Budhanada - Stat

Regression - Midsem
Time series - Endsem

Endsem exam (RE + TS)

Prediction
↓
with less error
↓
Absolute distance (L_1 norm) or squared distance (L_2 norm) } Assump
↓
Dataanalysis (Given)
↓
model Building
↓
Prediction

Simple Leniar Regression

$$Y = B_0 + B_1 X + \epsilon$$

Y ↓ Dependent

X ↓ Independent of Y; B_0, B_1 and non-random/Non-stochastic

{ value of the variable is final & no noise/Error associated with x, whatever you know is final }

ϵ ↓ $\epsilon \sim N(0, \sigma^2)$
σ^2 unknown

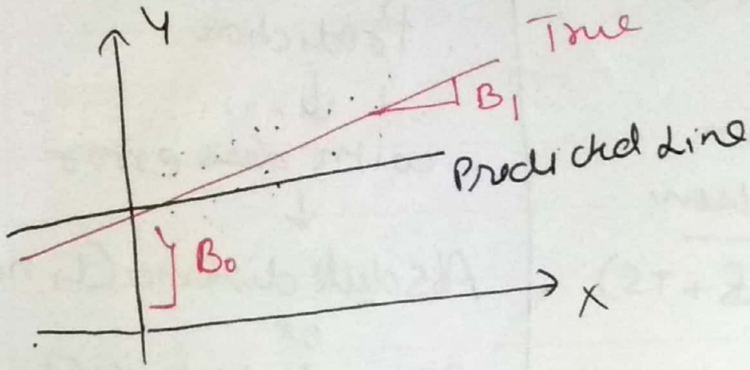
Leniar model - means - Leniar in parameter not in variable

Eg:

$Y = B_0 + B_1 x$ is a leniar

$Y^* = B_0^* + B_1^* x + B_2^* x^2$ is also a leniar mode

$$Y^* = \begin{pmatrix} 1 & x & x^2 \end{pmatrix} \begin{pmatrix} B_0^* \\ B_1^* \\ B_2^* \end{pmatrix}$$



True $\quad Y = B_0 + \beta_1 x + \epsilon$

Estimate $\quad \hat{Y} = \hat{B}_0 + \hat{B}_1 x_{new}$

Error : (Squared error)
or L_2 (Norm)

$\underset{\sim}{y} = (y_1, y_2 \cdots y_n)^T$ then the length of y is

$$\|y\|_2^2 = \sum_{i=1}^{n} y_i^2 = y \cdot y = y^T y$$

$$\underset{\sim}{Z} = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} \quad \underset{\sim}{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

$$\|\underset{\sim}{y}\|_2 = \sqrt{\sum_{i=1}^{n} y_i^2}$$
$$\|\underset{\sim}{y}\|_1 = \sum_{i=1}^{n} |y_i|$$

then

$$\|z - y\|_2^2 = \sum_{i=1}^{n} (y_i - z_i)^2 = (\underset{\sim}{y} - \underset{\sim}{z})^T (\underset{\sim}{y} - \underset{\sim}{z})$$
$$= (\underset{\sim}{y} - \underset{\sim}{z}) \cdot (\underset{\sim}{y} - \underset{\sim}{z})$$

<u>we</u> want to <u>minimize</u>

$$\cancel{S = \sum_{i=1}^{n} [y_i - (B_0 + B_1 x)}$$

$$S = \sum_{i=1}^{n} \left[y_i - (B_0 + B_1 x_i) \right]^2$$

Scanned by CamScanner

Then the Estimate we get for B_0 & B_1 are the known as least square estimates of B_0, B_1.

$$\hat{B}_{0_{LS}}, \hat{B}_{1_{LS}}$$

given data = $\{(y_1, x_1), (y_2, x_2) \dots (y_n, x_n)\}$

$$\left.\begin{aligned} \frac{\partial S}{\partial B_0} &= 0 \\ \frac{\partial S}{\partial B_1} &= 0 \end{aligned}\right] \text{normal equations}$$

Solve the normal equations for $\hat{B}_{0_{LS}}$ & $\hat{B}_{1_{LS}}$

$$\hat{B}_{0LS} = \bar{y} - \hat{B}_{1LS}\bar{x}$$

$$\hat{B}_{1LS} \quad \frac{\Sigma y_i x_i - n\bar{x}\bar{y}}{\Sigma x_i^2 - n(\bar{x})^2}$$

$\Rightarrow$ we can predict

$$\hat{y}_{new} = B_{0LS}$$

we can estimate the error

$$\epsilon^2 = \sum_{i=0}^{n} (\hat{y}_i - y_i)^2$$

Testing we can't do as both $\hat{B}_{0\,LS}$, $\hat{B}_{1\,LS}$ are Random variable

Res

Results from Linear Algebra

Data (x_1, y_1) (x_2, y_2) . . . (x_n, y_n)

$$y_i = B_0 + B_1 x_i + \epsilon_i \quad \forall\, i = 1, 2 \ldots n$$

$$\underset{\sim}{Y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \quad \underset{\sim}{B} = \begin{pmatrix} B_0 \\ B_1 \end{pmatrix}$$

$$\underset{\sim}{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

① ϵ_i iid $N(0, \sigma^2)$

As mean would automatically get adjusted in the constant

② B_0, B_1, σ^2 are unknown

③ x_i's are non random

$$\underset{\sim}{Y} = X\underset{\sim}{B} + \underset{\sim}{\epsilon}$$

$\underset{n\times 1}{\underset{\sim}{Y}} \in \mathbb{R}^n$ $\quad \underset{n\times 2}{X}$, $\underset{\sim}{B} \in \mathbb{R}^2$, $\underset{\sim}{\epsilon} \in \mathbb{R}^n$

Defination Vector Space

A vector Space V over $\mathbb{R}$, denoted by $(V, +, \cdot, \mathbb{R})$ has the following proporty $\forall \alpha, \beta \in \mathbb{R}$ & $x, y, z \in V$

(1) (a) Closure (a) $+ : V \times V \longrightarrow V$

Asso (b) $(x + y) + z = x + (y + z)$

Identity (c) $\exists\ 0 \in V$ Such that

$$0 + x = x + 0 = x \quad \forall x \in V$$

Inverse (d) $\exists\ -x\ \forall x \in V$ such that

$$-x + x = x + (-x) = 0$$

Commutative (e) $x + y = y + x \quad \forall x, y\ V$

(2) $\cdot : \mathbb{R} \times V \to V$

(3) $\alpha \cdot (\beta \cdot \underset{\sim}{x}) = (\alpha \cdot \beta) \cdot \underset{\sim}{x}$

(4) $1 \cdot \underset{\sim}{x} = \underset{\sim}{x}$

(5) $(\alpha + \beta) \cdot x = \alpha \cdot x + \beta \cdot x$

(6) $\alpha \cdot (\underset{\sim}{x} + \underset{\sim}{y}) = \alpha \underset{\sim}{x} + \alpha \underset{\sim}{y}$

Example

(1) $\mathbb{R}^n$ is a vector spaces

(2) C^n -- -- --

(3) P_n : {all polynomials with degree $\leq n$} is also a vector space

$$P_n(x) = a_0 + a_1 x \text{ ---- } a_n x^n$$

$$= (a_0 \cdots\cdots a_n) \begin{pmatrix} 1 \\ x \\ x^2 \\ \vdots \\ x^n \end{pmatrix}$$

Subspace

Let $(V, +, \cdot, \mathbb{R})$ is a vector space & $S \subseteq V$ such that $(S, +, \cdot, \mathbb{R})$ is also a vector space then S is known as the subspace of V

A subset which is a vector space under the same operation

Examble

(1) If $\mathbb{R}^2$ is a vector space with usual vector additio & scalar multiplication

then (1) x axis is a Subspace
(2) y axis is Subspace
(3) Any line passing through origin)

(2) In $\mathbb{R}^3$, (a) (0, 0, 0)
(b) All the lines passing through (0,000)
(c) All planes passing through (0, 0 0)
are Subspaces

(3) If P_7 is the vector space
$S = P_r$ $r \le 7$ is a subspace

$P_n \rightarrow$ n+1 dimensional vector space

Independent vectors : A set of vectors $\underset{\sim}{v_1}, \underset{\sim}{v_2} \ldots \underset{\sim}{v_n}$ are said to be (linearly) independent iff

$$\sum_{i=1}^{n} c_i v_i = \underset{\sim}{0} \iff c_1 = c_2 \ldots c_n = 0$$

$\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ by $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$

$$\Leftrightarrow c_1 \begin{pmatrix} 0 \\ 1 \end{pmatrix} + c_2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Span : Let $\{ \underset{\sim}{v_1}, \underset{\sim}{v_2} \dots \underset{\sim}{v_K} \}$ be a collection of k vectors in R^n, the span of this collection is defined to be <u>all possible linear combinations</u> of $\{ v_1, v_2 \dots v_K \}$

$$Sp \{ \underset{\sim}{v_1}, \underset{\sim}{v_2} \dots \underset{\sim}{v_K} \}$$

$$= \left\{ \sum_{i=1}^{K} c_i v_i \;\middle|\; c_i \in R, v_i \in R^n \right\}$$

Ex 1) $\underset{\sim}{v} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ $\underset{\sim}{v_2} \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

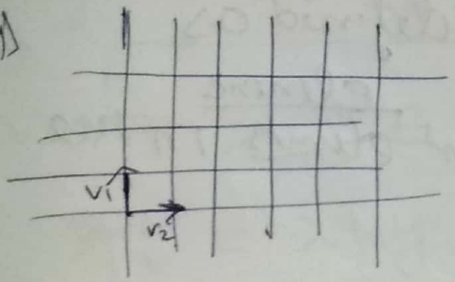
$$Sp \{ \underset{\sim}{v}, \underset{\sim}{v_2} \} = \left\{ c_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} \;\middle|\; c_1, c_2 \in R \right.$$

$$= \mathbb{R}^2$$

2) $w_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ $w_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

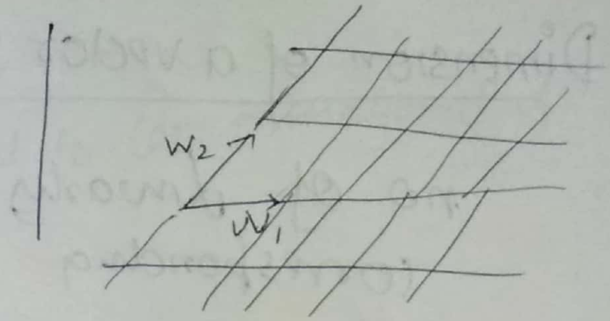
$$Sp \{ \underset{\sim}{w_1}, \underset{\sim}{w_2}) = \mathbb{R}^2$$

then what is the diff b/w ~~w_1~~ 1 & ~~w_2~~ 2

1) 

R^2

$\begin{pmatrix}1\\0\end{pmatrix}\begin{pmatrix}0\\1\end{pmatrix}$

ii) 

R^2

$\begin{pmatrix}1\\0\end{pmatrix}\begin{pmatrix}1\\1\end{pmatrix}$

Same space can be Represented by 2 diff vect

<u>Basis</u> : A basis of vector space $(S \subseteq V)$ ~~of~~ or a subspace $(S \subseteq V)$ defined as a set of <u>linearly independent vectors</u> such that the <u>span of the set is the vector space</u> V or the subspace $(S \subseteq V)$ respectively

earlier 2 example of basis of R^2

=> <u>Basis may ~~may~~ not be unique</u>

$\left\{\begin{pmatrix}1\\0\end{pmatrix}\begin{pmatrix}0\\1\end{pmatrix}\right\}$ is a basis of $\mathbb{R}^2$

$\left\{\begin{pmatrix}1\\0\end{pmatrix}\begin{pmatrix}1\\0\end{pmatrix}\right\}$ is a basis of $\mathbb{R}^2$

$\left\{\begin{matrix}1\\0\end{matrix}\right\}$ cannot generate $\mathbb{R}^2$ but it generates x axis which is a subspace of $\mathbb{R}^2$

Dimension of a vector space is defined as no of linearly independent elements in the corresponding basis

$V = Sp\{B\}$ then $|B|$ is the dimension of V if B is the basis

$\dim(V) = |B|$
when B is basis of V

$A = \{(x, y, 0), x \in R, y \in R\}$

$A = Sp\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\} = \left\{ x \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + y \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, x, y \in R \right\}$

Among $A \subset R^2$, $\dim(A) = 2$

If $V_1, V_2 \ldots V_K \in R^n$ then

i) $K > n$ then the collection is always linearly dependent

$$\left\{ \underbrace{\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}}_{R^3}, \underbrace{\begin{pmatrix} a \\ b \\ c \end{pmatrix}}_{K} \right\}$$

Scanned by CamScanner

Orthogonal vectors

vector $\underset{\sim}{x}\ \underset{\sim}{y} \in V$ are said to be orthogonal if

$$x \cdot y = x^T y = y^T x = 0$$

Notation : $\underset{\sim}{x} \perp y$

(A) $\underset{\sim}{0}$ is orthogonal to any vector by itself

Orthogonal basis

If the vector of the basis set B are orthogonal to each other then the Basis is known as orthogonal basis.

* helps in making calculation & Analysis easier

Orthogonal Compliment

If S is a subspace of V then the orthogonal compliment of S is defined by

$$S^{\perp} = \{ \underset{\sim}{v} \mid v^T u = 0 \quad \forall\, \underset{\sim}{u} \in S \}$$

$$S^{\perp} \cap S = \underset{\sim}{0} \neq \phi$$

ex

$\mathbb{R}^3 = V$

$S = sp\left\{ \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\}$

$S^{\perp} = sp\left\{ \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$

Scanned by CamScanner

$\dim(S) + \dim(S^{\perp}) = \dim(V)$

★ These dimentions will give us the degrees of free dom in statistical analysis

---

Projection Matrix : If S is a subspace of a vector space then the projection matrix P_S for $S \subseteq V$ is satisfie the following condition

① If $\underset{\sim}{v} \in S$, ~~Project~~ $P_S \underset{\sim}{v} = \underset{\sim}{v}$

② $\forall v \in V$, $P_S \underset{\sim}{v} \in S$

$\underset{\sim}{v} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \in R^3 = V$, $S = \left\{ (x, y, 0) \mid \begin{matrix} x \in R \\ y \in R \end{matrix} \right\}$

$P_S = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ $\quad P_S \underset{\sim}{v} = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$

$v_1 = \begin{pmatrix} 3 \\ 4 \\ 0 \end{pmatrix} \in S$ $\quad P_S \underset{\sim}{v_1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{pmatrix} 3 \\ 4 \\ 0 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \\ 0 \end{pmatrix}$

★ If P_S is a projection matrix for $S \subseteq V$ and $(I_n - P_S)$ is the projection matrix for $S^{\perp} \subseteq V$ then P_S and $I_n - P_S$ are orthogonal projection matrices of S & $S^{\perp}$ respectively

$P_S \underset{\sim}{V} = V_1$ $\quad$ V_1, V_2 are orthogonal matrix

$(I_n - P_S) \underset{\sim}{V} = V_2$

$$P_S = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad I_3 - P_S = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$P_S \underset{\sim}{V} = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \quad (I_3 - P_S) V = \begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix}$$

$$\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix} = 0$$

$$(P_S \underset{\sim}{V})^T (I_3 - P_S) \underset{\sim}{V}$$

$$= \underset{\sim}{V}^T (P_S^T I_3 - P_S^T P_S) \underset{\sim}{V}$$

$$= \underset{\sim}{V}^T (P_S - P_S P_S) \underset{\sim}{V} \quad - \quad P_S^T = P_S, \; P_S^2 = P_S$$

$$= \underset{\sim}{V} \; O_{3\times 3} V = 0$$

Idempotent

(Q)

<u>Imp Theorem</u> Projection matrix is an Idempotent matrix

Let P_S be the projection of $S \subseteq V$

Only when $\Rightarrow \boxed{\forall \underset{\sim}{V} \in V}$ $\quad - \quad P_S \underset{\sim}{V} = P_S (P_S \underset{\sim}{V})$

$$\Rightarrow P_S = P_S^2$$

Th: Projection matrix has eigen values 1 & 0 only

Pf

Let λ be an eigen value of P_S with nonzero eigen vector $\underset{\sim}{v}$ then

① $\quad P_S \underset{\sim}{v} = \lambda \underset{\sim}{v}$ ①

$$P_S \underset{\sim}{v} = P_S^2 \underset{\sim}{v} = P_S \cdot (P_S v) = P_S(\lambda v) = \lambda P_S v = \lambda^2 \underset{\sim}{v} \quad ②$$

① − ②

$\Rightarrow \underset{\sim}{0} = \lambda(1-\lambda) \underset{\sim}{v}$

$\underset{\sim}{v}$ is a non zero vector

$\Rightarrow 0 = \lambda(1-\lambda) \Rightarrow \lambda = 1$ or 0

Th If $\{v_1, v_2 \dots v_k\}$ is an orthonormal basis (norm = 1 & orthogonal to each other)

$$P_S = \sum_{i=1}^{k} \underset{\sim}{v_i} v_i^T$$

then is an orthogonal projection matrix for S

Proof HW

<u>Column space</u> : The column space of $A = [\underset{\sim}{a_1}, a_2 \cdots \underset{\sim}{a_n}]$

$$C(A) = Sp\{\underset{\sim}{a_1}, \underset{\sim}{a_2} \cdots \underset{\sim}{a_n}\}$$
$$= \left\{ \sum_{i=1}^{n} x_i a_i \mid \underset{\sim}{x} \in R \right\}$$
$$= \{ A \underset{\sim}{x} \mid x \in R^n \}$$

Row space A is $R(A) = C(A^T)$

<u>Prop</u>

① $C(A : B) = C(A) + C(B)$
② $C(AB) \subseteq C(A)$
③ $\dim(C(A)) = \text{Rank}(A)$
④ $C(AA^T) = C(A) \Rightarrow \text{Rank}(AA^T) = \text{Rank}(A)$
⑤ If A has n column then
$\dim(C(A)^{\perp}) = n - \text{Rank}(A)$

<u>Proof</u> ④

<u>Part 1</u> $C(AA^T) \subseteq C(A)$ by ②

<u>Part 2</u> Let $\underset{\sim}{L} \in (C(AA^T))^{\perp}$

$\Leftrightarrow \underset{\sim}{L}^T A A^T = \underset{\sim}{0}^T$ ①

$\underset{\sim}{L}^T A A^T \underset{\sim}{L} = \underset{\sim}{0}^T \underset{\sim}{L}$

A'

$\therefore (A^T L)^T (A^T L) = 0$

$\Rightarrow A^T L = 0$

$\Rightarrow \alpha^T A = 0^T$ (2)

$\Leftrightarrow \alpha \in (C(A))^T$

$(C(AA^T))^{\perp} \subseteq (C(A))^{\perp}$

For any matrix A the projection matrix for C(A) is (AA^-) and the orthognal projection matrix is $A(A^T A)^- A^T$

(1) Positive definite matrix

$A_{n \times n}$ is said to be a p.d matrix

if $\underset{\sim}{x}^T A \underset{\sim}{x} > 0 \;\forall\; \underset{\sim}{x} \neq \underset{\sim}{0}$

(2) Positive semi definite matrix

if $\underset{\sim}{x}^T A \underset{\sim}{x} \geq 0 \;\forall\; x \neq \underset{\sim}{0}$

$\Leftrightarrow |A| > 0$ $\Leftrightarrow$ pd. matrix

$|A| \geq 0$ $\Leftrightarrow$ p.s.d matrix

Generalized inverse of a matrix A usualy denoted by G or A^- with prop that

$$AGA = A$$
$$AA^-A = A$$

① Need not be unique
② True for non square matrix also

Multivariate Analysis

Def 1 $\underset{\sim}{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = (x_1, x_2 \ldots x_n)^T$

$$E(\underset{\sim}{x}) = \begin{pmatrix} E(x_1) \\ E(x_2) \\ \vdots \\ E(x_n) \end{pmatrix} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} = \underset{\sim}{u} = (u_1, u_2 \ldots u_n)^T$$

if $E(x_i) < \infty \ \forall \ i = 1, 2 \ldots n$

Df 2 : Dispersion matrix of $\underset{\sim}{x}$ / variance & covariance matrix of $\underset{\sim}{x}$

$$D(\underset{\sim}{x}) = ((\mathrm{cov}(x_i, x_j)))_{ij}$$
$$= ((E(x_i x_j) - E(x_i)E(x_j)))_{ij}$$
$$= ((E(x_i x_j) - u_i u_j))_{ij}$$

Scanned by CamScanner

Representation : Σ or Σ_x
$n \times n \longrightarrow$ dimension

Diagonal elements are the variances of the corresponding random variables. [put $i=j$ Note - you can see it's variance]

NB [other elements are normal covariance

Properties

① $cov(\underset{\sim}{U}_p, \underset{\sim}{V}_q) = ((cov(U_i, V_j)))_{ij}$

dim : $p \times q$

Colum Row

$= E(\underset{\sim}{U}_p \underset{\sim}{V}_q^T) - \underset{\sim}{\mu}_U \mu_V^T$

matrix

$= cov(\underset{\sim}{U}_p, \underset{\sim}{V}_q)$

② $E(\underset{\sim}{x} + \underset{\sim}{b}) = E(\underset{\sim}{x}) + \underset{\sim}{b}$

where $\underset{\sim}{b}$ is constant vector

③ $D(\underset{\sim}{x} + \underset{\sim}{b}) = D(\underset{\sim}{x})$ where $\underset{\sim}{b}$ is constant vecto

Scanned by CamScanner

④ $cov(\underset{\sim}{x}+\underset{\sim}{b}, \underset{\sim}{y}+\underset{\sim}{c}) = cov(\underset{\sim}{x}, \underset{\sim}{y})$ where

$\underset{\sim}{b}, \underset{\sim}{c}$ are constant vector

Product Result

① $E(\underbrace{\underset{\sim}{l}^T \underset{\sim}{x}}_{\text{dot product}}) = \underset{\sim}{l}^T \underset{\sim}{\mu}$ where l is constant vector

② $D(\underset{\sim}{l}^T \underset{\sim}{x}) = l^T \Sigma l = Var(\underset{\sim}{l}^T \underset{\sim}{x})$

③ $E(\underset{\downarrow}{A} \underset{\downarrow}{\underset{\sim}{x}}) = A \underset{\sim}{\mu}$

gen

matrix × vector = vector

④ $D(A\underset{\sim}{x}) = A \Sigma A^T$

⑤ $Cov(\underset{\sim}{U_p}, \underset{\sim}{V_q}) = \Gamma = E(\underset{\sim}{U_p} \cdot V_q^T) - \cancel{U_p V_q^T}\ \underset{\sim}{U_u}\ \underset{\sim}{U_v^T}$

generalized then $Cov(A\underset{\sim}{U}, B\underset{\sim}{V}) = A \Gamma B^T$

$\Rightarrow \quad cov(A\underset{\sim}{x}, B\underset{\sim}{x}) = A \Sigma B^T$

Proof ③

$E(A\underset{\sim}{x})$

$= E\left[\underset{\sim}{a_1}\ \underset{\sim}{a_2} \quad \cdots \quad \underset{\sim}{a_n}\right]\begin{pmatrix} x_1 \\ x_1 \\ \vdots \\ x_n \end{pmatrix}$

$= E\left(\sum_{i=1}^{n} x_i \underset{\sim}{a_i}\right) = \sum_{i=1}^{n} u_i \underset{\sim}{a_i} = A\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} = A\underset{\sim}{u}$

Th: Show that $D(\underset{\sim}{x})$ is positive semi definite matrix

$D(\underset{\sim}{x}) = \Sigma$

Let $l \neq 0$ and we need to

prove that $\underset{\sim}{l}^T \Sigma l \geq 0$

$\Leftrightarrow \operatorname{var}(\underset{\sim}{l}^T \underset{\sim}{x}) \geq 0$

because the variance of a scalar random variable is non negative

Th: let $E(\underset{\sim}{x}) = \underset{\sim}{u}$ $\quad D(\underset{\sim}{x}) = \Sigma$

$$P\big((\underbrace{\underset{\sim}{x} - \underset{\sim}{u}}_{\text{random}}) \in \underbrace{C(\Sigma)}_{\text{deterministic}}\big) = 1$$

linear combination

$(\text{In } Y = X - \underset{\sim}{u}$

$\Rightarrow X = U + IY$

Th To prove that, it is enough to show that any element is in $(C(\Sigma))^{\perp}$ is orthogonal to $(\underset{\sim}{X} - \underset{\sim}{\mu})$

Let $\underset{\sim}{l} \in (C(\Sigma))^{\perp}$

(from earlier proof) $\Leftarrow$ $l^T \Sigma = 0^T$

$$l^T \Sigma l = 0^T l = 0$$

$$\Rightarrow \text{Var}(l^T \underset{\sim}{X}) = 0$$

$$\Rightarrow \text{var}(l^T(\underset{\sim}{X} - \underset{\sim}{\mu})) = 0 \quad \text{--- (1)}$$

& we know $E(l^T(\underset{\sim}{X} - \underset{\sim}{\mu})) = 0$ — (2)

$$\Rightarrow P(\underset{\sim}{l}^T(\underset{\sim}{X} - \underset{\sim}{\mu}) = 0) = 1$$

$\Rightarrow \underset{\sim}{l}$ is orthogonal to $(\underset{\sim}{X} - \underset{\sim}{\mu})$

Th: Let X be a random vector with n components such that

$E(\underset{\sim}{X}) = U$, $D(\underset{\sim}{X}) = \Sigma$ and

rank $(\Sigma) = r \leq n$ -

If we assume that $\Sigma = BB^T$

~~such~~ where B is a $(n \times r)$ matrix

& C is a left inverse of B

is $\underset{(r \times n)}{C} \underset{(n \times r)}{B} = I_r$ then defining

$Y = C(\underset{\sim}{X} - \underset{\sim}{U})$ we get

① $E(\underset{\sim}{Y}) = 0$

② $D(\underset{\sim}{Y}) = I_r$

③ $X = U + BY$. with prob 1

Scanned by CamScanner

$$\Sigma_{n\times n} = \Sigma^T$$

$$\Sigma = P D P^T$$

where P is an orthogonal matrix

$$D = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix} \qquad \lambda_i \geq 0$$

$$\Sigma \underset{\sim}{P_i} = \lambda_i P_i$$

$$\Sigma^{\frac{1}{2}} = P D^{\frac{1}{2}} P^T$$

<u>Multivariate Normal</u>: A random vector X_n is said to follow multivariate normal if it has pdf

$$f(x) = \frac{e^{\left\{-\frac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\mu})^T \Sigma^{-1} (\underset{\sim}{x}-\underset{\sim}{\mu})\right\}}}{(\sqrt{2\pi})^n \sqrt{|\Sigma|}}$$

$$E(\underset{\sim}{X}) = \underset{\sim}{\mu} \qquad D(\underset{\sim}{X}) = \Sigma_{n\times n}$$

univar

$$\frac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sqrt{2\pi}\,\sigma}$$

Consider. $\underset{\sim}{X} \in \mathbb{R}^2$

$$E(\underset{\sim}{X}) = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \qquad D(X) = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} = \underset{2\times 2}{\Sigma}$$

$$f\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{\exp\left\{-\frac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\mu})^T \Sigma^{-1}(x-\mu)\right]}{(\sqrt{2\pi})^2\sqrt{|\Sigma|}}$$

$$= \frac{e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{x-\mu_2}{\sigma_2}\right)\right]}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$

Let X be random vector in R^n $\qquad A \simeq$ matrix $n\times n$

$$E(X) = \mu \quad \& \quad D(\underset{\sim}{X}) = \Sigma$$

$$E(X^T A X) = tr(A\Sigma) + \underset{\sim}{\mu}^T A \underset{\sim}{\mu}$$

$$E(\underset{\sim}{X}^T A \underset{\sim}{X}) \Rightarrow E[tr(\underset{\sim}{X}^T A \underset{\sim}{X})$$

As this is scalar
so $X^T A X = tr[X^T A X$

$= E\left[\mathrm{tr}(\underset{\substack{\downarrow \\ \text{scalar multiplier}}}{A}\ \underset{\substack{\downarrow \\ \text{random par}}}{XX^T})\right]$

$\underset{\sim}{x}^T A \underset{\sim}{x} \Rightarrow$ quadratic form

$\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\, x_i x_j$

$= \mathrm{tr}[A\ E(XX^T)]$

$D(\underset{\sim}{x}) = \Sigma = E(XX^T) - \underset{\sim}{\mu}\underset{\sim}{\mu}^T$

$= \mathrm{tr}[A(\Sigma + \underset{\sim}{\mu}\underset{\sim}{\mu}^T)]$

$= \mathrm{tr}(A\Sigma) + \mathrm{tr}(A\underset{\sim}{\mu}\underset{\sim}{\mu}^T)$ as done earlier in $\mathrm{tr}(AXX^T)$

$= \mathrm{tr}(A\Sigma) + \mathrm{tr}(\underset{\sim}{\mu}^T A \underset{\sim}{\mu})$

$= \mathrm{tr}(A\Sigma) + (\underset{\sim}{\mu}^T A \underset{\sim}{\mu})$ — ①

---

when $X_1, X_2 - - X_n$ iid $N(0,1)$

$X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \sim N(\underset{\sim}{0}, I_n)$

$E(\underset{\substack{\downarrow \\ \text{summation}}}{\sum} X_i^2) = E(\underset{\sim}{X}^T I_n \underset{\sim}{X})$

$= \underset{\substack{\downarrow \\ n}}{\mathrm{tr}(I_n I_n)} + \underset{\sim}{0}^T I_n \underset{\sim}{0}$ from eq ①

$= n + 0$

$= n$

$\sum E(x_i^2) = n$
$i=n$

$x_i \overset{iid}{\sim} N(u, 1) \Rightarrow \Sigma \; I_n$

$A = I_n$

$\underset{\sim}{u} = u \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$

$E(\Sigma x_i^2)$

$= tr(A\Sigma) + \underset{\sim}{u}^T A \underset{\sim}{u}$

$= tr(I_n I_n) + u(\underset{\sim}{1}^T) I_n u(\underset{\sim}{1})$

$= n + u^2 n = (1+u^2)n$

If $x_i \sim N(u_i, 1)$ & independent

then $E(\Sigma x_i^2) = n + \underset{\sim}{u}^T I_n \underset{\sim}{u}$

$= n + \sum_{i=1}^{n} u_i^2$

If $x_i \overset{iid}{\sim} N(0, 1)$ then $\sum_{i=1}^{n} x_i^2 \sim \chi_n^2$

$x_i \sim N(u_i, 1)$ then

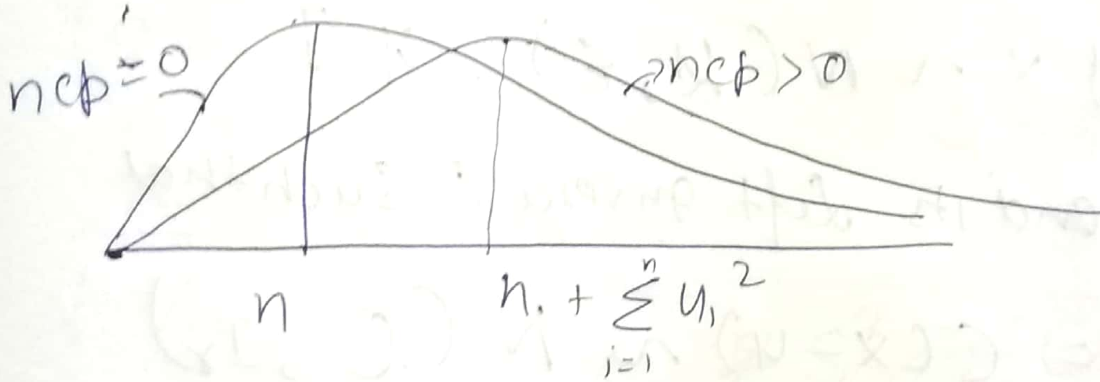
and $x_i \sim N(u_i, 1)$ then $\sum_{i=1}^{n} x_i^2 \sim \chi^2$ df $= n$, ncp $= \sum_{i=1}^{n} u_i = \underset{\sim}{u}^T \underset{\sim}{u}$

non centrality parameter $\Rightarrow$ ncp

Scanned by CamScanner

$$n \leq n + \underset{\sim}{u}^T u$$

$$E\left(\chi^2_{\substack{df = n \\ ncp = 0}}\right) \leq E\left(\chi^2_{\substack{df = n \\ ncp = \underset{\sim}{u}^T \underset{\sim}{u}}}\right)$$



(★) If $\underset{\sim}{x} \sim N(\underset{\sim}{u}, I_n)$ then

$$\underset{\sim}{x}^T A \underset{\sim}{x} \sim \chi^2_{\substack{df = \text{rank}(A) \\ ncp = \underset{\sim}{U}^T A \underset{\sim}{U}}}$$

iff A is an idempotent matrix

— Eq (2).

Result

① If $X \sim N(\underset{\sim}{\mu}, \Sigma)$

$$A\underset{\sim}{x} \sim N(A\mu, A \Sigma A^T)$$

$Var(AX) = A \Sigma A^T$
$E(AX) = A\underset{\sim}{\mu}$

② If $X \sim N(\underset{\sim}{\mu}, \Sigma)$, if $\exists$

B and its left inverse C such that

$$Y = C(X-\mu) \sim N(\underset{\sim}{0}, I_r)$$

$X = \underset{\sim}{\mu} + BY$ with probability 1

③ If A_1, A_2 symmetric & idempotent matrix such that $Q = A_1 - A_2$ is p.s.d then

$\underset{\sim}{x}^T Q \underset{\sim}{x}$ & $\underset{\sim}{x}^T A_2 \underset{\sim}{x}$ are independent

$A_1 = Q + A_2$, $|Q| \geq 0$

for normally distributed $\underset{\sim}{x}$

④ If $A = A^T$ (symmetric) and $CA = 0$ matrix

then $X^T A X$ & CX are independent

VV imp result

$X \sim N(0,1)$
$Y \sim \chi_n^2$ > independent

$$\frac{X}{\sqrt{Y/n}} \sim t_n$$

$\bar{x}$ & $S^2 = \sum (x_i - \bar{x})^2$

are happen to be independent

26/7/19

$x_1, x_2 \cdots\cdots x_n$ iid $N(\mu, \sigma^2)$

Naive prediction $= \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$

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

why

(don't worry about this just a scaling factor?)

we are interested for the independence of $\bar{x}$ & $S^2 = \sum_{i=1} (x - \bar{x})^2$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \left(\frac{1}{n} \cdots \frac{1}{n}\right)\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

$$\underset{\sim}{l} = \frac{1}{n}\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \quad \underset{\sim}{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \Rightarrow \bar{x} = \underset{\sim}{l}^T \underset{\sim}{x}$$

$$\underset{\sim}{l} = \frac{1}{n}\underset{\sim}{1} \qquad \underset{\sim}{l}^T \equiv \underset{\sim}{C}$$

$$S^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 \Leftrightarrow \underset{\sim}{x}^T A \underset{\sim}{x} \quad ??$$

$$P = \left(I_n - \frac{1}{n}\underset{\sim}{1}\,\underset{\sim}{1}^T\right)$$

$$x^T P x = \underset{\sim}{x}^T\left(I_n - \frac{1}{n}\underset{\sim}{1}\,\underset{\sim}{1}^T\right)x$$

$$= \underset{\sim}{x}^T I_n \underset{\sim}{x} - \frac{1}{n}(\underset{\sim}{x}^T \underset{\sim}{1})(\underset{\sim}{1}^T \underset{\sim}{x})$$

$$= \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum x_i\right) \qquad \boxed{n\bar{x} = \sum_{i=1}^{n} x_i}$$

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

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

① $\ell^T P = 0^T$ ② $P^T = P$ ③ $P^2 = P \rightarrow$ as $\underset{\sim}{X} \sim N(\underset{\sim}{\mu}, I_n)$

$\Downarrow$

$X^T A \underset{\sim}{X} \sim \chi^2_{df=rank,\ ncp}$

$\downarrow$

if A is idempotent

① $= \frac{1}{n} \underset{\sim}{1}^T \left(I_n - \frac{1}{n} \underset{\sim}{1} 1^T \right)$

$= \frac{1}{n} \underset{\sim}{1}^T \left(1^T - 1^T \right) = 0^T$

② $P^T = \left(I_n - \frac{1}{n} \underset{\sim}{1} \underset{\sim}{1}^T \right)^T$

$= \left(I_n - \frac{1}{n} (\underset{\sim}{1}^T)^T (\underset{\sim}{1})^T \right) = P$

③ $P^2 = P$

$\underset{\sim}{X} \sim N(\mu, \Sigma)$

$P\underset{\sim}{X} \sim N(P\mu, P\Sigma P^T)$ — from Result ① in eq①

$\Downarrow$

from ① ② ③

① As $\ell^T P = 0^T$ & $P^T = P$

$\underset{\sim}{\ell}^T \underset{\sim}{X} = \bar{X}$ & $X^T P X = S^2$

are independently distributed

② $\bar{X} = \underset{\sim}{l}^T \underset{\sim}{x} \sim N(\underset{\sim}{l}^T \underset{\sim}{u}, \sigma^2 l^T I_n l)$

$(A \Sigma A^T \quad A = l^T$

$\boxed{\begin{array}{l} \underset{\sim}{u} = u \underset{\sim}{1} \\ \underset{\sim}{l} = \frac{1}{n} \underset{\sim}{1} \end{array}}$

$\equiv N\left(u, \frac{\sigma^2}{n}\right)$

$\left(\frac{1}{n} \underset{\sim}{1}^T \frac{1}{n} \underset{\sim}{1}\right)$

$= \left(\frac{1}{n}\right)^2 \cancel{1^T 1}$

$= \frac{1}{n}$

③ $S^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2$

$= \underset{\sim}{x}^T P \underset{\sim}{x}$

$\Leftrightarrow \frac{S^2}{\sigma^2} = \frac{\underset{\sim}{x}^T}{\sigma} P \frac{\underset{\sim}{x}}{\sigma}$

$= \underset{\sim}{y}^T P y$

$\frac{S^2}{\sigma^2} \sim \chi^2_{df = rank(P)}$
$\quad ncp = (u^*)^2 \underset{\sim}{1}^T P \underset{\sim}{1}$

How

$\underset{\sim}{x} \sim N(u \underset{\sim}{1}, \sigma^2 I_n)$

$x_i \sim N(u, \sigma^2)$

to get Result ~~of~~
acc to eq 2
we need scale down
as varia is I_n in
eq ③

$y_i = \frac{x_i}{\sigma}$

$\underset{\sim}{y} \sim \left(\left(\frac{u}{\sigma}\right) \underset{\sim}{1}, I_n\right)$

$\therefore \sim N(u^* \underset{\sim}{1}, I_n)$

$\text{rank}(P) = n-1 \rightarrow$ How

$(U^*)^2 \, \underline{1}^T P \underline{1}$

$= \underline{1}^T \left(I_n - \frac{1}{n} \underline{1}\, \underline{1}^T \right) \underline{1}$

$= \underline{1}^T \underline{1} - \frac{1}{n} (\underline{1}^T \underline{1}) \cdot (\underline{1}^T \underline{1})$

$= n - \frac{1}{n} n \, n = 0$

$\frac{S^2}{\sigma^2} \sim \chi^2_{df = n-1,\ ncp = 0} \Rightarrow S^2 \sim \sigma^2 \chi^2_{n-1}$

(4) $t = \frac{\sqrt{n}(\overline{X} - u)}{\sqrt{S^2/n-1}} = \frac{\left(\frac{\overline{X} - u}{\sigma/\sqrt{n}}\right)}{\sqrt{\frac{S^2}{\sigma^2 (n-1)}}}$

$\overline{X} \sim N\left(u, \frac{\sigma^2}{n}\right) \Rightarrow \left(\frac{\overline{X} - u}{\sigma/\sqrt{n}}\right) \sim N(0,1)$

$\frac{S^2}{\sigma^2} \sim \chi^2_{n-1,\, ncp=0}$ and they are independent

~~indepent~~ $t \sim t_{n-1,\, ncp=0}$

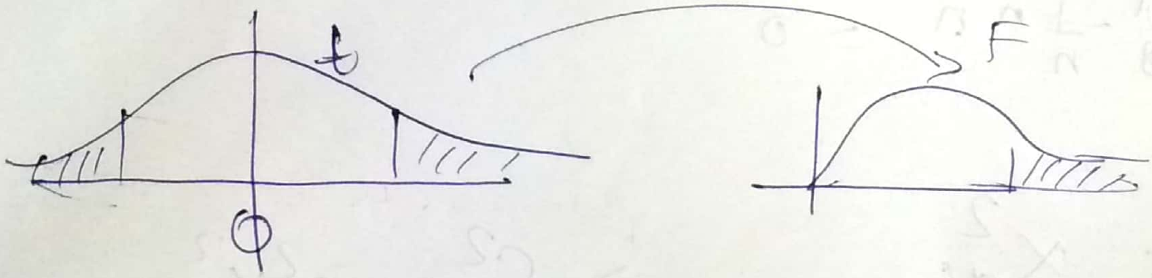
---

t is used to do testing about u when σ is unknown

$X \sim \chi^2_n$
$Y \sim \chi^2_m$ > independent

$$t^2 = \left(\frac{N(0,1)}{\chi^2_{n-1}/n-1} \right)^2 = \frac{\chi^2_1/1}{\chi^2_{n-1}/n-1}$$

$= F_{1, n-1}$



Both tail test

Th

Let $\underset{\sim}{X} \sim N(\underset{\sim}{\mu}, I_n)$ & $\underset{\sim}{X}^T A \underset{\sim}{X} = \sum_{i=1}^{k} \underset{\sim}{X}^T A_i \underset{\sim}{X}$

when $A_i^T = A_i$ & $A^2 = A$ then

$\underset{\sim}{X}^T A_i \underset{\sim}{X} \sim \chi^2_{df = \text{rank}(A_i)}$
$ncp = \mu^T A_i \mu$

& Each A_i are independent

unknown → known

$$y_i = \beta_0 + \beta_i x_i + \epsilon_i$$

y_i → Random

β_0, β_i → non random (but unknown)

ϵ_i → Random

ϵ_i iid $N(0, \sigma^2)$

Gauss markov model : unknown parameters $\beta_0, \beta_1, \sigma^2$

we can get a neumerical value from data an estimate of $\hat{\beta}_0, \hat{\beta}, \sigma^2$, but to do testing we need their distribution

Eg Hypothesis : $H_0 : \beta_0 = 0$
$H_1 = \beta_0 \neq 0$

Simple Leniar regression

$$y_i = \beta_0 + \beta_i x_i + \epsilon_i \quad \forall = i = 1 \text{ to } n$$

$D = \{(y_i, x_i), i = 1 \dots n\}$ known data set

unknown : B_0, B_1, σ^2 where
$\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

If ϵ_i follows $N(0, \sigma^2)$ then it is known as Gauss markov model

$$\underset{\sim}{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \quad \underset{\sim}{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}, \quad \underset{\sim}{B} = \begin{pmatrix} B_0 \\ B_1 \end{pmatrix} \quad X = \begin{bmatrix} 1 & x_1 \\ \vdots & \\ 1 & x_n \end{bmatrix}$$

$$\underset{\sim}{Y} = X\underset{\sim}{B} + \epsilon$$

$$\epsilon \sim N(\underset{\sim}{0}, \sigma^2 I_n)$$

$$\boxed{\underset{\sim}{Y} \sim N(X\underset{\sim}{B}, \sigma^2 I_n)}$$

☆ Y_i is independent but not identically distributed

Estimation of B_0, B_1, σ^2

① Least square method

② maximum likelihood method

Estimated Value of B_0, B_1 same in Both case but not σ^2

Least square method

LS condition is to ~~maximum~~ minimize

$$S(B_0, B_1) = \sum_{i=1}^{n} (y_i - B_0 - B_1 x_i)^2$$

$$= \| \underset{\sim}{y} - X \underset{\sim}{B} \|_2^2$$

$$= (\underset{\sim}{y} - X\underset{\sim}{B})^T (\underset{\sim}{y} - X\underset{\sim}{B})$$

$$\begin{pmatrix} \hat{B_0} \\ \hat{B_1} \end{pmatrix} = \underset{(B_0, B_1)}{\arg\min} \; S(B_0, B_1)$$

$$\frac{\partial \underset{\sim}{b}^T A}{\partial \underset{\sim}{b}} = A^T$$

$$\frac{\partial \underset{\sim}{b}^T A \underset{\sim}{b}}{\partial \underset{\sim}{b}} = b^T (A + A^T)$$

$$= (\underset{\sim}{y}^T - B^T \underset{\sim}{X}^T)(y - XB)$$

$$= y^T y - y^T X B - B^T X^T y - B^T \underset{\sim}{X}^T X B$$

$$\frac{\partial S}{\partial B} = 0 \Rightarrow X^T \underset{\sim}{y} = X^T X \hat{B}$$ | Normal Equation

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

(*) $(X^T X)$ will not be invertible if all x_i values are same or (2 colum of X are linear combination of each other)

Since one column is $\begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix}$ the other must be

Scalar multiple of other $\begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix}$ $\Rightarrow$ all x_i are

same

~~We~~ either solve ① ~~②~~ ② or ③ we get

$$\hat{B_0} = \bar{y} - \hat{B_1}\bar{x}$$

$$\hat{B_1} = \frac{\sum_{i=1}^{n} \cancel{b}y_i x_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} = \frac{S_{xy}}{S_{xx}}$$

<u>Ref Bhe</u> $S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$

$$= \sum_{i=n}^{n}(x_i - \bar{x})y_i - \bar{y}\underbrace{\sum_{i=1}^{n}(x_i - \bar{x})}_{0}$$

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

Prediction

$\hat{y}_i = \hat{B}_0 + \hat{B}_1 x_i \quad \forall i = 1 - - - n$

$\hat{y} = X\hat{B}$

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

$\hat{y} = P_X Y$

$\hat{y}$ is the orthogonal of $\underset{\sim}{y}$ on the column space of X ie $C(\underset{\sim}{X})$

The error vector in this LS prediction

is $\underset{\sim}{e} = \underset{\sim}{y} - P_X Y$

$= (I_n - P_X)\underset{\sim}{y}$

$\Rightarrow$ 1 $\underset{\sim}{e}$ is orthogonal to $\hat{y}$

2 Although ϵ cov matrix $I_n\sigma^2$

e doesnot have independent component

$y \sim N(X\underset{\sim}{\beta}, \sigma^2 I_n)$

$\underset{\sim}{e} = (I - P_x)\underset{\sim}{y} \sim N(\underset{\sim}{0}, \sigma^2(I_n - P_x))$

$\hat{y} = P_x y \sim N(P_x X\underset{\sim}{\beta}, \sigma^2 P_x^T P_x)$

$\sim N(X\underset{\sim}{\beta}, \sigma^2 P_x)$

$= P_x^T P_x = P_x$

as $= X(X^TX)^- X^T X(X^TX)^- X^T$

A G A

$= X(X^TX)^- X^T = P_x$

$\underset{\sim}{e} = (I - P_x)\underset{\sim}{y} \qquad \underset{\sim}{y} \sim N(X\beta, \sigma^2 I_n)$

$e \sim N((I - P_x)X\underset{\sim}{\beta}, \sigma^2 (I - P_x)^T I_n (I - P_x))$

$= N((X - X)\beta$

$P_x x = x$

SLR

$S = (\underset{\sim}{y} - X\underset{\sim}{B})^T (\underset{\sim}{y} - X\underset{\sim}{B})$

$= \sum_{i=1}^{n} (y_i - B_0 - B_1 x_i)^2$

$\hat{\underset{\sim}{B}} = \text{Arg min } S(\underset{\sim}{B})$

$\boxed{\hat{B} = (X^T X^-) X^T Y}$

Normal equation
$X^T Y = X^T X B$

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

$\hat{y} \sim N(X\underset{\sim}{B}, \sigma^2 P_X)$

$\underset{\sim}{e} = y - \hat{y} = (I_n - P_X)\underset{\sim}{Y}$

$\underset{\sim}{e} \sim N(\underset{\sim}{0}, (I_n - P_X)\sigma^2)$

---

* $(X^T X)$ is invertible

$\hat{B} \sim N((X^T X)^{-1} X^T X \underset{\sim}{B},$
$(X^T X)^{-1} X^T I_n X (X^T X)^{-1} \sigma^2)$

• $Y \sim N(X\underset{\sim}{B}, \sigma^2 I_n)$

$Z \sim N(\mu, \Sigma)$
$AZ \sim N(A\mu, A\Sigma A^T)$

$\hat{B} \sim N(\underset{\sim}{B}, (X^T X)^{-1} \sigma^2)$

~~Normal form~~ Normal equation we know

$$\hat{B}_1 = \frac{S_{xy}}{S_{xx}}$$

$$\hat{B}_0 = \bar{y} - \hat{B}_1 \bar{x}$$ 4N

$$\hat{B}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{S_{xx}}$$

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

Let $\left(\frac{x_i - \bar{x}}{S_{xx}}\right) = \alpha_i$

when $\alpha_i = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$

$$= \sum_{i=1}^{n} y_i \alpha_i$$

$\hat{B}_1 = \underset{\sim}{\alpha}^T y \longrightarrow$ Linear estimator so B_1 has linear estimator B_1

as we $\underset{\sim}{y} \sim N(X\underset{\sim}{B}, \sigma^2 I_n)$

$$\hat{B}_1 \sim N(\alpha^T X \underset{\sim}{B}, \sigma^2 \alpha^T I_n \underset{\sim}{\alpha})$$

$\boxed{\cancel{\sum_{i=1}^{n} \alpha_i = 0}}$ - ~~as $\sum x_i - n\bar{x} = 0$ / S_{xx}~~

$$\Rightarrow \sum_{i=1} \alpha_i = 0$$

$$\Rightarrow \underset{\sim}{1}^T \underset{\sim}{\alpha} = \underset{\sim}{\alpha}^T \underset{\sim}{1} = 0$$

$$\underset{\sim}{\alpha}^T \underset{\sim}{x}$$

$$= \sum_{i=1}^{n} \alpha_i x_i$$

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

$$\hat{B}_1 \sim N\left(\alpha^T X \underset{\sim}{B},\ \sigma^2 \alpha^T I_n \underset{\sim}{\alpha}\right)$$

$$\sim N\left(\left(\underset{\sim}{\alpha}^T \underset{\sim}{1} \beta_0, \alpha^T \underset{\sim}{x}\right), \sigma^2 \underset{\sim}{\alpha}^T \underset{\sim}{\alpha}\right)$$

$$\sim N\left(B_1,\ \sigma^2 \frac{S_{xx}}{S_{xx}^2}\right) \qquad \alpha^T \alpha = \sum \frac{(x_i - \bar{x})^2}{S_{xx}^2}$$

$\Rightarrow \hat{B}_1$, is an unbiased linear estimator B_1

$$B_0 = \bar{y} - \hat{B}_1 \bar{x} = \underset{\sim}{C}^T \underset{\sim}{y} = \sum_{i=1}^{n} c_i y_i$$

$\Rightarrow \hat{B}_0$ is an unbiased, linear estimator of $\hat{B}_0$

(HW)

hw

(A) $\hat{B}_0$ & $\hat{B}_1$ may be dependent
Find the $cov(\hat{B}_0, \hat{B}_1)$ hw
which the ~~...~~ (1,2) or (2,1) elements of
$\sigma^2 (X^T X)^{-1}$

Error $\underset{\sim}{e} = (y - \hat{y}) \sim N(0, (I_n - P_x)\sigma^2)$

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

$e = (y - \hat{y}) = (I - P_x)\underset{\sim}{y}$

Proper

(1) $P_x^T = P_x$ (Just proved earlier)

(2) $P_x^2 = P_x$

Sum and error terms

P_x has rank 2

$I - P_x$ has rank $(n-2)$

$$\sum_{i=1}^{n} e^2 = \underset{\sim}{e}^T \underset{\sim}{e} = [(I_n - P_x)\underset{\sim}{y}]^T [(I - P_x)\underset{\sim}{y}] = y^T (I - P_x)\underset{\sim}{y}$$

$$\frac{ESS}{\sigma^2} = \frac{\underset{\sim}{y}^T (I_n - P_x) \underset{\sim}{y}}{\sigma^2}$$

$$\sim \chi^2 \quad df = \text{rank}\left((I_n - P_x) I_n \right)$$

$$ncp = \left(\frac{XB}{\sigma}\right)^T (I_n - P_x) \left(\frac{XB}{\sigma}\right)$$

If A is an idempotent matrix then

$$Z \sim N(\mu, I_n)$$

$$\underset{\sim}{Z}^T A \underset{\sim}{Z} \sim \chi^2_{r(A\Sigma)},\ \mu^T A \mu$$

$$\underset{df}{\Downarrow}$$

$$df = n - 2$$

$$ncp = \left(\underset{\sim}{B}^T \underset{\sim}{X}^T (I_n - P_x) \underset{\sim}{X} \underset{\sim}{B} \right) \frac{1}{\sigma^2}$$

$$= \underset{\sim}{B}^T \left(X^T X - X^T X (X^T X)^{-1} X^T X \right) B \left(\frac{1}{\sigma^2} \right)$$

$$ncp = \underset{\sim}{B}^T\, 0\, \underset{\sim}{B} \frac{1}{\sigma^2} = 0$$

$$\frac{ESS}{\sigma^2} \sim \chi^2_{n-2,\,0}$$

$$ESS \sim \sigma^2 \chi^2_{n-2},\ ncp=0$$

$$\hat{B}_1 \sim N\left(B_1, \frac{\sigma^2}{S_{xx}}\right)$$

$$\hat{B}_0 \sim N\left(B_0, \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)\sigma^2\right) \text{ HW}$$

$$ESS \sim \sigma^2 \chi^2_{n-2}\ ncp=0$$

$$\hat{\sigma}^2 = \frac{ESS}{n-2} = \frac{\sum (y_i - \hat{y}_i)^2}{n-2}$$

A $\hat{\sigma}^2$ is an unbiased estimator of σ^2

$$E(\hat{\sigma}^2) = \frac{E[ESS]}{n-2} = \sigma^2 \frac{(n-2)}{(n-2)} = \sigma^2$$

Q (1) $\hat{\underset{\sim}{e}}$ and $\hat{y}$ are orthogonal to each to each other and distribution they are indident

$$\hat{y} \sim N(XB, \sigma^2 P_X)$$
$$\hat{\underset{\sim}{e}} \sim N(0, \sigma^2(I_n - P_X))$$

If the covariance b/w 2 Normal distribution is zero they will be independent (completely - Not only linear)

$cov(\hat{e}, \hat{y})$,

use
$cov(AX, BX) = A\Sigma B^T$

$\hat{y} = P_x y$
$\hat{e} = (I - P_x) y$

$y \sim N(X\beta, \sigma^2 I_n)$

$cov(Ay, By)$
$= (A I B^T)\sigma^2$

$\Rightarrow cov(\hat{y}, \hat{e}) = P_x I_n (I_n - P_x)^T \sigma^2$
$= P_x I_n (I_n - P_x)\sigma^2$
$= (P_x - P_x^2)\sigma^2$
$= (P_x - P_x)\sigma^2 = 0$ matrix

independent

$$(P_X y)^T (I - P_X y) = y^T P_X (I - P_X) y$$

$$= 0$$

$\Rightarrow$ orthogonal

---

we have Data $\{(x_i, y_i) \mid i = 1 \ldots n\}$

From these we get $\hat{\beta}_0, \hat{\beta}_1$ & LS estimate

So the prediction line or regression line

· Now for a new value x_0 under the same model, we can predict the value

value of y_0 as follows

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

Although $y_0 \sim N(\beta_0 + \beta_1 x_0, \sigma^2)$

But we only know x_0, β_0, β_1 are only estimated

$\hat{y}_0 = \hat{B}_0 + \hat{B}_1 x_0$ as $\hat{B}_0 = \bar{y} - \hat{B}_1 \bar{x}$

$= \bar{y} \cancel{+} \hat{B}_1 \bar{x} + \hat{B}_1 x_0$

$\boxed{\begin{aligned} &\hat{B}_1 = \alpha^T y \\ &\bar{y} = \tfrac{1}{n} 1^T y \end{aligned}}$

$E(\hat{y}_0) = (B_0 + B_1 \bar{x}) + \hat{B}_1 (x_0 - \bar{x})$

$= B_0 + B_1 x_0$

$\downarrow$

So it is an unbiased estimator of Ey_0

$var(\hat{y}_0) = var(\bar{y}) + (x - \bar{x}_0)^2 \, var\, B_1$

$+ 2\, cov(\bar{y}, \hat{B}_1)(x - x_0)$

$\Downarrow$

$= \frac{\sigma^2}{n} + (x_0 - \bar{x}) \frac{\sigma^2}{S_{xx}} + 0$

$\hat{y}_0 \sim N\left(B_0 + B_1 x_0, \sigma^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)\right)$

$E(\bar{y}) = E\left(\frac{1}{n} \Sigma y_i\right) = \frac{1}{n} \Sigma (B_0 + B_1 x_i)$

Hypothesis problem

$H_0: B_0 = b_0$

$H_1: B_0 \neq b_0$

$$\hat{B}_0 \sim N\left(B_0, \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)\right)$$

Testing mean

$$\frac{\hat{B}_0 - B_0}{\sqrt{\sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}} \sim N(0,1)$$

Under H_0

$$\frac{\hat{B}_0 - b_0}{\sigma\sqrt{\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}} \sim N(0,1)$$

but when who put the unbiased estimated of σ we get a t estimator

Look it is really beautiful

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2$$ → you are estimating & then predicting the error

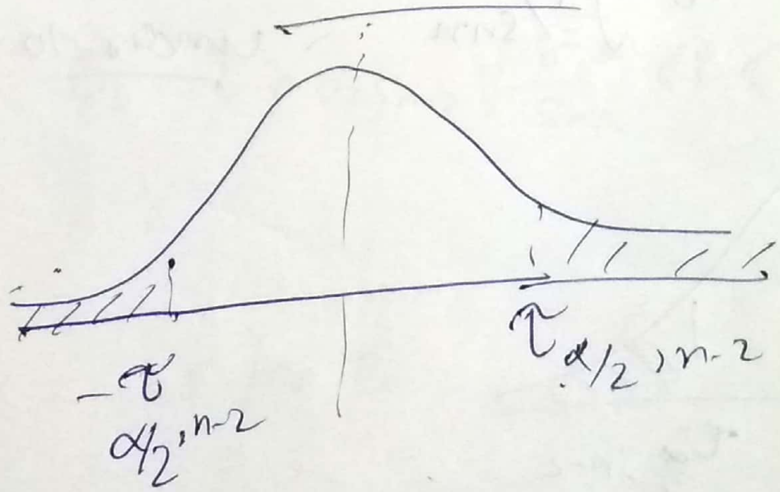
$$= S_{yy} - \frac{S_{xy}^2}{S_{xx}}$$ → here you are getting error without even estimating line

Estimated average error

$$T = \frac{\hat{B}_0 - b_0}{\hat{\sigma}\sqrt{\frac{1}{n} + \frac{\bar{x}}{S_{xx}}}} \sim t_{n-2}$$

under H_0

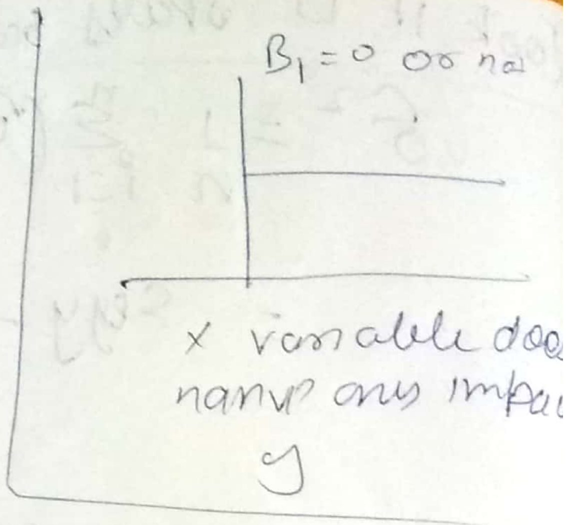
$H_1 = B_0 \neq b_0$



$|T_{obs}| > \tau_{\alpha/2, n-2}$, we reject the H_0 in favor of H_1 at level α

95% CI for B_0 ??

Testing



$H_0 : \beta_1 = b_1$

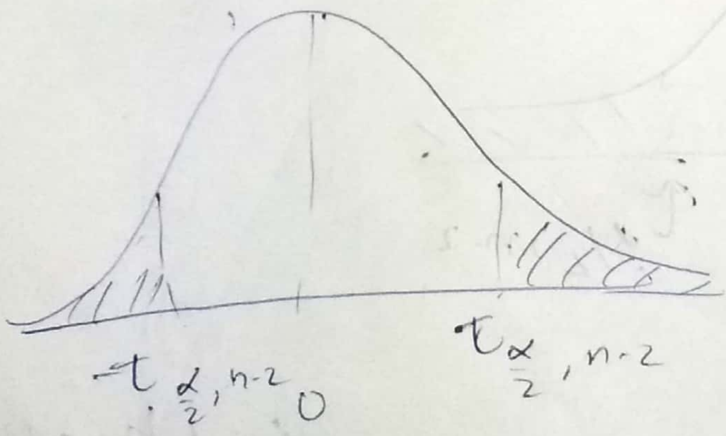
$H_1 : \beta_1 = b_1$

$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\sigma^2 / S_{nn}}} \sim N(0,1)$$

Under H_0, $$\frac{\hat{\beta}_1 - b_1}{\sqrt{\sigma^2 / S_{nn}}} \sim N(0,1)$$

But σ^2 is unknown and $\hat{\sigma}^2 = \frac{ESS}{n-2}$

Hence $$T = \frac{\hat{\beta}_1 - b_1}{\hat{\sigma}\sqrt{1/S_{nn}}} \sim t_{n-2}$$ under H_0



If $|T_{ob}| > T_{\alpha/2, n-2}$ then we reject H_0 in favour of H_1

95% CI for B_1

$$T = \frac{\hat{B}_1 - B_1}{\sqrt{\hat{\sigma}^2 / S_{xx}}} \sim t_{n-2}$$

$$P\left(-t_{0.025, n-2} < T < t_{0.025, n-2}\right) = 0.95$$

$$P\left(-t_{0.025, n-2} < \frac{\hat{B}_1 - B_1}{\sqrt{\hat{\sigma}^2 / S_{xx}}} < t_{0.025, n-2}\right) = 0.95$$

$$P\left(\hat{B}_1 - t_{0.025, n-2}\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} < B_1 < \hat{B}_1 + t_{0.025}\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}\right) = 0.95$$

95% CI for B_0

$$P\left(\hat{B}_0 - t_{0.025, n-2}\sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)} < B_0 < \hat{B}_0 + t_{0.025, n-2}\sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}\right)$$

$$= 0.95$$

Scanned by CamScanner

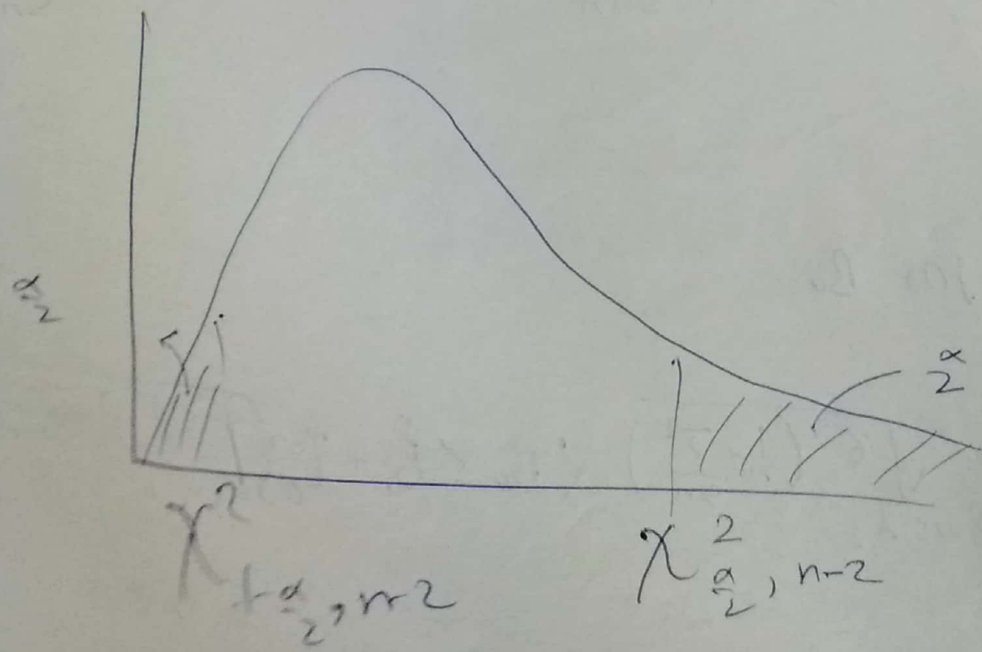
95% CI for σ^2

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2} = \frac{ESS}{n-2}, \quad \frac{ESS}{\sigma^2} \sim \chi^2_{n-2}$$

$$P\left(\chi^2_{1-\frac{\alpha}{2}, n-2} < \frac{ESS}{\sigma^2} < \chi^2_{\frac{\alpha}{2}, n-2}\right) = 1-\alpha$$

$\alpha = 0.05$

$$P\left(\frac{ESS}{\chi^2_{\frac{\alpha}{2}, n-2}} < \sigma^2 < \frac{ESS}{\chi^2_{1-\alpha/2, n-2}}\right) = 1-\alpha$$



Prediction Interval of y_0 for

some no under same model from where the data are coming

why do we calcultate it name it somehow else

y_0 is not a parameter, but it is a random variable

$$\text{True } y_0 \sim N(\beta_0 + \beta_1 x_0, \sigma^2)$$

$$\hat{y}_0 \sim N\left(\beta_0 + \beta_1 x_0, \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)\sigma^2\right)$$

So also y_0 & $\hat{y}_0$ are independent

$$T = \frac{y_0 - \hat{y}_0}{\sqrt{\sigma^2_{y_0}}} \sim N(0,1)$$

$$\Rightarrow \frac{y_0 - \hat{y}_0}{\sqrt{\hat{\sigma}^2_{y_0}}} \sim t_{n-2} \qquad \sigma^2 \text{ is replaced by } \hat{\sigma}^2_{y_0}$$

$$P\left(-t_{\alpha/2, n-2} < \frac{y_0 - \hat{y}_0}{\sqrt{\hat{\sigma}^2_{y_0}}} < t_{\alpha/2, n-2}\right) = 1-\alpha$$

PI

Prediction interval

$$P\left(\hat{y}_0 - T_{\alpha/2, n-2}\sqrt{\hat{\sigma}^2_{y_0}} < y_0 < \hat{y}_0 + T_{\alpha/2, n-2}\sqrt{}\right.$$

$$= 1-\alpha$$

Least square estimate of in SLR

Maximum Likelihood estimate

$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

$\Rightarrow y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$ are independently distributed

$$(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2) = \underset{\beta_0\, \beta_1\, \sigma^2}{\arg\max} \prod_{i=1}^{n} \frac{e^{-\frac{1}{2}\left(\frac{y_i - \beta_0 - \beta_1 x_i}{\sigma}\right)^2}}{\sqrt{2\pi}\,\sigma}$$

as independent

$$= \underset{\beta_0\, \beta_1\, \sigma^2}{\arg\max} \frac{e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2}}{(\sqrt{2\pi}\,\sigma)^n}$$

now taking log & diff to max gives minim of $\sum(y_i - \beta_0 - \beta$

$\Rightarrow$ $\hat{\beta}_{0_{LS}} = \hat{\beta}_{0_{MLE}}$

& $\hat{\beta}_{1_{LS}} = \hat{\beta}_{1_{MLE}}$

$$\boxed{\begin{array}{ll} \hat{\sigma}^2_{LS} = \dfrac{ESS}{n-2} & \text{Unbiased} \\ \sigma^2_{MLE} = \dfrac{ESS}{n} & \text{Biased} \end{array}}$$

Now earlier x_i was non stochastic & y_i was stochastic as e_i was stochastic

Now (X, Y) is a paired random variable with bivariate normal distribution

$$N(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$$

Given data $D_i = \{(x_i, y_i), \; i = 1 \dots n\}$

The Regression model

$E(Y|X=x)$ or $E(X|Y=y)$

$$f(x,y) = \frac{e^{-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)\right\}}}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}$$

$$= \left[\left(\frac{y-\mu_y}{\sigma_y}\right) - \rho\left(\frac{x-\mu_x}{\sigma_x}\right)\right]^2 + (1-\rho^2)\left(\frac{x-\mu_x}{\sigma_x}\right)^2$$

Scanned by CamScanner

$$f(x,y) = \frac{e^{-\frac{1}{2}\frac{1}{(1-\rho^2)}\left[\left(\frac{y-\mu_y}{\sigma_y}\right) - \rho\left(\frac{x-\mu_x}{\sigma_x}\right)\right]^2}}{\sqrt{2\pi}\sqrt{1-\rho^2}\,\sigma_y} \cdot \frac{e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2}}{2\pi\sigma_x}$$

$$E(Y/x) = \mu_y + \rho\frac{\sigma_y}{\sigma_x}(x-\mu_x), \quad V(Y/x) = \text{varian}\left(\sqrt{1-\rho^2}\,\sigma_y\right)^2 = (1-\rho^2)\sigma_y^2$$

$$E(X/Y) = \mu_x + \rho\frac{\sigma_x}{\sigma_y}(Y-\mu_y), \quad var(X/Y) = (1-\rho^2)\sigma_x^2$$

$$var(X/Y) = (1-\rho^2)\sigma_x^2 \le \sigma_x^2$$
$$var(Y/X) = (1-\rho)\sigma_y^2 \le \sigma_y^2$$

* conditional variances less or $\underset{\rho=0}{\text{equal}}$ as compared to uncondition variance

$E(Y|X=x) = 0$

$B_0 = \mu_y - \rho \frac{\sigma_y}{\sigma_x} \mu_x$

2 So to check regression

$H_0: B_1 = 0 \iff \text{'} \rho = 0 \text{'}$ as σ_x, σ_y are $\neq 0$

$H_1: B_1 \neq 0 \iff \text{'} \rho \neq 0 \text{'}$

$H_0: \rho = 0$ vs $\rho \neq 0$ in Bivariate ~~normama~~ normal

$$\hat{\rho} = r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Under H_0: $\rho = 0$

$$T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{n-2, \, ncp=0}$$

But life is much more complicated when

$H_0: \rho = \rho_0$, vs $H_1: \rho \neq \rho_0$

H_0 Define: $Z = \frac{1}{2} \log\left(\frac{1+r}{1-r}\right) = \tanh^{-1}(r)$

$$u_z = \frac{1}{2} \log\left(\frac{1+\rho_0}{1-\rho_0}\right) = \tanh^{-1}(\rho_0)$$

$$\sigma_z^2 = (n-3)^{-1}$$

$$\frac{Z - u_z}{\sigma_z} \sim N(0,1) \text{ when } n \uparrow \infty$$

Variance stabilization formula

$$T_n \xrightarrow{p} \theta \qquad \text{as } n \uparrow \infty, \text{ p} \to 1$$

$$\sqrt{n}(T_n - u) \xrightarrow{d} N(0, \sigma^2)$$

then $\sqrt{n}(g(T_n) - g(u)) \xrightarrow{d} N(0, \sigma^2 (g'(u))^2)$

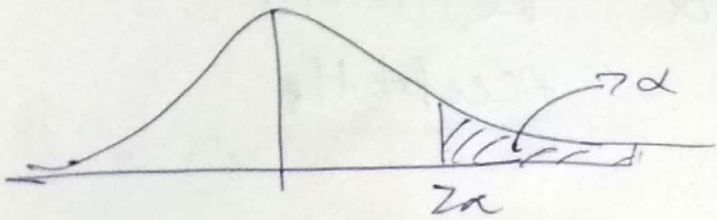
P value

H: $u = u_0$ vs $u > u_0$

test at level $\alpha = 0.1$, or $0.05, 0.01$

$Z = \left(\frac{\bar{X}-\mu_0}{\sigma}\right)\sqrt{n}$ σ known

$Z \sim N(0,1)$ under H_0



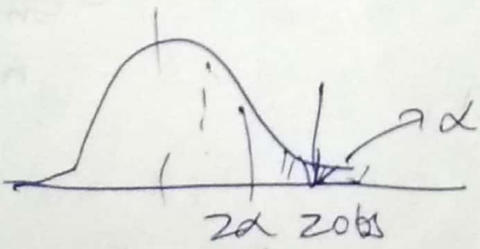
(7) we reject H_0 in favour of H_1 if

$Z_{obs} > Z_\alpha$

| we need to know Z_α value according to α

if α changes Z_α so conclusion of test may change

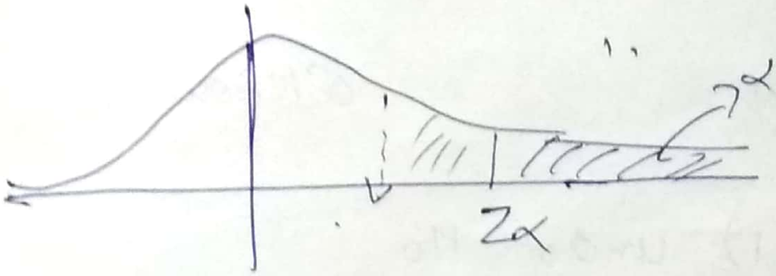
Suppose



Calulate $P_{H_0}(Z > Z_0)$ because it an upper tail test

check $P_{H_0}(Z > Z_{obs}) < \alpha$ to Reject H_0

pvalue $< \alpha$ to Reject.



$p < \alpha$ Reject H_0

$p > \alpha$ Accept H_0

Multiple Linear Regression

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}_{n\times 1}, \quad X = \begin{pmatrix} 1 & x_{11} & x_{21} & \cdots & x_{n1} \\ 1 & x_{12} & x_{22} & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{1n} & x_{n2} & & x_{kn} \end{pmatrix}_{n\times(k+1)}$$

$$\underset{\sim}{B} = \begin{pmatrix} B_0 \\ \vdots \\ B_k \end{pmatrix}_{(k+1)\times 1} \qquad \epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}_{n\times 1}$$

$$\underset{\sim}{Y} = X\underset{\sim}{B} + \underset{\sim}{\epsilon} \qquad \underset{\sim}{\epsilon} \sim N(0, \sigma^2 I_n)$$

LS condition, $S^2 = (\underset{\sim}{Y} - X\underset{\sim}{B})^T (Y - XB)$

$S^2(B)$

$$\hat{\underset{\sim}{B}} = \arg\min \left(S^2(\underset{\sim}{B}) \right)$$

$$\underset{\sim}{B} \in R^{k+1}$$

$$\hat{B} = (X^T X^{-1}) X^T Y \qquad \text{Assm } |X^T X| \neq 0$$

$$\hat{B} \sim N(\underset{\sim}{B}, (X^T X)^{-1} \sigma^2)$$

H_0 : $B_1 = B_2$ —— $B_k = 0$

H_1 : H_0 is not true

$$y_i \overset{iid}{\sim} N(B_0, \sigma^2)$$

$$\underset{\sim}{B} = \begin{pmatrix} B_0 \\ \vdots \\ 0 \end{pmatrix}$$

whether we should build model or not

If we replace the prediction by $\bar{y}$

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

$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 =$ Error Sum of square

$= \underset{\sim}{y}^T (I_n - P_x) \underset{\sim}{y}$

Scanned by CamScanner

$$SST = \underset{\sim}{y}^T \left(I_n - \frac{1}{n} \underset{\sim}{1} \underset{\sim}{1}^T \right) y$$

$$SS\,model = \sum_{i=1}^{n} (\hat{y}_i - \bar{y}) = \underset{\sim}{y}^T \left(P_x - \frac{1}{n} \underset{\sim}{1} \underset{\sim}{1}^T \right) \underset{\sim}{y}$$

$$SST = SS\,error + SS\,model$$
$$= ESS + SS\,model$$

$$SST = \underset{\sim}{y}^T \left(I_n - \frac{1}{n} \underset{\sim}{1} \underset{\sim}{1}^T \right) \underset{\sim}{y}$$

$$= \underset{\sim}{y} \left(I_n - P_x + P_x - \frac{1}{n} \underset{\sim}{1} \underset{\sim}{1}^T \right) \underset{\sim}{y}$$

$$= \underset{\sim}{y}^T (I_n - P_x) \underset{\sim}{y} + y^T \left(P_x - \frac{1}{n} \underset{\sim}{1} \underset{\sim}{1}^T \right) \underset{\sim}{y}$$

$$= SSE + SS\,model$$

$SST \sim \sigma^2 \chi^2_{n-1}$, $ncp = \lambda$

$SS\,error \sim \sigma^2 \chi^2_{n-(k+1)}$, $ncp = 0$ $\longrightarrow$ Rank of (P_x)

$SS\,model \sim \sigma^2 \chi^2_{k}$, $ncp = \lambda$

NCP for SS Model

NCP for $\underset{\sim}{y}^T \left(P_x - \frac{1}{n}\underset{\sim}{1}1^T\right)\underset{\sim}{y}$

$$\underset{\sim}{y} \sim N(X\beta, \sigma^2 I_n) \Rightarrow \frac{y}{\sigma^2} \sim N(X\beta, \sigma^2)$$

$u^T A u$

$$\Rightarrow (X\beta)^T \left(P_x - \frac{1}{n}1 1^T\right)(X\underset{\sim}{\beta})$$

$$= \underset{\sim}{\beta}^T X^T \left(P_x - \frac{1}{n}1 1^T\right) X\underset{\sim}{\beta}$$

Devide by σ^2

$$N = \frac{1}{\sigma^2}\underset{\sim}{\beta}^T X^T \left(P_x - \frac{1}{n}1 1^T\right) X\underset{\sim}{\beta}$$

$$\underset{\sim}{\beta} = \begin{pmatrix}\beta_0 \\ \beta_R\end{pmatrix} = \begin{pmatrix}\beta_0 \\ \beta_1 \\ \vdots \\ \beta_K\end{pmatrix} \qquad X = (1 : X_R)$$

$$\Rightarrow \frac{1}{\sigma^2} \beta^T \left(X^T P_x X - \frac{1}{n} X^T \underset{\sim}{1} 1^T X\right)\underset{\sim}{\beta}$$

$$= \frac{1}{\sigma^2} \beta^T \left(X^T X - \frac{1}{n} X^T \underset{\sim}{1} 1^T X\right)\underset{\sim}{\beta}$$

$$= \frac{1}{\sigma}\beta^T\left(\begin{bmatrix} n & 1^T X_R \\ X_R^T 1 & X_R^T X_R\end{bmatrix} - \frac{1}{n}\begin{pmatrix}1^T \\ X_R^T\end{pmatrix}\left(\underset{\sim}{1}\underset{\sim}{1}^T\right)(1 X_R)\right)\beta$$

$$= \frac{1}{\sigma}\beta^T\left(\begin{matrix} n & 1^T X_R \\ X^T 1 & X_R^T X_R\end{matrix} - \frac{1}{n}\begin{pmatrix} n \\ X_R^T \underset{\sim}{1}\end{pmatrix}(n \; 1^T X_R)\right)\beta$$

$$\frac{1}{\sigma^2}\beta\begin{pmatrix} n & \cdot\ \cancel{\beta}\ \underset{\sim}{1}^T X_R \\ X_R^T \underset{\sim}{1} & X_R^T X_R \end{pmatrix} - \begin{pmatrix} n & \underset{\sim}{1}^T X_R \\ X_R^T \underset{\sim}{1} & \cancel{\ }\frac{1}{n} X_R^T \underset{\sim}{1}\,\underset{\sim}{1}^T X_R \end{pmatrix} \underset{\sim}{\beta}$$

$$= \frac{1}{\sigma^2}\underset{\sim}{\beta}^T \begin{bmatrix} 0 & \underset{\sim}{0}^T \\ \underset{\sim}{0} & X_R^T X_R - \frac{1}{n} X_R^T \underset{\sim}{1}\,\underset{\sim}{1}^T X_R \end{bmatrix} \underset{\sim}{\beta}$$

$$= \frac{1}{\sigma^2}\begin{pmatrix} \beta_0 \\ \beta_R \end{pmatrix}^T \begin{bmatrix} 0 & \underset{\sim}{0}^T \\ \underset{\sim}{0} & X_R^T X_R - \frac{1}{n} X_R^T \underset{\sim}{1}\,\underset{\sim}{1}^T X_R \end{bmatrix} \begin{pmatrix} \beta_0 \\ \beta_R \end{pmatrix}$$

$$= \frac{1}{\sigma^2}\underset{\sim}{\beta}_R^T \left(X_R^T X_R - \frac{1}{n} X_R^T \underset{\sim}{1}\,\underset{\sim}{1} X_R \right) \underset{\sim}{\beta}_R$$

then $\boxed{ncp \Rightarrow \lambda = 0}$ iff $\underset{\sim}{\beta}_R = 0$

which is nothing But

$$\boxed{\beta_1 = \beta_2 \ \cdots\cdots = \beta_R = 0}$$

which was H_0

HW

Express $\left(X_R^T X_R - \frac{1}{n} X_R^T \underset{\sim}{1}\,\underset{\sim}{1}^T X_R \right)$ in the form of centered regression variable

$SST \sim \sigma^2 \chi^2_{n-1}$, $ncp = \lambda$

$SSE \sim \sigma^2 \chi^2_{n-k-1}$ $ncp = 0$

$SS_{model} \sim \sigma^2 \chi_k$, $ncp = \lambda$

Total $H_0 : B_1 = B_2 \text{ --- } B_k = 0$

vs H_1: H_0 is not true

$$\frac{SSmodel}{\sigma^2} \sim \chi_k \text{ , } ncp = \frac{\lambda}{\sigma^2}$$

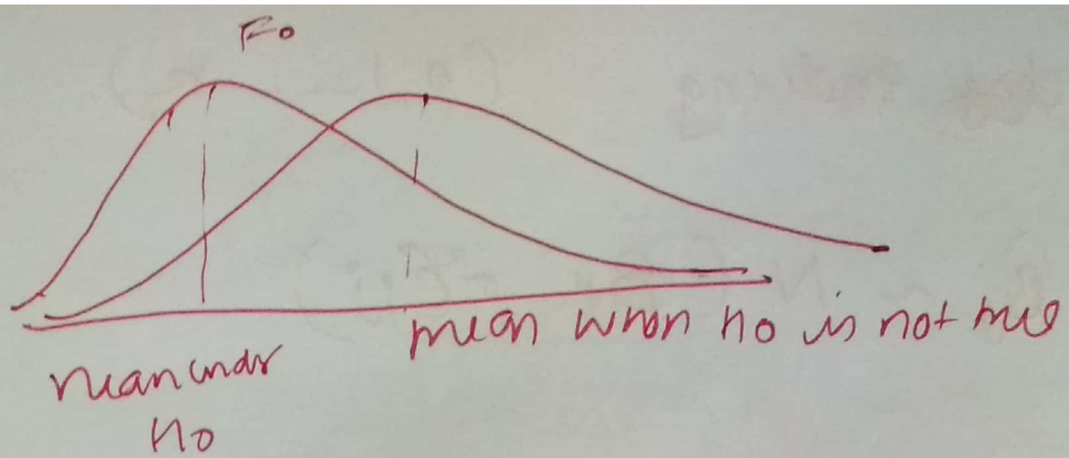
under H_0

$$\frac{SSmodel}{\sigma^2} \sim \chi^2_k \quad ncp = 0$$

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

$$F = \frac{SSmodel/k}{SS\ Error/n-k-1} \sim F_{k, n-k-1}$$

under H_0



So we reject H_0 in favour of H_1 if

$$F_{observe} > F_{\alpha, k, n-k-1}$$

☆ Analysis of variance (ANOVA) for Regression model

$$\hat{\underset{\sim}{B}} \sim N\left(\underset{\sim}{B}, (X^T X)^{-1} \sigma^2\right)$$

$H_0 : B_j = b_j$ vs $H_1 : B_j \neq b_j$

$j = 0 \dots k$

$$\underset{\sim}{B} = \begin{pmatrix} B_0 \\ \vdots \\ B_k \end{pmatrix}$$

$C_{\cancel{k \times k, k}} = (X^T X)^{-1}$

$(k+1 \times k+1)$ $(k+1 \times k+1)$

Scanned by CamScanner

By in digg indexing $(0, 1 \ldots k)$

$$\Rightarrow \hat{B}_j \sim N(B_j, \sigma^2 C_{jj})$$

$$\Rightarrow \frac{\hat{B}_j - B_j}{\sqrt{\sigma^2 C_{jj}}} \sim N(0, 1)$$

$$\Rightarrow T = \frac{\hat{B}_j - b_j}{\hat{\sigma}\sqrt{C_{jj}}} \sim t_{n-k-1}$$

undersno

$$\hat{\sigma}^2 = \frac{SS\,Error}{n-k-1}$$

We Reject H_0 in favour of H_1 if

$$|T_{ob}| > t_{\alpha/2, n-k-1}$$

$$F = T^2 > F_{\alpha, 1, n-k-1}$$

$\downarrow$

How? Next Page

$$T = \frac{Z}{\sqrt{Y/m}} \qquad \left.\begin{array}{l} Z \sim N(0,1) \\ Y \sim \chi^2_m \end{array}\right\} \text{independent}$$

$$Z^2 \sim \chi^2_1$$

$$Y \sim \chi^2_m$$

$$T^2 = \frac{Z^2}{Y/m} = \frac{Z^2/1}{Y/m} \sim F$$

Test for

$H_0 : \beta_j - \theta\beta_i = 0$

vs $H_1 : \beta_j - \theta\beta_i \neq 0$

$$\hat{\beta}_j - \theta\hat{\beta}_i \sim ?$$

If $\underset{\sim}{Z} \sim N(\mu, \Sigma)$

$$l^T Z \sim N(l^T\mu, l^T\Sigma l)$$

$\hat{B} \sim N(\underset{\sim}{B}, \sigma^2 C)$

$l^T = (0 \, \_\_\_ \, \underset{j}{1} \, \_\_\_ \, \underset{-i}{-\theta} \, \_\_\_ \, 0)$

$l^T \underset{\sim}{B} \sim N(B_j - \theta B_i, \sigma^2 (C_{jj} + \theta^2 C_{ii} - 2C_{ij}))$

How to get this

$(1 \; 0 \; -1) \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$

$(a + f) - 2c \, \_\_$

$\hat{B} \sim N(\underset{\sim}{B}, (X^T X)^{-1} \sigma^2)$

If $|(X^T X)| \neq 0$ then any linear parametric function of $\underset{\sim}{B}$ in the form $\underset{\sim}{p}^T \beta$ is estimable

---

i.e we have linear combination of $\underset{\sim}{Y}$ eg $\underset{\sim}{l}^T \underset{\sim}{Y}$ such that $E(\underset{\sim}{l}^T \underset{\sim}{Y}) = p^T \underset{\sim}{B}$

Linear parametric function (LPF)

A LPF of $\underset{\sim}{\beta}$ is a linear combination of β i.e

$\underset{\sim}{p}^T \underset{\sim}{\beta}$

Linear unbiased estimator (LUE)

$E(\underset{\sim}{l}^T \underset{\sim}{y}) = \underset{\sim}{p}^T \underset{\sim}{\beta}$ then $\underset{\sim}{l}^T y$ is said to LUE of $p^T \beta$

if it holds for all $\beta \in \mathbb{R}^{k+1}$

Linear Zero Estimator (LZE)

if $E(\underset{\sim}{l}^T \underset{\sim}{y}) = 0 \quad \forall \ \beta \in \mathbb{R}^{k+1}$ then $\underset{\sim}{l}^T \underset{\sim}{y}$ is

known as a linear zero estimator

---

① Show that $l^T y$ is an LUE of $\underset{\sim}{p}^T \underset{\sim}{\beta}$

iff $x^T \underset{\sim}{l} = \underset{\sim}{p}$

$l^T y$ is an LUE of $p^T \beta$

iff $E(\underset{\sim}{l}^T \underset{\sim}{y}) = \underset{\sim}{p}^T \underset{\sim}{\beta} \quad \forall \ \beta \in \mathbb{R}^{k+1}$

we know, $y \sim N(X\beta, \sigma^2 I_m)$

$l^T X \beta = p^T \beta \quad \forall \ \beta \in \mathbb{R}^{k+1}$

$\Rightarrow \underset{\sim}{l}^T X = p^T$

$X^T \underset{\sim}{l} = \underset{\sim}{p}$

(2) $\underset{\sim}{l}^T \underset{\sim}{y}$ is a LZF iff $X^T l = 0$

from def^n

$$E(l^T y) = 0 \quad \forall \underset{\sim}{\beta} \in R^{k+1}$$

$$\Leftrightarrow l^T X\beta = 0$$

$$\Leftrightarrow l^T X = \underset{\sim}{0}^T$$

$$\Leftrightarrow X^T \underset{\sim}{l} = \underset{\sim}{0}$$

$$\Rightarrow l \in (C(X))^\perp$$

$$l \in C(I_n - P_x)$$

$\Rightarrow \exists\, m \in R^n$ such that

$$l = (I - P_x)m$$

$$l^T X = \underset{\sim}{m}^T (I - P_x)X = \underset{\sim}{0}^T$$

$\Rightarrow$ Every LZF can be written as $\boxed{\underset{\sim}{m}^T (I - P_x)\underset{\sim}{y}}$

$$E(\underset{\sim}{m}^T (I - P_x)\underset{\sim}{y}) = \underset{\sim}{m}^T (X\underset{\sim}{\beta} - P_x X\underset{\sim}{\beta})$$

$$= \underset{\sim}{m}^T (X - X)\beta = 0$$

A linear parametric function of $p^T\beta$ is said said to be estimable if there exist $l^T y$ such that

$$E(l^T y) = p^T \beta \quad \forall \ \beta \in R^{k\times 1}$$

③ An (iii) iff condition for $p^T\beta$ to be estimable if $p \in C$

$$p \in C(X^T)$$

HW

$$E(l^T Y) = p^T\beta$$

$$l^T X\beta = p^T\beta$$

$$X^T l = p$$

p is a linear combination of X^T

$$p \in C(X^T)$$

Best linear unbiased estimator (BLUE)

An unbiased estimator of $p^T\beta$ is said to the ~~Blue~~ BLUE of $p^T\beta$ if it has minimum variance among all linear unbiased estimators of $p^T\beta$

$$U = \{ \underset{\sim}{l}^T \underset{\sim}{Y} \mid E(\underset{\sim}{l}^T \underset{\sim}{Y}) = p^T \beta \; \beta \in R^{k+1} \}$$

$$Var(l_0^T Y) \leq V(\underset{\sim}{l}^T \underset{\sim}{Y}) \quad \forall \; l, l_0 \in U$$

Th: A linear function of $\underset{\sim}{Y}$ is BLUE of its expection iff it is uncorrelat with all linear zero function/Estimator (LZF)

$$E(\underset{\sim}{l}^T \underset{\sim}{Y}) = \underset{\sim}{p}^T \underset{\sim}{B}$$

iff $cov(\underset{\sim}{l}^T \underset{\sim}{Y}, Z^T \underset{\sim}{Y}) = 0$

where $Z^T Y$ is LZF

Ex 1: If $l^T Y$ is LUE of $p^T B$ then corresponding BLUE is $\underset{\sim}{l}^T P_x \underset{\sim}{Y}$

$$\underset{\sim}{l}^T \underset{\sim}{Y} = \underset{\sim}{l}^T I \underset{\sim}{Y} = \underset{\sim}{l}^T [P_x + (I - P_x) Y]$$

$$\Rightarrow \; = \underset{\sim}{l}^T P_x Y + \underbrace{l^T (I_n - P_x) Y}_{LZF}$$

from Earlier result

① $E(l^T (I_n - P_x) Y) = 0$

(ii) $E(\underset{\sim}{l}^T \underset{\sim}{Y}) = \underset{\sim}{p}^T \underset{\sim}{\beta} = E(l^T P_x Y) = l^T P_x X \beta$

$= \underset{\sim}{l}^T X \beta$ from earlier result $\boxed{l^T X = p^T}$

$= p^T \beta$

same

$l^T Y$ is an LUE of $p^T \beta$

$E(\cancel{l^T Y = l^T P_x Y}$ $\quad E(l^T Y) = p^T \beta$

$\cancel{E(p^T \beta) =}$

$E(l^T P_x Y) = p^T \beta$

This is an LUE & $l Z F$

such $l^T P_x Y$ & $l^T (I_n - P_x) Y$ are uncorrelated $\Rightarrow l^T P_x Y$ is the BLUE

Ex 5 every estimable LPF has unique BLUE

Let's assume $\exists$ $l_1^T Y$ and $l_2^T Y$ which is BLUE of $p^T \beta$

$$l_1^T \underset{\sim}{Y} = l_2^T Y + (l_1 - l_2)^T Y$$

$$E(l_1^T Y) = E(l_2^T Y) + E(l_1 - l_2^T Y)$$

$$\Rightarrow \quad E((l_1 - l_2)Y) = 0$$

$$\Rightarrow (l_1 - l_2)^T Y \text{ is LZF}$$

$$\Rightarrow V(\underset{\sim}{l_1}^T Y) = V(\underset{\sim}{l_2}^T Y) + V((l_1 - l_2)^T \underset{\sim}{Y}) + 0$$

$\downarrow$
cov
b/w
BLUE &
LZF = 0

Case 1 if $V((l_1 - l_2)^T Y) > 0$

then $V(l_1^T Y) > V(l_2^T Y)$ contradiction

Case 2 $V((l_1 - l_2)^T Y) = 0$.

with $E((L_1 - L_2)^T Y) = 0$

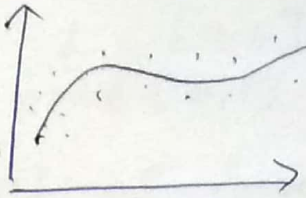
$$P((l_1 - l_2)^T Y = 0) = 1$$

$$P(l_1^T Y = l_2^T Y) = 1$$

$\therefore$ Hence proved $l_1^T Y =$

$l_1 = l_2$

Polynomial Regression



(A) Polynomial Regression is Linear Regression

$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \epsilon_i \qquad \epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

$= f(x_i) + \epsilon_i$

$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i} + \beta_4 x_{1i}^2 + \beta_5 x_{2i}^2 + \epsilon_i$

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

$= g(x_{1i}, x_{2i}) + \epsilon_i$

* f & g both are linear in parameters

for model 1 $\quad \underset{\sim}{X} = \begin{pmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ \vdots & \vdots & & \\ 1 & x_n & x_n^2 & x_n^3 \end{pmatrix}$

$Y = X\underset{\sim}{B} + \epsilon \qquad \epsilon_i \sim N(0, \sigma^2 I_n)$

Note

① If all x_i are equal then $(X^T X)$ is not invertible

② If $\sum_{i=1}^{n} |x_i| < \epsilon$, small number, with higher degree of x_i's will become numerically zero

③ We cant fit at most $(n-1)$ degree polynomial
⟺ Interpolation problem
It will lead to over fitting

$$\underset{\sim}{y} = \begin{pmatrix} 1 & x_1 & \cdots & x_1^{n-1} \\ \vdots & \vdots & & \\ 1 & x_n & \cdots & x_n^{n-1} \end{pmatrix}_{n \times n} \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_n \end{pmatrix}$$

now when all x_i are different then

$0 \neq |A|$

4. If we fit $(n-1)$ degree polynomial estimation of σ^2 will not be possible

5. In a polynomial regression, heirarchy may not be maintained.
   ⇒ may be B_4 is significant but B_2 or B_3 may be insignificant

$$y = B_0 + B_1 x + B_4 x^4$$

6. $y_i = g(x_{1i}, x_{2i}) + \epsilon_i$ (Model 2)

Can be used as response surface model in design of experiment

$y = X\underset{\sim}{B} + \epsilon$

$$\underset{(K+1)}{\hat{\underset{\sim}{B}}} = \underset{(K+1)\times(K+1)}{(X^T X)^{-1}} X^T \underset{\sim}{y}$$

Orthogonal Polynomia Regresson

$x \cdot y = x^T \underset{\sim}{y} = 0 \quad \underset{\sim}{x}, \underset{\sim}{y} \in \mathbb{R}^n$

$\Sigma x_i y_i = 0$

$x \perp y$

Scanned by CamScanner

$\langle f(t), g(t) \rangle = 0$ $\qquad t \in [a,b]$
or $t \in [0,1]$

then also we say f & g are orthogonal function

$$\int_a^b f(t) g(t)\,dt = 0 \quad , \quad \int_0^1 \sin(2\pi t)\cos(2\pi t)\,dt = 0$$

If $y_i = \sum_{j=0}^{K} \alpha_j P_j(x_i)$ $\qquad i = 1, 2 \ldots n$

$[\ P_j(x_i) =$ orthogonal polynomial of x_i

$p_0(y_i) = 1 \ \forall i$

$$\sum_{i=1}^{n} P_j(x_i) P_K(x_i) = 0 \quad \forall\ j \neq K$$