

INDIAN INSTITUTE  
OF  
TECHNOLOGY  
KHARAGPUR

NPTEL  
National Programme  
on  
Technology Enhanced Learning

Applied Multivariate Statistical Modeling

Prof. J. Maiti  
Department of Industrial Engineering and Management  
IIT Kharagpur

Lecture – 22

Topic

MLR- Sampling Distribution of Regression  
Coefficients

Today our discussion is.

(Refer Slide Time: 00:26)

Sampling distribution of regression coefficients

$$\hat{\beta} = (X^T X)^{-1} X^T y$$
$$X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix}_{(n \times p)}$$
$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n \times 1}$$
$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{bmatrix}$$

$\hat{\beta}$  is a random variable

Sampling distribution of regression coefficients in last class we have estimated  $\beta$  coefficient like this  $(X^T X)^{-1} X^T y$  where  $X$  is the design matrix first column is all 1 second column will be the first variable data  $X_{11} X_{21} \dots X_{n1}$  similarly last one will be the  $p$ th variable data  $X_{1p} X_{2p} \dots X_{np}$  and  $y$  is the  $n$  observations this is  $n$  cross  $1$  this is  $n$  into  $p + 1$  and the this one  $\beta$  this will be definitely  $p + 1$  cross  $1$ .

So if you collect one sample you get some value for  $\hat{\beta}$  cap if you go for another sample your  $\hat{\beta}$  cap will be different for example, sample one  $\hat{\beta}$  cap maybe 1 sample two  $\hat{\beta}$  cap maybe for sample two like this if we go for  $n$  sample so  $\hat{\beta}$  cap  $n$  so you will be getting series of beta values. And please keep in mind that  $\beta$  is or  $\beta$  cap is  $p + 1$  x  $1$  vector that is  $\beta_0$  cap  $\beta_1$  cap like this  $\beta_p$  cap so  $\hat{\beta}$  cap is a random variable is a random variable and we want to know its distribution and that is the sampling distribution of regression coefficient and then using the distribution of  $\hat{\beta}$  cap we will we will derive confidence interval and other things.

(Refer Slide Time: 03:21)

The image shows a handwritten derivation on a blue background. The equations are as follows:

$$\begin{aligned}
 * E(\hat{\beta}) &= E[(X^T X)^{-1} X^T y] \\
 &= E[(X^T X)^{-1} X^T (X\beta + \epsilon)] \quad \text{as } y = X\beta + \epsilon \\
 &= E\left[\underbrace{(X^T X)^{-1} X^T X}_{I} \beta + \underbrace{(X^T X)^{-1} X^T}_{X^T} \epsilon\right] \quad X = \\
 &= E[\beta] + (X^T X)^{-1} X^T E(\epsilon) \quad \beta: \text{Regression parameter.} \\
 &= \beta + (X^T X)^{-1} X^T \cdot 0 \quad E(\hat{\beta}) = \beta \leftarrow \text{Unbiased estimator} \\
 &= \beta
 \end{aligned}$$

So for a distribution you require to know what will be the expected value of  $\hat{\beta}$  cap we want distribution of this so we first required to know the expected value of  $\hat{\beta}$  cap this will be expected

value of  $\hat{\beta}$  is  $(X^T X)^{-1} X^T y$  so we can write this as expected value of  $(X^T X)^{-1}$  now you know  $y$  is  $X\beta + \epsilon$  as  $y$  equal to this is the regression equation so if we do little more manipulation this will become  $(X^T X)^{-1} X^T X\beta + (X^T X)^{-1} \epsilon$  equal to now  $(X^T X)^{-1} (X^T X)^{-1}$  the symmetric matrix inverse symmetric matrix we get identity matrix  $I$  so this will become  $I$  so ultimately expected value of  $I$  into  $\beta$  is  $\beta$  as  $x$  data already collected.

So  $(X^T X)^{-1} X^T$  this is a fixed quantity so we will keep it out from the expectation operator then we are writing expected value of this now expected value of  $\hat{\beta}$  is  $\beta$  because  $\beta$  is the regression coefficients these are regression parameters or coefficients from the population point of view that is constant plus expected value of error is 0 that is our assumption so you will get  $(X^T X)^{-1} X^T$  into 0 so this will become  $\beta$  so it says that expected value of  $\hat{\beta}$  is the  $\beta$  that is unbiased estimation okay so in addition now first we want we have taken this one, that expected value of what is the  $\hat{\beta}$  cap we want to know also the covariance matrix of covariance matrix of  $\hat{\beta}$ .

(Refer Slide Time: 06:26)

$$\hat{\beta} = \begin{bmatrix} \cdot \\ \cdot \\ \cdot \end{bmatrix}_{(p+1) \times 1}$$

$$\text{Cov}(\hat{\beta}) = \begin{bmatrix} \cdot \\ \cdot \\ \cdot \end{bmatrix}_{(p+1) \times (p+1)}$$

$$= \frac{E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T]}{n}$$

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

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

$$(\hat{\beta} - \beta)^T = [(X^T X)^{-1} X^T \epsilon]^T$$

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

*as  $(X^T X)$  is symmetric & square.*

Cap so we want to know covariance of  $\hat{\beta}$  cap this will be expected value of all of you know that  $\hat{\beta}$  cap - expected value of  $\hat{\beta}$  cap whole this is  $p$  cross  $1$  now this into  $\hat{\beta}$  cap - expected value of  $\hat{\beta}$  cap this transpose because covariance matrix of  $\hat{\beta}$  means it will be  $\hat{\beta}$  cap is a  $p + 1 \times 1$  so our

covariance matrix of  $\hat{\beta}$  cap that will be a matrix of  $p + 1 \times p + 1$  okay so now we can write that this one expected value of now  $\hat{\beta}$  cap minus expected value of  $\hat{\beta}$  cap is  $\hat{\beta}$  into  $\hat{\beta}$  cap -  $\beta$  transpose okay so let us find out what is the value of  $\hat{\beta}$  cap -  $\beta$  so what the value of  $\hat{\beta}$  cap if you see here when we have described this one.

For expected value of this one is this final means ultimately you got  $\beta$  plus this quantity this is  $\beta +$  this so I can write this one using this that  $\beta + (X^T X)^{-1} X^T \epsilon$  that is we can write for  $\hat{\beta} - \beta$  this transpose this will become  $(X^T X)^{-1} X^T \epsilon$  transpose so when you next transpose it will be just the reverse order so  $\epsilon^T (X^T X)^{-1} X^T$  that will remain because symmetric matrix that will remain this so  $\epsilon^T (X^T X)^{-1} X^T$  is symmetric and definitely square also so then if I use this if you use now this operator expectation operator here.

(Refer Slide Time: 09:39)

$$\begin{aligned}
 & E \left[ (\hat{\beta} - \beta)(\hat{\beta} - \beta)^T \right] \\
 &= E \left[ (X^T X)^{-1} X^T \epsilon \epsilon^T X (X^T X)^{-1} \right] \\
 &= (X^T X)^{-1} X^T E(\epsilon \epsilon^T) X (X^T X)^{-1} \quad E(\epsilon \epsilon^T) = \sigma^2 I \\
 &= (X^T X)^{-1} X^T \cdot \sigma^2 \cdot X (X^T X)^{-1} \\
 &= \sigma^2 (X^T X)^{-1} X^T X (X^T X)^{-1} \\
 &= \sigma^2 (X^T X)^{-1} = \text{Cov}(\hat{\beta}) = \sigma_e^2 (X^T X)^{-1} \\
 &\underline{\sigma^2} = \underline{\sigma_e^2} = \frac{SSE}{n - p - 1} \quad n = p + 1
 \end{aligned}$$

So we can take one more page so expected value of  $\hat{\beta}$  cap -  $\beta$   $\hat{\beta}$  cap -  $\beta^T$  this is nothing but expected value of that is coming  $(X^T X)^{-1} X^T \epsilon$  then for this one  $\epsilon^T (X^T X)^{-1}$  this is coming so for this we have written this portion for the other one written this portion now again what we will do then we will just bring out the fixed values like  $(X^T X)^{-1} X^T$  then expected value of  $\epsilon^T$  then  $\times (X^T X)^{-1}$  so what is the  $\times$  this one covariance matrix of error

term now covariance matrix of error term we say that equality of variances across the  $x$  observations so that will for  $y$  and that will go to the place so ultimately expected value of epsilon,  $\epsilon^T$  this will become  $\sigma^2$  square  $x^T I +$  there are  $n$  observations for.

Same that is from the assumption it is coming so this quantity will become  $(X^T X)^{-1} X^T \sigma^2$  square  $I \times$  into  $x$  into  $(X^T X)^{-1} X$  so now  $\sigma^2$  is constant  $I$  is the identity matrix that will go so finally it will be like this  $\sigma^2 (X^T X)^{-1}$  now here is  $X^T$  there is one more  $x$ , then  $(X^T X)^{-1}$  that may ok so what is this quantity then  $(X^T X)^{-1} X^T X$  and this is again  $I$  so  $\sigma^2 I$  is  $\sigma^2$  so this will become  $\sigma^2 (X^T X)^{-1}$  this is what is covariance of  $\beta$  cap so then how do we get the value of  $\sigma^2$   $\sigma^2$  is also not known so  $\sigma^2$  will be  $Se^2$  that mean the from the error whatever value you get from there you will calculate the variance component.

And this will be SSE by degrees of freedom for SSE is number minus parameters estimated so we can write then that covariance of  $\beta$  cap is  $Se^2 (X^T X)^{-1}$  so you will not get this  $\sigma^2$  value these are population value so from sample error you are able to calculate  $Se^2$  this  $Se^2$  is nothing but SSE by  $n - p - 1$  and this  $n - p - 1$  is coming because there are  $n$  is the sample size number of parameter to be estimated is  $p + 1$  so this degrees of freedom is lost when you are calculating the errors so this is the case then we can say that.

(Refer Slide Time: 13:53)

$\beta$  cap is a random variable is a random variable with mean vector  $\beta$  and covariance matrix we can write this way  $\sigma^2 (X^T X)^{-1}$  this is correct what will be the distribution this distribution will be multivariate normal this is the assumption we have started with that data comes from multivariate normal and the way we have estimated  $\beta$ ,  $\beta$  cap is  $(X^T X)^{-1} X^T Y$  so here what happen because of this multivariate normality assumption this will also become multivariate normal now this multivariate normal what will be the so N actually in the regression the error is the key point key issue y is definitely multivariate y is here one variable so normal actually what is happening here.

In this particular in this regression multiple regression case so you have a different sets of x values then you are calculating y values and everywhere there are there is error this error is N error will be there so that error component also will go for multivariate N cross N so we are here we are estimating  $p + 1$  parameters so  $\beta$  is, we are assuming that this is multivariate normal with  $p+1$  and this is what is the distribution this is the distribution of  $\beta$  cap now if this is the case now we will go back to our some earlier lectures where we say if x this is not one to one relation but to get some clue for related to  $\beta$ .

For some more derivation we have seen that suppose if  $x$  is multivariate normal with  $\mu$  and  $\sigma$  then we calculated  $\bar{x}$  and that one is also multivariate normal  $\mu$  and  $\sigma$  by  $n$  so similarly this we have taken this data from multivariate normal and we have calculated some statistic  $\bar{x}$  is also statistics here also we calculated some statistics and for that statistics we found out the mean vector and the covariance matrix and like  $\bar{x}$  covariance matrix and mean is also, calculated so in the same manner it is  $\beta$ .

Is multivariate normal now let what happened we have also described that what is the confidence region can you remember confidence region for  $\mu$  where  $\mu$  equal to  $\mu_1 \mu_2$  to  $\mu_p$  that we have discussed earlier and with the help of  $\bar{x}$  we found out where  $\bar{x}$  is  $\bar{x}_1 \bar{x}_2 \bar{x}_p$  we have found out this confidence region of  $\mu$  and if you can remember you will find out that, we have found like this that  $n(\bar{x} - \mu)^T S^{-1}(\bar{x} - \mu)$  this will become less than equal to  $n - 1$  by  $n - p$  this one probability of this equal to  $1 - \alpha$ .

This is the confidence region, we have found out with respect to that population mean vector so can we not find out the similar thing here with respect to  $\beta$  getting me so then it is not derivation what I am giving it is not derivation it is just understanding that why how this confidence regions can also be calculated with respect to beta cap which is multivariate normal.

(Refer Slide Time: 18:36)

CR for  $\beta$

$$P \left\{ (\hat{\beta} - \beta)^T (X^T X) (\hat{\beta} - \beta) \leq \lambda_c^{(n)} F_{p+1, n-p-1}^{(n)} \right\}$$

$\sigma^2 (X^T X)^{-1} = \lambda_c^{(n)} (X^T X)^{-1}$

SCI

Bonferroni approach

$\hat{\beta}_2 = 1 - \alpha$

$\hat{\beta}_1$

So can we not write like this then  $\hat{\beta} - \beta$  that is  $\bar{x} - \mu$  similar type of things this transpose there was S inverse so S in S is the covariance matrix so our covariance matrix is what our covariance matrix is  $\sigma^2 (X^T X)^{-1}$  which is we can write  $Se^2 (X^T X)^{-1}$  so what do we want inverse of this that means x transpose by  $Se^2$  so if I write like this  $(X^T X)$  okay I am not writing  $Se^2$  here later on we will write then  $\hat{\beta} - \beta$  What distribution it will follow that is important that also follows here that this  $Se^2$  it will come again I am telling you it is not derivation maybe something some different way people will come.

But this is the way we will understand for our application so this is  $Se^2 p + 1 p + 1 n - p - 1 \alpha$  probability that this quantity will be less than this is  $1 - \alpha$  so this  $Se^2 E$  square what is coming out here by this we are keeping here and the rest of the things are like this but there are some modification with respect to the parameter number of parameters to be estimated so this is the confidence region for  $\beta$  but we are not we what we will do with the confidence region so parallel then what you require to do.

You require to go for simultaneous confidence interval understood an may what I am saying that  $\hat{\beta}$  cap is multivariate normal  $\hat{\beta}$  cap is a statistic earlier we have seen  $\bar{x}$  multivariate normal that

statistic with  $\bar{x}$  we created the confidence region for  $\mu$  here with  $\beta$  cap we want to find out the confidence region for  $\beta$  now this is the formulation which can be used then from confidence region we have gone to simultaneous confidence interval for two variable case you have found out that when multivariate normality will be coming that something like this you got confidence ellipse.

So, from this ellipse to you want to go to the sides what is this interval confidence what will be this side that is true for  $\beta$  cap also suppose this is  $\beta_1$  cap and this side it is  $\beta_2$  cap same thing nature s same okay so for simultaneous confidence interval I think you have we have seen two approaches one of them is Bonferroni approaches so this Bonferroni approach is approach is easier for us and it gives good result also we can use Bonferroni approach and in regression whatever it is found that individual parameters when you test and you will find out that.

(Refer Slide Time: 22:27)

$\hat{\beta}_j \sim \text{Univariate normal.}$   
 $\bar{x} \rightarrow \frac{\bar{x} - E(\bar{x})}{SE(\bar{x})} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim t_{n-1}$   
 $\frac{\hat{\beta}_j - E(\hat{\beta}_j)}{SE(\hat{\beta}_j)} = \frac{\hat{\beta}_j - \beta}{SE(\hat{\beta}_j)} = \frac{\hat{\beta}_j - \beta}{\sqrt{\sigma_e^2 c_{jj}}}$   
 $\text{Cov}(\hat{\beta}) = \sigma_e^2 (X^T X)^{-1} = \sigma_e^2 C$   
 $SE(\hat{\beta}_j) = \sqrt{\sigma_e^2 c_{jj}}$

Interestingly this situation will arise that if I go instead of  $\beta$  cap If I go for  $\beta_j$  cap what will be the distribution of this will be univariate normal one variable so this will be univariate normal will okay now depending on sample size what will happen we will go for when we have taken  $\bar{x}$  if you can remember we created something like this that  $\bar{x} - \text{expected value of } \bar{x} \text{ divided by}$

standard error of  $\bar{x}$  can you remember this one is nothing but  $\bar{x} - \mu$  by  $s \sqrt{n}$  when  $\sigma$  is not known so this follows t distribution with  $n - 1$  degrees of freedom.

Now for  $\beta$  cap let us do the similar thing that if I write that  $\beta_j$  cap - expected value of  $\beta_j$  cap divided by standard error of  $\beta_j$  cap what will be this value, this value will be  $\beta_j$  cap -  $\beta$  divided by what will be the SE standard error of  $\beta_j$  cap that you require to find out how do we find out this one see you know that covariance structure of  $\beta$  cap is already given to you this one is nothing but  $Se^2 (X^T X)^{-1}$  so if I write  $(X^T X)^{-1}$  as  $C$  then this is  $Se^2$  into  $C$  so this can be written like this that  $Se^2$  will remain then how many what is the size, size is  $p + 1 \times p + 1$ .

First one is 0 then 1 like this up to  $p$  here also 0 1 up to  $p$  first one is  $C_{00}$  then  $C_{01}$  like this  $C_{0p}$  then  $C_{01}$   $C_{11}$ ,  $C_{1p}$  so then here it will be, it will be this is 2 so  $C_{02}$  so like this  $C_{0p}$  then  $C_{12}$  like  $C_{1p}$  so in this manner you will calculate then  $C_{pp}$  somewhere the  $j$ th one will be  $C_{jj}$  understood. so what I am saying that  $(X^T X)^{-1}$  is a square matrix of the order  $p + 1 \times p + 1$  and that we are writing in terms of capital  $C$  and each element is  $C_{jk}$  now  $j$  stands from 0 to  $p$   $k$  stands from 0 to  $p$  so the diagonal elements of this when multiplied by  $Se^2$  will give you standard error of  $\beta_j$  so this will be  $Se^2$  into  $C_{jj}$ .

And off diagonal will give you the covariance between the parameters estimates so you can write now that  $\beta_j$  cap -  $\beta$  by you can write this  $Se^2$  this is standard error as it is standard error we have to give the square root. This is standard error so we have to give the square root and this component is the variance component off diagonal elements are variance so  $Se^2 C_{jj}$  okay now under the assumption of null hypothesis  $H_0$ .

(Refer Slide Time: 26:55)

→  $H_0: \beta_j = 0$ ,  $x_j$  - variable.  
 $H_1: \beta_j \neq 0$

---

$$y = \beta_0 + \beta_1 x_1 + \beta_j x_j + \dots + \beta_p x_p + \epsilon$$

The image shows handwritten text on a blue background. At the top, it says '→ H0: βj = 0, xj - variable.' followed by 'H1: βj ≠ 0'. A horizontal line is drawn below these hypotheses. Below the line is the regression equation 'y = β0 + β1 x1 + βj xj + ... + βp xp + ε'. An arrow points from the βj term in the equation to the text 'xj - variable.' in the hypothesis above.

That  $\beta_0$  our null hypothesis that there is no effect of the  $x_j$  variable now what do I what do you mean our regression equation is like this  $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \epsilon$  somewhere there is suppose this is our  $\beta_j x_j$  what we are saying if  $x_j$  has no relation with  $y$  other way  $x_j$  does not contribute in explaining the variability of  $y$  then we can assume that  $\beta_j$  value is 0 so this is our null hypothesis when you test the individual regression parameters then alternative hypothesis is  $\beta_j$  not equal to 0 so if we cleared this.

(Refer Slide Time: 27:55)

normal.

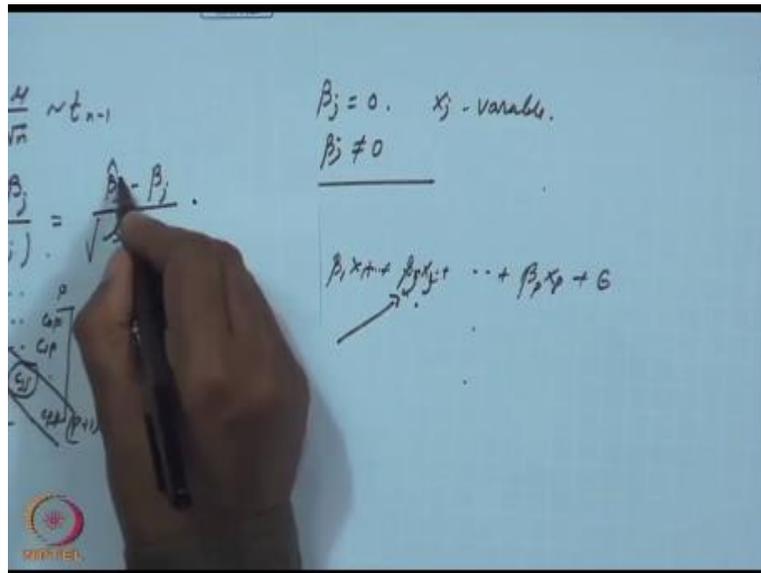
$$\frac{\bar{x} - E(\bar{x})}{SE(\bar{x})} = \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t_{n-1}$$
$$\frac{E(\hat{\beta}_j)}{SE(\hat{\beta}_j)} = \frac{\hat{\beta}_j - \beta}{\sqrt{s_e^2 g_j}} \quad \beta_j = 0$$

$x_p + 6$

$s_e^2 C = \begin{bmatrix} 0 & 1 & \dots & p \\ C_{01} & C_{11} & \dots & C_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ C_{p1} & C_{p2} & \dots & C_{pp} \end{bmatrix}$

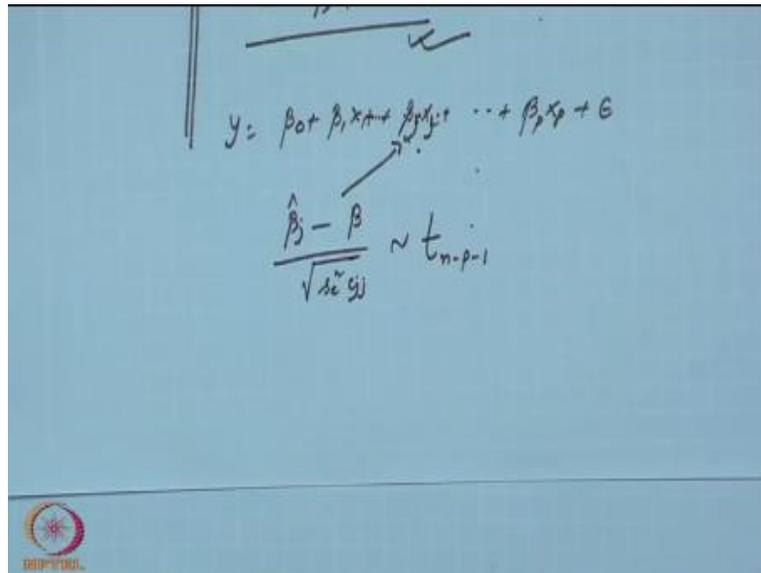
NPTEL

(Refer Slide Time: 28:09)



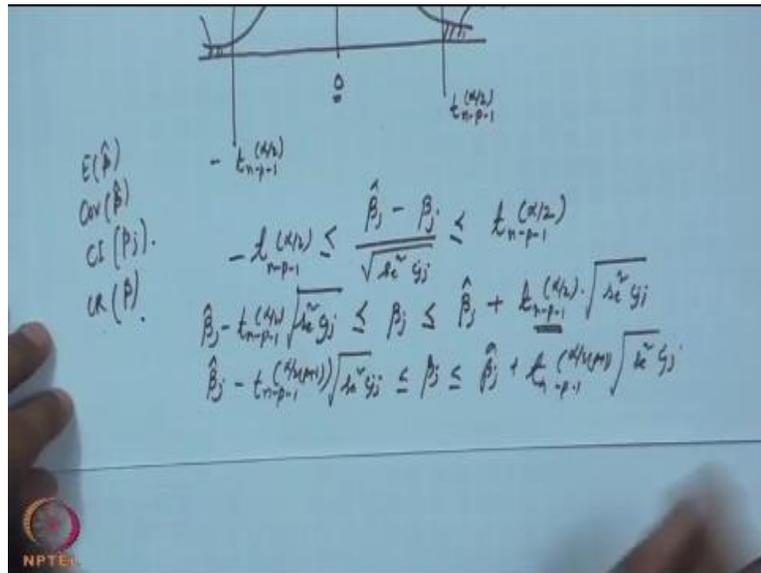
What will happen ultimately then under null hypothesis this quantity that  $\hat{\beta}_j - \beta_j$  - this will be 0  $\hat{\beta}_j$  equal to 0 so then by square root of  $SE(\hat{\beta}_j)$  getting me difficult now Sir  $\hat{\beta}_j - \beta_j$  that one  $\hat{\beta}_j$  cap -  $\beta_j$  all are  $\hat{\beta}_j$  here all are  $\hat{\beta}_j$ .

(Refer Slide Time: 28:29)


$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \epsilon$$
$$\frac{\hat{\beta}_j - \beta}{\sqrt{se^2 c_{jj}}} \sim t_{n-p-1}$$

Because we are talking about the  $j$ th variable only now I will come back to this estimate that test part later on now let us see that we will just what we are developing that confidence interval part so  $\hat{\beta}_j - \beta$  by square root of  $se^2 C_{jj}$  this will follow T distribution with  $n - p - 1$  degrees of freedom okay.

(Refer Slide Time: 29:13)



So if this true then we can write like this, this is my T distribution this is 0 this is  $t_{n-p-1}$   $\alpha$  by 2 this side is this is  $-t_{n-p-1}$   $\alpha$  by 2 this value and please remember that we are talking about  $\beta_j$  only under null hypothesis this is null hypothesis case we will not discuss now so then can we not write down now with respect to this quantity will be  $\beta_j$  cap -  $\beta$  by root over  $Se^2 C_{jj}$  less than equal to  $t_{n-p-1}$   $\alpha$  by 2 and here-  $t_{n-p-1}$   $\alpha$  by 2 student sir  $\beta_j$  okay so now manipulate little more so what you require to do then you want  $\beta_j$  here this side and this side less than equal to and less than equal to this quantity will come here as  $\beta_j$  cap +  $t_{n-p-1}$   $\alpha$  by 2 into  $Se^2 C_{jj}$  and this side it will be  $\beta_j$  cap -  $t_{n-p-1}$   $\alpha$  by 2 root over of  $Se^2 C_{jj}$ .

This is from usual individual the t distribution but again you see that it is  $n - p - 1$  degrees of freedom is much less compared to  $n - 1$  in our original  $\bar{x}$  case so what we have discussed so far we have discussed ultimately we have found out that expected value of  $\beta$  cap we found out covariance of  $\beta$  cap also we have found out the individual confidence interval for  $\beta_j$  also you know now that.

What is the confidence region for  $\beta$  with the help of  $\beta$  cap and from there using Bonferroni approach you can using yes you can use this equation only thing is that this will not be  $\alpha$  by 2 it

will be  $\alpha$  by 2 into number of parameters  $p + 1$  okay so that mean this will be  $\beta_j - t p - 1 \alpha$  by 2 into  $p + 1$  square root of  $Se^2 C_{jj}$  less than equal to  $\beta_j$  less than equal to  $\beta_j$  cap +  $t n - p - 1 \alpha$  by 2  $p + 1$  square root of  $Se^2 C_{jj}$  okay so now let us solve one problem with this whatever we have discussed so far.

(Refer Slide Time: 33:04)

CityCan Data.

<u>l</u>	<u>y = Sales.</u>	<u>x<sub>1</sub> = % Absenteeism</u>	<u>x<sub>2</sub> = Breakdown hrs</u>
1	180	9	62
2	110	8	58
3	105	7	64
4	94	14	60
5	95	12	63
6	99	10	<del>57</del>
7	104	7	55
8	108	4	56
9	<del>100</del> 105	6	59
10	<del>98</del> 98	5	61
11	<del>100</del> 105	7	57
12	110	6	60

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

We will solve one problem we will solve the problem earlier we have discussed that city can data we will consider the sales volume y this is the data set 12 months data we have collected and x1 absentee is x2 breakdown hours these two independent variables we are considering we are not considering m ratio for this explanation because the computation will be much more but you have to collect all possible data for all possible relevant variables and then you have to develop the regression equation okay with respect to this what I want to know I want to calculate  $\beta$  cap so first work is  $\beta$  cap you calculate this will be  $X^T y$  so we will not go for calculation now because already we have seen.

(Refer Slide Time: 34:07)

Handwritten notes on a whiteboard showing the calculation of the least squares regression coefficients. The notes include the following equations and matrices:

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

$$(X^T X)^{-1} = \begin{bmatrix} 40.894 & 0.114 & -0.703 \\ 0.114 & 0.012 & -0.004 \\ -0.703 & -0.004 & 0.012 \end{bmatrix}$$

$$X^T y = \begin{bmatrix} 1231 \\ 9622 \\ 72960 \end{bmatrix}$$

$$\hat{\beta} = \begin{bmatrix} 130.22 \\ -1.24 \\ -0.30 \end{bmatrix}$$

Sales volume =  $130.22 - 1.24 \text{ Abs.} - 0.30 \text{ BH} + \text{Error}$

$$y = 130.22 - 1.24 x_1 - 0.30 x_2 + \epsilon$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

$$\hat{\beta}_0 = 130.22 \quad \hat{\beta}_1 = -1.24 \quad \hat{\beta}_2 = -0.30$$

That this is the data matrix  $X^T X$  the conversion from  $X^T X$  these are the values  $(X^T X)^{-1}$  is also computed and this is the values and  $X^T y$  is this value and if you calculate this then your values are - 130.22 - 1.24 - 0.30 so that means your sales volume equal to 130.22 - 1.24 absentee is - 0.30 breakdown hours let it be BH +some error will be there so other way we can write y equal to 130.22 - 1.24 x 1 - 0.30 x2 + epsilon now what do we want we want that whether this 130.22 is really 130.22 it is not 0 you want to see that what about 1.24 what about this and two ways we have seen that one thing is that we will go for hypothesis testing.

Then using t test we will see that they are significant or not second one we have said that you create the confidence interval for each of the origin population here the population parameter is like this  $\beta_0 \beta_1 x_1 + \beta_2 x_2 + \epsilon$  this 130  $\beta_0$  cap is 130.22  $\beta_1$  cap is - 1.24 and  $\beta_2$  cap is - 0.30 so whatever we have developed so far now that we have seen here with respect to this what I want to do now we want to find out the interval confidence interval for let for  $\beta$ .

(Refer Slide Time: 36:36)

$\alpha = 0.05$ ,  $k = n - p - 1$   
 $\hat{\beta}_j \pm t_{n-p-1, \alpha/2} \sqrt{Se^2 C_{jj}}$   
 $(X^T X)^{-1}$   
 $X^T X = \begin{bmatrix} 12 & 95 & 712 \\ 95 & 845 & 5663 \\ 712 & 5663 & 42334 \end{bmatrix}$   
 $(X^T X)^{-1} = \begin{bmatrix} 40.894 & 0.114 & -0.703 \\ 0.114 & 0.012 & 0.004 \\ -0.703 & 0.004 & 0.012 \end{bmatrix}$   
 $X^T y = \begin{bmatrix} 1231 \\ 9622 \\ 72960 \end{bmatrix}$   
 $\hat{\beta} = (X^T X)^{-1} X^T y = \begin{bmatrix} 130.22 \\ -1.24 \\ 0 \end{bmatrix}$   
 $y = 130.22 - 1.24 x_1$   
 Sales Volume =

0 so you find out CI  $\beta_0$  for  $\alpha$  equal to 0.05 so what are the things you require for this if you want to calculate this what are the things you require as I written here obtain confidence interval  $\beta_j$   $j$  equal 0 to 1 for the problem what is given now  $\alpha$  equal to 0.05 you require to know this quantity you have already seen you have already seen the formulation that  $\hat{\beta}_j \pm t_{n-p-1, \alpha/2} \sqrt{Se^2 C_{jj}}$  so here our  $\hat{\beta}$  cap is  $\hat{\beta}_j$  cap is known we want to know this value  $Se^2$  we have to compute  $C_{jj}$  we also require to know and like this.

Now how do we get this values okay so if you require to calculate  $c_j$  what is this  $C_{jj}$  you require to know  $(X^T X)^{-1}$  and I have already given you that  $(X^T X)^{-1}$  is 40.894 this matrix you will be computing this matrix so that mean the diagonal elements are your variance part first one is for  $\beta_0$  second one is for  $\beta_1$  third one is for  $\beta_2$ .

(Refer Slide Time: 38:29)

Obtain CI for  $\beta_j, j=0,1,2$  for the problem given

$$\alpha = 0.05, \quad t_{n-p, \alpha/2} = t_{12-3, 0.025} \quad t_{9(0.025)} = 2.262$$

$$t_{9(0.025/2)} = t_{9(0.0125)} = 2.933$$

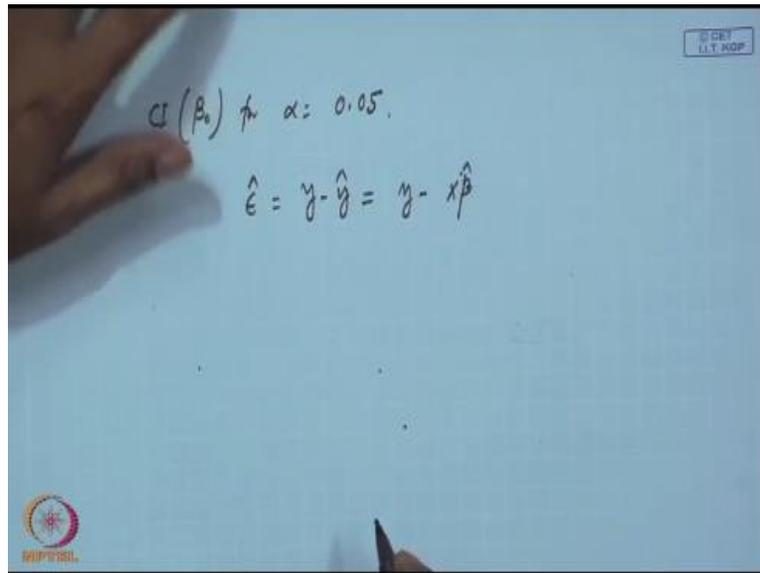
$$\hat{\beta}_j - t_{n-p, \alpha/2} \sqrt{\hat{\sigma}_e^2 c_{jj}} \leq \beta_j \leq t_{n-p, \alpha/2} \sqrt{\hat{\sigma}_e^2 c_{jj}}$$

$$(X^T X)^{-1} \hat{\beta}_0$$

$$C_{00} = 40.894$$

So if I go for  $\beta_0$  then  $\beta_0$  cap - we will write this 1 but ultimately for beta 0 that  $C_{00}$  will be taken  $C_{00}$  is 40.894 this is our  $C_{00}$  so second one will be your  $C_{00}$  third one will be your  $C_{22}$  so you will take form this matrix this values but you also require to know  $Se^2 E$  square you are knowing this value fine but what will be the  $Se^2$  value correct so  $Se^2 E$  you have to find out error that error you require to find out so for this you require to find out.

(Refer Slide Time: 39:29)



CI ( $\beta_0$ ) for  $\alpha = 0.05$ .

$$\hat{\epsilon} = y - \hat{y} = y - x\hat{\beta}$$

Epsilon cap epsilon cap is  $y - \hat{y}$  which is  $y - x\hat{\beta}$  I think  $\beta$  cap getting me.

(Refer Slide Time: 39:51)

CityCan Data.

$y = \text{Sales.}$

$X_1 = \% \text{ Absenteeism}$

$X_2 = \text{Breakdown hr}$

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

	$y = \text{Sales.}$	$X_1 = \% \text{ Absenteeism}$	$X_2 = \text{Breakdown hr}$
1	100	9	62
2	110	8	58
3	105	7	54
4	94	14	60
5	95	12	63
6	99	10	57
7	104	7	55
8	108	4	56
9	105	6	59
10	98	5	61
11	105	7	57
12	110	6	60

Now if we go to the original data original data so this is y.

(Refer Slide Time: 39:57)

$$CI(\beta_0) \text{ for } \alpha = 0.05$$
$$\hat{e} = y - \hat{y} = y - X\hat{\beta}$$
$$= \begin{bmatrix} 100 \\ 110 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ 110 \end{bmatrix}_{12 \times 1} - \begin{bmatrix} 1 & 9 & 62 \\ 1 & 8 & 58 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & 6 & 60 \end{bmatrix}_{12 \times 3} = \begin{bmatrix} 130.22 \\ -1.24 \\ -0.30 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}_{3 \times 1}$$

Now using the formula  $x \beta$  cap you have to find out  $y$  cap now then this will be nothing but if we write like this I am not writing all values some values I am writing 110 so like this finally 110, this value -  $x$  value you have to write what is your  $x$  value when you compute your  $x$  value will be all 1 then 9 8 like this then 6, 62, 58 like this 60 this is your  $x$  value so this one is 12 cross 1 this one also 12 cross 3 then your  $\beta$  cap value already is there what are those values  $\beta$  cap values you found out estimated values 130.22 - 1.24 - 0.30 so this is 3 x 1 so the resultant value will be like this  $\beta$  cap will be like this.

(Refer Slide Time: 41:13)

③

$$H = X(X^T X)^{-1} X^T$$
$$\hat{E}_0 = \begin{bmatrix} -0.64 \\ 7.12 \\ 2.68 \\ -0.82 \\ -1.41 \\ -10.69 \\ -1.02 \\ -0.45 \\ -0.07 \\ -7.71 \\ -1.42 \\ 5.23 \end{bmatrix}$$
$$E^T E = 155$$
$$= \frac{y^T (I - H) y}{155}$$

That is  $12 + 1$  I required to compute.

(Refer Slide Time: 41:31)

Handwritten notes on a whiteboard:

$$H = X(X^T X)^{-1} X^T$$

$$E = \begin{bmatrix} -0.44 \\ 7.12 \\ 2.68 \\ -0.82 \\ -1.41 \\ -1.69 \\ -1.02 \\ -0.45 \\ -0.07 \\ -7.71 \\ -1.42 \\ 5.23 \end{bmatrix}$$

$$E^T E = 155$$

$$= Y^T (I - H) Y$$

$$SSE = E^T E = 155$$

$$S_e^2 = \frac{SSE}{n - p - 1} = \frac{155}{12 - 3} = \frac{155}{9}$$

This all those things we compute and the difference will come like this so once you get this that means you know SSE, SSE is epsilon transpose epsilon so it will be  $1 \times n$ ,  $n \times 1$  that one value this value if you compute it will be coming around 155 because of rounding error there might be here and there some error but it will be around 155 then what is  $Se^2$   $Se^2$  is SSE divided by degrees of freedom  $n - p - 1$  so we can write 155 by  $n$  is our 12  $p + 1$  is 3 so 155 by 9 so 155 this 9 is known that means what are the things known to you know what you require to compute.

Here you require this so our  $Se^2$  is 155 by 9  $C_{jj}$  is known  $Se^2$  is known we require to know what will be the t value  $t_{12 - 3}$  into  $\alpha$  let it be 0.05 by 2 so that mean  $t_{9, 0.025}$  this value if we see in the table it will be 2.262 now let us put all the values here so our  $\beta_0$  our  $\beta_0$ .

(Refer Slide Time: 43:45)

The image shows handwritten mathematical derivations for confidence intervals of three regression coefficients,  $\beta_0$ ,  $\beta_1$ , and  $\beta_2$ . The derivations are as follows:

$$\hat{\beta}_0: 130.22 - 2.262 \times \sqrt{\frac{155}{9} \times 40.894} \leq \beta_0 \leq 130.22 + 2.262 \times \sqrt{\frac{155 \times 40.894}{9}}$$

$$\hat{\beta}_1: -1.24 - 2.262 \sqrt{\frac{155}{9} \times 0.012} \leq \beta_1 \leq -1.24 + 2.262 \sqrt{\frac{155 \times 0.012}{9}}$$

$$\hat{\beta}_2: -0.30 - 2.262 \sqrt{\frac{155}{9} \times 0.012} \leq \beta_2 \leq -0.30 + 2.262 \sqrt{\frac{155 \times 0.012}{9}}$$

Is  $130.22 - t_{9, 0.025} - 2.262$  into square root of  $155$  by  $9$  into  $C_{00}$  is  $40.894$  that less than equal to  $\beta_j$  less than equal to you will be getting some  $130.22 + 2.262$  into square root of  $155$  into  $40.894$  divided by  $9$  now you have to see that what is this value how much it will come some range we will be getting similarly, if you want to use it for the second coefficient  $\beta_1$  it will be - this is for  $\beta_0$  cap  $\beta_0$  this is for  $\beta_1$  confidence interval that  $- 1.24 - 2.262$  square root of  $155$  by  $9$  into  $C_{11}$  now  $C_{11}$  value is how much  $C_{11}$  value is  $0.012$  less than equal to here it is  $\beta_0$  the second one is  $\beta_1$  less than equal to  $- 1.24 + 2.262$  square root of  $155$  into  $0.012$  by  $9$ .

So  $\beta_2$  for  $\beta_2$  what will happen  $\beta$  will be  $- 0.30 - 2.262$  square root of  $155$  by  $9$  into I think the  $C_{22}$  also same value here  $0.12$  less than equal to  $\beta_2$  less than equal to  $- 0.30 + 2.262$  into  $155$  by  $9$  into  $0.012$  you will be getting all the intervals now if any interval content  $0$  then that parameter is not significant this is in nutshell what is basically what we will talk about that sampling distribution of sampling distribution of  $\beta$  cap sampling distribution of  $\beta$  cap but similarly what will happen suppose you.

(Refer Slide Time: 46:47)

$$y = x\beta + \epsilon$$

$$\hat{\epsilon} = y - \hat{y} = y - x\hat{\beta}$$

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

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

$$= y \left[ I - X(X^T X)^{-1} X^T \right]$$

$$= y(I - H)$$

Hat matrix

Are you know that  $y$  equal to  $x\beta + \epsilon$  now also we know that this one is  $y - \hat{y}$  that means  $y - x\hat{\beta}$  you may be interested to know what is the distribution of this how do I find out can we not get this now the is the this distribution similar to epsilon please when we define in the population domain we say this is epsilon if it is  $x_i$  then epsilon  $i$  the distribution now you are finding out epsilon  $i$  cap for this case so similarly that is why I am saying that epsilon in general that cap will it be same then as the population epsilon it will not be the same so here because this distribution.

The distribution will be governed by this beta cap what you are estimating here so little bit some clue I am giving you here this is  $y$  minus  $\hat{y}$  equal to  $y - x\hat{\beta}$  we have used earlier  $y = x\beta$  and  $x\hat{\beta}$  suppose I want to write down instead of  $\hat{\beta}$  cap I will write down  $(X^T X)^{-1} X^T y$  then what will happen this one will become  $y - (X^T X)^{-1} X^T y$  but one  $x$  is there so you write  $x$  here please go through  $y - x\hat{\beta}$  cap so I am writing the same  $y - x$  into  $\hat{\beta}$  cap is  $(X^T X)^{-1}$  now this one I can write like this  $I - X(X^T X)^{-1} X^T$  this matrix is known as hat matrix.

So hat matrix is denoted as H, H equal to  $X(X^T X)^{-1} X^T$  hat matrix is very popular in regression and in diagnostics of MLR multiple regression this hat matrix will be used it is the projection property basically this one is the project done from the that direct planes x plane and y plane so these are the projected ones so that mean we can write this one as  $y(I - H)$  so you require to your  $I - H$  into  $I$  think  $I - H$  into  $y$  you have to see that the this one matrix multiplication that comfort ability so we can write  $I - H$  epsilon cap equal to  $I - H$  into  $y$  I think this is comfortable one.

(Refer Slide Time: 50:57)

$$\hat{\epsilon} = (I-H)y.$$

$$E(\hat{\epsilon}) = (I-H)(X\beta + \epsilon)$$

$$= E[(I-H)X\beta + (I-H)\epsilon]$$

$$= E[(I-H)X\beta] + E[(I-H)\epsilon]$$

$$Cov(\hat{\epsilon}) = \sigma^2(I-H).$$

$$E[X\beta - X(X^T X)^{-1} X^T X\beta]$$

$$= E[X\beta - X\beta] = 0$$

$(I-H)^T(I-H) = I-H$   
 $AA = A$

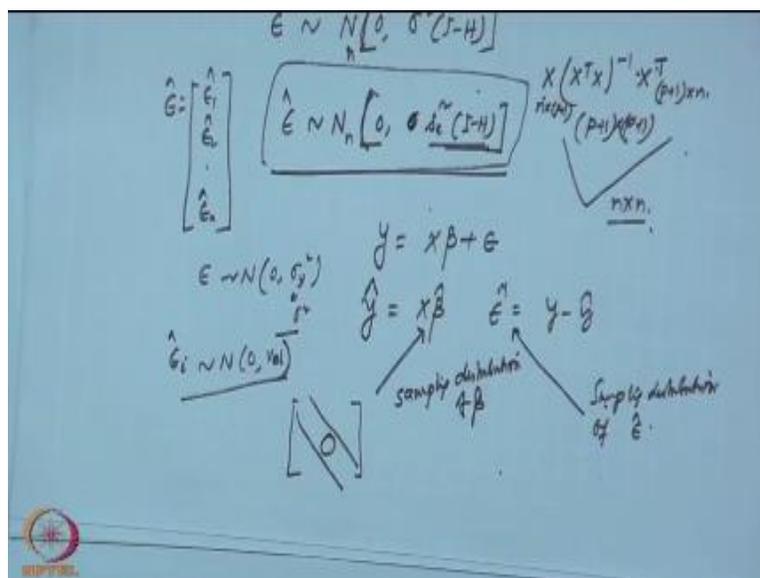
This is comfortable one now what will be the expected value of this then  $I - H$  I will write again  $x\beta + \epsilon$  here see this cap and this is not same so you can write this  $I - H$   $x\beta$   $I - H$  epsilon you require to give an expectation operator here now H is your  $(X^T X)^{-1}$  now x I think what we will define by as  $X(X^T X)^{-1} X^T$  that is our H these are all fixed values so that means  $I - H$  is fixed values x is also fixed value now  $\beta$  is constant correct so I can write this one now straight way I can bring it to expected value of this beta of this  $I - H$   $x\beta$  this is the these are the fixed values and expected value of  $I - H$  epsilon.

What will be the this what will be this one, expected value of epsilon 0 that will be 0 what will be this one it will be the fixed value fixed value so, that means some value will be there so

similarly you require to find out the covariance matrix of this the resultant covariance matrix will be I am straight away righting the resultant part this will be  $\sigma^2 I - H$  so  $I - H$   $I - H$  transpose those things will also come so  $I - H$  suppose transpose  $I - H$  this will become  $I - H$  only because the property of this  $I - H$  this is this is a matrix which is known as idempotent matrix.

That  $A$  into  $A$  equal to  $A$  so the resultant value will be this so you know the this value I think we have to this one is  $x\beta$  if we just do little bit manipulation also because  $I - H$  is there because  $\beta$  is there  $\beta$  will come this also will become 0 you just do manipulation what will happen this is  $x\beta - H$  into  $x\beta$  so I will write  $x\beta$  here so that means  $X^T$  that is  $I$  that will be  $x\beta$  so that  $I - H$   $x\beta$  what will happening this quantity will become  $E$  i am again writing  $x\beta - H$  is  $x(X^T X)^{-1} X^T \beta$  this quantity what I so that will become  $x\beta - x\beta$  so that so it is a fixed value definitely but ultimately that fixed value is 0.

(Refer Slide Time: 54:45)



So our now the distribution of this one is normal distribution with 0 and  $\sigma^2 I - H$  will it be multivariate or univariate why what is what is  $x(X^T X)^{-1}$  this is  $p + 1 \times p + 1$  if we multiply it with, what is this  $x$  is there then  $X^T$  is there what is  $x$ ,  $x$  is  $n$  into  $p + 1$  and this one is  $p + 1$  into  $n$  so what will be the resultant quantity  $m \times n$   $m \times n$  so this is multivariate normal with  $n$  so  $n$  you

are getting  $\epsilon$  this you will be getting  $\epsilon_1$   $\epsilon_2$  like this  $\epsilon_n$  all values so it is  $n$  you can write this is multivariate normal with  $N(0, \sigma^2 I)$  I can write  $Se^2$  I think you now.

With respect to this  $y$  equal to  $x\beta + \epsilon$  we have computed that  $y$  this one is this one and  $\epsilon$  is our  $y - \hat{y}$  so with related to this sampling distribution of sampling distribution of  $\beta$  also you now you now know what are the mean and covariance and this also this is the distribution of  $\epsilon$  so sampling distribution of what is the learning here apart from this beta all those things that  $\epsilon$  its distribution is not the distribution of  $\epsilon$  we have assumed the error terms.

In our population domain this we say it is normally distributed each error is normally distributed  $0$  mean and  $\sigma^2$  and this  $\sigma^2$  we have written like this but you are getting when you are calculating the error terms, that we are getting in multivariate domain you are getting a multivariate normal distribution for the error term now if you take a particular suppose  $\epsilon_i$  what will happen this definitely univariate normal with  $0$  and what will be mean  $0$  and what will be the variance component. Variance component will be this one so those particular variance components you have to find out because it will be a matrix of you have seen already.

You have seen it is a matrix of  $n$  cross  $n$  so you will be having  $n$  cross  $n$  matrix so is the term if I if I can write variance  $\text{ve}$   $\text{vei}$ ,  $\text{vei}$  that is the term I am just term I am writing here. I think the basic statistics part is very important that is why everywhere you are coming to some point when the statistics to be, to be used and that the sampling distribution of the statistics is to be known otherwise you cannot do this because many a times you will be using why  $t$  distribution why not other distribution and how regression parameters are becoming following  $t$  distribution similar things will be there. So we have covered up to sampling distribution next I think we will go for goodness of it okay next class I think tomorrow if possible.

**NPTEL Video Recording Team**  
**NPTEL Web Editing Team**  
**Technical Superintendents**

**Computer Technicians**

**A IIT Kharagpur Production**

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

**Copyrights Reserved**