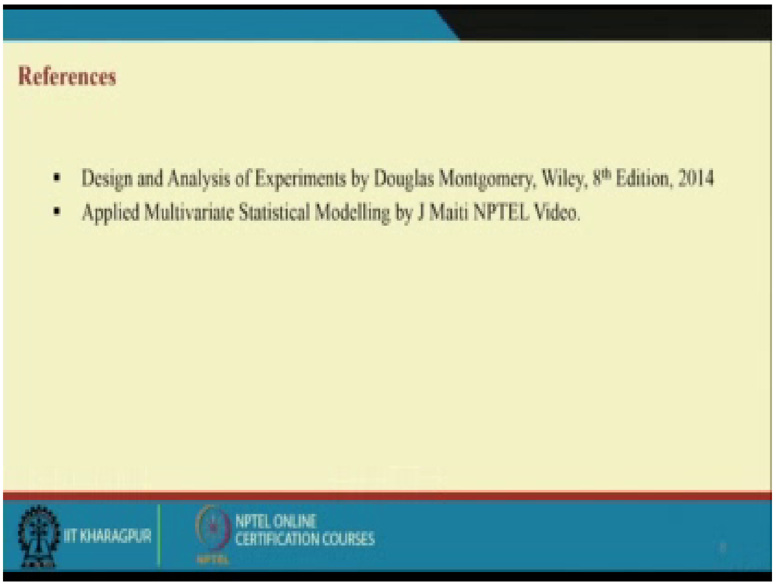
So, that is what this is basically variability sigma square, sigma square, sigma square, sigma square. So, we are saying, then the first assumption is errors are normally distributed with mean mu equal to 0, and sigma square is the variance. This sigma square is the variability of the observed y variable. Now another second one is there that common variance for all population; that is this one common variance. What do I mean? We are mean common variance means, here first one is errors are normally distributed second one is common variance.

So, for x 1 the y variability x 2 y variability, when x equal to x i y variability they will be same. So, everywhere it will be sigma square; that is known as homoscedasticity. So, ys variability across all values of x will be constant and that will be sigma square. Third one is that whether when you are drawing simple sample the observations. So, every observation they will be independent. So, you have i equal to how many observation 1 2 n observations, x is there, x 1, x 2 like this. Finally, y is also there. So, this is first observation, second observation.

So, these observations will be independent. Suppose if I consider y 1 to yn. So, all those y 1 is no affecting y 2. Another they are independent and they are coming from the same population that is normal population. So, they, each of them are also normally distributed with the corresponding population distribution. Here we will go for 1 population and we withdraw in n number of samples or a sample of n observations, then each of the observations will be coming from that population with parameter may be normal parameter mu and sigma square, then there will be this one.

Apart from these, what will happen, because if this y is no independent, every observation then the errors will also become independent, and this ys variability as the fixed portion is coming out that the variability portion is going to the error that is why error is having, epsilon is having n 0 sigma square variability. And other one is there should not be any autocorrelation; that means, independent. If they are independent there is no autocorrelation means y 1 is not dependent on y 2 or y 2 not dependent on y 1, as such y i is not dependent on maybe y i minus k may be at any lag, may be 1 lag or 2 lag, but it they will not be autocorrect ok.

(Refer Slide Time: 06:59)

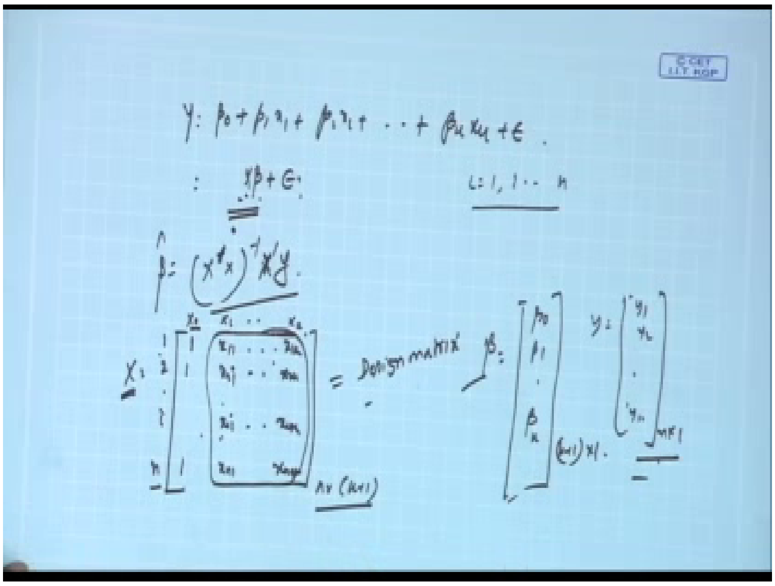


References

- Design and Analysis of Experiments by Douglas Montgomery, Wiley, 8th Edition, 2014
- Applied Multivariate Statistical Modelling by J Maiti NPTEL Video.

IIT KHARAGPUR
NPTEL ONLINE CERTIFICATION COURSES

(Refer Slide Time: 17:11)



$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \epsilon.$$

$$= X\beta + \epsilon \qquad i = 1, 2 \dots n$$

$$\hat{\beta} = (X^T X)^{-1} X' y.$$

$$X = \begin{array}{c|cccc} & x_0 & x_1 & \dots & x_k \\ 1 & 1 & x_{11} & \dots & x_{1k} \\ 2 & 1 & x_{21} & \dots & x_{2k} \\ \vdots & & \vdots & & \\ i & 1 & x_{i1} & \dots & x_{ik} \\ n & 1 & x_{n1} & & x_{nk} \end{array}_{n \times (k+1)} = \text{Design matrix} \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}_{(k+1) \times 1}. \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n \times 1}$$

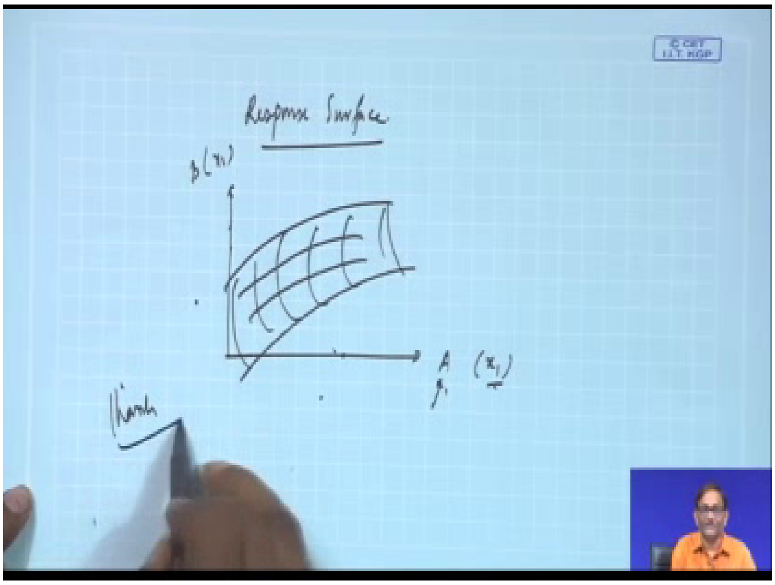
So, these are the few assumptions, and then what I want to say, that next class we will see elaborately, but suppose we say that y is beta 0 plus beta 1 x 1 plus beta 2 x 2. So, like this, this, this beta kxk plus epsilon. So, this can be written like this x y equal to x beta plus epsilon, and we will see that how, what is the x n, what is the beta when i equal to 1 to n. So, how this x will be computed x will be derived beta will derive all those things.

Usually what we will see that when i will equal to x will be n cross k, and beta will be beta plus 1 into k and like this. And finally, using ols regression, we will compute beta cap equal to x transpose x x transpose x inverse, if x transpose y where x is nothing, but we have n observations. So, x 0 x 1 like this xk. So, we have constant term for everywhere 1 1 1 1 1, and this will be x 1 1 x 2 1. Suppose there is i x i 1 then n 1 and like this x 1 k x 2 k then x i k and x n. Then suppose this is x i k and x n k.

So, then this is 1 0 to k. So, that and it is n. So, n cross k plus 1, this is the design, this is known as design matrix and beta will be your beta 0, beta 1, 2 beta k. This will be k plus 1 cross 1 parameter to be estimated, and y is nothing, but that y 1 y 2 to y n. This is n cross 1, this is the, this is what is basically the dependent variable observations.

So, here in the design matrix you are getting an interceptum where all 1 a every row content 1, and the remaining this portion. This is basically the independent variable, but these are the observered or fixed values for x 1, x 2 and x k at different observation bringing, or the different settings, basically with reference to experiment. So, fine, please keep in mind multiple linear regression is a very important topic for designing of experiment, because we will be later using the concept called response surface, response surface way response surface.

(Refer Slide Time: 19:58)



So, this response surface will give you the behavior of the y. For example, if I have two factors. So, A or I can say x 1 or B if I say x 2. So, depending on when i you want to change A and B with a purpose, then the y will be the, y values may create something like this, this kind of surface.

Now, what happened for a particular value of A and B or particular value of x 1 and x 2. You will be having the y value and if that y value is desirable one. So, you will run the process at that x 1 and x 2 controllable values, or other I can say factors so.

Thank you next class we will see more details.