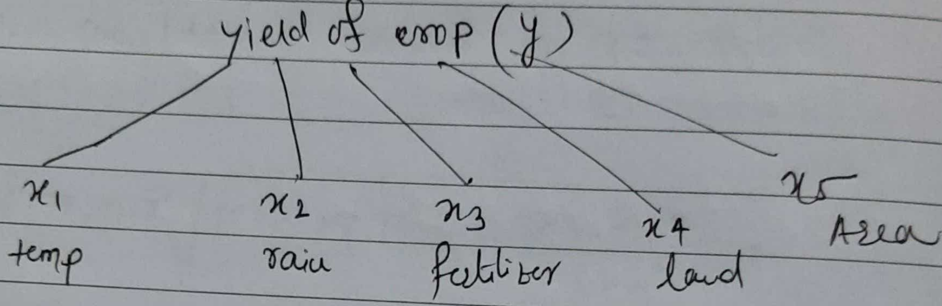
Multiple linear Regression Analysis

19th M

$y = f(x)$ Simple linear Regression

$y = \beta_0 + \beta_1 x$ — only one independent variable 'x'

ex.



$y = f(x_1, x_2, \ldots x_5)$

Multiple independent variables $x_1, x_2, \ldots x_p$

ex. $\boxed{y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_p x_p}$

linear wrt parameters.

$y = \beta_1 x + \beta_2 x^2 + \beta_3 x_1 x_3 + \ldots + \beta_n x_n x_{n-2}$

eg $y = \dfrac{\beta_1 x_1 + \beta_2 x_2}{\beta_1^2 x_2 + \beta_2^2 x_3}$ Not linear

* If $\dfrac{\partial y}{\partial \beta_i}$ = independent of β_i then the functional relationship is linear

eg $y = \beta_1 x_1 + \beta_2 x_2$

$\frac{\partial y}{\partial \beta} = x_1 + x_2$ independent of β.

eg $y = \beta_1 x_1 + \beta_2^2 x_2$

$\frac{\partial y}{\partial \beta} = x + 2\beta_2$ dependent of β_2.

So not linear Relationship.

* Multiple LR model can be treated as
→ intercept model OR
→ without intercept model

$y = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$

General model.

$\cancel{\beta_1 x_1 = x}$ $\underline{x_1 = 1}$

$y = \beta_1 + \beta_2 x_2 + \cdots + \beta_n x_n$.

$\downarrow$ intercept

* $y = \beta_0 + \beta_1 x + \varepsilon$ ⟶ Statistical Model

classmate
Date
Page

$$y = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon$$ — Statistical model

$y \downarrow$ Response Variable

$\varepsilon \downarrow$ Error.

$x_1, x_2, \ldots, x_p$: Regression / Independent variables

$$= (x_1, x_2 \ldots x_p) \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} + \varepsilon$$

$$\boxed{y = x^T \beta + \varepsilon}$$ where $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix}_{p \times 1}$; $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}_{p \times 1}$

No. of para to estimate = 'p'

1st year data $(y_1, x_{11}, x_{12}, \ldots, x_{1p})$

2nd yr data $(y_2, x_{21}, x_{22}, \ldots, x_{2p})$

$\vdots$

n^{th} year data $(y_n, x_{n1}, x_{n2}, \ldots, x_{pp})$

classmate
Date
Page

* **fitted model ⇒**

$$y_1 = \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_p x_{1p} + \varepsilon_1$$

$$y_2 = \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_p x_{2p} + \varepsilon_2$$

$$\vdots$$

$$y_n = \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_p x_{np} + \varepsilon_n.$$

$Y \cancel{y} = X\beta + \varepsilon$ where

(2) $Y \cancel{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}_{n\times 1}$, $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}_{p\times 1}$, $\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}_{n\times 1}$

Vector of Response variables. | vector of parameters | Vector of Error

(3) $X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & & & \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}_{n\times p}$. *

Vector of coefficient terms.

: Measure of independent variables.

(4) fit all (1), (2), (3) data in $\hat{y} = x^T \hat{\beta}$

Assumptions $\Rightarrow$

① $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

② $E(\varepsilon_i) = 0$, $cov(\varepsilon_i, \varepsilon_j) = \sigma^2 = var(\varepsilon_i)$

③ $Rank(X) = p$ ($x_1, x_2, - x_p$ are independent variables)

* find $\beta_1, \beta_2 -- \beta_p$ and model para(σ^2), total ($p+1$) parameters based on dataset

* Least Square Estimation Method $\Rightarrow$

find parameters $\beta_1, \beta_2 - \beta_p, \sigma^2$ so that sum of the sq^r of error term is min.

$\sum \varepsilon_i^2$ is min

$$S(\beta) = \sum_{i=1}^{n} \varepsilon_i^2$$

$$S(\beta) = \varepsilon^T \varepsilon$$

find β so that $S(\beta) = \varepsilon^T \varepsilon$ is minimum

Multivariate min problem

find β.

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

$$\varepsilon = y - X\beta.$$

$$\therefore \; S(\beta) = \varepsilon^T \varepsilon.$$

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

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

$$S(\beta) = y^T y - 2\beta^T X^T y + \beta^T X^T X \beta.$$

$$\frac{\partial S}{\partial \beta} = 0 \Rightarrow \beta = ?$$

$$\frac{\partial^2 S}{\partial \beta} = (\quad) \text{ semidefinite matrix.}$$

$$\boxed{\hat{\beta} = (X^T X)^{-1} X^T Y}$$ ① estimation of β

where, $$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_p \end{pmatrix}$$

fitted Regression model

$$\boxed{\hat{y} = X^T \hat{\beta}}$$

$$y = X\beta + \varepsilon$$ such that Error term is min

fitted regn^n model. $$\boxed{\hat{y} = X^T\hat{\beta}}$$ $X\hat{\beta}$

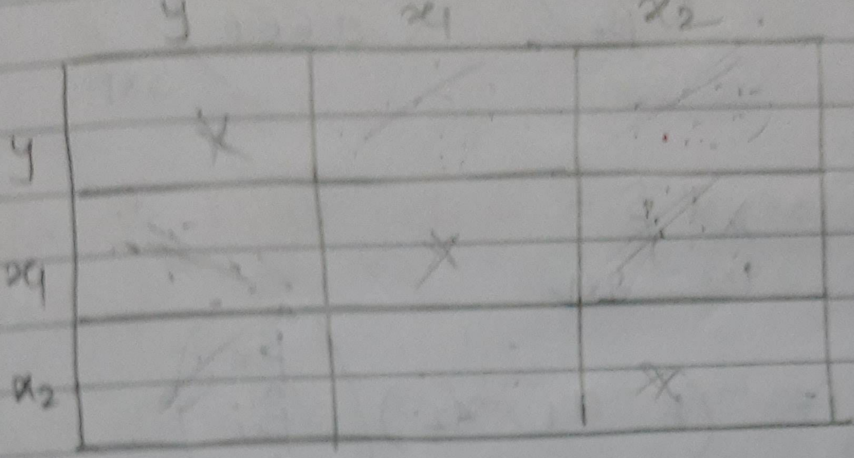
$$\hat{y} = (x_1 \; x_2 \; \cdots \; x_p)_{1\times p} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_p \end{pmatrix}_{p\times 1}$$

$$\hat{y} = x_1\hat{\beta}_1 + x_2\hat{\beta}_2 + \cdots + x_p\hat{\beta}_p$$

Ex1)

Market price (y) \$1000	Total no of sq^r feet (x_1)	Age of house (x_2)
63	1605	35
65.1	2489	45
69.9	1553	20
76.8	2404	32
73.9	1884	25
77.9	1558	14
74.9	1748	8
78	3105	10
79	1682	28
83.4	2470	30
79.5	1820	2
83.9	2143	6

Matrix Scatter plot



$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon.$$

$y = x^T\beta + \varepsilon$ where, $x = \begin{pmatrix} 1 \\ x_1 \\ x_2 \end{pmatrix}$ and

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}$$

fitted linear Regression model

$$\hat{y} = x^T \hat{\beta}$$

$$\hat{\beta} = (X^TX)^{-1} X^T Y_{23\times1}$$

$$X = \begin{array}{c} \begin{array}{ccc} & x_1 & x_2 \end{array} \\ \begin{pmatrix} 1 & 1605 & 35 \\ 1 & \cdot & \cdot \\ 1 & \cdot & \cdot \\ 1 & \cdot & \cdot \\ 1 & 2417 & 6 \end{pmatrix}_{23\times3} \end{array}, \quad Y = \begin{pmatrix} 63 \\ 63.1 \\ \vdots \\ 83.9 \end{pmatrix}$$

$$\hat{\beta} = \left(X^T_{3\times23} X_{23\times3}\right)^{-1} X_{3\times23} Y_{23\times1} \quad ?$$

$3\times3 \quad 3\times23$

$3\times$

$$\cancel{Y} = \hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} 57.5 \\ 0.177 \\ 0.666 \end{pmatrix}_{3\times 1}$$

Vector of fitted

$$\hat{y} = \underset{1\times 3}{X^T} \underset{3\times 1}{\hat{\beta}}$$

$$= \begin{pmatrix} 1 & x_1 & x_2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix}$$

$$\hat{y} = 57.4 + 0.177\, x_1 + 0.666\, x_2$$

$$\hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x_2$$

$$\boxed{Y = X\beta + \varepsilon.}$$

vector of fitted values

$$\boxed{\hat{y} = \cancel{X\hat{\beta}}}\ X\hat{\beta}$$

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

* fitted Regression model.

$$\hat{y} = x^T \hat{\beta}$$

Vector of fitted values $\hat{y}_i$ corresp. to observed values y_i is

$$\hat{y} = X\hat{\beta} = \underbrace{X(X^TX)^{-1}X^T}_{H} y. = Hy$$

$$y_i = x_i^T \beta + \varepsilon$$

$$\Downarrow$$

$$Y = X\beta + \varepsilon.$$

$y = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon$ then

20th M our model $y = x^T \beta + \varepsilon$

$$\Downarrow$$

fitted model $\hat{y} = x^T \hat{\beta}$

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

$$\Downarrow$$

$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$

$\hat{\beta} = (X^TX)^{-1}X^TY$ where our data pts were

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & & & \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

$$\hat{\beta} = (X^TX)^{-1}X^TY = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_p \end{bmatrix}$$

$\therefore$ Based on data we got parameters $\hat{\beta}$ (using LSE)

$\therefore \boxed{\hat{y} = x^T \hat{\beta}}$ is fitted Regression model of

Vector of fitted values

$$\boxed{\hat{Y} = X\hat{\beta}}$$

where $y_i = x_i^T \beta + \varepsilon_i$

$\downarrow$

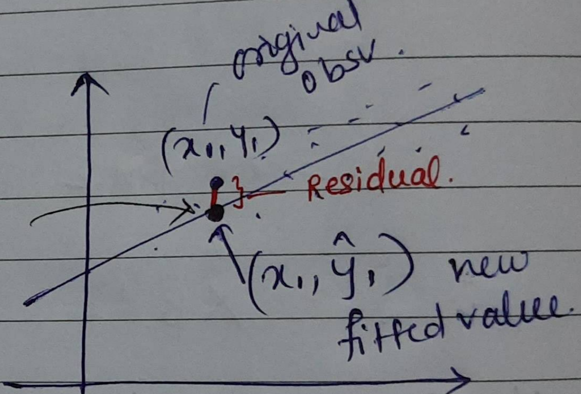
vector notatn

$\downarrow$

$$Y = X\beta + \varepsilon$$

$$\hat{Y} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix}$$

$$\hat{Y} = X\hat{\beta}$$



original obsv

fitted Response values

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & & & \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$$

$$\hat{Y} = X\underbrace{\hat{\beta}}_{\text{estimate using LSE}} = \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{pmatrix}$$

$$= X(X^TX)^{-1}X^TY$$

* fitted values take crucial Role in order to determine many imp things in Regression Analysis

classmate
Date
Page

$\Rightarrow$ $\hat{y} = X\hat{\beta}$

$$\hat{Y} = X(X^TX)^{-1}X^TY$$

$$\hat{Y} = HY \quad \text{where } H = X(X^TX)^{-1}X^T.$$

hat matrix.

* Residual of ith observation $\Rightarrow$

$$e_i = y_i - \hat{y}_i$$

$$e = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix} = \begin{pmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ \vdots \\ y_n - \hat{y}_n \end{pmatrix}$$ Residual Vector.

$$e = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} - \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{pmatrix}$$

$$\therefore \boxed{e = Y - \hat{Y}}$$

$$e = Y - HY$$

$$e = (I - H)Y$$

$$e = \bar{H}Y \qquad \bar{H} = I_n - H.$$

$\downarrow$ response variable

classmate
Date
Page

* Properties ⇒

1) $E(\hat{\beta}) = \beta$, $\hat{\beta}$ is an unbiased estimator of β

when all parameters on avg are taking the true values.

$$\begin{pmatrix} E(\hat{\beta}_1) \\ E(\hat{\beta}_2) \\ \vdots \\ E(\hat{\beta}_p) \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}$$

LSE is v good estimator as it satisfies unbiasness property.

2) The variance property of $\hat{\beta}$ -

$cov(\hat{\beta}) = \sigma^2 C$, where $C = (X^T X)^{-1}$

// $cov(x_i, x_i) = var(x_i)$

C_{22} - coefficient of β_1 & x_1

$$\begin{array}{ccc} & 1 & x_1 \\ 1 & & \\ x & & \end{array}$$

$$\begin{pmatrix} var(\hat{\beta}_1) & cov(\hat{\beta}_1, \hat{\beta}_2) & \cdots & cov(\hat{\beta}_1, \hat{\beta}_p) \\ cov(\hat{\beta}_2, \hat{\beta}_1) & var(\hat{\beta}_2) & \cdots & cov(\hat{\beta}_2, \hat{\beta}_p) \\ \vdots & & & \vdots \\ cov(\hat{\beta}_p, \hat{\beta}_1) & cov(\hat{\beta}_p, \hat{\beta}_2) & \cdots & var(\hat{\beta}_p) \end{pmatrix} = \sigma^2 \begin{pmatrix} C_{11} & C_{12} & \cdots & C_{1p} \\ C_{21} & C_{22} & \cdots & C_{2p} \\ \vdots & \vdots & & \\ C_{p1} & C_{p2} & \cdots & C_{pp} \end{pmatrix}$$

sigma

for testing β_1

$var(\hat{\beta}_j) = C_{jj}$, jj^{th} elem of the matrix $C = (X^T X)^{-1}$

ClASSMATE
Date
Page

(2) Estimation of $\sigma^2 \Rightarrow$

variance quantity of error.

β

i.e. For each obs how much error ~~you are~~ ~~doing~~ can vary.

→ Estimating σ^2 using Sum of the sq^r of the residual.

$$SS_{Res} = \sum_{i=1}^{n} e_i^2 = e^T e \quad \text{where } e = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

$$\therefore SS_{Res} = (y - \hat{y})^T (y - \hat{y})$$

$$\boxed{SS_{Res} = y^T y - \hat{\beta}^T X^T y} \; *$$

* **Result** ⇒

$$\boxed{\frac{SS_{Res}}{\sigma^2} \sim \chi^2_{n-p}}$$ usual chi-sq^r

we know if $X \sim \chi^2_n$ then $E(X) = n$.

∴ here

$$E\left(\frac{SS_{Res}}{\sigma^2}\right) = n-p.$$

$$\therefore E\left(\frac{SS_{Res}}{n-p}\right) = \sigma^2$$

$$\therefore \boxed{\hat{\sigma}^2 = \frac{SS_{Res}}{n-p} = MS_{Res}.}$$

Date 20th A

Gen. model $\Rightarrow y = \beta_1' x_1' + \beta_2' x_2' + \dots + \beta_p' x_p'$

* consider intercept model $(x_1 = 1)$

particular case of Gen model

$$y = x^T \beta + \varepsilon$$

$$= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_K x_K.$$

where

$$x = \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_K \end{pmatrix} \qquad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{pmatrix}$$

$(K+1 = p)$

$$\begin{pmatrix} x_1' = 1 \\ x_2' = x_1 \\ \vdots \\ x_p' = x_K \end{pmatrix}$$

General model

$y = \beta_1 x_1 + \beta_2 x_2 + \dots \beta_p$

$\beta_1' = \beta_0$

$\beta_2' = \beta_1$

$\beta_p' = \beta_K$.

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

where given data is

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_K \end{pmatrix}, \quad X = \cancel{\begin{pmatrix} x_{11} & x_{12} & \dots & x_{1K} \\ x_{21} & x_{22} & \dots & x_{2K} \\ \vdots & & & \\ x_{n1} & x_{n2} & \dots & x_{nK} \end{pmatrix}}$$

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1K} \\ 1 & x_{21} & x_{22} & \dots & x_{2K} \\ 1 & \vdots & & \vdots & \vdots \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nK} \end{pmatrix}$$

*

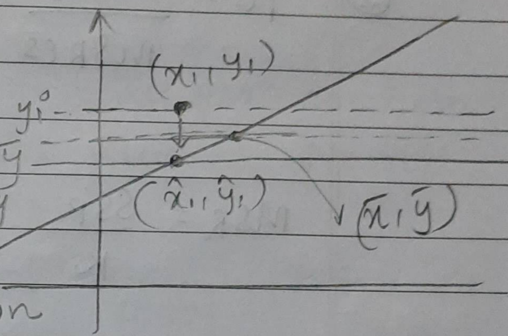
1

ANOVA for intercept model $\Rightarrow$

Total variation is given by Sum of the sq^r of the total variation $(y_i - \bar{y})^2$.

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

total variability in the ith observation



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

$$SST = \cancel{y^T}\, Y^T Y - n\bar{y}^2$$

$SST = SSR + SSRes$. where ~~$SSRes = X^T X - \hat{\beta} X^T Y$~~

$SSRes = Y^T Y - \hat{\beta}^T X^T Y$

$$SSR = SST - SSRes$$

$$= \cancel{Y^T Y} - n\bar{Y}^2 - \cancel{Y^T Y} + \hat{\beta}^T X^T Y$$

$$\boxed{SS_R = \hat{\beta}^T X^T Y - n\bar{Y}^2}$$

* Result $\Rightarrow$

(1) $\dfrac{SSRes}{\sigma^2} \sim \chi^2_{n-k-1}$

$k+1 = p$

$n-p$

$n-(k+1) = \underline{n-k-1}$

classmate
Date
Page

(1) $\frac{SSRes}{\sigma^2} \sim \chi^2_{n-k-1}$

(2) $\frac{SS_R}{\sigma^2} \sim \chi^2_K$ ← non-central chi sqr distribution (λ)

(3) $F_0 = \frac{MS_R}{MSRes} \sim F_{K, n-K-1}$ ⟷ non-central F-distribution

$MS_R = \frac{SS_R}{K}$, $MSRes = \frac{SSRes}{n-K-1}$

$y = x^T \beta + \varepsilon$

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_K x_K$

out of these K regession variables, which of the variables are actually responsible for determining value of y (response variable).

i.e. if you reject Null

$y = x^T \beta + \epsilon.$

$= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_K x_K$ $\quad x = \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_K \end{pmatrix}, \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{pmatrix}$

* Test for significance of Regression :-

→ $H_0: \beta_1 = \beta_2 = \dots = \beta_K = 0$

vs $H_1: \beta_j \neq 0$ for atleast one $j = 1, 2, \dots k.$ } Global Test.

↳ If you accept Null (H_0) means (all para are zero) there is No roll of independent variables ($x_1, x_2 \dots x_K$)

$y = \beta_0$ model will be const.

(i.e. if you reject Null) ↳ If you accept your alternative it means one of b_j is ~~is~~ 1 non-zero. i.e. atleast one of the variable ~~of~~ out of K-variables has crucial Role in determining value of y.

↳ Once you accept your alternative (H_1) → Go for. ~~G~~ Lobal Test. Case by Case Testing. → checking which variable is crucial.

$H_0: \beta_j = 0$ vs $H_0: \beta_j \neq 0$ — for each j to check which independ variable has a role in determining y.

↳ Local test

* Once you accept your alternative (H_1) then Go for local Test → Case by case test.

eg $H_0: \beta_j = 0$ vs $H_1: \beta_j \neq 0$ for each j.

* Global F-test ⇒ Once we accept alternative (H_1) using Global Test then go for local test.

* Global F-Test :-

$H_0: \beta_1 = \beta_2 = \dots = \beta_K = 0$

vs

$H_1: \beta_j \neq 0$ for atleast one $j = 1, 2, \dots k$

level-α-test ⇒

Reject H_0 if

Non-central F-dist^n

$$F_0 = \frac{MSR}{MRes} > F_{\alpha, K, n-K-1}$$

Intuition ⇒

Reject H_0 for large value of F_0 and

F_0 will be large when $MS_R >> MSRes$.

Total variation $\quad SS_T = SS_R + SS_{Res}$

$$MSRes = \frac{SSRes}{n-K-1}$$

$SS_T = SS_R + SS_{Res}$

SS_R ↓ how much variation is explained by the model

SS_{Res} ↓ how much variation is left

Now if $MS_R >> MS_{Res}$

means $SS_R > SS_{Res}$ means, most of the variation is actually explained & there is very less variation which is left

atleast one of the

i.e. The model is fitted very well, means the coefficient parameters $\beta_1, \beta_2 \dots \beta_K$ are non-zero.

Imp

i.e. Accepting the alternative, OR Rejecting the null.

ANOVA table ⇒

Source of Variation	Sum of squares	Degrees of Freedom	Mean Sq^r	F_0
Regression	$SS_R = \beta^T X^T Y - n\bar{y}^2$	K	$MS_R = \frac{SS_R}{K}$	$F_0 =$
Residual	$SS_{Res} = Y^T Y - \hat{\beta}_0^T X^T Y$	$n-K-1$	$MS_{Res} = \frac{SS_{Res}}{n-K-1}$	$F_0 = \frac{MS_R}{MS_{Res}}$
Total	SS_T	$n-1$		

previous example ⇒

$$\hat{y} = \underset{\hat{\beta}_0}{57.4} + \underset{\hat{\beta}_1}{0.0177}\, x_1 + \underset{\hat{\beta}_2}{0.666}\, x_2$$

Regression fitted model obtained from data.

Now testing problem ⇒

$H_0: \beta_1 = \beta_2 = 0$ vs $H_1:$ atleast one of the β_j non-zero $\forall j = 1, 2$

Do F-test → ANOVA table.

$\cancel{y = \beta_0 + \beta_1 x_1 + \beta_2 x_2}$

$\cancel{DOF = 2}$ ↑

~~ANOVA table.~~

~~Regression~~ SSR =   DOF =   Mean Residual

$(k = 2)$

$n-k-1 = 23-2-1$

$\Rightarrow 20$

MSR =

n-K Ru

$$\hat{\beta} = (X^T X)^T X^T Y = \begin{pmatrix} 57.5 \\ 0.0177 \\ 0.666 \end{pmatrix} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix}$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$$

Testing problem ⇒ Out of two variables x_1 & x_2 whether they have any role in determining market price (y) or not

$H_0: \beta_1 = \beta_2 = 0$ vs H_1: atleast one of the $B_j \neq 0$ (for $j = 1, 2$)

24th Oct afternoon

Do Global F-test :

ANOVA Table ⇒

Source of variation   Sum of sqrs.

ANOVA Table ⇒

Source of Variation	Sum of squares	Degrees of freedom	Mean square	Fo
① Regression	SSR = 8189.7	K = 2	MSR = 4094.9	
② Residual	SSRes = 2861.9	n-K-1 ⇒ 23-2-1 ⇒ 20.	MSRes = 143.1	Fo = 28.63

SST = 8189.7 + 2861.9
SST ⇒ 11050.7

To Reject Ho compare with Fo $(\alpha = 0.05)$.

∴ Reject Ho if

$$F_0 = \frac{MSR}{MSRes} > F_{\alpha, K, n-K-1}.$$

$$F_0 = 28.63 > F_{0.05, 2, 20}.$$

$$F_0 = 28.63 > F_{0.05, 2, 20} = 3.49.$$ ✓

∴ As per level-α- test, we should Reject Ho
i.e. we should accept alternative (H_1)
So atleast one of β_1 & β_2 is non-zero.

when all coefficients are zero

* If we fail to reject H_0 then $y = \beta_0$ i.e. none of the regression variables x_1, x_2 has any role in determining Response variable (y).

i.e if atleast one of β_1 & β_2 non-zero it means atleast one of x_1 and x_2 has imp role in determining the market price (Y)

* check whether both of the variables are imp OR one of them is imp.

22/10/14

Intercept model ⇒

$$y = x^T \beta + \varepsilon$$

$$= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K + \varepsilon$$

$$X = \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_K \end{pmatrix} \qquad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{pmatrix}$$

Our dataset →

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \qquad X = \begin{array}{cccc} & & x_1 & x_2 \\ \end{array}$$

$$X = \begin{pmatrix} 1 & x_{11} & x_{21} \\ 1 & x_{12} & x_{22} \\ \vdots & \vdots & \vdots \\ 1 & x_{1K} & x_{2K} \end{pmatrix}$$

~~We rejected~~

We accepted H_1 means atleast one of β_1 and β_2 is non-zero.

$\therefore$ Global f-test assured you that atleast one of variable has crucial role to determine y. (x_1 or x_2)

* Then which variable has imp in determining y or both has importance

Marginal Test / T-test $\Rightarrow$

As you ↑s no. of variables SSR ↑ so our model is fitted very well.

But SSRes will dec.

So while adding another Regression Variable check whether SSR is significant or not.

We discussed, $\quad MSRes = \hat{\sigma}^2 = \frac{SSRes}{n-k-1}$

$E(\hat{\beta}_j) = \beta_j$,

$var(\hat{\beta}_j) = \sigma^2 C_{jj}$, where C_{jj} is jth elem of matrix $C = (X^T X)^{-1}$

Result ⇒

$$T = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} \overset{a}{\sim} t_{n-k-1}$$

* For each $j = 1, 2, \ldots k$

$H_0: \beta_j = 0 \quad vs \quad H_1: \beta_j \neq 0$

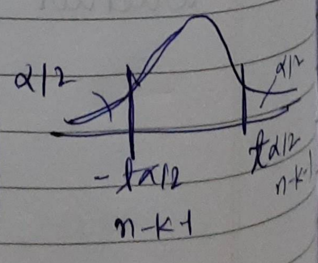
level-α-test ⇒

Reject H_0 if $|T| = \left| \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} \right| > t_{\alpha/2, n-k-1}$

Partial or marginal test →

Under H_0 $(\beta_j = 0)$

$T = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} \sim t_{n-k-1}$



Using this t-test to check whether $\beta_1 \neq 0$ or $\beta_2 \neq 0$ for prev problem

classmate
Date
Page

(1) $H_0: \beta_1 = 0$ vs $H_1: \beta_1 \neq 0$.

$$|T| = \left| \frac{\hat{\beta}_1}{\sqrt{\hat{\sigma}^2 C_{22}}} \right| = \frac{0.177}{\sqrt{143.1 \times 0.0000069}} = \underline{\underline{5.63}}$$

$\underline{\underline{C_{11}}}$

$var(\hat{\beta}_0) = \sigma^2 C_{11}$

$var(\hat{\beta}_1) = \sigma^2 C_{22}$ ← Here we are testing β_1 so use this as var quantity.

$var(\hat{\beta}_2) = \sigma^2 C_{33}$.

$$\hat{\sigma}^2 = \frac{SSRes}{n-k-1} = 143.1$$

$$C_{jj} = (\underset{3\times 23}{X^T} \underset{23\times 3}{X})^{-1}$$

$$C_{jj} = (X^T X)^{-1}_{3\times 3}$$

if → then use C_{11}

$$\begin{array}{cccc} & \beta_0 & \beta_1 & \beta_2 \\ \beta_0 & C_{00} & C_{01} & C_{02} \\ \beta_1 & C_{10} & C_{11} & C_{12} \\ \beta_2 & C_{21} & C_{22} & C_{22} \end{array}$$

$$C = \begin{array}{cccc} & 1 & \beta_1 & \beta_2 \\ 1 & & & \\ \beta_1 & & C_{22} & \\ \beta_2 & & & \end{array}$$

Take it → 2nd diagonal elem

$\therefore C_{22} = 0.00000690$.

$n-k-1$

$\alpha/2$

$t_{\alpha/2, n-k-1}$

$\alpha = 0.05$

$t_{\alpha/2, n-k-1} \Rightarrow t_{0.025, 20} \Rightarrow 2.086.$

$|T| = 5.63 > t_{0.025, 20} = 2.086$ ✓

∴ Reject H_0 i.e. Accept H_1 i.e. coeff $\beta_1 \neq 0$

i.e. x_1 has crucial role in determining y

∴ ~~y~~ =

② $H_0: \beta_2 = 0$ vs $H_1: \beta_2 \neq 0$

$$|T| = \left| \frac{\hat{\beta}_2}{\sqrt{\hat{\sigma}^2 \cancel{C_{33}}}} \right| = \frac{0.666}{\sqrt{143.1 \times C_{33}}} = 2.99$$

C_{22} change

$C_{33} = 0.00336$

$\alpha = 0.05$

$|T| = 2.99 > t_{\alpha/2, n-k-1} = 2.086$ ✓

Accept H_1 Reject H_0

∴ Var x_2 also has role in det y.

House of market price depends on on both x_1, x_2

F - test $\Rightarrow$ Global Test

T - test $\Rightarrow$ Local Test.

$$C = (X^T X)^T = \begin{pmatrix} C_{00} & C_{01} & C_{02} \\ C_{10} & C_{11} & C_{12} \\ C_{20} & C_{21} & C_{22} \end{pmatrix}$$

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

classmate
Date
Page

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

22nd_A

Coefficient of determination ⇒

Imp

$$R^2 = \frac{SSR}{SS_T} = 1 - \frac{SS_{Res}}{SS_T}.$$

↳ Total variation in data

$SS_T = SS_R + SS_{Res}.$

R^2 = measure of Goodness of fit. i.e. How Good you have fitted the model.

For multiple LRA

$SS_T = Y^T Y - n\bar{y}^2$

$SS_R = \hat{\beta}^T X^T y - n\bar{y}^2$

$SS_{Res} = y^T y - \hat{\beta}^T X^T y.$

~~Intercept~~ model ⟶ $y = \beta^T x + \varepsilon = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$

$$X = \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \vdots \\ \beta_k \end{pmatrix}$$

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & \vdots & & & \\ 1 & \vdots & & & \\ \vdots & & & & \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix}$$

$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$ SSR ⇒ variation captured by the model

SSRes ⇒ Variation not captured or variation left or Remaining variation.

→ Imp → R^2 is actually defined for intercept model.

* Observations ⇒

① $R^2 = 1$ iff $SSRes = 0$
↳ means all variation in data is explained by model. i.e., all data pts lie on line. ∵ $SST = SSR + SSRes$ (0).
i.e. model is fitted very well

② $R^2 = 0$ iff $SSRes = SST$. ⇒ $SSR = 0$

i.e. not even single variation in data is expressed by model. everything is left so model is fitted very worst.

③ $0 \leq R^2 \leq 1$ Bounded by 0 to 1.

④ if $R^2 = 0.74$, 74% variation explained by the model.

* Drawbacks of R^2 ⇒

① R^2 ↑ as you ↑s no. of independent variables in model, means you can ↑s no. of variables & claim that model is fitted very well.

$R^2 = 1 - \frac{SSRes}{SST}$

Sol^n ⇒ Adjusted R^2 ⇒

$$R^2 adj = 1 - \left(\underbrace{\frac{SSRes}{n-K-1}}_{MSRes} \Big/ \underbrace{\frac{SST}{n-1}}_{\text{MS of total sum. of sqr.}} \right) = 1 - \frac{MSRes}{MS_T}$$

$$SS_T = \sum_{i=1}^{n} (y_i - \bar{y})^2 .$$

SS_T will ~~you~~ ↑s as you add ~~to~~ more no of variables

$\frac{SS_T}{n-1}$ ⇒ standardizing the ~~avg~~ variation
ie. on an avg how much variation is there in data.

So even if you add more no. of variables

$\frac{SS_T}{n-1}$ won't change. & will behave like constant

So now $R^2 adj$ will ↑s only if MSRes decrease.

$$R^2 adj = 1 - \frac{MSRes}{MST}$$

∴ $R^2 adj$ will ↑s only when MSRes ↓ ie.

$\frac{SSRes}{n-K-1}$ ↓ OR $\frac{SSR}{K}$ ↑ cz $SS_T = \uparrow SS_R + SSRes \downarrow$

$$\frac{SSRes}{n-K-1}\downarrow \Leftrightarrow \frac{SSR}{K}\uparrow$$

So if Mean sq^r due to Regression ↑ (MSR↑) means your model is fitted well.

∴ R^2 adj will ↑s only when you add a variable if ~~it~~ has significant Role in determining Resp. var

If you add var then $\frac{SSRes}{n-K-1}\downarrow$ and $\frac{SSR}{K}\uparrow$.

MSRes ↓

$$R^2_{adj} = 1-\left(\frac{n-1}{n-K-1}\right)\left(\frac{SSRes}{SST}\right)$$

$$\boxed{R^2_{adj} = 1-\left(\frac{n-1}{n-K-1}\right)(1-R^2)}$$

K = 1

for simple linear regression analysis.

$$R^2_{adj} = 1-\left(\frac{n-1}{n-2}\right)(1-R^2)$$

drawback ⇒

i) R^2_{adj} can take -ve values as well.

$n=10,\ K=2,\ R^2=0.16$.

$R^2adj = 1 - \frac{9}{7}(1-0.16) = -0.08 < 0.$

* $0 \leq R^2 \leq 1$

Example continued ...

→ $R^2 = 1 - \frac{SSRes}{SST} = 1 - \frac{2861.9}{110.1..} = 0.741$

$R^2adj = 1 - \frac{23-1}{20}(1-0.741)$

$= 0.7151$

∴ If you use R^2adj as Goodness of fit then model is fitted 71%.

$R^2 - R^2adj = 0.741 - 0.7151 = \underline{0.26}$

↑

This diff will ↑s if you add more no. of variable. which are unnecessary

* If you add more var$^{'}$ R^2 ↑ But R^2 adj will ↑s only when var has imp role to det y

∴ If you add some var which doesn't have any role $R^2 - R^2adj$ diff will ↑s

classmate

Date

Page

② Relⁿ betⁿ R^2 and F-statistic (F_0) :-

$H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0.$

vs $H_1: \beta_j \neq 0$ for $k = 1, 2, \ldots$

Reject H_0, $F_0 = \frac{MSR}{MSRes} > F_{\alpha, k, n-k-1}$.

$$= \frac{SSR / k}{SSRes / n-k-1}$$

$$F_0 = \frac{n-k-1}{k} \times \frac{SSR}{SSRes} \quad (SST = SSR + SSRes)$$

$$= \frac{n-k-1}{k} \times \frac{SST - SSRes}{SSRes}$$

$$= \frac{n-k-1}{k} \times \frac{1 - \frac{SSRes}{SST}}{\frac{SSRes}{SST}}$$

$$\boxed{F_0 = \frac{n-k-1}{k} \times \frac{R^2}{1-R^2}}$$

If $R_k^2 = 1$, $F_0 = \infty$ then we Reject H_0

i.e. atleast one var has imp role model fit well

If $R = 0$, $F_0 = 0 \Rightarrow$ Will accept H_0 i.e. model not fitted vell as $y = \beta_0$ is constant No var role in det y i.e. model fitted worst.

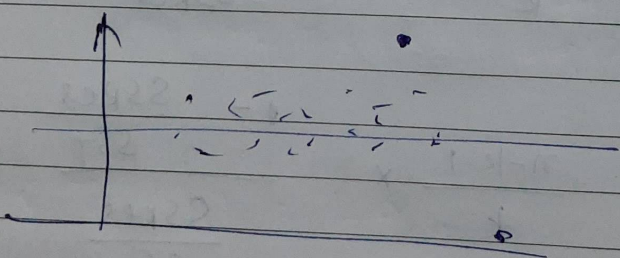
$R = 1$ $\quad F_0 = \infty$ $\quad$ ✓ Best fit

$R = 0$ $\quad F_0 = 0$ $\quad$ Worst fit.

③ If no intercept term

$$y = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K$$ then R^2 cannot be defined.

④ R^2 too much depend on outliers.



i.e. R^2 ↑s & it seems model fitted well but originally its not fitted well

We should avoid such drawbacks then R^2 is best goodness of fit

(9) Confidence Interval on Regression Coeff.

$$y = x^T \beta + \varepsilon = \beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K.$$

$$x = \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_K \end{pmatrix} \qquad \hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_K \end{pmatrix}$$

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

$$X = \begin{pmatrix} 1 & x_{11} & x_{1K} \\ 1 & x_{21} & x_{2K} \\ \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{nK} \end{pmatrix} \qquad Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

interval estimation $\Rightarrow$

Result $\Rightarrow$

$$T_j = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} \sim t_{n-K-1}$$

$se(\hat{\beta}_j) \Rightarrow$

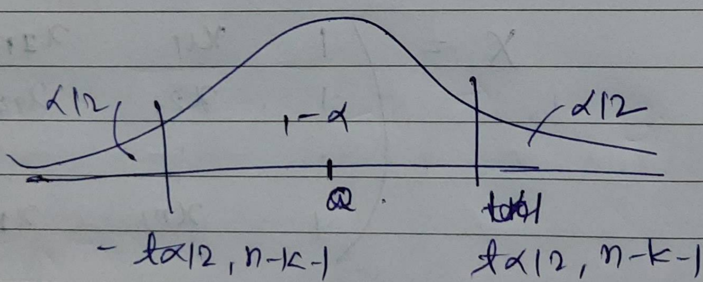
$$\therefore T_j = \frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)}, \quad se = \sqrt{\hat{\sigma}^2 C_{jj}}$$

Std estimated error.

$$E(\hat{\beta}_j) = \beta_j$$

$$se(\beta_j) = \sqrt{\hat{\sigma}^2 C_{jj}}$$
$$= \sqrt{MSRes\, C_{jj}}$$

where C_{jj} = jth elem of mat $C = (X^T X)^{-1}$



$$P(a < T_j < b) = 1 - \alpha.$$

$$P\left(-t_{\alpha/2, n-k-1} < \frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} < t_{\alpha/2, n-k-1}\right) = 1 - \alpha$$

$$P\left(\hat{\beta}_j - t_{\alpha/2, n-k-1}\, Se(\hat{\beta}_j) < \beta_j < \hat{\beta}_j + t_{\alpha/2, n-k-1}\, Se(\hat{\beta}_j)\right)$$

this is $100(1-\alpha)\%$ CI for given T_j.

* CI on Regression Coefficients

(1) Two sided CI

$$\left(\hat{\beta}_j - t_{\alpha/2, n-k-1}\, Se(\hat{\beta}_j),\ \hat{\beta}_j + t_{\alpha/2, n-k-1}\, Se(\hat{\beta}_j)\right)$$

(2) Upper CI

$$\left(\hat{\beta}_j - t_{\alpha, n-k-1}\, Se(\hat{\beta}_j),\ \infty\right)$$

(3) Lower CI

$$\left(-\infty,\ \hat{\beta}_j + t_{\alpha, n-k-1}\, Se(\hat{\beta}_j)\right)$$

Point estm for $\beta_1 \Rightarrow$

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

95% Interval estimatr for $\beta_1 \Rightarrow$

$$\left(\hat{\beta}_1 - t_{\alpha/2, n-k-1}\, se.(\hat{\beta}_1)\ ,\ \hat{\beta}_1 + t_{\alpha/2, n-k-1}\, , se(\hat{\beta}_1)\right)$$

point estimation

$\Rightarrow$ $0.0177 - 2.086 \times$

$t_{0.025,20} = 2.086$

$$se(\hat{\beta}_1) = \sqrt{MSRes\ C_{11}} \Rightarrow \sqrt{148.1 \times 0.00006}$$

$\Rightarrow 0.00$

$C = (X^T X)^{-1} \Leftarrow$ C_{11} elem

$$C = \begin{array}{ccc} & \beta_1 & \beta_2 \\ \end{array}$$
$$\begin{pmatrix} 1 & & \\ 1 & \bigcirc & \\ 1 & & \end{pmatrix}$$

calulate $\Rightarrow$

Interval estimate for $\hat{\beta}_2 \Rightarrow$

Same as above. (just take C_{22})

26 M

* Application of Multiple linear Reg Analysis ⇒

(prediction of future obs, prediction and mean response)

① predicting future obs. [once model is fitted] and predicting Mean Response ↗

$$y = \beta\, x^T \beta + \varepsilon = \beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K + \varepsilon$$

where $x = \begin{pmatrix} 1 \\ x_1 \\ \vdots \\ x_K \end{pmatrix}$ $\beta = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_K \end{pmatrix}$

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1K} \\ 1 & & & \\ \vdots & & & \\ 1 & x_{n1} & \cdots & x_{nK} \end{pmatrix}$$

Estimation ⇒ $\hat{y} = x^T \hat{\beta} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots$

$$\hat{\beta} = (X^T X)^{-1} X^T Y = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_K \end{pmatrix}$$

point estimation

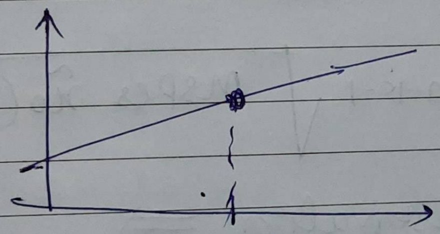
CI of Mean Response ⇒

* If $\tilde{x} = \tilde{x}_0$ ⇒ $x_1 = x_{01}, x_2 = x_{02} \ldots$
$x_k = x_{0k}$.

Once you fit this value then $E(y/\tilde{x} = x_0)$?

on avg what is market price when $\tilde{x} = x_0$?

$$\tilde{x} = x_0 \Rightarrow \begin{pmatrix} 1 \\ x_{01} \\ x_{02} \\ \vdots \\ x_{0k} \end{pmatrix}$$



$x = x_0$

Unbiased point estimator for $E(y/x = \tilde{x}_0)$ is

point estimation $\boxed{\hat{\mu}_{y/\tilde{x}_0} = \tilde{x}_0^T \hat{\beta}_0}$ $= \hat{\beta}_0 + \hat{\beta}_1 x_{01} + \ldots + \hat{\beta}_k x_{0k}$.

This value y will take on avg.

classmate

Date

Page

* CI for Mean Response ⟹

Interval Estimation ⟹

$$T = \frac{\hat{\mu}_{y/\tilde{x}_0} - \mu_{y/\tilde{x}_0}}{\sqrt{MSRes\ \tilde{x}_0 (X^T X)^{-1} \tilde{x}_0^T}}$$

Interval $P(a < T < b) = 1 - \alpha.$

$$P\left(\underline{\hat{\mu}_{y/\tilde{x}_0}} - t_{\alpha/2, n-k-1} \sqrt{MSRes\ \tilde{x}_0 (X^T X)^{-1} \tilde{x}_0^T},\right.$$

$$\left.\underline{\hat{\mu}_{y/\tilde{x}_0}} + t_{\alpha/2, n-k-1} \sqrt{MSRes\ \tilde{x}_0 (X^T X)^{-1} \tilde{x}_0^T}\right)$$

* Prediction of New Observation ⟹

* Prediction interval ⟹ Generally larger than CI ↗

If $x = \tilde{x}_0$, $y = ?$

point estimate ⟹ of y for given $x = \tilde{x}_0$

$$\boxed{\hat{y} = \tilde{x}_0^T \hat{\beta}}$$

classmate

Date
Page

prediction interval ⇒

Result ⇒ $T^* = \frac{\hat{y}_0 - y_0}{\sqrt{MSRes(1 + \tilde{x}_0 (X^T X)^{-1} \tilde{x}_0^T)}}$

added.

i.e. SD is slighty larger
i.e. variance is slightly larger so interval will be slightly bigger.

$100(1-\alpha)\%$ predictⁿ interval.

$$\left(\hat{y}_0 - t_{\alpha/2, n-k-1} \sqrt{MSRes(1 + \tilde{x}_0 (X^T X)^{-1} \tilde{x}_0^T)}, \right.$$

$$\left. \hat{y}_0 + t_{\alpha/2, n-k-1} \sqrt{MSRes(1 + \tilde{x}_0 (X^T X)^{-1} \tilde{x}_0^T)} \right)$$

(slightely bigger than Mean Response

⇒ Ex. If $x_0 = (\underset{x_1}{29}, \underset{x_2}{2800})$ then market price ?

$$\hat{\mu}_{y|\tilde{x}_0} = \tilde{x}_0^T \hat{\beta}$$

$$= (\underset{\text{imp}}{1}, 29, 2800) \begin{pmatrix} 57.4 \\ 0.0177 \\ -0.015 \end{pmatrix}$$

Calc $MS_{Res}\, \tilde{x}_0 \, (X^TX)^{-1}_{3\times3} \, \tilde{x}_0^T$ & then find interval estm.

& do same for predict^n interval by adding (+1)

written check ✓. in formula

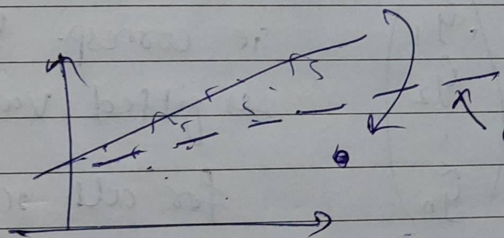
* Model Adequecy Checking → Chpt 4

Assumptions

1. $var(\varepsilon_i) = \sigma^2$, $E(\varepsilon_i) = 0$
2. $cov(\varepsilon_i, \varepsilon_j) = 0 \quad \forall i \neq j$
3. $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$.

If this doesn't hold then it may yield unstable model.

① **Outlier →** Extreme obs which doesn't fit with rest of data.

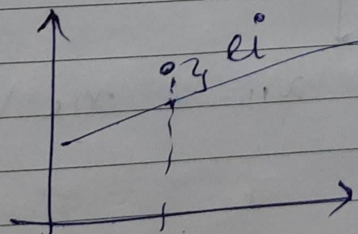


outlier ~~pooled~~ pulled the model towards it

If it has some importance

Identify outlier & then check.

* Detecting Outlier → Residual Analysis ⇒

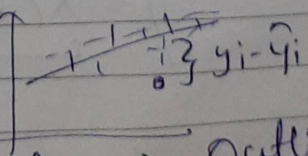


Residual can be treated as Realized or obs value of Error.

$\hat{\varepsilon}_i = e_i \quad \therefore$ error can be estimated by Residual

① Ordinary Residual

$\hat{\varepsilon}_i = e_i = y_i - \hat{y}_i$



If Residual quantity is large then → there is Outlier

$$y = x^T\beta + \varepsilon$$

$$x = \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \end{pmatrix} \qquad \beta = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_K \end{pmatrix}$$

* Residual Analysis ⇒

(1) $\hat{\varepsilon}_i = e_i = y_i - \hat{y}_i$

~~res~~ $SS_{Res} = \sum_{i=1}^{n} e_i = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$

$MS_{Res} = \dfrac{SS_{Res}}{n-k-1}$

* fitted Reg line

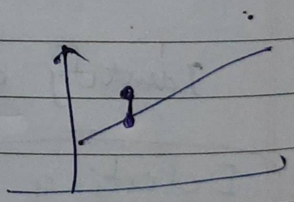
$\hat{y} = x^T\hat{\beta}$

* Vector of fitted ~~model~~ values

$$\hat{Y} = \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{pmatrix}$$

i.e. corresp. to x_1 what is fitted value & so on for all x_K

$$\hat{y} = \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{pmatrix} = X\hat{\beta}$$



$$X\hat{\beta} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1K} \\ 1 & & \vdots & \\ 1 & & \vdots & \\ 1 & x_{n1} & \cdots & x_{nK} \end{pmatrix} \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_K \end{pmatrix}$$

$$\hat{y} = \underbrace{X(X^TX)^{-1}X^T}_{H} y = Hy$$

$H \leftarrow$ hat matrix.

$$e = \begin{pmatrix} e_1 \\ e_2 \\ | \\ e_n \end{pmatrix} = \begin{pmatrix} y_1 - \hat{y}_1 \\ | \\ | \\ y_n - \hat{y}_n \end{pmatrix} = y - \hat{y}$$

$$= Y - HY$$

$$= Y(I - H)$$

(2) $E(e) = 0. \Rightarrow \begin{pmatrix} E(e_1) \\ | \\ E(e_n) \end{pmatrix} = \begin{pmatrix} 0 \\ | \\ 0 \end{pmatrix}$

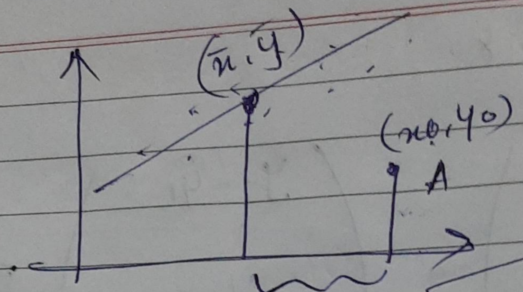
(3) $\boxed{Var(e_i) = \sigma^2 (1 - h_{ii}) \approx MS_{Res}(1 - h_{ii})}$

h_{ii} - ith diagonal elem of hat matrix H.

$$H = \begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1n} \\ h_{21} & h_{22} & \cdots & h_{2n} \\ | & & & \\ h_{n1} & h_{n2} & \cdots & h_{nn} \end{pmatrix} = X(X^TX)^{-1}X^T$$

$h_{ii} \rightarrow$ measure of loc of ith point.

Date
Page



$h_{ii} = \boxed{x_0 - \bar{x} = h_{00}}$

$h_{ii} \rightarrow$ distance of the point from centre of the data in x-space.

if h_{ii} is large then it means pt is far away.

(2) * Studentized Residual $\Rightarrow$

$$r_i = \frac{e_i}{\sqrt{\underbrace{MSRes\,(1-h_{ii})}_{}}}$$ exact SD of e_i

$var(r_i) = 1 \Rightarrow$ when model is correct

If pt is outlier $h_{ii} \uparrow$ then $1-h_{ii}$ will be smaller then $\frac{e_i}{\sqrt{MSRes(1-h_{ii})}}$ will be larger. So for larger values of studentized (r_i) residual that ith pt is outlier.

finding outlier is imp → if useful include
else exclude / delete

classmate
Date
Page

Standardized Residual

$$d_i = \frac{e_i}{\underbrace{\sqrt{MS_{Res}}}_{\text{appx SD of } e_i}}$$

(Intercept model)

$$MS_{Res} = \frac{SS_{Res}}{n-k-1}$$

$$= \frac{\sum_{i=1}^{n} e_i^2}{n-k-1}$$

$$\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

$$MS_{Res} = \frac{\sum_{i=1}^{n}(e_i-\bar{e})^2}{n-k-1} \approx Var(e_i)$$

$\bar{e}=0$ as $E(e_i)=0$

$MS_{Res} =$ e_i

Large value of d_i represents ith obs is outlier

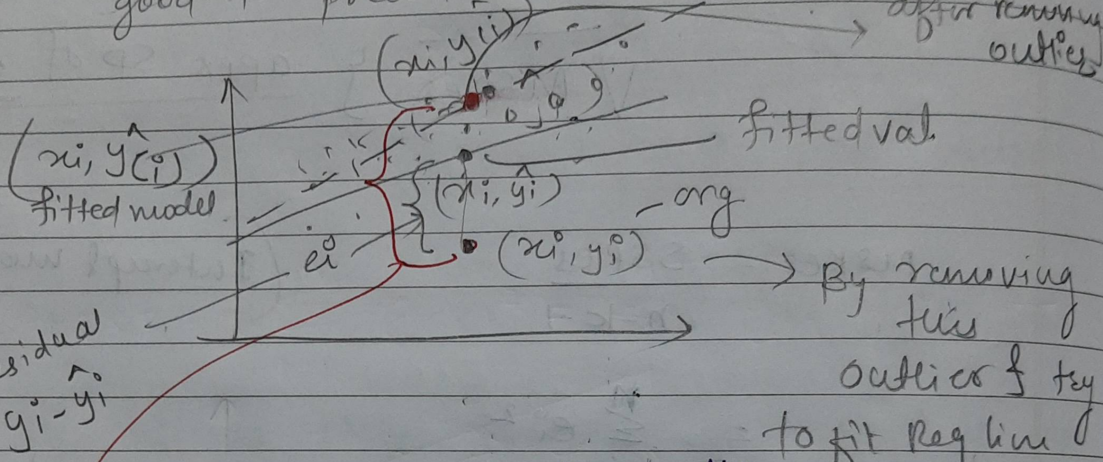
will

be

(r_i)

④ PRESS Residual – Predictⁿ Error Sum of Sqʳ

(also gives Test stat to check whether model is good to predict New observatⁿ)



$e_{(i)} = y_i - \hat{y}_{(i)}$ ← Residual quantity of ith obs when you fit Reg model using $(n-1)$ pts leaving ith pt

is called as PRESS Residual

PRESS Residual for ith obs

$$e_{(i)} = y_i - \hat{y}_{(i)}$$

↳ fitted value of ith observation by not considering (x_i) ie leaving ith obs.

Originary Residual

$e_i = y_i - \hat{y}_i$

for each $i = 1, 2, \ldots n$)

If PRESS $e(i)$ is large means it has influencial Role in determining ith observation i.e. whether its imp or not

* If x_i is influential it will pull regression line towards it & if x_i is not influencial it won't pull Reg line towards it

for each. $i = 1, 2, \ldots n$

for
$$e(i+1) = y_{i+1} - \hat{y}(i+1)$$ (i+1) PRESS Residual

∴ To cal PRESS resid, fit diff reg model & then another line will come for each new value

$$\boxed{\begin{aligned} e(i) &= y_i - \hat{y}(i) \\ e(i) &= \frac{e_i}{1 - \underbrace{h_{ii}}_{\text{ith diagonal elem of } H}} \end{aligned}}$$

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

classmate
Date
Page

If ith $\overbrace{\text{PRESS res}}^{e(i)}$ is large then the

Result ⟹ corres ith obs is ~~Resid~~ outlier.

how Good model is for New obs.

PRESS Statistics ⇒ PRESS

$$PRESS = \sum_{i=1}^{n} e(i)^2 = \sum_{i=1}^{n} \left(y_i - \hat{y}(i)\right)^2$$

$$= \sum_{i=1}^{n} \left(\frac{e_i}{1-h_{ii}}\right)^2$$

sum of sqr of PRESS residual called as PRESS

* PRESS has huge role in determining whether given model will predict the New Observation *

$$SSRes = \sum_{i=1}^{n} e_i^2$$

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

sum of sqr of ordinary residual.

R² for prediction ⇒

$$R^2_{pred} = 1 - \frac{PRESS}{SST}$$

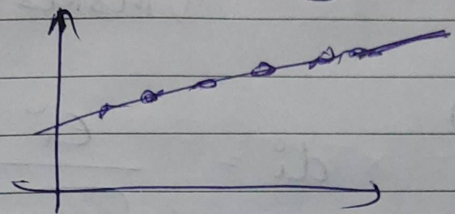
* How good is the model to predict future observation

① $0 \leq R^2_{pred} \leq 1$

② $R^2_{pred} = 1$ when PRESS $= 0$ means each of PRESS residual is zero meaning $y_i \equiv \hat{y}(i)$

If you remove pt 4 try to fit Regression model then check predicted value that will be same as original valu

ie all pts are lying on same line.



$R^2 pred = 1 - \frac{455}{5784}$ — RRESS, ← SST.

$R^2 pred = 0.9209$.

$R^2 = 1 - \frac{SSRes}{SST} \Rightarrow 0.9596$.

↓ 95.96% variabil: explained by fitted Reg model

* If you use given model for predⁿ of New obs then 92.09% of time predⁿ will be correct.

ordinary -- e_i

Student → r_i

Std Res → d_i

29/M

ordinary ⇒ $e_i = (y_i - \hat{y}_i)$

Student ⇒ $r_i = \dfrac{e_i}{\underbrace{\sqrt{MSRes(1-h_{ii})}}_{\text{std dev of } e_i}}$

Std Resi ⇒ $d_i = \dfrac{e_i}{\sqrt{MSRes}}$

PRESS ⇒ $y_i - \hat{y}_{(i)}$

To detect outlier :-

New app- to detect outlie

⑤ R-student measure.

$t_i = \dfrac{e_i}{\sqrt{S_{(i)}^2 (1-h_{ii})}}$

$S_{(i)}^2 = \hat{\sigma}^2 = \dfrac{(n-k-1) MSRes - \dfrac{e_i^2}{(1-h_{ii})}}{n-k-2}$

estimate of ~~m~~ σ^2 by removing ith obsv

lev pt may any

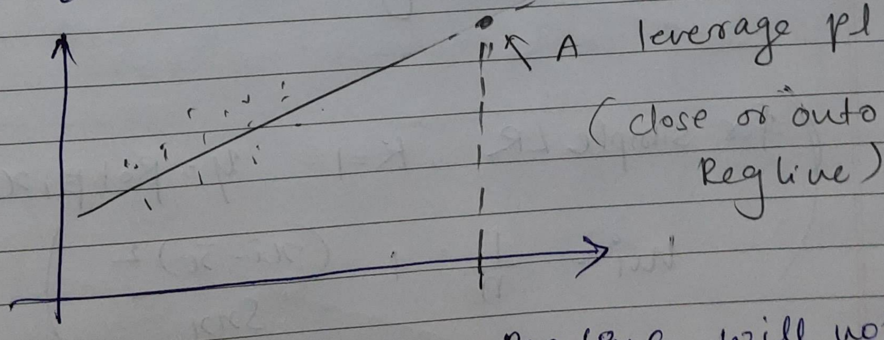
If ith obs is not outlier then not much diff bet Studentized Res & R-Student meas.

for large value of ti, ith obs is outlier.

(8) Leverage Point (chptr 6) ⇒

A pt is said to be lev pt if it has unusual x-values (but it lies almost on the Regression line passing through rest of sample pts.) → possibility

lev pt lie may anywhere



Effect of lev pt on Reg line will not be there

But for other estimates it affects.

eg (SST = ~~$\Sigma y_i - \bar{y}$~~ Total variation will affect.

∴ Lev pt will affect overall summary eg
how good model fits ✓

Detecting leverage pt ⇒

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

$$= \begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1n} \\ h_{21} & h_{22} & & \\ \vdots & & & \\ h_{n1} & h_{n2} & \cdots & h_{nn} \end{pmatrix}$$

h_{ii} → measures dist of ith obv from centre of the data in x-space

(for simple LR, $K=1$ $y = \beta_0 + \beta_1 x_1$

$$h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}}$$
)

* Guiding principal. ⇒

If $h_{ii} > \frac{2(k+1)}{n}$ then ith obs is

called Leverage pt.

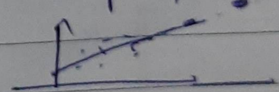
$n = 25, \quad k = 2$

$\frac{2 \times 3}{25} = 0.24$ if $h_{ii} > 0.24$ then

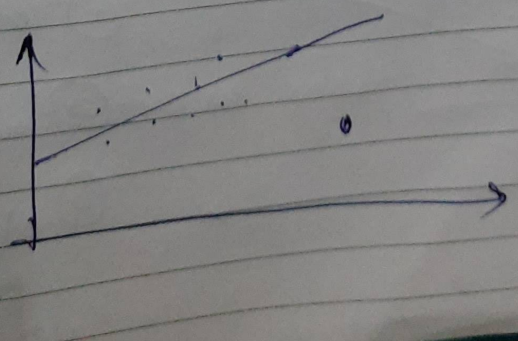
ith pt is leverage pt

2cases for Lev pt ⇒

case ① Lev pt ~~lies~~ almost lies on Reg line passing through Rest of sample pts.



case ② Lev pt does not lie on fitted Reg line obt by Rest of the sample pts.



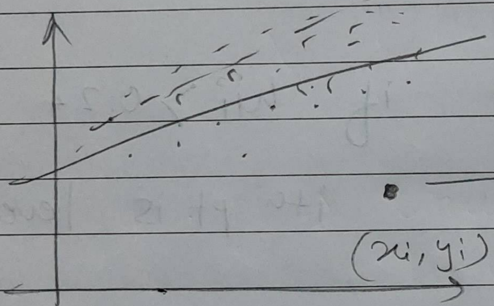
unusual x-value

* Diff betn Outlier & Lev pt ⇒

some of the outlier can be Lev pt but not all outliers are lev pt

All Lev pts are outliers

But all outliers are not lev pts.

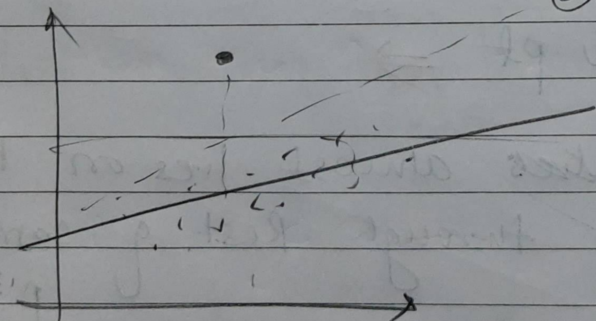
1. 

(x_i, y_i)

① outlier due to extreme x-value

② Leverage pt (x-value)

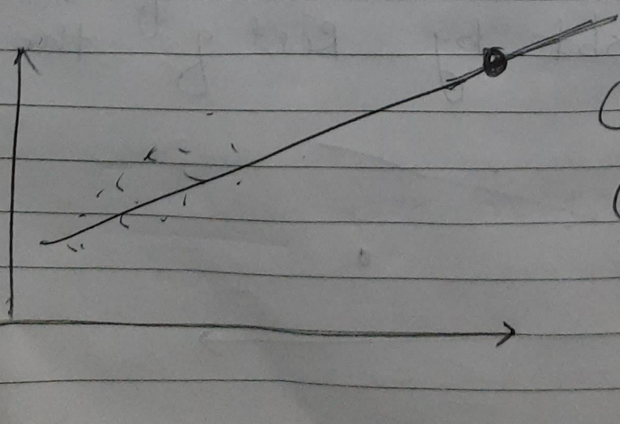
③ Infl pt

2. 

① outlier (Unusual y-value)

② Not lev pt (as no unusual x-value y-value is unusual)

③ Inf pt

3. 

① outlier (for unus x & y value)

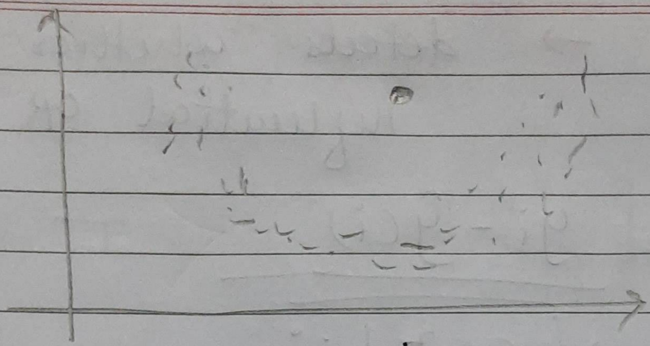
② lev pt

③ Not influential pt

would affect coeff of Reg line

classmate
Date
Page

(4)

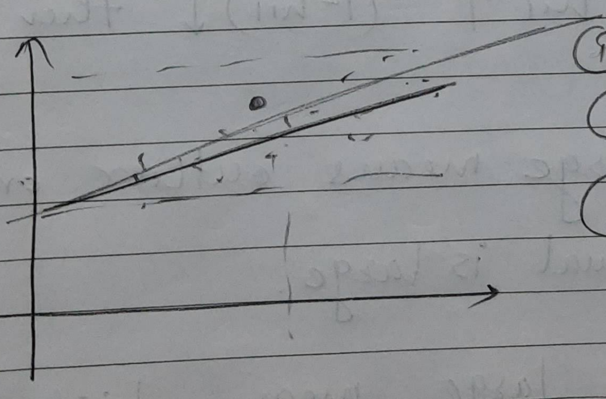


(1) outlier
(doesn't lie in trend of rest of data)

Influential pt ⇒

An obs that, if removed it significantly impacts coefficients & other summary of model on fitted Reg line.

(5)



(1) not outlier
(2) not lev pt
(3) Influential pt
Slight influence on coeff of Regline

* Lev pt may/may not impact Reg coefficient

* B

Date
Page

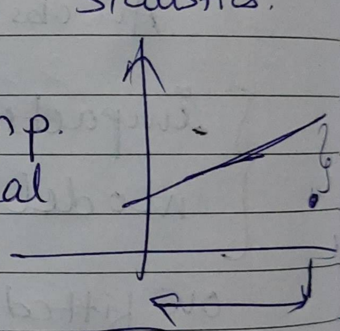
① DFFITSi → detects whether pt is influential OR not.

$$DFFITS_i = \frac{y_i - \hat{y}(i)}{\sqrt{S(i)^2 h_{ii}}}$$

$$= \sqrt{\frac{h_{ii}}{1-h_{ii}}} \times t_i$$

t_i — R-Student Statistics.

Influen pt ⇒ Unusual x-comp.
higher Residual



then $h_{ii} \uparrow$ $(1-h_{ii}) \downarrow$ then $\sqrt{\frac{h_i}{1-h_{ii}}}$ large

if t_i large means outlier or when Residual is large.

* if t_i large mean high residual

* if $\sqrt{\frac{h_i}{1-h_i}}$ large means large x-comp.

~~Guidt~~ Guiding principle ⇒

$$|DFFITS_i| > 2\sqrt{\frac{(k+1)}{n}}$$

then ith obs is an influencial pt

(2) DFBETAS ⇒

$$DFBETAS_{\beta_j, i} = \frac{\hat{\beta}_j - \hat{\beta}_{j(i)}}{\sqrt{S_{(i)}^2 C_{jj}}}$$

$\hat{\beta}_j$ = estimate of para by taking all pts.

$\hat{\beta}_{j(i)}$ = estimate without taking ith pt

if ~~diff~~ diff $\hat{\beta}_j - \hat{\beta}_{j(i)}$ is large means ith has dramatic effect on coefficients.

(whether ith pt is affecting jth coeff)

If diff less means ith obs has NO effect in jth coeff.

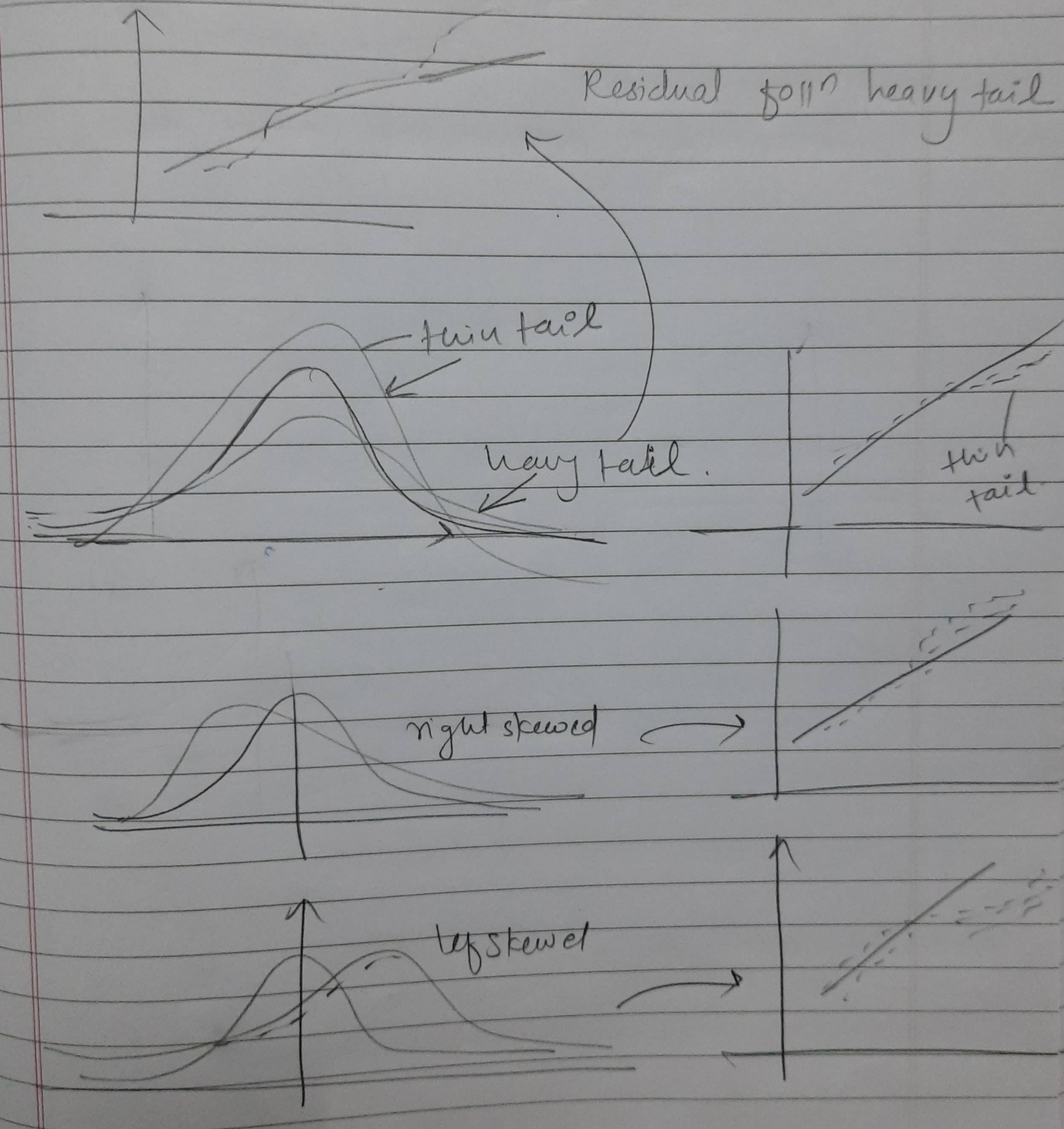
Guiding pri

If $|DFBETAS_{j,i}| > \frac{2}{\sqrt{n}}$ then

ith obs is an influential pt ie - it is affecting j coefficients (β_j)

* Normality plot ⇒

Checking whether $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$.



(Assumpt^n Error has const var)

$$\hat{\varepsilon_i} = e_i$$