Week 8

Multiple Regression [MLR]

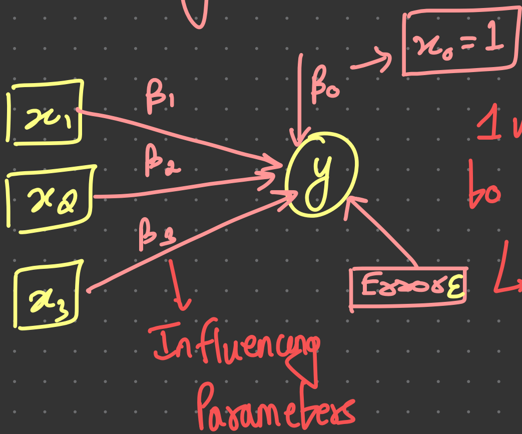
* Conceptual model
* Estimation of parameters ($\hat{\beta}$)
* Sampling distribution of $\hat{\beta}$
* Sampling distribution of $\hat{\varepsilon}$
* Adequacy of regression model
* Test of individual regression parameters (β)
* Test the assumption
* Dignostic issues
* Prediction using MLR

Data

Dependent Variable [DV]

Independent Variable [IDV]

Try to identify if IDV affects DV one at a time



1 unit of change in x_1 corresponds to β_1 change

Similar for x_2, x_3

$\boxed{\beta_0}$ is taken as a constant term so that if x_1 ... x_n's don't contribute, we will still have a defined term

Also called intercept

$$\boxed{y = \beta_0 x_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon}$$

$\Downarrow$

$$\boxed{y = \beta_0 + \beta_1 x_1 - \quad + \beta_n x_n + \varepsilon}$$

General regression equation

Generalised Case :

$$X = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}_{(p+1)\times 1} \qquad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}_{(p+1)\times 1}$$

$p \times 1$ variable vector

For n-data points to be collected,

DV $\qquad\qquad$ IV

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_i \\ y_n \end{bmatrix}_{n\times 1} \qquad X_{n\times p} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{i1} & & & x_{ip} \\ x_{n1} & & - & x_{np} \end{bmatrix}_{n\times p}$$

$$y_i = \underbrace{\beta_0 + \beta_1 X_{11} + \quad \beta_p X_{1p}}_{\text{Weighted linear combination}} + \varepsilon_1$$

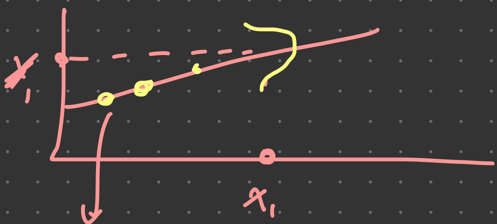
Weighted linear combination
(Variate)
↳ Represents expected value of y.

$$y_1 = \underbrace{E(y_1 | X_1 = x_{12}, X_{12} \quad x_{1p})}_{\hat{y}_1} + \varepsilon_1$$

$$\hat{y}_1 \Rightarrow \boxed{\varepsilon_1 = y_1 - \hat{y}_1}$$

Can arise around a point

$\mu_{y_1|x_1}$ is taken



All these points correspond the expected value

$$\underset{n\times 1}{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}} = \underset{n\times(p+1)}{\begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_p \\ & \vdots & & & \\ 1 & x_{n1} & \cdots & & x_{np} \end{bmatrix}} \underset{(p+1)\times 1}{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}} + \underset{n\times 1}{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}}$$

$n\times(p+1)$ $(p+1)\times 1$ → $n\times 1$

$n\times(p+1)$ Data/Design Matrix

$(p+1)\times 1$ Regression coefficients

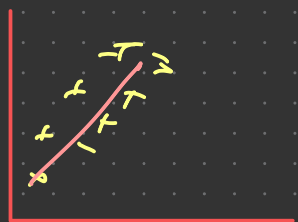
$$n\times 1 \leftarrow \boxed{y = X\beta + \varepsilon} \rightarrow n\times 1 \text{ error terms}$$

Lecture 2

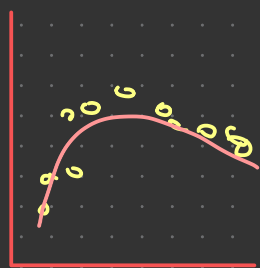
Assumptions:

1. Linearity
2. Homoskedasticity or equal Y variance across the values of X (IV's)
3. Uncorrelated error terms
4. Normality of error terms

①



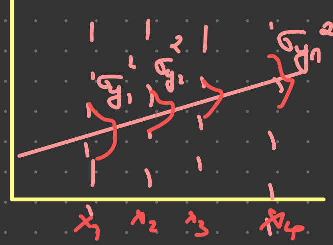
Linear



Not applicable as its not linear

→ Usually transform x-variable

②



→ Usually transform y-axis

Several y-values for a particular X

Then variability observed of Y must be equal

$\sigma_{y_1}^2 = \sigma_{y_2}^2 \ldots = \sigma_{y_n}^2$

→ If violated, y is not homogenous from variability point of view

③ Uncorrelated error terms

$$\varepsilon_i \sim N(0, \sigma_y^2)$$

$$\sigma^2 = \sigma_y^2 = \sigma_{y_1}^2 = \sigma_{y_2}^2 \cdots$$

for any 2 σ_i, σ_k

$$Cov(\sigma_i, \sigma_k) = 0$$

↓

Uncorrelated

$$\begin{bmatrix} \ddots & & 0 \\ & \ddots & \\ 0 & & \ddots \end{bmatrix}_{n \times n}$$

Estimation of model parameters:

$$y = X\beta + \varepsilon \longleftarrow \text{MLR}$$

$$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} \quad \beta_p X_{ip} + \varepsilon_i$$

$$\left.\varepsilon_i = y_i - \sum_{j=0}^{p} \beta_j X_{ij} \quad \right\} \text{ for } i^{th} \text{ variable}$$

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left[y_i - \sum_{j=1}^{p} \beta_j X_{ij} \right]^2$$

We try to minimise the squares as possible

↳ Chose beta (β) such that SSE is minimised

$$\frac{\partial(SSE)}{\partial\beta} = 0 \quad \text{subject to} \quad \frac{\partial^2(SSE)}{\partial\delta\beta_i\,\partial\beta_j} > 0$$

$\begin{bmatrix} \quad \end{bmatrix}$ → positive mat[r]x → $x^T A x = 0$

$$\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$$\boxed{SSE = \varepsilon^T \varepsilon \cdot y}$$

$$\frac{\partial SSE}{\partial \beta} = -2x^T(y - x\beta) = 0$$

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

$$x^T x \beta = x^T x$$

$$\underbrace{(x^T x)^{-1}(x^T x)}_{\text{Gives identity matrix}}\beta = (x^T x)^{-1} x^T y$$

$x^T x$
↳ SSCP

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

→ Random Variable

Population Parameter

$Y = X\beta + \varepsilon$

$E(\hat{\beta}) = \beta$ → Constant & Unknown

If only 1 independent variable, it is called simple regression ($p+1=2$)

$y = f(x)$

$y = \beta_0 + \beta_1 x + \varepsilon$

For $p+1 \geq 3$, multiple regression

Example in 23

Lecture 3

Sampling distribution of regression coefficients:

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

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ & & & & \\ 1 & x_{n1} & \cdots & \cdots & x_{np} \end{bmatrix}_{n \times (p+1)}$$

$$\hat{\beta}^{(1)}$$

$$\hat{\beta}^{(2)}$$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n \times 1}$$

$$\hat{\beta}_{(p+1)\times 1} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{bmatrix}$$

→ $\hat{\beta}$ is a random variable

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

$$E[\hat{\beta}] = E\left[(X^T X)^{-1} X^T (X\beta + \varepsilon)\right]$$

$$'' = E\left[\underbrace{(X^T X)^{-1} X^T X}_{I}\beta + (X^T X)^{-1} X^T \varepsilon\right]$$

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

$$= \beta + (X^T X)^{-1} X^T 0$$

$$= \beta$$

$$\boxed{E(\hat{\beta}) = \beta}$$ → Unbiased estimation

To find covariance of $\hat{\beta}$,

$$Cov(\hat{\beta}) = E\left[\{\hat{\beta} - E(\hat{\beta})\}\{\hat{\beta} - E[\hat{\beta}]\}^T\right]$$

$$Cov(\hat{\beta}) = E\left[(\hat{\beta}-\beta)(\hat{\beta}-\beta)^T\right]$$

$$\hat{\beta} - \beta = \cancel{\beta} + (X^TX)^{-1}X^T\varepsilon - \cancel{\beta}$$
$$= (X^TX)^{-1}X^T\varepsilon$$

$$Cov(\hat{\beta}) = E\left[\left[(X^TX)^{-1}X^T\varepsilon\right]\underbrace{\left[(X^TX)^{-1}X^T\varepsilon\right]^T}\right]$$

$$Cov(\hat{\beta}) = E\left[(X^TX)^{-1}X^T\varepsilon\varepsilon^TX(X^TX)^{-1}\right]$$

$\hookrightarrow (\hat{\beta}-\beta)^T = \left[(X^TX)^{-1}X^T\varepsilon\right]^T$

$(\hat{\beta}-\beta)^T = \varepsilon^TX\underbrace{(X^TX)^{-1}}$

$\hookrightarrow$ Symmetric

$$Cov(\hat{\beta}) = (X^TX)^{-1}X^T\sigma^2 I X(X^TX)^{-1}$$

$$Cov(\hat{\beta}) = \sigma^2(X^TX)^{-1}X^TX(X^TX)^{-1}$$

$$\boxed{Cov(\hat{\beta}) = \sigma^2(X^TX)^{-1}} = \boxed{s_e^2(X^TX)^{-1}}$$

Not known

$$\sigma^2 = s^2 = \frac{SSE}{n-(p+1)}$$

$\hookrightarrow$ This many degrees of freedom is lost

$\hat{\beta}$ is a RV with $\hat{\beta} \sim N_{p+1}(\beta, \sigma^2(X^TX)^{-1})$

$$X \sim N_p(\mu, \Sigma)$$
$$\bar{X} \sim N_p(\mu, \Sigma/n)$$

We know that

$$P\left\{n(\bar{x}-\mu)^T S^{-1}(\bar{x}-\mu) \leq \frac{(n-1)p}{n-p} F_{n,np}\right\} = 1-\alpha$$

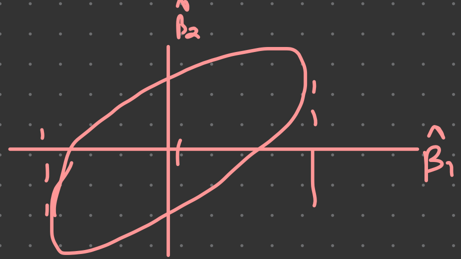
$$P\left\{(\hat{\beta}-\beta)^T(X^TX)(\hat{\beta}-\beta) \leq s_e^2(p+1)F_{p+1,\,n-(p+1)}(\alpha)\right\} = 1-\alpha$$

↓

Confidence region

⇓

Simaltaneous CI [With $\hat{\beta}$ we find CI for β



$\hat{\beta}_j$ = Univariate normal

↳ Found using Bonferroni Approach

$$\bar{x} \longrightarrow \frac{\bar{x}-E(\bar{x})}{SE(\bar{x})} = \frac{\bar{x}-\mu}{s/\sqrt{n}} \sim t_{n-1}$$

$$\frac{\hat{\beta}_j - E(\hat{\beta}_j)}{SE(\hat{\beta}_j)} = \frac{\hat{\beta}_j - \beta}{SE(\hat{\beta}_j)} = \frac{\hat{\beta}_j - \beta}{\sqrt{s_e^2 C_{jj}}}$$

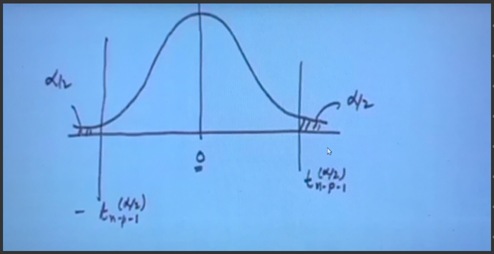
$$Cov(\hat{\beta}) = s_e^2(X^TX)^{-1} = s_e^2 C = s_e^2 \begin{array}{c|ccc|l} & 0 & 1 \cdots & p & \\ \hline & c_{00} & \cdots & c_{0p} & 0 \\ & \vdots & & & \vdots \\ & c_{p0} & \cdots & c_{pp} & p \end{array}_{(p+1)\times(p+1)}$$

$$C \rightarrow (X^TX)^{-1}$$

$$\boxed{SE(\hat{\beta}_j) = \sqrt{s_e^2 C_{jj}}}$$

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

$$\longrightarrow \frac{\hat{\beta}_j - \beta}{\sqrt{s_e^2 C_{jj}}} \sim t_{n-(p+1)}$$



$$-t_{n-p-1}^{(\alpha/2)}$$

$$-t_{n-p-1}^{(\alpha/2)} \le \frac{\hat{\beta}_j - \beta_j}{\sqrt{s_e^2 g_j}} \le t_{n-p-1}^{(\alpha/2)}$$

$$\hat{\beta}_j - t_{n-p-1}^{(\alpha/2)}\sqrt{s_e^2 g_j} \le \beta_j \le \hat{\beta}_j + t_{n-p-1}^{(\alpha/2)}\cdot\sqrt{s_e^2 g_j}$$

$$\hat{\beta}_j - t^{(\alpha/2(p+1))}\sqrt{s_e^2 g_j} \le \beta_j \le \hat{\beta}_j + t_{n-p-1}^{(\alpha/2(p+1))}\sqrt{s_e^2 g_j}$$

Lecture-4

Gby Can Data example in the start
↳ Find $\beta_0, \beta_1, \beta_2$

$$\varepsilon^T\varepsilon = y^T(I-H)y \quad \text{where } H = X(X^TX)^{-1}X^T$$

$$SSE = \varepsilon^T\varepsilon$$

$$s_e^2 = \frac{SSE}{n-(p+1)}$$

$$\hat{\varepsilon} = y - \hat{y} = y - X\hat{\beta}$$

$$\hat{\varepsilon} = y - X(X^TX)^{-1}X^Ty$$

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

$$\hat{\varepsilon} = y[I-H]$$ ↳ Hat matrix

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

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

$$E(\hat{\varepsilon}) = E\left[(I-H)X\beta + (I-H)\varepsilon\right]$$

$$\boxed{E[\hat{\varepsilon}] = E\left[(I-H)X\beta\right]} + \underbrace{E\left[(I-H)\varepsilon\right]}_{0}$$

$$\underbrace{X(X^TX)^{-1}X^T}_{X}X\beta$$

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

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

$$E(\hat{\varepsilon}) = E\left[X\beta - X(X^TX)^{-1}X^TX\beta\right]$$

$$E(\hat{\varepsilon}) = E\left[\cancel{X\beta} - X\beta\right] = 0$$

$$\boxed{E(\hat{\varepsilon}) = 0}$$

$$\begin{bmatrix} \hat{\varepsilon}_1 \\ \hat{\varepsilon}_2 \\ \vdots \\ \hat{\varepsilon}_n \end{bmatrix} \qquad \hat{\varepsilon} \sim N_n\left[0, \sigma^2(I-H)\right] \sim N_n\left(0, s_e^2(I-H)\right)$$

Thus a multivariate normal with 'n' as

$$\underbrace{\underset{n\times(p+1)}{X}\,\underset{(p+1)\times(p+1)}{(X^TX)^{-1}}\,\underset{(p+1)\times n}{X^T}}_{n\times n}$$

Finally, $y = X\beta + \varepsilon$

$\hat{y} = X\hat{\beta} \longrightarrow$ Sampling distribution of β

$\hat{\varepsilon} = y - \hat{y} \longrightarrow$ Sampling distribution of ε

Taking particular $\hat{\varepsilon}_1 \sim N(0, v_{e_1})$ → Univariate normal

v_{e_1} get from matrix $s_e^2(I\,1)$

Lecture-5

Multiple Linear Regression Model Adequacy Tests

$y = X\beta + \varepsilon$

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

$E[\hat{\beta}] = \beta$

$Cov(\hat{\beta}) = s_e^2 (X^T X)^{-1}$

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

$E(\hat{\varepsilon}) = 0$

$Cov(\hat{\varepsilon}) = s_e^2 (I - H)$

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

* Goodness of Fit Test

$R^2 = \dfrac{\text{Explained variance of } Y}{\text{Total variance of } Y}$

R^2 → Coefficient of determination

$y \rightarrow s_y^2$

SST [Sum Square Total]

→ $(n-1) s_y^2$

→ n = No of observations

This SST divided into ① Explained
② Unexplained

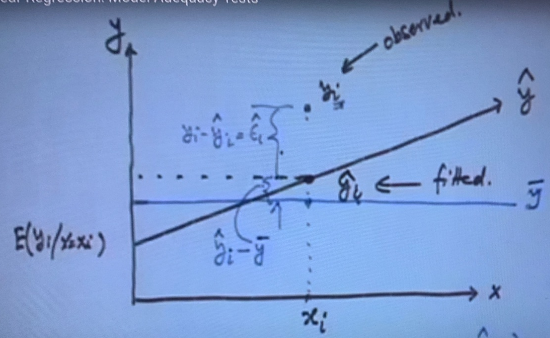
$SST = SSR + SSE$

Sum square total

Sum square Regression [Explained]

Sum square error [Unexplained]

e Linear Regression: Model Adequacy Tests



$$y_i = \bar{y} + (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i).$$

$$y_i - \bar{y} = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i)$$

: Multiple Linear Regression: Model Adequacy Tests

Watch later Share

$$\sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$SST = SSR + SSE$$

$$n-1 = p + n-p-1$$

i	y
1	y_1
2	y_2
.	
.	

$$s_y^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2 = \frac{SST}{n-1}$$

$$R^2 = \frac{SSR}{SST} \geq .90$$

for engineering applications or Experiments

< 0.90 \ Social / Administrative / Management Sciences

YouTube

5:38 / 23:24

n = max{30, 5p}

egression: Model Adequacy Tests

$$R^2 = \frac{SSR}{SST} = \frac{SST - SSE}{SST}$$

$$= 1 - \frac{SSE}{SST} \qquad s_e^2 = \frac{SSE}{n-p-1}$$

$$= 1 - \frac{s_e^2 (n-p-1)}{s_y^2 (n-1)}$$

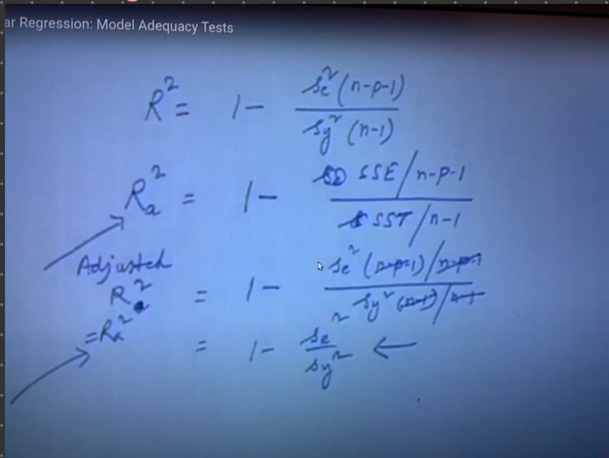
Cond^n: $\boxed{n = p+1}$ $\quad R^2 = 1 - \frac{s_e^2 [\overset{=0}{n-n}]}{s_y^2 (n-1)} = 1$.

R^2 is sensitive to n and p.

$n >> p$

Generally $5p \leq n$ and desirable $n = 10p$ [too much is also not good] Always min $n = 30$

To avoid dependency on n, p

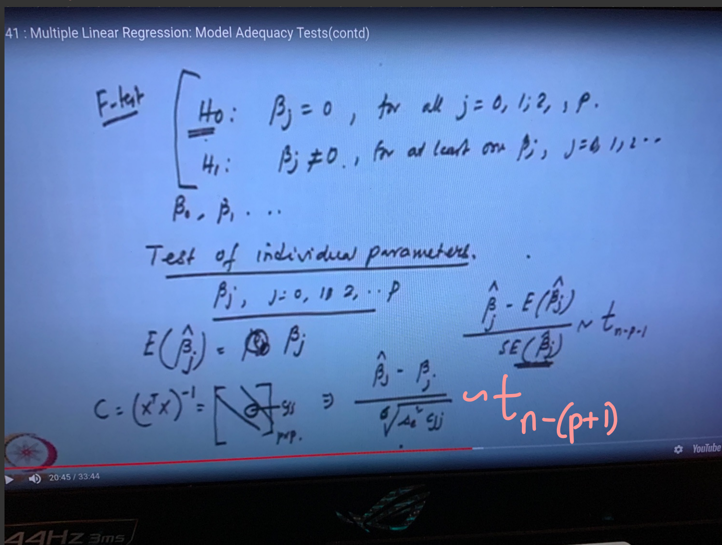


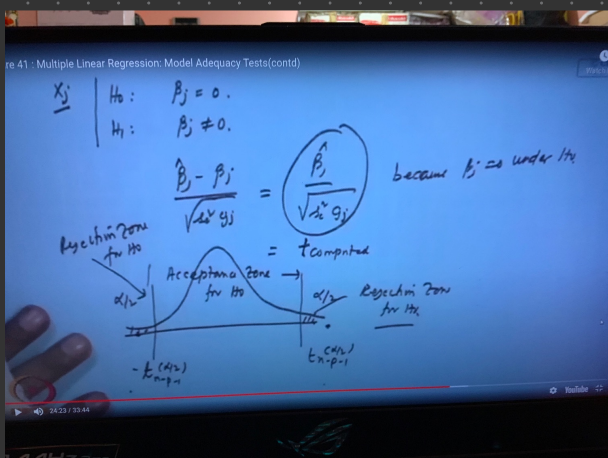
Lecture-6

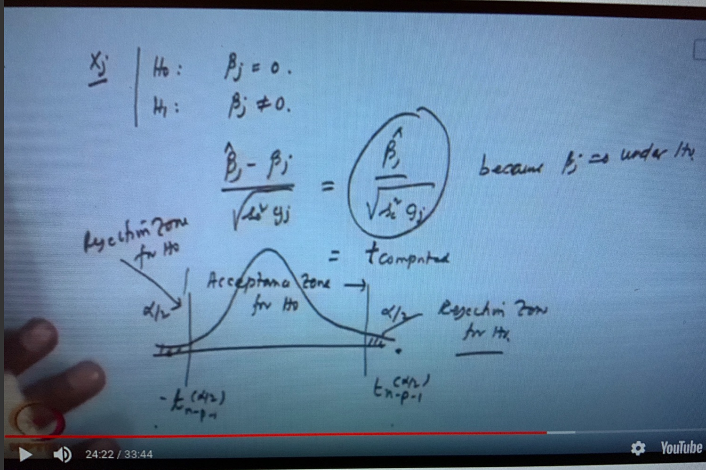
Example in the start

$$\boxed{R_e^2 \leq R^2} \rightarrow \text{Always}$$

Creating ANOVA type table





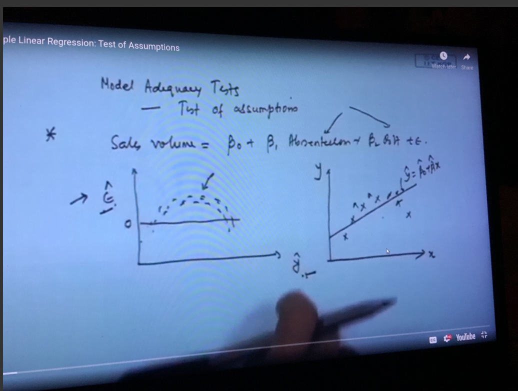


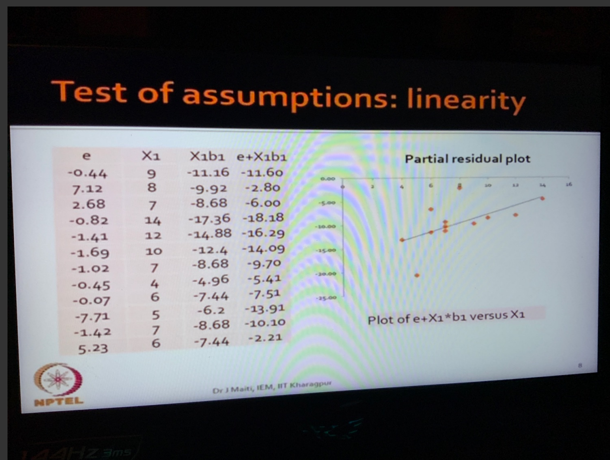
Ask a question

Lecture 7

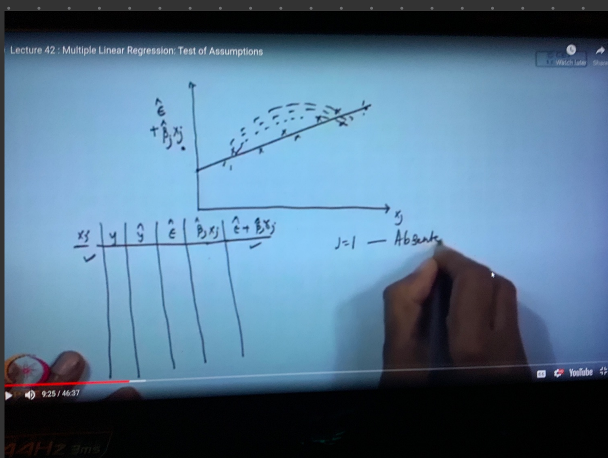
Model Adequacy Tests [Tests of Assumptions]

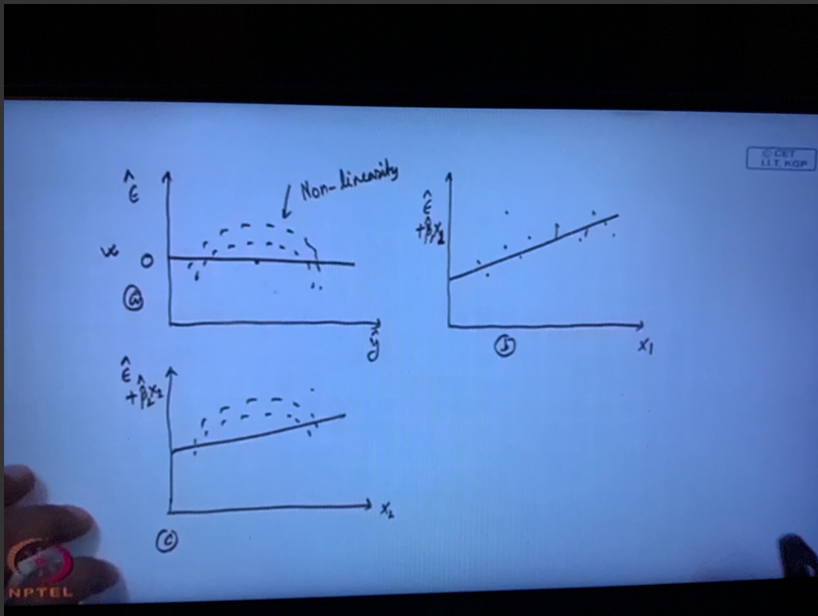
① Linearity :





If diff from linearity its like form curve around the regression [residuals]





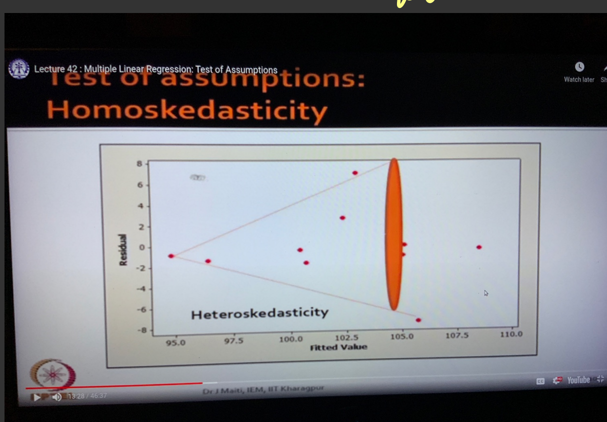
ⓐ These is non-linearity

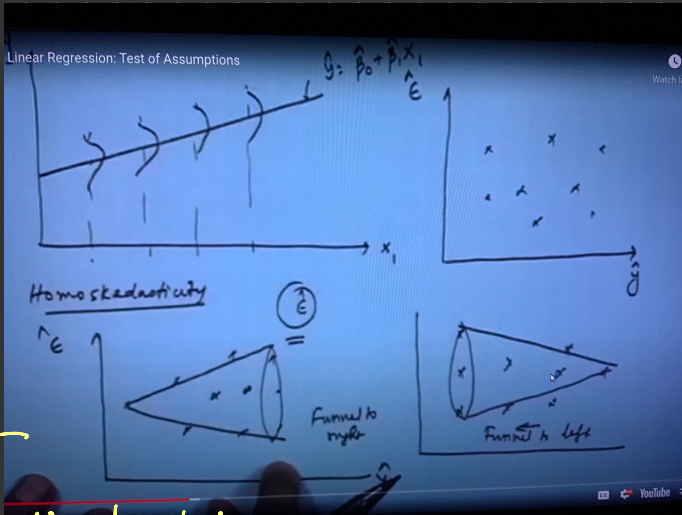
ⓑ There is non-linearity but not due to X_1

ⓒ There is non-linearity due to X_2

② Homoskedasticity :

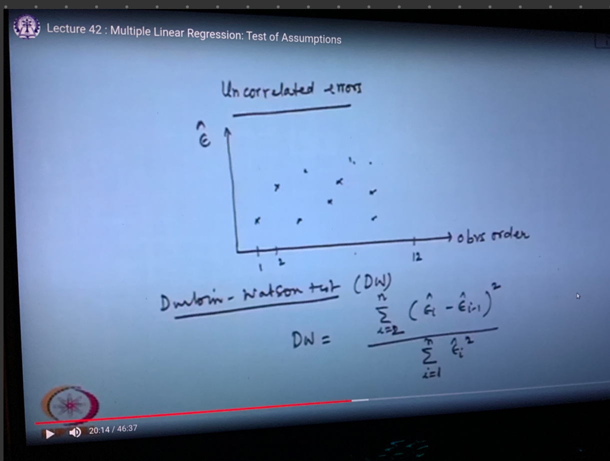
Equal error variance

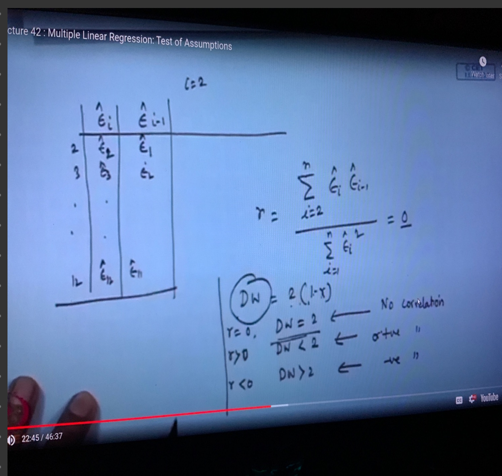




Problem with dependent variable (y)

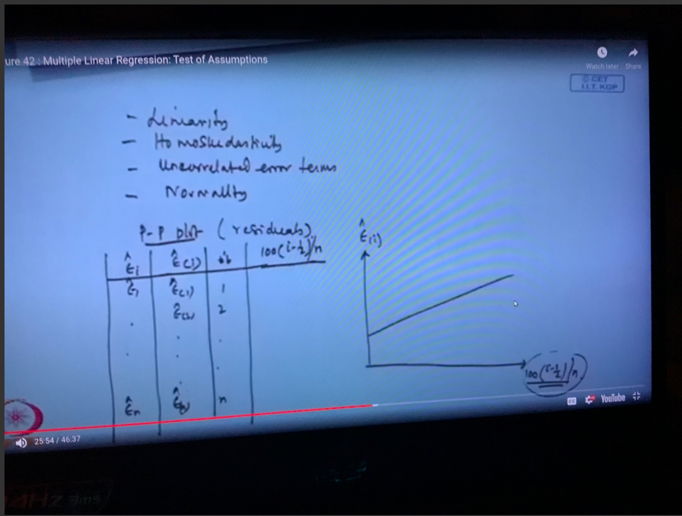
3 Uncorrelated errors



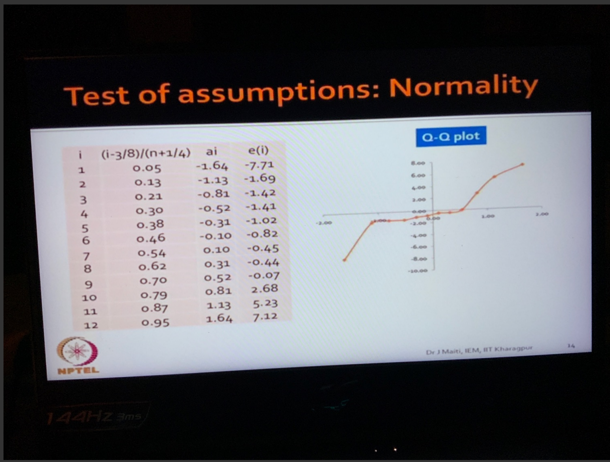


④ Normality

p-p plot is related to residuals



Usually when heteroskedasticity, normality is usually ~~the~~ further



i	(i-3/8)/(n+1/4)	ai	e(i)
1	0.05	-1.64	-7.71
2	0.13	-1.13	-1.69
3	0.21	-0.81	-1.42
4	0.30	-0.52	-1.41
5	0.38	-0.31	-1.02
6	0.46	-0.10	-0.82
7	0.54	0.10	-0.45
8	0.62	0.31	-0.44
9	0.70	0.52	-0.07
10	0.79	0.81	2.68
11	0.87	1.13	5.23
12	0.95	1.64	7.12

Lecture 42 : Multiple Linear Regression: Test of Assumptions

Remedy against violation of assumptions

- Heteroskedascity: Transform y, Box-Cox method
- Linearity: Transform y, x or both
- Normality: Box-Cox Method

Dr J Maiti, IEM, IIT Kharagpur

Lecture 42 : Multiple Linear Regression: Test of Assumptions

Remedy: Heteroskedasticity

Relationship of σ^2 to $E(y)$	Transformation
$\sigma^2 \alpha$ constant	$y' = y$
$\sigma^2 \alpha\ E(y)$	$y' = \sqrt{y}$
$\sigma^2 \alpha\ E(y)[1-E(y)]$	$y' = \sin^{-1}(\sqrt{y}) \quad (0 \le y_i \le 1)$
$\sigma^2 \alpha\ [E(y)]^2$	$y' = \text{In(y)}$
$\sigma^2 \alpha\ [E(y)]^3$	$y' = y^{-1/2}$
$\sigma^2 \alpha\ [E(y)]^4$	$y' = y^{-1}$

(Montgomery et al., 2003)

Like binomial case

watch

~~Read~~ from 32

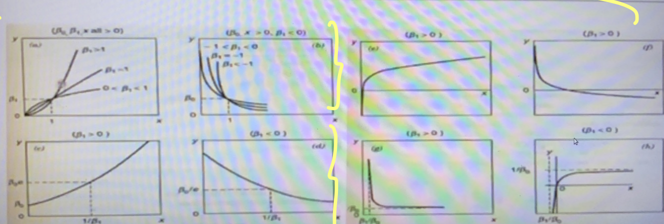
Lecture 42 : Multiple Linear Regression: Test of Assumptions

Remedy: Linearity

e,f

a,b

c,d



(Montgomery et al., 2003)

g,h

Dr J Maiti, IEM, IIT Kharagpur

Lecture 42 : Multiple Linear Regression: Test of Assumptions

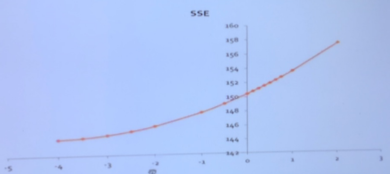
Remedy: Linearity

Figure	Linearizable Function	Transformation	Linear form
a & b	$y = \beta_0 x^{\beta_1}$	$y' = \log y,\ x' = \log x$	$y' = \log \beta_0 + \beta_1 x'$
c & d	$y = \beta_0 e^{\beta_1 x}$	$y' = \ln y,$	$y' = \ln \beta_0 + \beta_1 x$
e & f	$y = \beta_0 + \beta_1 \log x$	$x' = \log x$	$y' = \beta_0 + \beta_1 x'$
g & h	$y = \dfrac{x}{\beta_0 x - \beta_1}$	$y' = \dfrac{1}{y},\ x' = \dfrac{1}{x}$	$y' = \beta_0 - \beta_1 x'$

(Montgomery et al., 2003)

Dr J Maiti, IEM, IIT Kharagpur

Remedy: Normality & heteroskedasticity (Box-Cox method)



Choose the λ for which SSE(λ) is the minimum

Dr J Maiti, IEM, IIT Kharagpur

Lecture 42 : Multiple Linear Regression: Test of Assumptions

Remedy: Normality & heteroskedasticity (Box-Cox method)

$$y^{(\lambda)} = \begin{cases} \dfrac{y^{\lambda} - 1}{\lambda \dot{y}^{\lambda - 1}}, & \lambda \neq 0 \\ \dot{y} \ln y, & \lambda = 0 \end{cases}$$

$$y^{(\lambda)} = X\beta + \varepsilon$$

$$\dot{y} = \ln^{-1}[(1/n \Sigma_{i=1}^{n} \ln y_i)]$$

Choose the λ for which SSE(λ) is the minimum

(Montgomery et al., 2003)

Dr J Maiti, IEM, IIT Kharagpur

Lecture 42 : Multiple Linear Regression: Test of Assumptions

Box-Cox Method.

$$\dot{y} = \ln^{-1}\left[\frac{1}{n}\sum_{i=1}^{n} \ln y_i\right]$$

Transformed $y \longrightarrow y^{(\lambda)} = \dfrac{y^{\lambda} - 1}{\lambda \dot{y}^{\lambda - 1}}, \quad \lambda \neq 0.$

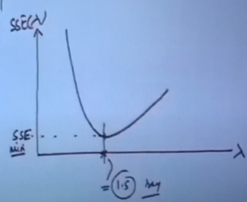
$\dot{y} \ln y, \quad \lambda = 0.$

λ	-2	-1	0	0.25	1.50	1.75	 2
y'	$y_1^{(-2)}$						
	$y_n^{(-2)}$						

Lecture 42 : Multiple Linear Regression: Test of Assumptions

$y^{(\lambda)} = X\beta + \epsilon.$

λ	ANOVA
-2	SSE = $(SSE)_1$
-1.5	$(SSE)_2$
-1.00	
-.75	
-.50	
-.25	
0	
.	
.	



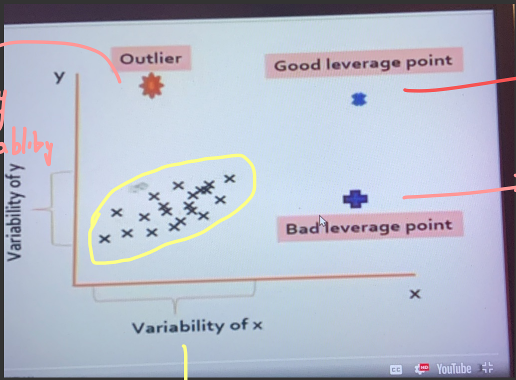
Lecture - 8

Model Diagnostics :

Issues :

1) Finding out leverages
2) Finding out influential points
3) Detecting and Remedying multicolinearity

Lies significantly away but lies outside variablity of y but within variablity of x



Possibly in regression line

Low variablity of y but high for x

with respect to mass of datapoints

Leverage Point : Lies outside x-variablity's general mass

Outlier : Lies outside y-variablity

All these have influence of MLR parameters

↳ Called influential observations

↳ Generally outliers won't affect much but leverage point will make a difference

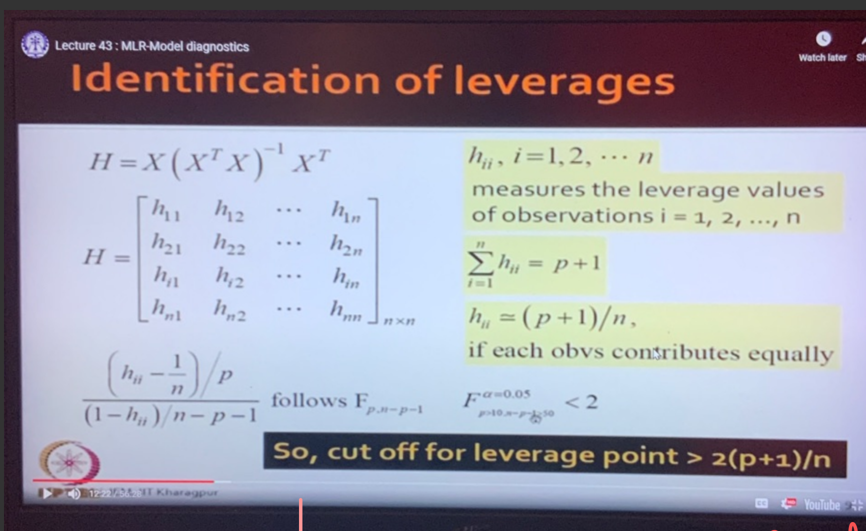
$H = X(X^TX)^{-1}X^T \leftarrow$ X-space

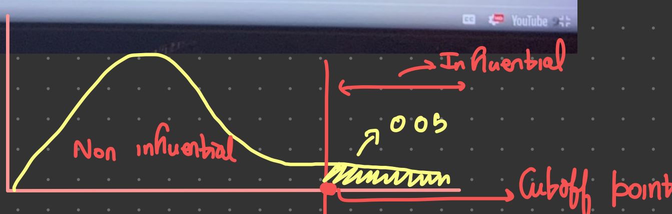
$$H = \begin{bmatrix} h_{11} & h_{12} & & h_{1n} \\ h_{21} & h_{22} & & \\ & & \ddots & \\ & & & h_{nn} \end{bmatrix}$$

Leverage values

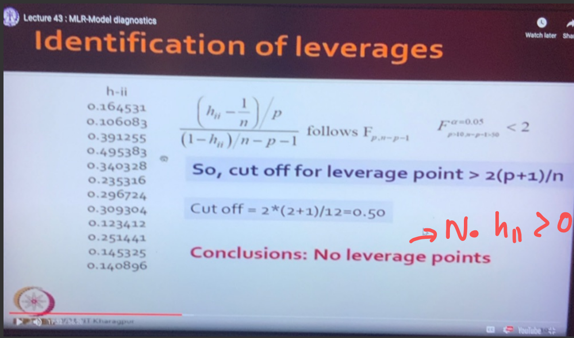
$$\begin{array}{c|c} 1 & h_{11} \\ \hline 1 & h_{11} \\ 2 & h_{22} \\ & \vdots \\ n & h_{nn} \end{array}$$

Cutoff value for h_{ii} [To decide if influential or not



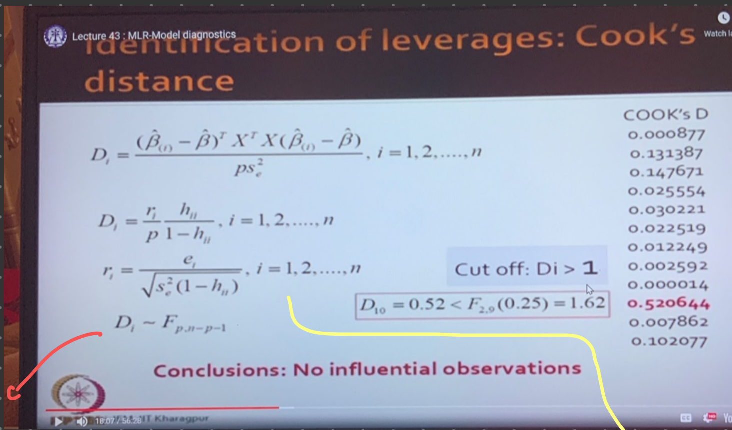


Example:

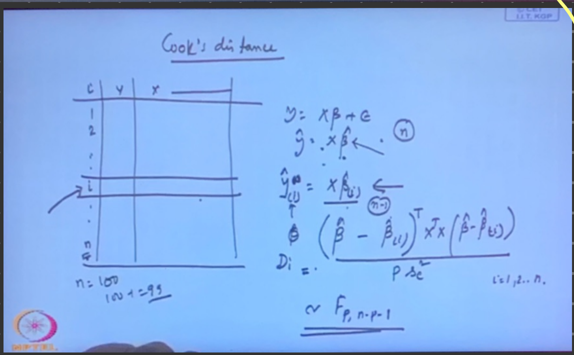


→ No. $h_{ii} > 0.50$

A point won't affect if in general mass, but if leverage point there will be effects



Cutoff



If n is high, then we can't use previous method formula

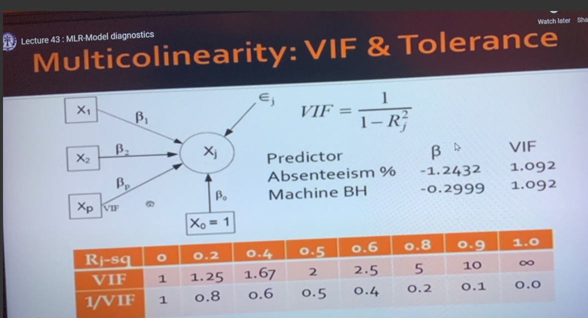
So, we use

Multicollinearity

↳ Not truly independent

① Variation Inflation factors (VIF)
② Tolerance Statistic
③ Eigen value structure
④ Multicolinearity condition numbers

① Consider one independent as dependent



$x_j = \beta_0 + \beta_1 x_1 + \beta_p x_p + \varepsilon$

[doesn't include x_j & y]

$$R_j^2 = \frac{SSR_j}{SST_j}$$

$$\boxed{\text{Tolerance} : \frac{1}{VIF}}$$

$$VIF = \frac{1}{1 - R_j^2}$$

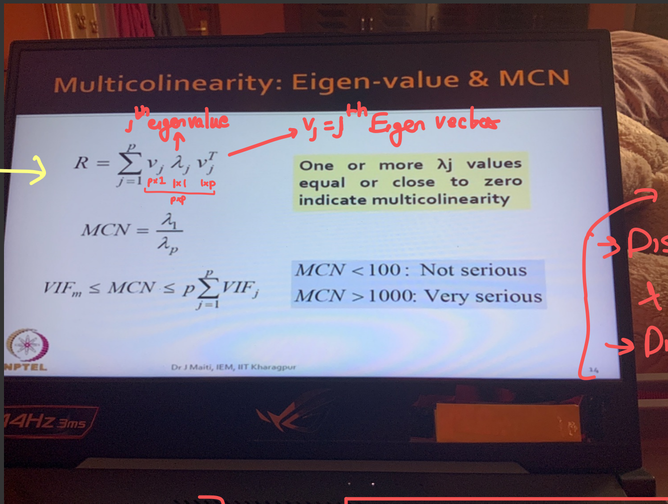
Cutoff : VIF = 10 [5 = warning limit]

[Shouldn't be more]

If $R_j = 0$, VIF = 1 [No correlation]

[Desired value]

Spectral Decomposition



Only one dimension is needed

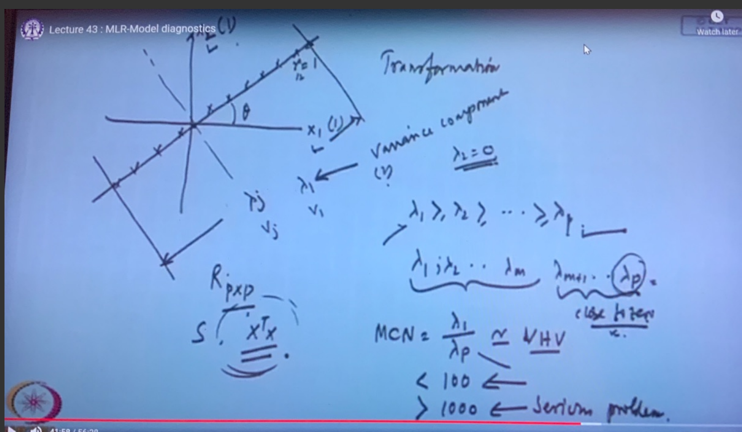
→ Distance captured by λ

→ Direction by v_j

$$R = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}_{p\times p}$$

$\boxed{\text{Eigen value purpose in 37:00}}$

$\lambda_1 \rightarrow$ Variance component



$\boxed{\text{Example given at end 45:00}}$

→ Method to solve multicolinearity – PCA [Principle Component Analysis] [Transforming X]

[Will estimate same regression parameters, but it will allow IV to covary amongst themselves]

If interest is prediction

→ If we don't want to lose nature of data → Use path analysis