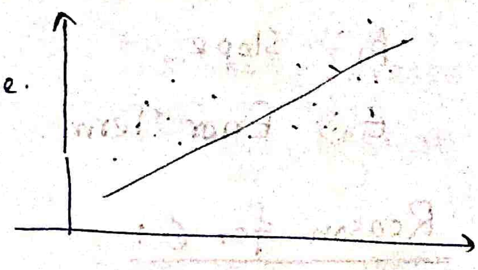
Regression Techniques

Revise : ① Matrix Theory and Linear Algebra.

A statistical Techniques for investing and modelling the relationship between variable.

Km	Minutes
1	3
2	4
5	10
⋮	⋮
20	45



$y \to$ one way travel time.

$x \to$ distance

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

Let, the perpendicular distance between the observed value of y and the straight line $\beta_0 + \beta_1 x$ be ϵ ; Then a more appropriate model

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

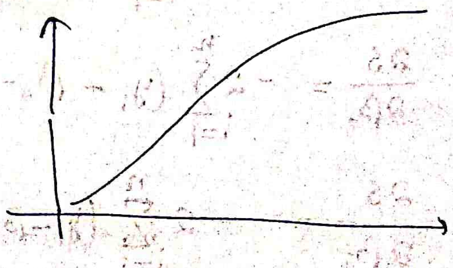
$\hookrightarrow$ Linear Reg. Model.

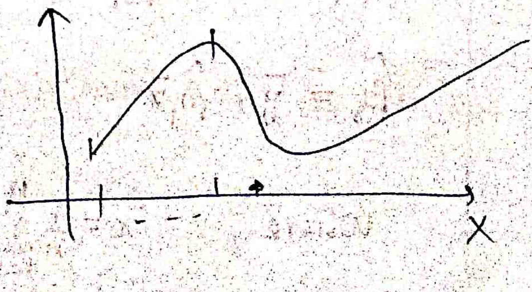
y: response variable (dependent)

x: Independent variable (Regression).

Simple Linear Regression Model:

- One independent variable
  : Simple LRM
- More than one independent variable : $\Rightarrow$ Multiple LRM





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

β_0, β_1 : parameters of the model ~~of the~~ usually unknowns.

β_0 : intercept.

β_1 : Slope

ϵ : Error Term.

Reason for ϵ :

Assume that ϵ are uncorrelated and have mean zero and constant variance.

Consider x as non-stochastic and y as random variable.

$$E(y) = \beta_0 + \beta_1 x ; \quad V(y) = \sigma^2$$

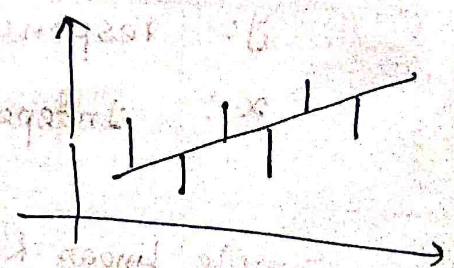
$$\left[E(\epsilon) = 0, \quad V(\epsilon) = \sigma^2 \right]$$

Least Sq. Estimation:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \qquad i = 1, 2, \ldots, n$$

LSE: Minimise $S(\beta_0, \beta_1)$ w.r.t. β_0 & β_1

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$



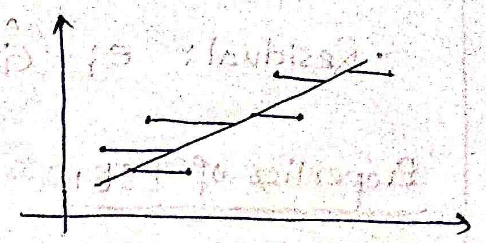
(vertical distance)

$$\left.\begin{aligned} \frac{\partial S}{\partial \beta_0} &= -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) = 0 \\ \frac{\partial S}{\partial \beta_1} &= -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) x_i = 0 \end{aligned}\right\} \text{(normal eq}^n\text{)}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad ; \quad \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$

Where, $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \quad ; \quad S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$

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



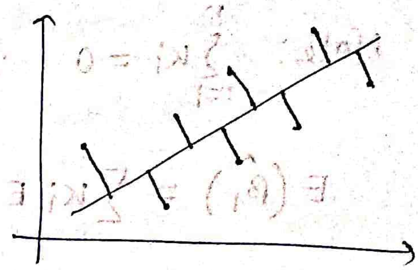
(Horizontal Distance ↓)
[Reverse / Inverse Regression]

$$\frac{\partial^2 S}{\partial \beta_0^2} = 2n \quad ;$$

$$\frac{\partial^2 S}{\partial \beta_1^2} = 2\sum_{i=1}^{n} x_i^2 \quad ;$$

$$\frac{\partial^2 S}{\partial \beta_0 \partial \beta_1} = 2n\bar{x}$$

$$\frac{\partial^2 S}{\partial \beta_1 \partial \beta_0} = 2n\bar{x}$$



[Orthogonal Reg. or Major axis Reg.]

The Hessian matrix :
$$\begin{bmatrix} \frac{\partial^2 S}{\partial \beta_0^2} & \frac{\partial^2 S}{\partial \beta_0 \partial \beta_1} \\ \frac{\partial^2 S}{\partial \beta_0 \partial \beta_1} & \frac{\partial^2 S}{\partial \beta_1^2} \end{bmatrix}$$

$$= \begin{bmatrix} 2n & n\bar{x} \\ n\bar{x} & 2\sum_{i=1}^{n} x_i^2 \end{bmatrix}$$

$$|H| = 2^{2} \left[n\sum x_i^2 - n^2\bar{x}^2 \right]$$
$$= 4\, n \sum (x_i - \bar{x})^2 \geq 0$$

If $\sum_{i=1}^{n} (x_i - \bar{x})^2 = 0$ then x_i's are constant.

$\Rightarrow$ $S(\beta_0, \beta_1)$ has a global minimum of $(\hat{\beta}_0, \hat{\beta}_1)$

∴ The fitted line : $y = \hat{\beta}_0 + \hat{\beta}_1 x$

• The predicted values : $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$, $i = 1, 2, \ldots, n$.

• Residual: $e_i = \hat{\epsilon}_i = y_i - \hat{y}_i$; $i = 1, 2, \ldots, n$

Properties of LSE:

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \quad ; \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}.$$

$$\hat{\beta}_1 = \sum_{i=1}^{n} K_i y_i \quad ; \quad K_i = \frac{x_i - \bar{x}}{S_{xx}}$$

Note: $\sum_{i=1}^{n} K_i = 0$; $\sum_{i=1}^{n} K_i x_i = 1$

$$E(\hat{\beta}_1) = \sum K_i E(y_i) = \sum K_i (\beta_0 + \beta_1 x_i) = \beta_1$$

$$E(\hat{\beta}_0) = E(\bar{y} - \hat{\beta}_1 \bar{x}) = \beta_0 + \beta_1 \bar{x} - \beta_1 \bar{x} = \beta_0$$

Variance:

Assume, $E(\epsilon_i) = 0$, $V(\epsilon_i) = \sigma^2$

$Cov(\epsilon_i, \epsilon_j) = 0 \quad \forall i \neq j.$

$var(y_i) = \sigma^2$; $Cov(y_i, y_j) = 0 \quad \forall i \neq j.$

$$Var(\hat{\beta}_1) = \sum_{i=1}^{n} K_i^2 V(y_i) + \sum_{\substack{i,j \\ i \neq j}}\sum K_i K_j Cov(y_i, y_j)$$

$$= \sigma^2 \cdot \frac{\sum (x_i - \bar{x})^2}{S_{xx}^2}$$

$$= \frac{\sigma^2}{S_{xx}}$$

$$Var(\hat{\beta}_0) = var(\bar{y}) + \bar{x}^2 var(\hat{\beta}_1) - 2\bar{x}\, Cov(\bar{y}, \hat{\beta}_1)$$

$$Cov(\bar{y}, \hat{\beta}_1) = Cov\left(\frac{1}{n}\sum Y_i, \sum K_i Y_i\right)$$

$$= \frac{1}{n} \sum K_i\, var(y_i) = 0 \quad [\text{as } \sum K_i = 0]$$

$$Var(\hat{\beta}_0) = \sigma^2/n + \frac{\bar{x}^2\sigma^2}{S_{xx}} = \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)$$

$$Cov(\hat{\beta}_0, \hat{\beta}_1) = Cov(\bar{y}, \hat{\beta}_1) - \bar{x}\, var(\hat{\beta}_1)$$

$$= -\frac{\bar{x}\sigma^2}{S_{xx}}$$

Residual Sum of squares: (RSS, SS_{Res}).

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

$$= \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

$$= \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

$$= \sum_{i=1}^{n} (y_i - \bar{y} + \hat{\beta}_1\bar{x} - \hat{\beta}_1 x_i)^2$$

$$= \sum_{i=1}^{n} (y_i - \bar{y})^2 + \hat{\beta}_1^2 \sum_{i=1}^{n} (x_i - \bar{x})^2 - 2\hat{\beta}_1 \sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})$$

$$= S_{yy} + \hat{\beta}_1^2 S_{xx} - 2\hat{\beta}_1$$

Estimation of σ^2:

$$\hat{\sigma}^2 = s^2 = \frac{RSS}{(n-2)}$$

Estimate of variances of $\hat{\beta}_0$ & $\hat{\beta}_1$

$$\hat{Var}(\hat{\beta}_0) = s^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right) \; ; \; \hat{Var}(\hat{\beta}_1) = \frac{s^2}{S_{xx}}$$

Useful of LS fit :

For a given sample, $\{e_i\}$ satisfies the following

ⓐ $\sum_{i=1}^{n} e_i = 0$, ⓑ $\sum_{i=1}^{n} x_i e_i = 0$

ⓒ $\sum_{i=1}^{n} \hat{y}_i e_i = 0$

ⓓ The LS line always passes through $(\bar{x}, \bar{y})$ of the data.

ⓔ $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{y}_i$

ⓒ $\sum \hat{y}_i e_i = \sum (\hat{\beta}_0 + \hat{\beta}_1 x_i) e_i$

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

$$= \bar{y} + \hat{\beta}_1 (x - \bar{x})$$

Centered Model:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad , i = 1, 2, \ldots, n.$$

$$= \beta_0 + \beta_1 (x_i - \bar{x}) + \beta_1 \bar{x} + \epsilon_i$$

$$= \beta_0^* + \beta_1 (x_i - \bar{x}) + \epsilon_i$$

$$; \beta_0^* = \beta_0 + \beta_1 \bar{x}$$

$$\hat{\beta}_0^* = \bar{y} \; ; \; \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$

Calculating: $E(\hat{\beta}_0^*)$

$E(\hat{\beta}_1)$

$Cov(\hat{\beta}_0^*, \hat{\beta}_1)$

In matrix Notation:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad , \; i = 1, 2, \dots, n.$$

$$\underset{\sim}{y} = (y_1, y_2, \dots, y_n)'$$

$$\underset{\sim}{\epsilon} = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)'$$

$$\underset{2\times 1}{\underset{\sim}{\beta}} = (\beta_0, \beta_1)'$$

$$\overset{n\times 2}{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \qquad \underset{\sim}{y} = X \underset{\sim}{\beta} + \underset{\sim}{\epsilon}$$

Books:

(i) Linear Reg. Analysis [Matrix]
(Seber & Lee.)

(ii) Applied Reg. Analysis
(Draper & Smith)

(iii) Introduction to Linear Reg. Analysis
Montogomery, Peek & Vinning.

Linear model:

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

$$y = \beta_0 + \beta_1 \log x + \epsilon$$

$y_1, \ldots, y_n \qquad y_i = \beta_0 + \beta_1 x_i + \epsilon_i \;, \quad i = 1, 2, \ldots, n$

$\underline{Y} = X\underline{\beta} + \underline{\epsilon} \qquad ; \quad \underset{n\times 1}{\underline{Y}} = (Y_1, \ldots, Y_n)'$

$\qquad , \quad \underset{n\times 1}{\epsilon}$

$\underline{\beta} = (\beta_0, \beta_1)'$

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_n \end{pmatrix}.$$

Assumption: $\quad E(\underline{\epsilon}) = \underline{0} \quad , \quad var(\underline{\epsilon}) = \sigma^2 I_n$

LSE : $\quad$ Minimise $S(\underline{\beta}) = \underline{\epsilon}'\underline{\epsilon} = (\underline{Y} - X\underline{\beta})'(Y - X\underline{\beta})$

$$= \| \underline{Y} - X\underline{\beta} \|^2$$

Vector Differentiation :

$Q(\underline{\theta}) \quad , \quad Q : \mathbb{R}^p \longrightarrow \mathbb{R}$

$\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_p)'$

$$\frac{\partial Q(\underline{\theta})}{\partial \underset{\sim}{\theta}} = \left(\frac{\partial Q(\underline{\theta})}{\partial \theta_1}, \ldots \frac{\partial Q(\underline{\theta})}{\partial \theta_p} \right)'$$

$$\frac{\partial^2 Q(\underline{\theta})}{\partial \underline{\theta} \partial \underline{\theta}^T} = \begin{bmatrix} \frac{\partial^2 Q(\underline{\theta})}{\partial \theta_1^2} & \frac{\partial^2 Q(\underline{\theta})}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial^2 Q(\underline{\theta})}{\partial \theta_1 \partial \theta_p} \\ \vdots & & & \\ \frac{\partial^2 Q(\underline{\theta})}{\partial \theta_p \partial \theta_1} & & \cdots & \frac{\partial^2 Q(\underline{\theta})}{\partial \theta_p^2} \end{bmatrix}$$

If $S: \mathbb{R}^p \longrightarrow \mathbb{R}^2$, $S(\underline{\theta}) = (S_1(\underline{\theta}), S_2(\underline{\theta}))'$

$$\frac{\partial S(\theta)}{\partial \underline{\theta}} = \left(\frac{\partial S_1(\theta)'}{\partial \underline{\theta}}, \frac{\partial S_2(\theta)}{\partial \underline{\theta}} \right)'$$

$2 \times p$ matrix

$$\frac{\partial}{\partial \underline{\beta}} (\underline{\beta}' \underline{a}) = \underset{\sim}{a}$$

$\underline{a} = (a_1, \ldots, a_p)'$,
$\underset{\sim}{\beta} = (\beta_0 \, \beta_1 \ldots \beta_{p-1})'$

$$\frac{\partial}{\partial \underline{\beta}} (\underline{\beta}' A \underline{\beta}) = 2A\underline{\beta}$$

$\sum\sum a_{ij} \beta_i \beta_j$

$$S(\underline{\beta}) = (\underline{Y} - \underline{X}\underline{\beta})'(\underline{Y} - X\underline{\beta})$$

$$= \underline{Y}'\underline{Y} - \beta' \underline{X}' \underline{Y} - Y'X'\underline{\beta} + \underline{\beta}' \underline{X}' \underline{X} \underline{\beta}$$

$$\frac{\partial S(\beta)}{\partial \underline{\beta}} = 0 \Rightarrow -2X'\underline{Y} + 2(X'X)\underline{\beta} = 0$$

$\hookrightarrow$ normal eq^n

x_i's are generally chosen so that the columns of X are LIN (rank 2) : X is of full rank.

$$\Rightarrow (X'X)\underline{\beta} = X'\underline{Y}$$

Multiple linear regression Model

$x_1, x_2, \ldots, x_{p-1}$: explanatory variable.

$\hookrightarrow$ Known constant (or by the ..)

& measured with negligible error

β_j, $j = 0, 1, \ldots, p-1$

unknown parameter to be estimated.

$$y = \eta + \epsilon$$

$$\eta = \beta_0 + \beta_1 x_1 + \ldots + \beta_{p-1} x_{p-1}$$

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_{p-1} x_{i\,p-1} + \epsilon_i$$

$i = 1, 2, \ldots n$

x_{ij} : ith value of jth variable.

$$\underset{n\times 1}{\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}} = \underset{n\times p}{\begin{pmatrix} x_{10} & x_{11} & \cdots & x_{1\,p-1} \\ x_{20} & x_{21} & & x_{2\,p-1} \\ \vdots & \vdots & & \vdots \\ x_{n0} & x_{n1} & & x_{n\,p-1} \end{pmatrix}} \underset{p\times 1}{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix}} + \underset{n\times 1}{\begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}}$$

$$\underline{Y} = X\underline{\beta} + \underline{\epsilon}$$

$$x_{10} = x_{20} = \cdots x_{n0} = 1$$

We consider the case where column of X are linearly independent, so rank$(X) = p$.

So, $X'X$ is pd. and non-singular.

$$\hat{\underline{\beta}} = (X'X)^{-1} X'Y \rightarrow \text{LSE of } \underline{\beta}$$

Where column of X are not LIN, then $\hat{\beta}$ is no longer unique. A sol^n can be given by

$$\hat{\beta} = (X'X)^{-} X'\underline{Y}, \quad (X'X)^{-} \rightarrow \text{g. inverse of } X'X$$

G is a g-inverse of A

if $AGA = A$.

Identity:

$$(\underline{Y} - X\underline{\beta})'(\underline{Y} - X\underline{\beta}) = (\underline{Y} - X\hat{\underline{\beta}})'(\underline{Y} - X\hat{\underline{\beta}}) + (\hat{\underline{\beta}} - \underline{\beta})' X'X (\hat{\underline{\beta}} - \underline{\beta})$$

minimised uniquely where $\underline{\beta} = \hat{\underline{\beta}}$

$$(\underline{Y} - X\hat{\underline{\beta}} + X\hat{\underline{\beta}} - X\underline{\beta})' (\underline{Y} - X\hat{\underline{\beta}} + X\hat{\underline{\beta}} - X\underline{\beta})$$

$$= (\underline{Y} - X\hat{\underline{\beta}})'(\underline{Y} - X\hat{\underline{\beta}}) + (\hat{\underline{\beta}} - \underline{\beta})' X'X (\hat{\underline{\beta}} - \underline{\beta})$$

$$+ (Y - X\hat{\beta})' X (\hat{\beta} - \beta) + (\hat{\underline{\beta}} - \underline{\beta}) X' (\underline{Y} - X\hat{\underline{\beta}})$$

- Fitted : $\hat{\underline{Y}} = (\hat{y}_1, \ldots \hat{y}_n) = X\hat{\underline{\beta}} = X(X'X)^{-1}X'\underline{Y} = PY$
- Residual vector : $(Y - \hat{Y}) = \underline{Y} - X\hat{\underline{\beta}} = Y - PY = (\underline{I} - P)\underline{Y}$

$P = X(X'X)^{-1}X' \rightarrow$ Projection matrix

The minimum value of $\underline{\epsilon}'\underline{\epsilon} = \hat{\underline{\epsilon}}'\hat{\underline{\epsilon}} = \underline{e}'\underline{e}$

$$(\underline{Y} - X\hat{\underline{\beta}})' (\underline{Y} - X\hat{\underline{\beta}})$$

$$= \underline{Y}'\underline{Y} - 2\hat{\underline{\beta}}'X'\underline{Y} + \hat{\underline{\beta}}' X'X \underline{\beta}$$

$$= \underline{Y}'\underline{Y} - \hat{\underline{\beta}}'X'\underline{Y} + \hat{\beta}' [X'X\underline{\beta} - X'\underline{Y}]$$

$$= \underline{Y}'\underline{Y} - \hat{\underline{\beta}}'X'\underline{Y} = \underline{\underline{Y}'\underline{Y} - \hat{\underline{\beta}}'X'X\hat{\underline{\beta}}}$$

$\boxed{\hat{\underline{Y}} = P\underline{Y}}$

$\hookrightarrow$ residual sum of square (RSS)

$$RSS = (\underline{Y} - X\hat{\underline{\beta}})' (\underline{Y} - X\hat{\underline{\beta}}) = (\underline{Y} - P\underline{Y})' (\underline{Y} - P\underline{Y})$$

$$= \underline{Y}'(I - P)^2 \underline{Y} = \underline{Y}'(I - P)\underline{Y}$$

Example: ① Y_1 & Y_2 are independent r.v's with mean α & 2α respectively. Find LSE of α and RSS.

② Suppose $Y_1, \ldots, Y_n$ all have mean β. Write down it in matrix notation & find the LSE of β.

(1)

$$\hat{\alpha} = \frac{y_1 + 2y_2}{5}$$

$$RSS = y_1^2 + y_2^2 - \frac{(y_1 + 2y_2)^2}{5}$$

$$X = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

(2) $y_1, y_2, \ldots, y_n$ all have mean β.

$$X = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

Sol^n

(1)

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \alpha$$

$\therefore \quad \underline{Y} = \underset{\sim}{X} \underline{\beta}$

$\therefore \quad \underset{\sim}{X} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$

$$X'X = (1 \quad 2) \begin{pmatrix} 1 \\ 2 \end{pmatrix} = 1 + 4 = 5$$

$$\therefore \text{LSE} = \hat{\beta} = (X'X)^{-1} X' \underline{Y} = \frac{1}{5} (1 \quad 2) \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

$$= \frac{y_1 + 2y_2}{5} = \hat{\alpha}$$

Now, $\quad RSS = \underline{Y}'(I - P)\underline{Y}$

$$P = X(X'X)^{-1}X'$$

$$= \frac{1}{5}\begin{pmatrix} 1 \\ 2 \end{pmatrix}(1 \quad 2) = \frac{1}{5}\begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$$

$$I - P = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \frac{1}{5}\begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} = \begin{pmatrix} 4/5 & -2/5 \\ -2/5 & 1/5 \end{pmatrix}$$

$$= A \text{ (say)}$$

$\therefore \ \underline{y}'A\underline{y}$ = Quadratic form of A

$$= \frac{4}{5}y_1^2 + \frac{1}{5}y_2^2 - 2\cdot\frac{2}{5}y_1y_2$$

$$= \frac{4}{5}y_1^2 + \frac{1}{5}y_2^2 - \frac{4}{5}y_1y_2$$

(2) $\underline{Y} = (Y_1, Y_2, \dots Y_n)'$

mean vector = $\beta \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$

$$\therefore \ \underset{\sim}{Y} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \cdot \beta \ = X\beta$$

$$\therefore \ X = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

$\therefore X'X = n$

$$\therefore \text{LSE} = \hat{\beta} = (X'X)^{-1}X'y = \frac{1}{n}(1\ 1 \cdots 1)\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

$$= \frac{1}{n}\sum_{i=1}^{n} y_i = \overline{y}$$

Result:

Let $\underset{n\times m}{X}$ be any real matrix of rank r $(r \le m)$ and $y \in \mathbb{R}^n$ be any vector. Then ~~the eqn~~

(a) The eqn in unknown $\underline{\beta}_{m\times 1}$, $(X'X)\underline{\beta} = X'\underline{y}$ is always consistent (i.e. we can solve in all cases).

(b) $X\hat{\underline{\beta}}$ is always unique. where $\hat{\beta}$ is any solution of $X'X\beta = X'\underline{y}$

Proof:

(a) Let, $A = [X'X, X'y]$, then $r(A) \geq r(X'X)$

$A = X'[X, Y]$

rank $(A) = $ rank $(X'X) = r$

$\Rightarrow X'\underline{y} \in \mathcal{C}(X'X)$ → space generated by the columns of $X'X$ (column space of $X'X$)

$\Rightarrow X'X\beta = X'y$ is consistent.

(b) Let, β_1 and β_2 be any two solutions of $X'X\underline{\beta} = X'y$

$\Rightarrow X'X\beta_1 - X'X\beta_2 = X'\underline{y} - X'\underline{y} = 0$

$\Rightarrow X'X(\beta_1 - \beta_2) = 0$

$\Rightarrow (\underline{\beta}_1 - \underline{\beta}_2)'X'X(\underline{\beta}_1 - \underline{\beta}_2) = 0$ $\quad (\underline{w}'\underline{w} = 0)$

$\Rightarrow X(\underline{\beta}_1 - \underline{\beta}_2) = 0 \Rightarrow X\underline{\beta}_1 = X\underline{\beta}_2$

Theorem:

Suppose that X is $n \times p$ of rank p; so that $P = X(X'X)^{-1}X'$. Then the following statements hold:

ⓐ P and $I_n - P$ are symmetric and idempotent.

ⓑ $\text{rank}(I_n - P) = tr(I_n - P) = n - p$.

ⓒ $PX = X$

Proof:

ⓐ $P^2 = X(X'X)^{-1}X' \cdot X(X'X)^{-1}X'$

$= X(X'X)^{-1}X' = P$

$\therefore P$ is idempotent

$(I_n - P)^2 = I_n - 2P + P^2 = I_n - P$ [as P is idempotent]

ⓑ Since $(I_n - P)$ is symmetric and idempotent, then

$\text{rank}(I_n - P) = tr(I_n - P) = n - tr(P)$

Where $tr(P) = tr(X(X'X)^{-1}X')$ ~~…~~ $= p$

ⓒ $PX = X(X'X)^{-1}X' \cdot X = X$

Properties of LSE

$E(\underline{\epsilon}) = \underline{0}$ & columns of X are LIN.

$E(\hat{\underline{\beta}}) = (X'X)^{-1}X'E(Y) = (X'X)^{-1}X'X\underline{\beta} = \underline{\beta}$

$\hat{\underline{\beta}}$ is an unbiased for $\underline{\beta}$.

$Cov(\epsilon_i, \epsilon_j) = \delta_{ij}\sigma^2$, $\delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$

$Var(\underline{\epsilon}) = \sigma^2 I_n$

$\boxed{D(A\underline{Y}) = AD(\underline{Y})A'}$

$Var(\underline{y}) = Var(\underline{Y} - X\underline{\beta}) = V(\underline{\epsilon}) = \sigma^2 I_n$

$Var(\hat{\underline{\beta}}) = \sigma^2(X'X)^{-1} = Var((X'X)^{-1}X'\underline{Y})$

$= (X'X)^{-1}X' Var(Y) X (X'X)^{-1} = \sigma^2(X'X)^{-1}$

~~Unbiased~~

Unbiased Estimator of σ^2:

Theorem: If $E(\underline{Y}) = X\underline{\beta}$, X is $n \times p$ of rank p

$Var(\underline{Y}) = \sigma^2 I_n$

Then $$s^2 = \frac{(\underline{Y} - X\hat{\underline{\beta}})'(\underline{Y} - X\hat{\underline{\beta}})}{n-p} = \frac{RSS}{n-p}$$

is an unbiased estimator of σ^2.

Proof:

$$(n-p)s^2 = (\underline{Y} - X\hat{\underline{\beta}})'(\underline{Y} - X\hat{\underline{\beta}})$$

$$= \underline{Y}'(I-P)'(I-P)\underline{Y}$$

$$= \underline{Y}'(I-P)\underline{Y}$$

$$E[\underline{Y}'(I-P)\underline{Y}]$$

$$= tr[(I-P)\sigma^2 I_n] + \underline{\beta}'X'(I-P)X\underline{\beta}$$

$$= \sigma^2(n-p) + 0 \quad (PX = X)$$

$X = (x_i)$

$E(X) = \underline{\mu}$, $V(X) = \Sigma$

$E(X'AX)$

$= tr(A\Sigma) + \underline{\mu}'A\underline{\mu}$

$\Rightarrow E(s^2) = \sigma^2$

$\therefore$ s^2 is an unbiased estimator of σ^2.

The assumption made so far: $E(\underline{\epsilon}) = \underline{0}$ & $V(\underline{\epsilon}) = \sigma^2 I_n$

we further assume $\underline{\epsilon} \sim N_n(\underline{0}, \sigma^2 I_n)$

$\underline{Y} \sim N_n(X\underline{\beta}, \sigma^2 I_n)$

Theorem: If $\underline{Y} \sim N_n(X\underline{\beta}, \sigma^2 I_n)$, where X is $n \times p$ of rank p, then

ⓐ $\hat{\beta} \sim N_p(\beta; \sigma^2 (X'X)^{-1})$

ⓑ $(\hat{\underline{\beta}} - \underline{\beta})' X'X (\hat{\underline{\beta}} - \underline{\beta}) / \sigma^2 \sim \chi^2_p$

ⓒ $\hat{\beta}$ is independent of s^2

ⓓ $RSS/\sigma^2 = (n-p)s^2/\sigma^2 \sim \chi^2_{n-p}$

Proof: ⓐ $\hat{\beta} = (X'X)^{-1} X'Y = C\underline{Y}$, C is $p \times n$ matrix

rank(C) = rank$(X') = p$

using Theorem 1^*, $\hat{\underline{\beta}}$ has normal distⁿ &

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

$V(\hat{\beta}) = \sigma^2 (X'X)^{-1}$

So, we have

$$\hat{\underline{\beta}} \sim N_p(\underline{\beta}, \sigma^2 (X'X)^{-1})$$

[Theorem: 1^* $\underline{Y} \sim N_n(\underline{\mu}, \Sigma)$ $C_{m \times n}$ of rank m $d_{m \times 1}$, then $C\underline{X} + \underline{d} \sim N_m(C\underline{\mu} + d_1, C\Sigma C')$.]

ⓑ
$$\frac{(\hat{\underline{\beta}} - \underline{\beta})' (X'X) (\hat{\underline{\beta}} - \underline{\beta})}{\sigma^2}$$
$$= (\hat{\underline{\beta}} - \underline{\beta})' \left(\text{Var}(\hat{\underline{\beta}})\right)^{-1} (\hat{\underline{\beta}} - \underline{\beta}) \quad —(*)$$

[Theorem 2^*: $\underline{Y} \sim N_n(\mu, \Sigma)$ Σ is p.d. Then $Q = (\underline{Y} - \underline{\mu})' \Sigma^{-1} (\underline{Y} - \underline{\mu}) \sim \chi^2_n$]

Then using Th 2^* & ⓐ, we have

$$(*) \sim \chi^2_p.$$

(c) $\mathrm{Cov}(\hat{\underline{\beta}}, \underline{Y} - X\hat{\underline{\beta}})$

$= \mathrm{Cov}\left((X'X)^{-1}X'\underline{Y}, (I-P)\underline{Y}\right)$

$= (X'X)^{-1}X' \,\mathrm{Cov}(\underline{Y})\, (I-P)$

$= \sigma^2 (X'X)^{-1} X' (I-P)$

$= 0 \quad (PX = X)$

Theorem 3*:

$\underline{Y} \sim N_n(\underline{\mu}, \Sigma)$

Define $\underline{U} = A\underline{Y}$, $\underline{V} = B\underline{Y}$

Then $\underline{U}$ & $\underline{V}$ are independent

Iff $\mathrm{Cov}(\underline{U}, \underline{V}) = A\Sigma B' = \underline{0}$

So $\hat{\underline{\beta}}$ & $\underline{Y} - X\hat{\underline{\beta}}$ are independent.

$\Rightarrow$ $\hat{\underline{\beta}}$ & $(\underline{Y} - X\hat{\underline{\beta}})'(\underline{Y} - X\hat{\underline{\beta}})$ are independent.

$\Rightarrow$ $\hat{\underline{\beta}}$ & s^2 are independent.

(d) $\underline{Y}'(I-P)\underline{Y}$

$= (\underline{Y} - X\underline{\beta})'(I-P)(\underline{Y} - X\underline{\beta})$

[as $PX = X$]

$= \epsilon'(I-P)\epsilon$

Theorem 4*:

$\underline{Y} \sim N_n(0, \sigma^2 I_n)$

& Let A be symmetric

Then $\frac{1}{\sigma^2}\underline{Y}'A\underline{Y} \sim \chi^2_r$

iff A is idempotent of rank r.

Where $(I-P)$ is symmetric

& idempotent of rank $(n-p)$

Since, $\underline{\epsilon} \sim N_n(0, \sigma^2 I_n) \Rightarrow \frac{\underline{\epsilon}'(I-P)\underline{\epsilon}}{\sigma^2} \sim \chi^2_{n-p}$

Write $Q_1 = (\underline{Y} - X\underline{\beta})'(\underline{Y} - X\underline{\beta})$

$\frac{Q_1}{\sigma^2} \sim \chi^2_n$

$Q_1 = (\underline{Y} - X\hat{\underline{\beta}})'(\underline{Y} - X\hat{\underline{\beta}}) + (\hat{\underline{\beta}} - \underline{\beta})' X'X(\hat{\underline{\beta}} - \underline{\beta})$

$= Q + Q_2$

Result:

$$Q_i \sim \chi^2_{r_i}, \quad i = 1, 2, \ldots$$

& $Q = Q_1 - Q_2$ is independent of Q_2. Then

$$Q \sim \chi^2_r, \quad r = r_1 - r_2$$

Orthogonal columns of the Regression matrix:

Full rank model: $E(\underline{Y}) = X\underline{\beta}$

Write $X = (\underline{x}^{(0)}, \underline{x}^{(1)}, \ldots, \underline{x}^{(p-1)})$

- columns are mutually orthogonal.

$$\hat{\beta} = (X'X)^{-1} X'\underline{Y}; \quad \begin{bmatrix} \underline{x}^{(0)\prime}\underline{x}^{(0)} & & & 0 \\ & \underline{x}^{(1)\prime}\underline{x}^{(1)} & & \\ & & \ddots & \\ 0 & & & \underline{x}^{(p-1)\prime}\underline{x}^{(p-1)} \end{bmatrix}^{-1} = (X'X)^{-1}$$

$$X'\underline{Y} = \begin{bmatrix} \underline{x}^{(0)\prime}\underline{Y} \\ \underline{x}^{(1)\prime}\underline{Y} \\ \vdots \\ \underline{x}^{(p-1)\prime}\underline{Y} \end{bmatrix}$$

$$\Rightarrow \underset{\sim}{\hat{\beta}} = \begin{bmatrix} (\underline{x}^{(0)\prime}\underline{x}^{(0)})^{-1}\underline{x}^{(0)\prime}\underline{Y} \\ \vdots \\ (\underline{x}^{(p-1)\prime}\underline{x}^{(p-1)})^{-1}\underline{x}^{(p-1)\prime}\underline{Y} \end{bmatrix}, \quad \hat{\beta}_j = (\underline{x}^{(j)\prime}\underline{x}^{(j)})^{-1}\underline{x}^{(j)\prime}\underline{Y}$$

So, $\hat{\beta}_j$ is the LSE of β_j for the model $E(\underline{Y}) = \underline{x}^{(j)}\beta_j$

- LSE of β_j is unchanged if any of the other β_K $(K \neq j)$ is put equal to zero.

$$RSS = \underline{y}'\underline{y} - \hat{\underline{\beta}}X'\underline{y} = \underline{y}'\underline{y} - \sum_{j=0}^{p-1} \hat{\beta}_j \underline{x}^{(j)'}\underline{y}$$

$$= \underline{y}'\underline{y} - \sum_{j=0}^{p-1} \hat{\beta}_j^2 \left(\underline{x}^{(j)'}\underline{x}^{(j)}\right)$$

If we put $\beta_K = 0$, the change in RSS is just addition of the term $\hat{\beta}_K \underline{x}^{(k)'}\underline{y}$.

So the updated RSS

$$\underline{y}'\underline{y} - \sum_{\substack{j=0 \\ j \neq K}}^{p-1} \hat{\beta}_j \, x^{(j)'}\underline{y}$$

$$E(\underline{y}) = X\underline{\beta}, \quad Var(\underline{y}) = \sigma^2 I$$

$$E(\underline{y}) = X\underline{\beta} + Z\underline{\gamma}$$

$$X_{n \times p} \qquad Z_{n \times t}$$

$$\hat{\beta} = (X'X)^{-1}X'\underline{y} \qquad = (X \quad Z)\begin{pmatrix} \underline{\beta} \\ \underline{\gamma} \end{pmatrix} = W\underline{\delta}$$

Introducing Further Explanatory Variable :

Suppose that we already fitted the regression model $E(Y) = X\underline{\beta}$, $V(\underline{y}) = \sigma^2 I_n$

We decide to ~~all~~ take additional covariates in the model. Enlarged model , $G; E(\underline{Y}) = X\underline{\beta} + Z\underline{\gamma}$

$$= (X \; Z)\begin{pmatrix}\underline{\beta} \\ \underline{\gamma}\end{pmatrix}$$

$$= W\underline{\delta}$$

X: $n\times p$ of rank p

Z: $n\times t$ of rank t.

The columns of Z are LIN of the columns of X.

$\Rightarrow$ W is $n\times (p+t)$ of rank $(p+t)$

To find LSE $\hat{\delta}_G$ of $\underline{\delta}$

① Directly find $\hat{\underline{\delta}}_G = (W'W)^{-1} W'y$

$Var(\hat{\delta}_G) = (W'W)^{-1}\sigma^2$

② We can use all the calculation already done and reduce the complication.

Lemma: Write $R = I_n - P = I_n - X(X'X)^{-1}X'$, then $Z'RZ$ is p.d.

Pf: Let, $Z'RZ\underline{a} = \underline{0}$ $(\underline{a} = \underline{0})$

Then, $\underline{a}'Z'R'\underbrace{RZ\underline{a}}_{a} = \underline{a}'Z'RZ\underline{a} = 0$ [R: Projection]

$RZ\underline{a} = \underline{0} \Rightarrow (I-P)Z\underline{a} = 0$

$\Rightarrow \quad Z\underline{a} = X(X'X)^{-1} X'Z\underline{a} = X\underline{b}$, say

$\Rightarrow \quad \underline{a} = \underline{0}$ as the column of Z are linearly independent of the column of X.

Because $Z'RZ\underline{a} = 0 \Rightarrow \underline{a} = \underline{0}$

$Z'RZ$ has linearly independent column & is therefore PD.

$$\underline{a}'Z'RZ\underline{a} = (RZ\underline{a})'(RZ\underline{a}) \geq 0$$

Theorem:

Let, $R_G = I_n - W(W'W)^{-1}W'$, $L = (X'X)^{-1}X'Z$

$M = (Z'RZ)^{-1}$ and $\hat{\underline{\delta}}_G = \begin{pmatrix} \hat{\underline{\beta}}_G \\ \hat{\underline{\gamma}}_G \end{pmatrix}$

Then (a) $\hat{\underline{\gamma}}_G = (Z'RZ)^{-1} Z'R\underline{Y}$

(b) $\hat{\underline{\beta}}_G = (X'X)^{-1}X'(\underline{Y} - Z\hat{\underline{\gamma}}_G)$

$= \hat{\underline{\beta}} - L\hat{\underline{\gamma}}_G$

(c) $\underline{Y}'R_G\underline{Y} = (\underline{Y} - Z\hat{\underline{\gamma}}_G)R(\underline{Y} - Z\hat{\underline{\gamma}}_G)$

$= \underline{Y}'R\underline{Y} - \hat{\underline{\gamma}}_G Z'R\underline{Y}$

(d) $var(\hat{\underline{\delta}}_G) = \sigma^2 \begin{pmatrix} (X'X)^{-1} + LML' & -LM \\ -ML' & M \end{pmatrix}$

Proof:

$$\mathcal{C}(PZ) \subset \mathcal{C}(X)$$

$$X\underline{\beta} + Z\underline{\gamma} = X\underline{\beta} + PZ\underline{\gamma} + (I-P)Z\underline{\gamma}$$

$$= X\underline{\alpha} + RZ\underline{\gamma}$$

$$\underline{\alpha} = \beta + (X'X)^{-1}X'Z\underline{\gamma} = \underline{\beta} + L\underline{\gamma}$$

$$(X \quad RZ)\begin{pmatrix}\underline{\alpha} \\ \underline{\gamma}\end{pmatrix} = V\underline{\lambda}$$

$$\mathcal{C}(X) \perp \mathcal{C}(RZ)$$

So, $\text{rank}(RZ) = \text{rank}(Z'R'RZ) = \text{rank}(Z'RZ)$

$$= t$$

So V is a full rank matrix $\quad [r(A) = r(A'A) = r(AA')]$

$$E(\underline{y}) = V\underline{\lambda}, \quad \underline{\lambda} = \begin{pmatrix}\underline{\alpha} \\ \underline{\gamma}\end{pmatrix}$$

The LSE of λ

$$\tilde{\hat{\lambda}} = (V'V)^{-1}V'\underline{y}$$

$$= \begin{pmatrix} X'X & X'RZ \\ Z'R'X & Z'R'RZ \end{pmatrix}^{-1} \begin{pmatrix} X' \\ Z'R' \end{pmatrix} \underline{y}$$

$$\left[\begin{aligned} &X'RZ \\ &= X'(I-P)Z \\ &= X'Z - X'Z \\ &= 0 \end{aligned}\right].$$

$$= \begin{pmatrix} X'X & 0 \\ 0 & X'RZ \end{pmatrix}^{-1} \begin{pmatrix} X'\underline{y} \\ Z'R\underline{y} \end{pmatrix}$$

$$= \begin{pmatrix} (X'X)^{-1}X'\underline{y} \\ (Z'RZ)^{-1}Z'R\underline{y} \end{pmatrix} = \begin{pmatrix} \tilde{\alpha} \\ \tilde{\gamma} \end{pmatrix}$$

The relationship between $(\underline{\beta}, \underline{\gamma})$ to $(\underline{\alpha}, \underline{\gamma})$ is 1-1. So that the relationship exists between their LSEs.

$$\hat{\underline{\gamma}}_G = (Z'RZ)^{-1} Z'R\underline{Y} \longrightarrow (a)$$

(b) $$\hat{\underline{\beta}}_G = \tilde{\underline{\alpha}} - L\tilde{\underline{\gamma}} = \hat{\underline{\beta}} - L\hat{\underline{\gamma}}_G$$

$$= (X'X)^{-1}X'(\underline{Y} - Z\hat{\underline{\gamma}}_G)$$

(c) $$R_G \underline{Y} = (I - P_W)\underline{Y} \qquad [P_W\underline{Y} = W\hat{\underline{\delta}}_G = \hat{\underline{Y}}]$$

$$= \cancel{(X\beta + X\beta_G)}\ \underline{Y} - W\hat{\underline{\delta}}_G$$

$$= \underline{Y} - X\hat{\underline{\beta}}_G - Z\hat{\underline{\gamma}}_G \qquad [W\hat{\delta}_G = X\hat{\beta}_G + Z\hat{\gamma}_G]$$

$$= \underline{Y} - X(X'X)^{-1}X'(\underline{Y} - Z\hat{\underline{\gamma}}_G) - Z\hat{\underline{\gamma}}_G$$

$$= (I - X(X'X)^{-1}X')(\underline{Y} - Z\hat{\underline{\gamma}}_G)$$

$$= R(\underline{Y} - Z\hat{\underline{\gamma}}_G)$$

$$\underline{Y}'R_G\underline{Y} = \underline{Y}'R_G'R_G\underline{Y}$$

$$= (\underline{Y} - Z\hat{\underline{\gamma}}_G)'R'R(\underline{Y} - Z\hat{\underline{\gamma}}_G)$$

$$= (\underline{Y} - Z\hat{\underline{\gamma}}_G)'R(\underline{Y} - Z\hat{\underline{\gamma}}_G)$$

$$\underline{Y}'R_G\underline{Y} = \underline{Y}'R\underline{Y} - 2\hat{\underline{\gamma}}_G'Z'R\underline{Y} + \hat{\underline{\gamma}}_G'Z'RZ\hat{\underline{\gamma}}_G$$

$$= \underline{Y}'R\underline{Y} - \hat{\underline{\gamma}}_G'Z'R\underline{Y} \quad \text{, using } \hat{\underline{\gamma}}_G = (Z'RZ)^{-1}Z'R\underline{Y}$$

(d) $Var(\hat{\underline{\gamma}}_G) = (Z'RZ)^{-1} Z'R \, Var(\underline{y}) \, RZ \, (Z'RZ)^{-1}$
$= \sigma^2 (Z'RZ)^{-1} (Z'RZ) (Z'RZ)^{-1}$
$= \sigma^2 (Z'RZ)^{-1} = \sigma^2 M$

$Cov(\hat{\underline{\beta}}, \hat{\underline{\gamma}}_G) = Cov((X'X)^{-1} X'\underline{Y}, (Z'RZ)^{-1} Z'R\underline{Y})$

$[Cov(AX, BY) = A \, Cov(X,Y) B']$ $= 0$

$(Z'R = 0)$

$Cov(\hat{\underline{\beta}}_G, \hat{\underline{\gamma}}_G) = Cov(\hat{\underline{\beta}} - L\hat{\gamma}_G, \hat{\gamma}_G)$

$= Cov(\hat{\underline{\beta}}, \hat{\underline{\gamma}}_G) - L \, Var(\hat{\underline{\gamma}}_G)$

$= -\sigma^2 LM$

$Var(\hat{\underline{\beta}}_G) = Var(\hat{\underline{\beta}} - L\hat{\gamma}_G) = Var(\hat{\underline{\beta}}) + Var(L\hat{\gamma}_G)$

$= \sigma^2 ((X'X)^{-1} + LML')$

<u>One extra Variable</u>.

$\underline{x}^{(j)}$, jth column of X, $j = 0, 1, \dots, p-1$

$E(\underline{Y}) = (\underline{x}^{(0)}, \underline{x}^{(1)}, \dots, \underline{x}^{(p-1)}) \underline{\beta}$

$= \underline{x}^{(0)} \beta_0 + \underline{x}^{(1)} \beta_1 + \cdots + \underline{x}^{(p-1)} \beta_{p-1}$

want to include x_p, another variable

So, $Z\underline{\gamma} = \underline{x}^{(p)} \beta_p$

So the LSE of the p-variate model

$Z'RZ = \underline{x}^{(p)'} R \underline{x}^{(p)}$

$\hat{\beta}_{p,G} = \dfrac{\underline{x}^{(p)'} R\underline{Y}}{\underline{x}^{(p)'} R\underline{x}^{(p)}}$, $\hat{\underline{\gamma}}_G = (Z'RZ)^{-1} Z'R\underline{Y}$

$$\underline{\hat{\beta}}_G = \left(\hat{\beta}_{0,G}, \hat{\beta}_{1,G}, \ldots \hat{\beta}_{p-1,G}\right)'$$

$$= \underline{\hat{\beta}} - (X'X)^{-1} X' \underline{x}^{(p)} \hat{\beta}_{p,G}$$

$$\underline{Y}' R_G \underline{Y} = \underline{Y}' R \underline{Y} - \hat{\beta}_{p,G}\, x^{(p)\prime} R \underline{Y}$$

Estimation with Linear Restriction:

Effect on LSE where some hypothesized constraints are present.

Example:-

α, β, γ - angle

Y_1, Y_2, Y_3 - unbiased measurements of these angle (in radian)

$\alpha + \beta + \gamma = \pi$

To find the LSEs.

(1) use the constraints

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 - \pi \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{pmatrix}$$

↳ adhoc and not every with more complicated model.

(2) Alternative & general approach

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{pmatrix}$$

Minimize $(Y_1 - \alpha)^2 + (Y_2 - \beta)^2 + (Y_3 - \gamma)^2$

subject to the constraint $\alpha + \beta + \gamma = \pi$ using Lagrange's multipliers.

Method of Lagrange's Multipliers

Model : $\underline{Y} = X\underline{\beta} + \underline{\epsilon}$, X is $n\times p$ of full rank p.

To find the minimum of $\underline{\epsilon}'\underline{\epsilon}$ subject to the linear restriction $A\underline{\beta} = \underline{c}$

- A is known $q\times p$ matrix of rank q $(q<p)$
- $\underline{c}$ is a known $q\times 1$ vector

Use Lagrange's multiplier, one for each linear restriction $\underline{a}_i'\underline{\beta} = c_i$, $i = 1, 2, \ldots, q$ (1)

$\underline{a}_i' \longrightarrow$ ith row of A.

$$\sum_{i=1}^{q} \lambda_i \left(\underline{a}_i'\underline{\beta} - c_i\right) = \underline{\lambda}'\left(A\underline{\beta} - \underline{c}\right)$$

$$\underline{\lambda} = (\lambda_1 \ldots, \lambda_q)'$$

$$= \left(\underline{\beta}'A' - \underline{c}'\right)\underline{\lambda}$$

Write $Q = \underline{\epsilon}'\underline{\epsilon} + (\underline{\beta}'A' - \underline{c}')\underline{\lambda}$ (2)

Solve, $A\underline{\beta} = \underline{c}$ & $\dfrac{\partial Q}{\partial \underline{\beta}} = \underline{0}$ $\longrightarrow$ (A)

$$\frac{\partial Q}{\partial \underline{\beta}} = -2X'\underline{Y} + 2X'X\underline{\beta} + A'\underline{\lambda} = 0$$

Let, $\hat{\beta}_H$ & $\hat{\gamma}_H$ are the sol^n of (A)

$$\hat{\beta}_H = (X'X)^{-1}X'Y - \frac{1}{2}(X'X)^{-1}A'\hat{\lambda}_H$$

$$= \hat{\beta} - \frac{1}{2}(X'X)^{-1}A'\hat{\lambda}_H$$

from $A\underline{\beta} = \underline{c}$

$$\underline{c} = A\hat{\underline{\beta}}_H = A\hat{\underline{\beta}} - \frac{1}{2} A(X'X)^{-1} A'\hat{\lambda}_H$$

$X \longrightarrow$ full rank, $(X'X)$ is p.d., so is $(X'X)^{-1}$

A is $q \times p$ of rank q, so, $A(X'X)^{-1}A'$ is p.d.

& so n-s.

$$\frac{1}{2}\hat{\lambda}_H = \left[A(X'X)^{-1}A'\right]^{-1}(A\hat{\underline{\beta}} - \underline{c})$$

$$\hat{\underline{\beta}}_H = \hat{\beta} - (X'X)^{-1}A'\left[A(X'X)^{-1}A'\right]^{-1}(A\hat{\underline{\beta}} - \underline{c})$$

Q. $p = 5$

$\beta_1 = \beta_3$, $\beta_2 = \beta_4$

$A = ?$, $c = ?$

$$\underset{2\times 5}{A} = \begin{bmatrix} 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}, \quad \underline{c} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\hat{\underline{\beta}}_H = \hat{\underline{\beta}} + (X'X)^{-1} A' [A (X'X)^{-1} A'] (\underline{c} - A\hat{\underline{\beta}})$$

$$A\underline{\beta} = \underline{c}$$

$$\frac{1}{2} \hat{\underline{\lambda}}_H = [A(X'X)^{-1} A']^{-1} (A\hat{\underline{\beta}} - \underline{c})$$

We want to $\hat{\underline{\beta}}_H$ minimize $\underline{\epsilon}'\underline{\epsilon}$ under the constraint $A\underline{\beta} = \underline{c}$

$$\|X(\hat{\underline{\beta}} - \underline{\beta})\|^2 = (\hat{\underline{\beta}} - \underline{\beta})' X'X (\hat{\underline{\beta}} - \underline{\beta})$$

$$= (\hat{\underline{\beta}} - \hat{\underline{\beta}}_H + \hat{\underline{\beta}}_H - \underline{\beta})' X'X (\hat{\underline{\beta}} - \hat{\underline{\beta}}_H + \hat{\underline{\beta}}_H - \underline{\beta})$$

$$= (\hat{\underline{\beta}} - \hat{\underline{\beta}}_H)' X'X (\hat{\underline{\beta}} - \hat{\underline{\beta}}_H) + (\hat{\underline{\beta}}_H - \underline{\beta})' X'X (\hat{\underline{\beta}}_H - \underline{\beta})$$

$$+ 2 (\hat{\underline{\beta}} - \hat{\underline{\beta}}_H) X'X (\hat{\underline{\beta}}_H - \underline{\beta})$$

$$= (A) + (B) + (C).$$

$$(C) \Rightarrow 2 (\hat{\underline{\beta}} - \hat{\underline{\beta}}_H) X'X (\hat{\underline{\beta}}_H - \underline{\beta})$$

$$= \hat{\underline{\lambda}}_H' A (X'X)^{-1} X'X (\hat{\underline{\beta}}_H - \hat{\underline{\beta}})$$

$$= \hat{\underline{\lambda}}_H A (\hat{\underline{\beta}}_H - \underline{\beta})$$

$$= \hat{\underline{\lambda}}_H (\underline{c} - \underline{c}) = \underline{0}$$

$$\underline{\epsilon}'\underline{\epsilon} = \|Y - X\hat{\underline{\beta}}\|^2 + \|X(\hat{\underline{\beta}} - \underline{\beta})\|^2$$

$$= \|Y - X\hat{\underline{\beta}}\|^2 + \|X(\hat{\underline{\beta}} - \hat{\underline{\beta}}_H)\|^2 + \|X(\hat{\underline{\beta}}_H - \underline{\beta})\|^2$$

Then $\underline{\epsilon}'\underline{\epsilon}$ is minimum when $\|X(\hat{\underline{\beta}}_H - \underline{\beta})\|^2$.

$$\Rightarrow X(\hat{\underline{\beta}}_H - \underline{\beta}) = 0 \Rightarrow \underline{\beta} = \hat{\underline{\beta}}_H$$