(Refer Slide Time: 00:24)

Good morning. We will discuss now multivariate linear regression. Multivariate linear regression, we will be denoting it as MvLR, MvLR. There is another regression, which is known as multiple linear regression, which we have already completed that we have denoted like this, MLR. So, the difference between the two will be discussed.

(Refer Slide Time: 01:20)

Then the in the same order, just see this slide that we will go by the conceptual model. Then we will describe the assumptions of the model. Then we will estimate the parameters. Then I go for sampling distribution of the parameters. Finally, overall assessment of model fit, that will be for this lecture and then I think test of individual regression parameters, model diagnostics and other things will be will be discussed in the next lecture.

(Refer Slide Time: 01:59)
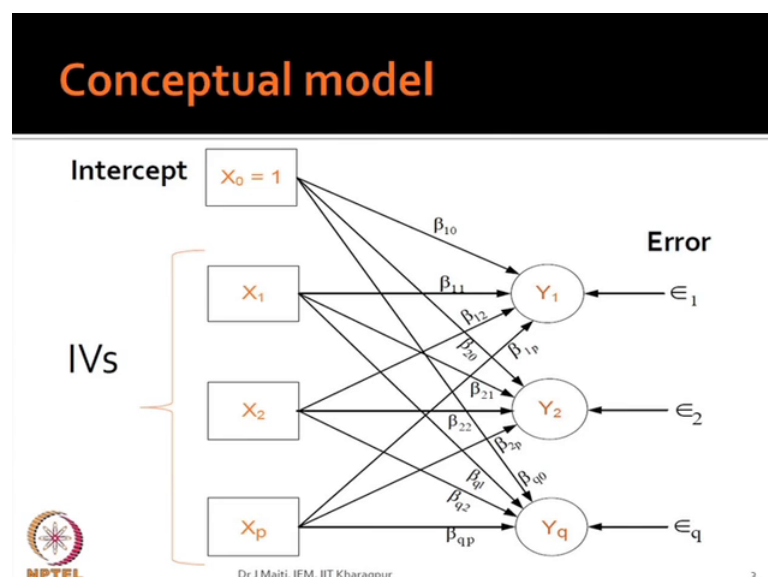
So, now let us see that what is the difference between MvLR and MLR? So, in this diagram, you see there are two parts. First one is this Y1, Y2, Yq, this side right hand side and left hand side, you see that X0 to Xp. This part, the right hand side part is known as the dependence part, left hand side part is known as the independent, independent part. What what do I mean by this? We are saying that the variables Y1, Y2, Yq are dependant variables and the variable X1 to Xp are independent variables. Now, all the dependent variables starting from Y1 to Yq, they are dependent on the same set of independent variables X1 to Xp. So, for example, let us consider onlY1 variable.

(Refer Slide Time: 03:15)
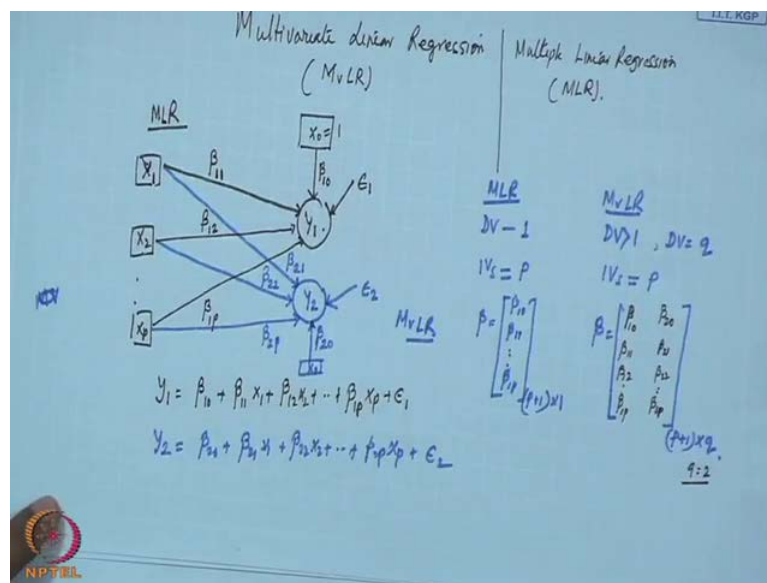
Y, Y1 and then let us think that there are set of p independent variables and obviously you have seen in multiple linear regression that is if you can recall, if you recall your MLR multiple linear regression, you can find out that we have pictorially depicted the multiple linear regression part like this, where there will be several coefficients. Those are known as regression coefficients that Y1 is affected by X1, so 1 related to dependent side followed by another subscript that 1, that 1 is related to the independent side, so beta 11.

So, similarly, second variable from the independent side is X2, which effects Y1, so 1 is affected by 2. So, similarly, 1 is affected by p, pth independent variable. We have also seen that there is a constant term X0, which we are saying that beta1 is affected by X0, X0 always takes value 1. So, this is what is your MLR? In MLR, the equation is Y1
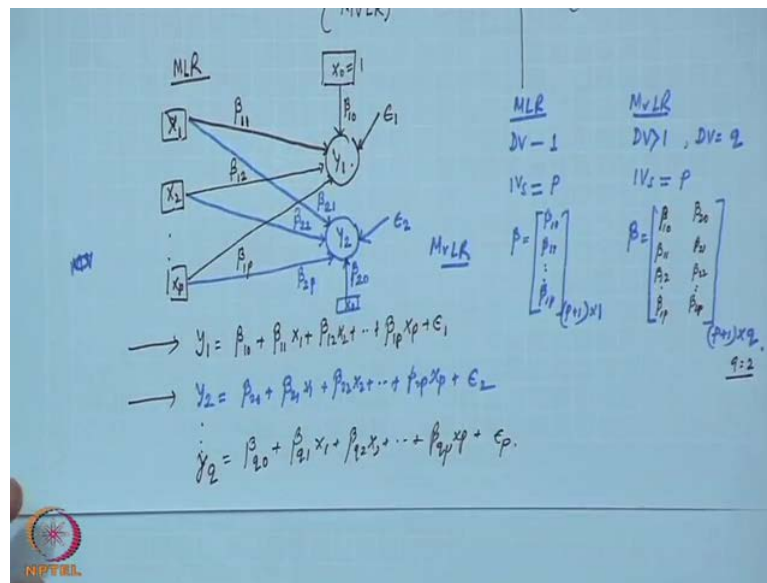
equal to beta10 plus beta11 X1 plus beta12 X2 like this beta 1p Xp plus epsilon 1. Now, let us add one more variable, which variable? Dependent side. So, we are saying the independent structure remains same X1 to Xp. You are adding one more dependent variable Y2.

Then, if I want to know how they are affecting, then 2 is affected by X1. So, it will be beta21. Then Y2 is affected by X2. It will be beta22. Then similarly, at the end, finally, 2 is affected by Xp 2p. So, it also has that X0 component. So, we can say that that I can, we can bring from here. I am writing here again. So, beta 2 is affected by 0.

So, if I write down this Y2 equal to beta20 plus beta21 X1 beta22 X2, so like this beta2p Xp plus epsilon2, so in this diagram, that means, then there is another component which is epsilon2. Similarly, in the in for the Y1 case, there will be another component which is epsilon1. So, this one, when we have added Y2, it becomes MvLR, multivariate linear regression. So, if I now think the difference, then if I write MLR, then DV will be 1 and if I write MvLR, your DV will be greater than 1, greater than 1, any value. Your IVs are p. Here, your IVs are also p, p number of IVs.

So, here in MLR, you are as dependent variable is 1, so beta is the parameter vector which is basically beta 10, beta 11 like beta 1p, which will be your p plus 1 cross 1. But, here what happen? If we assume that total DV equal to q, then your beta is no longer p plus 1 cross 1, it will become p plus 1 cross q because of q variable, q dependent variables. So, for this case, you can see that here what we have written here in this case. So, if I consider this particular example, then your this one, your beta is beta 10, beta 11, beta 12, like this beta 1p, then beta 20, beta 21, beta 22, beta 2p. Getting me? So, here that means q equal to 2.

(Refer Slide Time: 09:39)

This is what now you may be interested to add many more variables in the Y side; instead of 2q, maybe 10 and then you will be having ten simultaneous regression equations. This is for first variable, second variable, like this, you will be having equation for Yq, which is beta q is affected by the constant term plus beta q1 X1, beta q2 X2, so like this, beta qp Xp plus epsilon p.

(Refer Slide Time: 10:14)
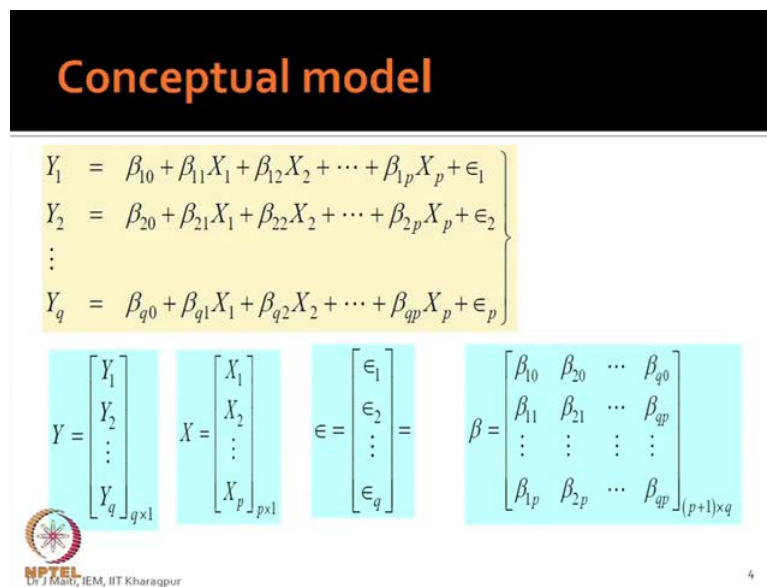
So, then we can write in matrix term. Matrix term, if we want to write, what we will get?

(Refer Slide Time: 10:27)
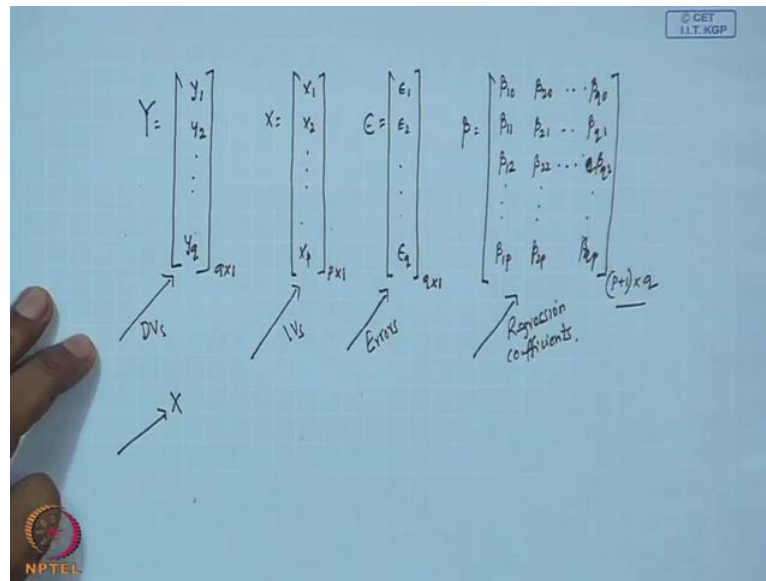
We will get Y variable, which is Y1, Y2 like this Yq, q cross 1. We have X independent variable side X1, X2, Xp, p cross 1. You also have your epsilon1, epsilon2, like this epsilon q, q cross 1 random that error variables. You have also beta that will be your beta10, beta11, beta12, then beta 1p, beta20, beta21, beta22, beta2p, beta q0, beta q1, beta, beta q2, so like this beta qp, this will be p plus 1 cross q. So, I am saying this is my DV. This one my IVs, this is my errors; this is my regression coefficients, regression coefficients. So, if you consider X as design matrix, suppose X is the design matrix, then see this X1 to Xp, these variables are there, but X0 will come in to consideration that we will see when we collect data. Getting me?

(Refer Slide Time: 12:50)
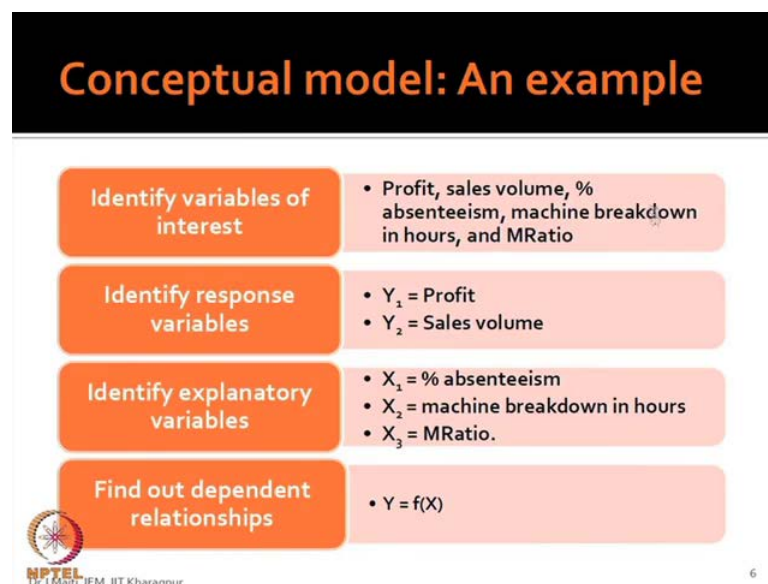

Conceptual model: An example

Sl. No.	Months	Profit in Rs million	Sales volume in 1000	Absenteeism in %	Machine breakdown in hours	M-Ratio
1	April	10	100	9	62	1
2	May	12	110	8	58	1.3
3	June	11	105	7	64	1.2
4	July	9	94	14	60	0.8
5	Aug	9	95	12	63	0.8
6	Sep	10	99	10	57	0.9
7	Oct	11	104	7	55	1
8	Nov	12	108	4	56	1.2
9	Dec	11	105	6	59	1.1
10	Jan	10	98	5	61	1.0
11	Feb	11	103	7	57	1.2
12	March	12	110	6	60	1.2

Dr J Maiti, IEM, IIT Kharagpur

5

Then, you see this example. I told you that profit, sales, absenteeism, these are the things. Now, here profit and sales, they can be dependent variables and absenteeism, machine, M ratio can be independent variables.
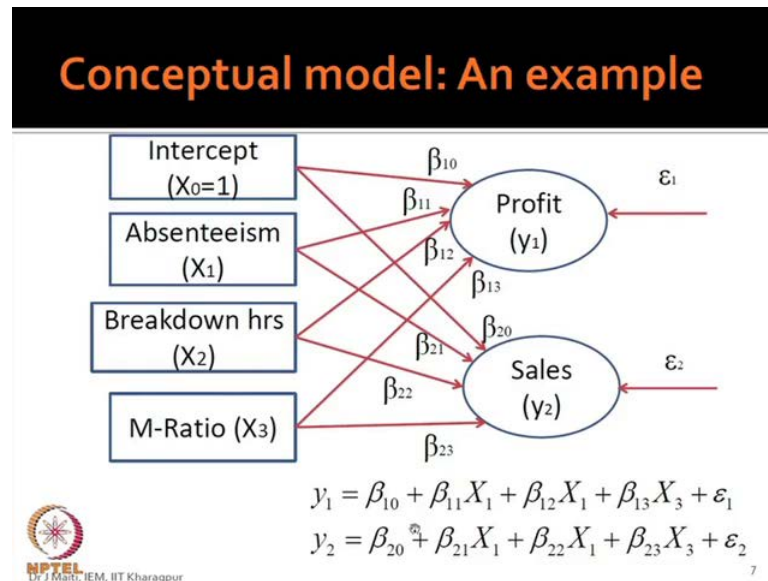
(Refer Slide Time: 13:14)

Conceptual model: An example

Identify variables of interest	• Profit, sales volume, % absenteeism, machine breakdown in hours, and MRatio
Identify response variables	• Y_1 = Profit • Y_2 = Sales volume
Identify explanatory variables	• X_1 = % absenteeism • X_2 = machine breakdown in hours • X_3 = MRatio.
Find out dependent relationships	• $Y = f(X)$

Dr J Maiti, IEM, IIT Kharagpur

6

Then, this is the example from this conceptual model point of view what we are discussing about the cities can. So, identify variables of interest. This is the first step. You have to, you identified that profits, sales, volume, percentage, absenteeism, machine breakdown, M ratio, and these are the variables of interest. Now, you have to identify
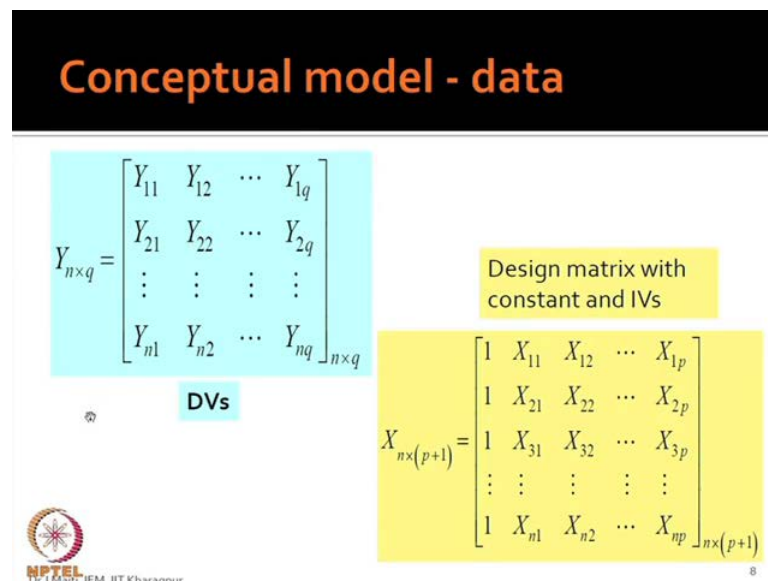
what are the dependent or response variables. So, in this case, Y1 is profit and Y2 is sales volume are the dependent variable. You have to identify the explanatory variables or independent variables of interest your X1, X2, X3, three independent variables. Then you have to find out the relationship, Y equal to function of X. In this case, we are talking about linear relationships.

(Refer Slide Time: 14:06)

Conceptual model: An example

$$y_1 = \beta_{10} + \beta_{11}X_1 + \beta_{12}X_1 + \beta_{13}X_3 + \varepsilon_1$$
$$y_2 = \beta_{20} + \beta_{21}X_1 + \beta_{22}X_1 + \beta_{23}X_3 + \varepsilon_2$$

Then, this is your diagram for multivariate linear regression model. So, in general, what will happen? There will be p and there will be as I said, there will be q. That is the case.

(Refer Slide Time: 14:29)

Conceptual model - data

$$Y_{n \times q} = \begin{bmatrix} Y_{11} & Y_{12} & \cdots & Y_{1q} \\ Y_{21} & Y_{22} & \cdots & Y_{2q} \\ \vdots & \vdots & \vdots & \vdots \\ Y_{n1} & Y_{n2} & \cdots & Y_{nq} \end{bmatrix}_{n \times q}$$

DVs

Design matrix with constant and IVs

$$X_{n \times (p+1)} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1p} \\ 1 & X_{21} & X_{22} & \cdots & X_{2p} \\ 1 & X_{31} & X_{32} & \cdots & X_{3p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix}_{n \times (p+1)}$$

Now, come to the data point, come to data. So, we require data for Y.

(Refer Slide Time: 14:42)
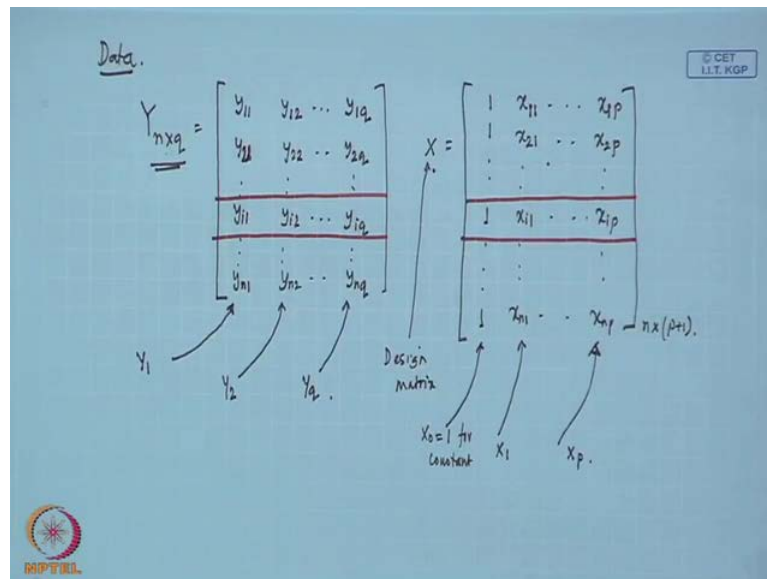
Suppose we will collect n data points on q Y variables that n data points on q Y variables will give you Y n cross q matrix. So, let us write down this y11, y12, y21 you write y11, y21, like this yi1, like this yn1. What is this? This data set, this data vector is on variable Y1, so n data points on variable Y1, so 1, 1 to n. So, I will go for variable 2, so first observation on variable 2, second observation on variable 2, so like this ith observation on variable 2, so like this yn2, this is my second observation on variable 2.

So, in the same manner, you will be finding out first observation on variable 1, second observation, first observation on variable q, second observation on variable q, so like this ith observation on variable q. When I am saying variable 1 or variable, variable 2 or variable q, these are all related to Y variable, so same manner n observation on Y variable q. So, this is my Yq. Now, this is what is data, this is related to the data. Now, so long, you are not collecting this data. These all data points are random and we, we, we will describe later that we assume multivariate normality. Then your another data is point is X, which we talk about as design matrix in this case.

So, if I say design matrix, then how many variables will be there, p variables plus constant, so all 1. So, this is the X0 part equal to 1, this for constant. Then I will come to the second one, x variable observation 1 on variable 1, observation 2 on variable 1, observation i on variable 1, then observation n on variable 1, this is my IV1, X1. So,

similarly, you will be finding out x variable observation 1 on x variable p, variable observation 2 on p, so like this i on p, so like this n on p. So, this is your Xp independent side.

So, then this matrix is n cross q Y matrix and the X matrix will be n cross p plus 1, p plus 1. So, if you take with this, if you take the ith observation, now, now, we will take the ith observation for all the cases, we will be taking the ith observation, so ith observation is this here. Here, it is ith observation is this. So, let us write the ith observation. What is the status?

(Refer Slide Time: 19:05)

So, if you write the ith observation, you the general equation is y. So, we will write ith observation for which one for the first variable, so ith observation first y. So, this can be written like this beta1 is affected by 0 plus, plus beta1 is affected 1, then x ith observation. So, in this manner, you will be getting beta1 is affected by p, x ip. So, similarly, you will be getting for y i2, beta20, beta 2 is affected 1 x i1, like this beta2 is affected by p, x ip. Same manner, you can if you write the kth variable, y ik equal to beta I think k 0 plus beta k is affected 1 x i1, like this beta k is affected by p, x ip. Then you go for y ip k iq, this will be beta q0, beta q1, x i1, like this beta qp, x ip.

So, this is from the x point of view, y and x, but x cannot capture y in totality. So, the error term, we also require to add everywhere that we have not added here. So, this will be your related to first variable related to second y variable related to third k y variable

related to q y variable, but these are all related to ith observation, so x i1, epsilon i1, epsilon i2, epsilon ik, epsilon iq. So, if you take this in general that for i equal to 1 to n observations because you have total n observations, so this set will be repeating n times. Getting me?

So, that is why, what happened? You are getting Y n cross q; that Y1, this i is changed from 1 to n, this one, that variable change from q. Then this one, this side, you see that X, that X will be your, it also will be n cross p plus 1. Then there will be beta. This beta will be p plus 1 cross q plus, you have epsilon, which is n cross q. This is a general equation in terms of data, the regression equation multiple, multivariate multiple regression. Here, we are, we have included the data points.

So, this is our model. This is our model. Now, see that what is required? Similar to MLR, you require to estimate beta, you require to estimate the epsilon also, but here it is not a one dependent variable case, q dependent variable case. Your estimation is now more in number. Getting me? The totality is more.

(Refer Slide Time: 23:57)

Data - example

$$
Y = \begin{pmatrix} 10 & 100 \\ 12 & 110 \\ 11 & 105 \\ 9 & 94 \\ 9 & 95 \\ 10 & 99 \\ 11 & 104 \\ 12 & 108 \\ 11 & 105 \\ 10 & 98 \\ 11 & 103 \\ 12 & 110 \end{pmatrix} \quad X = \begin{pmatrix} 1 & 9 & 62 & 1 \\ 1 & 8 & 58 & 1.3 \\ 1 & 7 & 64 & 1.2 \\ 1 & 14 & 60 & 0.8 \\ 1 & 12 & 63 & 0.8 \\ 1 & 10 & 57 & 0.9 \\ 1 & 7 & 55 & 1 \\ 1 & 4 & 56 & 1.2 \\ 1 & 6 & 59 & 1.1 \\ 1 & 5 & 61 & 1 \\ 1 & 7 & 57 & 1.2 \\ 1 & 6 & 60 & 1.2 \end{pmatrix}
$$

So, we will proceed with estimation and this is one the same example what we have discussed. So, we have considering profit and sales, Y1, Y2 and our absenteeism, breakdown hours, M ratio, so this is X. As I told you for X, there will be 1 and these are the X1, X2, X3, so this is 12 cross 1, 12 cross 1, 2, 3, 4, 4, 12 cross 4. This one is your 12 cross 2 matrix. So, we I will show you the result of this matrix basically this data set.

(Refer Slide Time: 24:34)
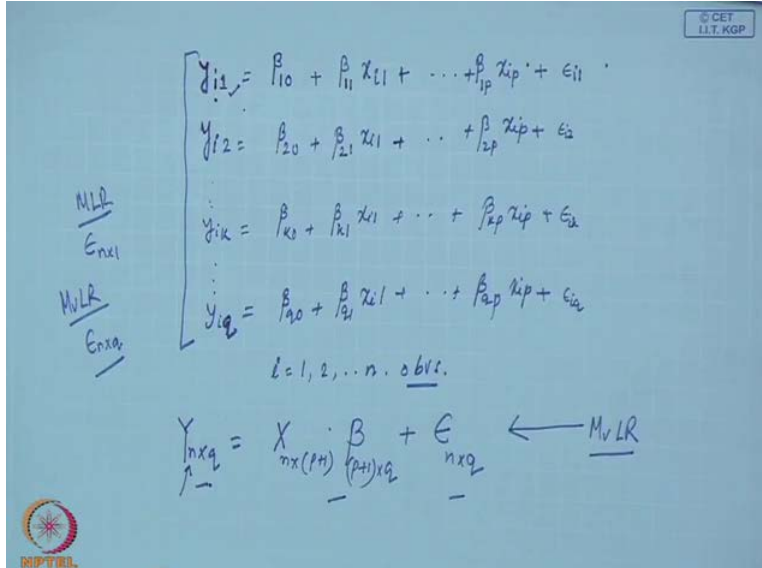

Now, what are the assumptions? This model, MvLR is also applicable under certain assumptions like MLR. The assumptions are; you see errors are multivariate normal. Error is now epsilon n cross q. In MLR, what is the error? It is n cross 1.

(Refer Slide Time: 25:08)

In MLR, error is n cross 1. In MvLR, error is n cross q. Keep in mind this difference. So, in MLR, what we have discussed that the error variance, error variance across all x observations, they will be same, which is known as homogeneity condition, so

homogenous, sorry, that we can say homoskedasticity. Homoskedasticity, homoskedasticity problem we have discussed.

(Refer Slide Time: 25:45)
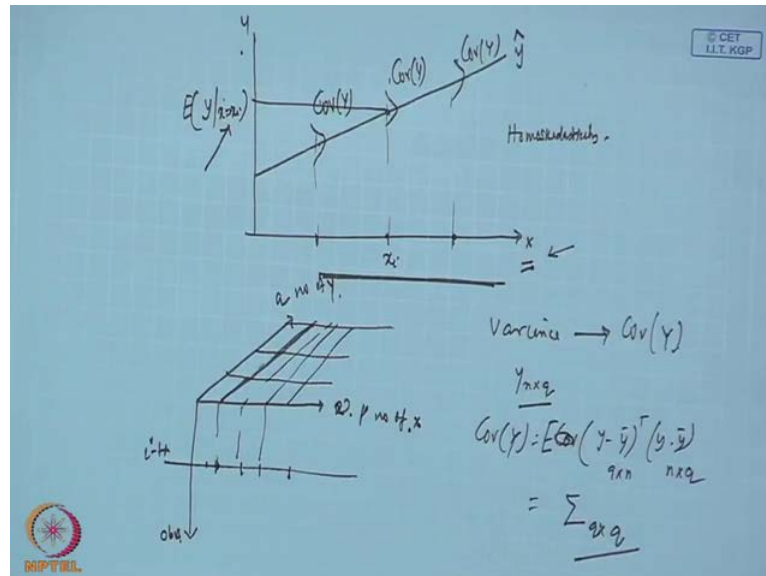
Homoskedasticity problem with respect to simple regression, we have discussed x and y. Then we say this is my regression line, which is y cap. Then when there are different observations, different observations, we say that the variance of y across x observations depends on values, this value, this value, this value, they are same that is homoskadasticity. So, here also, this remains same that error variances are equal across observations conditional on predictors. It is true, true because if this one is xi, this particular value is expected value of y given x equal to xi. This is conditional on predictor.

Now, errors have common covariance structure across observations. This is another assumption. Please understand that when we talk about MLR, we talk about across observations of X, we talk about error variances are equal, but here what is happening? Your xi remains same, but y side is now multiple means what I mean to say? This is my X. Suppose this is my observations, this is the X variables, but you have q number of this is p number of X, q number of Y.

So, if I say that p this is my ith observation is this ith observation on x, so fine, this is 1, 2, 3 like this, but for this ith observations as you have q number of y variables, so what will happen ultimately here? So, everywhere, you take any observation, you are getting

this q Y combinations. So, that is why, the third assumption is when we have q variable, so this variance will converted to covariance, variance will be leading to covariance. so covariance of y. So, that means here I will be talking about covariance of y. Here, I will be talking about covariance of Y. Here, I will be talking about covariance of Y. Understood?

So, covariance of Y as Y is n cross q. What will be covariance of Y that is covariance of we can you can write that y minus this mean and all those things. So, finally, you can write y minus y bar that transpose, this is the expected value, expected value of this y minus y bar this will be q cross n, n cross q. So, you will be getting something like this, which is q cross q. Go back to the first class first or second lecture that is what we have discussed. So, we say that there are p variables. Now, I can write down this covariance structure like this.

(Refer Slide Time: 29:34)
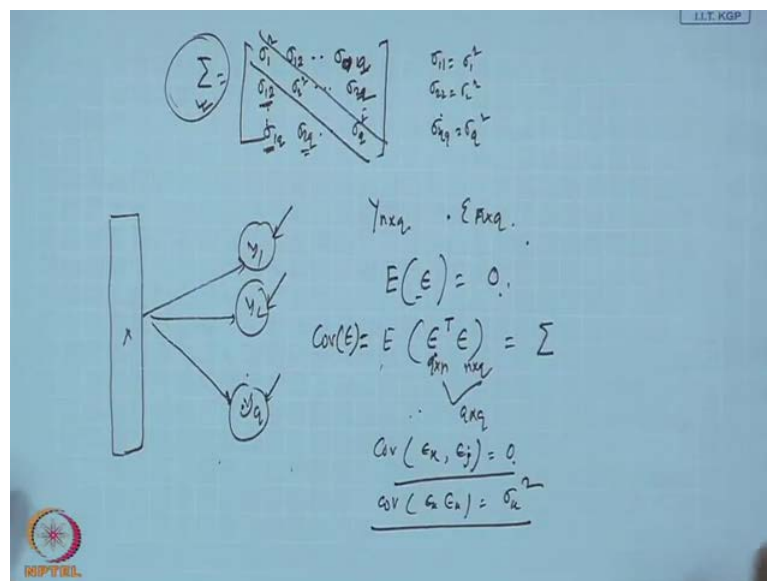
Now, I can write down this covariance structure like this, covariance sigma 1 square, sigma 12, sigma 1q here because we are taking q y variables sigma 12, sigma 2square, sigma 2q, now sigma 1q, sigma 2q, sigma q square. Actually, sigma 11 is sigma 1 square. So, sigma 22 is sigma 2 square, like this, sigma qq is sigma q square. This is the variance part.

So, in third assumption, we are saying error have covariance equal or common covariance structure, basically, common covariance structure means for this observation,

this observation, this observation, this observation, will be related to x, related to x. This is the ith observation and here I think, observation this side I think, we are saying n, basically these are all x variable for every observation, so ith observation. Similarly, for every observation what will happen ultimately? You will be finding sigma, capital sigma. This is the ys variability.

We are assuming here that see y, if you see the diagram, we are saying that y1, y2 like this, then yq and this side is a set of x and we are linking to everywhere here. Then we are saying this is my multivariate, but we are not talking about anything related to y1, y2, y to yq. We are assuming that are their correlation structure is insignificant or negligible basically. So, if this is the case, what will happen? These off diagonal elements become 0.

So, as all of you know that in the regression side that this variability part ultimately will be captured by error, so we can say this is the error covariance also. Your Y is n cross q. Similarly, error is also n cross q. So, if I want to find out what is the expected value of error, this will become 0. This is the assumption and expected value of error, transpose error. The reason is we want to find out the covariance of error. So, covariance of error is expected value of this minus mean transpose and this so as mean is 0, so this is the case. So, this one will be q cross n, this is n cross q, so ultimately this will be q cross q. So, this we are saying this.

So, that means the error, the regression error covariance matrix is this and we are saying that covariance from one dependent variable to other dependent variable that variability is like this the covariance between epsilon k and epsilon suppose j, two, two different variables this 0 and covariance epsilon k, epsilon k that is sigma k square. Then so you have to, you require to test this.
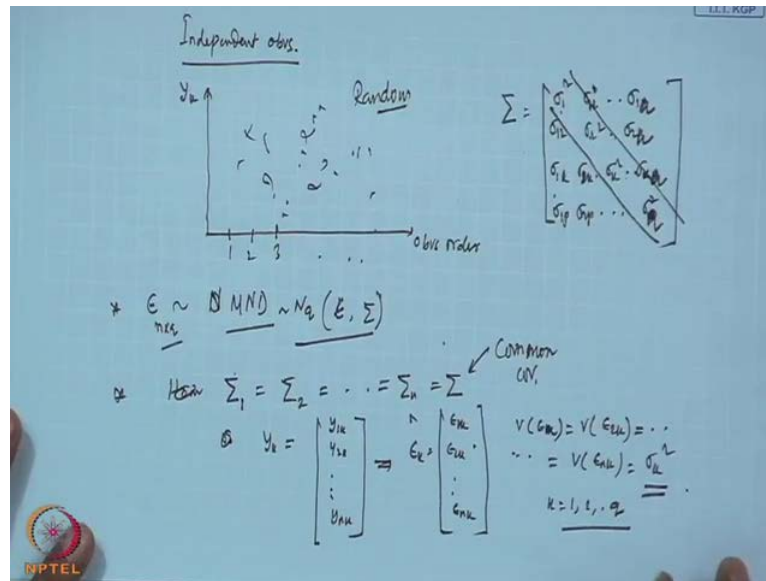
Then, the last one is independent observations, independent observations. How do we test independent observation? So, that means no auto correlation, independent observation means one observation Y, we are talking about related to Y. One observation is independent of others. It is not influenced by other observation. So, if you see the observation order vis a vis the plot for every variable, you can do this observation order. Then suppose this is Yk, then you will be getting a plot random, a random plot observation, first observation, second observation, third observation like this.

So, nutshell, what is there? The errors are very important component in regression and that too true in multiple, multivariate multiple regression also. So, error will ultimately tell you all the assumptions. First one is it will be your multivariate normal distribution. So, it will be MND. So, we say error is n cross q. So, definitely it will be q. That means, it will be n q that the mean value epsilon and you have to find out the variance, variance component, what will be the value? So, this you have to find out. Now, this is multivariate one. Second assumption is our homoskedasticity, equal covariance.

So, if you have observations, different observations, suppose n observations you have collected, then all will be equal to this. This is the common covariance matrix, your error variances are equal we say across observations, common covariance structure. That means another one is error variance means for every Y variable, if I say Yk, then if I take the observation related to Yk, then Y1k, Y2k like this Ynk observations will be there and

corresponding error will be epsilon k. This one will be epsilon1k, epsilon2k like this epsilon nk.

Then, what we mean by that errors are equal variance? That means variance of epsilon1k equal to variance of epsilon2k like this equal to variance of epsilon nk that will be sigma k square. Keep in mind that will be sigma k square. So, k stands from 1 to. Now, where from you will get the k? You will get the k from the common covariance matrix sigma1 square, sigma12, sigma1p, sigma12, sigma2 square, sigma2p, sigma1k, sigma2k, sigma k square, sigma kp, so like this, sigma1p, sigma2p, sigma p will give q. Why q, q variable? P, we are using for independent side.

So, this diagonal matrix, I think you understood fully. So, these are the assumptions, these assumptions. Now, actually what will happen? Later on, you will see that you cannot, you do not know this value, this matrix. You have to estimate, you have to estimate from data. So, our next issue is estimation of parameters.

(Refer Slide Time: 38:37)
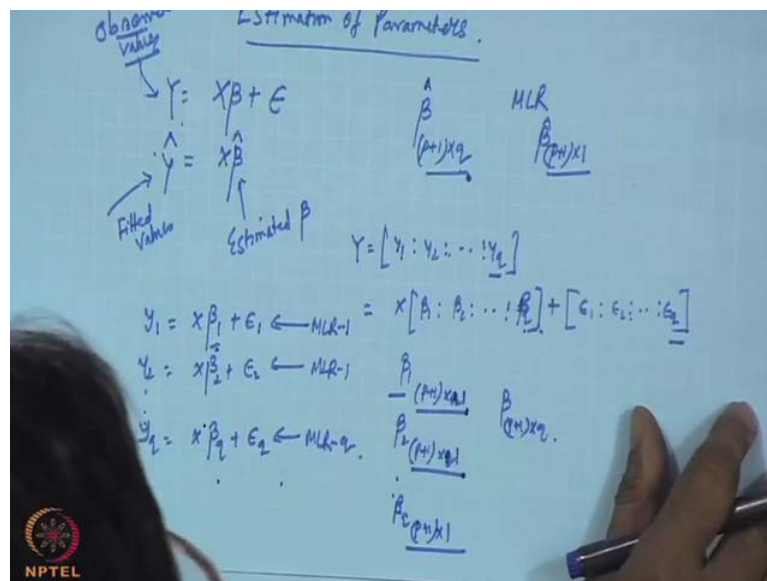
I think this estimation of parameters, we will complete this session by using, describing this estimation of parameters. So, our equation is Y is X beta plus epsilon. If we estimate, we say Y cap equal to X beta cap, this is our fitted value. This is the estimated parameter, estimated beta, these are observed values and this is known the error term. So, I have something to say here. I mean something more to say here. What I mean to say that please remember that Y is, there are q Y variables. Getting me?

So, similarly, and p X variables, so that means this beta, this beta, what you estimate that will be that will be how many parameters will be there. It will be matrix. What will be that matrix? That will be p plus 1 cross q. Now, in MLR, what will happen? MLR beta is p plus 1 cross 1. So, additionally, it is coming like this. So, that means if I write like this Y equal to something like this Y1, Y2, then it is Yq, this Y, then this equal to we say X beta. Can I not write like this beta 1, beta 2, beta q, so plus you will get what epsilon? This epsilon is also having epsilon1, epsilon2 like epsilon q, no, no q, this is q, this is q, q will be there.

So, it is this equation is nothing but as we have seen basically q MLR, q number of MLR means Y1 equal to X beta 1 plus epsilon1, Y2 equal to X beta 2 plus epsilon2, similarly Yq equal to x beta q plus epsilon q. So, this is MLR 1, MLR 2, MLR q. So, q MLR is there. Then what we will have to estimate them? What we have to estimate?

We have to estimate beta 1 to beta q and again beta 1 is p plus q into p plus 1 into q and everywhere beta 1 is also p plus 1 cross q vector beta 2 also p plus 1 cross q, sorry beta 1 p plus 1 vector. This is also one vector. Please remember beta in a nutshell, beta is p plus 1 cross q, but beta 1 p plus 1 cross 1 beta 2 p plus cross into 1, like this, beta q that also p plus 1 cross 1 vector. So, if you go, what is the procedure in MLR, what way? We suppose one variable, Y variable and we estimate, we estimate first the error term.

(Refer Slide Time: 42:47)

For example, if I recollect Y1 equal to X beta 1 plus epsilon1 that one MLR, what you have done? You have done, first find out epsilon1 that is Y1 minus X beta 1. So, if I forget about 1 because we are talking this subscript 1, we are talking about Y1 variable in this case. So, general term is Y minus X beta. Then you have found out YY transpose that is epsilon transpose epsilon that is Y minus X beta transpose Y minus X beta. Then you have this is what we say sum square errors, sum square errors.

Then, you have regress sum square errors. You have found out the derivatives of sum square SSE with respect to beta and then you put it to 0. Finally, you got the equation like this beta cap is X transpose X inverse X transpose Y, this was the estimation procedure there. This is I am talking in terms of MLR multiple linear regression. Now, if we consider Y1 in that, we want to keep that notation, then this will be beta 1 and this will be Y1. X part remains same because of no multivariate regression, X part is unchanged. Y part is having additional Y, that is the issue.

So, in the same manner, so what happen ultimately here? In this case, when you collected n into 1 data points and ultimately, this one is 1 into n, n into 1. Ultimately, SSE is 1 into 1, so it is a scalar quantity. SSE is a scalar quantity in case of MLR. Now, what will happen in this case? Suppose we are coming to MvLR. MvLR same thing, Y is X beta plus epsilon, but this one we are saying n cross q, this one n cross p plus 1, this one is n p plus 1 cross q, this is your n cross q. So, let us see what will happen. We want to see a errors terms. So, error epsilon is now Y minus X beta. So, error is n cross q; in sense also, Y minus X beta is n cross q.

Now, we want to find out the sum square errors like MLR that is to be minimized. So, if I want to do this, then if I will make like this epsilon transpose, this will be q cross n into epsilon, which is n cross q. So, what is happening here? You are getting a q cross q matrix. So, I cannot say it is a scalar quantity is a matrix. As a result, this will give you a structure like this and your this, this elements will be sum of i equal to 1 to n epsilon i 1 square. The second diagonal element will be i equal to 1 to n epsilon i 2 square. Similarly, the last diagonal will be i equal to 1 to n epsilon i q square and off diagonal element will be the covariance structure epsilon i 1, epsilon i 1 like this, so like this. Getting me?

So, your error now this is your sum squares error is no longer a scalar, it is a matrix. It has, it is symmetric, and it has diagonal elements, the variance component of the diagonal elements, the covariance component of this error terms although we assume that the covariance component will be 0 that is our assumption. So, that means this epsilon transpose epsilon is now sum square cross product. Getting me? Earlier, you have, we have described this sum square cross product epsilon E. We can write or you can say sum square epsilon, let it be E. E stands for error.

Now, we what we want? We want this matrix should be as minimum as possible because in MLR, we have seen ML that SSE the minimum. Then we derive, make derivatives, we get derivatives of SSE with respect to beta values, we put it to 0 and we got this equation. We have done the same thing here. We want this matrix to be as minimum as minimum as possible. Now, if I assume that the off diagonal elements are 0, then what we want? We do not want only that individual diagonal elements will be as minimum as possible. We want the total as minimum as possible, the total sum square across all the Y variables as minimum as possible.

(Refer Slide Time: 49:10)

So, what we do that is we take the trace of SSCPe or trace of SSCP e or you can say trace value of epsilon transpose epsilon. This is the diagonal values. So, this will be sum total of diagonal values epsilon i 1 square, epsilon i 2 square, like this epsilon i q square. So, we will minimize the trace. Your aim is minimize this. Getting me? You have to

minimize the trace. So, now how to minimize? What is your control here? Your control is you want the estimate in such a manner that this will minimise this.

So, your control is beta. Now, beta is not single here. Beta is basically; what is the dimension of beta, p plus 1 into q. So, this first one is for beta 1 that is p plus 1 into 1. Second one will be for beta 2, so like this. Getting me? So, what do we require? We require that we will find out derivative of trace, derivative of trace of this with respect to particular beta. Getting me? Now, by saying this, if I say beta k, then it is coming to the univariate case. So, beta k actually again, you have j equal to 1 to 0 to 1, 0 to p number of independent variables. So, ultimately beta k j will come. So, these things you will put into 0, put into 0.

So, actually what will happen? You will, you can write like this epsilon this one, this quantity cannot be, we cannot write like this that epsilon is Y minus X beta. So, epsilon transpose epsilon is Y minus X beta transpose Y minus X beta. So, when we are taking derivative trace that means you are taking trace of this means trace of this. That means only the diagonal elements are coming, so this.

Then, when you are taking the derivative of this with respect to this, when you are talking about the first beta 1 and that too beta 10, then ultimately all other things become 0. Only that quantity will be used for derivation. So, ultimately it will end into that. That means with it end into the individual Y variables, individual Y variables.

(Refer Slide Time: 52:41)

As a result, when you do this type of manipulation, ultimately you will be getting beta cap, which is nothing but X transpose X inverse X transpose. I can say y1, y2 like this yq. Getting me? Actually, this quantity will be multiplied to everyone. So, this totality when I write Y, then this one is X transpose X inverse X transpose capital Y. What is the difference these between these MLR beta? Here, Y is q number of variables, Y variables in MLR that will be one. So, this is our estimated case. So, essentially what we can say here? We can say that in this is the least square estimate and in least square estimate for MvLR, you will get beta cap like this, which is your beta 1 cap, beta 2 cap beta q is there q cap, beta q variables are there. This is nothing but here beta 1 cap means beta 10, beta 11, so like this beta 1 is affected by your p. This is your beta 20, and then beta 2 is affected by 1 like this. Beta 2 is affected by p. Similarly, beta q 0, beta q 1, then beta q is affected by p. This is the case.

This one is as we are saying that X transpose X, X transpose X inverse X transpose and we are saying that Y is there. So, can we not write like this? This one is similar to y1, y2 like this. This much is not required like this. Then y1 is again, y1, 1 to that n; all those things are there, so nutshell. What I mean to say that I mean to say that you, if you want to do this one is X transpose X inverse X transpose y1. Similarly, beta 2 cap is X transpose X inverse X transpose y2. So, like this, beta q cap is X transpose X inverse X transpose yq. Got it?

So, that means you may go for individual MLR and then whatever beta value you get and in multiple MvLR also, you will get the same thing. From beta point of view, there is absolutely no problem because here you are doing q number of calculations, computations. Here, in one computation, you are getting q beta vectors. So, both are same. Then where is the difference if I get the same beta value while going for q MLR? Why should I go for one MvLR? The error comes in, sorry not error, the difference comes in error terms because you have n errors. These errors are multivariate errors. Every observation is multivariate.

Earlier in MLR, your n errors are there, but from every univariate error, one observation per one error, but it is a multivariate observation now on q variables. That is the difference. In other way, there is no difference. So, that means as error plays the significant role in regression, so what do you want to do then? That means your all

treatment will be in the multivariate domain here that q, q every error observation is q data points, q number for y1 to yq.

So, test statistics differs. Test statistics will differ. It is similar to ANOVA and MANOVA. If you remember in ANOVA, ANOVA one response variable and several population in one way and two way that two factors and one response variable. Then you have used the ANOVA table. You have gone for f distribution, but in MANOVA, what happened? MANOVA, you have not got the straight cut distribution that is your SS, what I can say factor by your SS error. That means MS, MS factor that you did not get in MANOVA. Instead you got SSCP there and then you have gone for different test like Wilk's lambda test. It happened. Similar thing happened here. Getting me?

So, we will not go for like simple one coefficient of determination r square. We will go for; instead we will go for likelihood ratio test like Wilk's lambda. Getting me? That is the issue. So, ultimately your estimation part no problem, but model adequacy test will differ. Your test statistics will differ, but again the diagonal diagnostics case, you will find out the similarity with MLR. But, the beauty of this MLR is that may be that for the multivariate situation also, it can capture. So, many a times, we will not go for MvLR, but if MvLR, when there are more number of y variables, better to test through MvLR and because similar, it is not tough. Thank you very much. Next class, I will go for in detail the sampling distribution of beta, then model adequacy test and other things.