7/11/2024

$\beta^{LS} = \underline{(X^TX)^- X^Ty}$ is the least-square estimators of β.

For a model, $\underline{\beta_1}$ is estimable.

$l^T\beta$, l^TGy where G is a LS g-inverse of X.

$\beta_1 = e_1^T\beta$, e_1^TGy is the BLUE of β_1

$$e_1^T \underbrace{(X^TX)^- X^T}_{\text{LS g-inverse of } X} y \text{ is the BLUE of } \beta_1.$$

Also, this is the least square estimator of β_1.

- In general, least square estimators are not unique.

  Try to find an example.

$$\mathrm{Var}(\beta_1^{LS}) = \sigma^2 \left(e_1^T(X^TX)^-X^T\right)\left(e_1^T(X^TX)^-X^T\right)^T$$
$$= \sigma^2\left(e_1^T(X^TX)^-X^TX(X^TX)^-e_1\right)$$

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

$$\mathrm{RSS} \overset{\text{defn}}{=} (y - X\beta^{LS})^T(y - X\beta^{LS})$$

(Residual sum of squares)

$$= (y - P_Xy)^T(y - P_Xy)$$

$$E(\mathrm{RSS}) = E\left((y - P_Xy)^T(y - P_Xy)\right)$$
$$= E\left(y^Ty - y^TP_Xy - y^TP_X^Ty\right.$$

$$+ y^T (P_x^T) P_x y)$$

$$= E(y^T y - y^T P_x y - \cancel{y^T P_x y} + \cancel{y^T P_x y})$$

$$= E(y^T y - y^T P_x y)$$

$$= E(y^T (I_n - P_x) y)$$

$$= E((X\beta + \epsilon)^T (I_n - P_x)(X\beta + \epsilon)) \qquad y = X\beta + \epsilon$$

$$= E(\beta^T X^T (I_n - P_x) X\beta + \beta^T X^T (I_n - P_x)\epsilon + \epsilon^T (I_n - P_x) X\beta + \epsilon^T (I_n - P_x)\epsilon)$$

$$= E(\beta^T X^T (X - P_x X)\beta + 2\epsilon^T (X - P_x X)\beta + \epsilon^T (I_n - P_x)\epsilon)$$

$$= E(\epsilon^T \underbrace{(I_n - P_x)}_{Q} \epsilon)$$

$$= E(\sum a_{ij} \epsilon_i \epsilon_j)$$

$$= \sum_{i=1}^{n} a_{ii} \sigma^2$$

$$= \sigma^2 \sum_{i=1}^{n} a_{ii}$$

$$= \sigma^2 \, \mathrm{trace}(Q)$$

$$= \sigma^2 \, \mathrm{trace}(I_n - P_x)$$

$$= \sigma^2 (\mathrm{trace}(I_n) - \mathrm{trace}(P_x))$$

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

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

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

$\rho(P_x) \le \rho(X)$

Further $P_x X = X$

$\rho(X) \le \rho(P_x)$

$E(\epsilon_i \epsilon_j) = 0$
$\forall i \ne j$

$E(\epsilon_i^2) = \sigma^2$

$y = X\beta + \epsilon$

$E(\epsilon_i) = 0 \ \forall i$

$V(\epsilon_i) = \sigma^2 \ \forall i$

$Cov(\epsilon_i, \epsilon_j) = 0$
$\forall i \ne j$

$Var(\epsilon_i)$

$= \sigma^2 (n - \rho(X))$

$$E(RSS) = \sigma^2 (n - \rho(X))$$

$$E\left(\frac{RSS}{n-\rho(X)}\right) = \sigma^2 \quad (\text{assuming } \rho(X) < n)$$

For σ^2, $\left(RSS / n-\rho(X)\right)$ is an unbiased estimator.

$= E(\epsilon_i - E(\epsilon_i))$

$= E(\epsilon_i^2) = \sigma^2$

$Cov(\epsilon_i, \epsilon_j)$

$= E((\epsilon_i - E(\epsilon_i))(\epsilon_j - E(\epsilon_j)))$

$= E(\epsilon_i \epsilon_j) = 0$

Estimation under some restrictions

$$\underset{\text{known}}{y = X}\,\underset{\text{unknown}}{\beta} + \epsilon \qquad \text{where } E(\epsilon_i) = 0 \ \forall i$$

$$Var(\epsilon_i) = \sigma^2 \ \forall i$$

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

Restriction: $\underset{\text{known}}{L}\beta = \underset{\text{known}}{z}$

$$y_i = \mu + \alpha_i + \epsilon_i \qquad i = 1(1)n$$

H_0: $\alpha_i = \alpha_j \ \forall i \neq j$ $\qquad H_1$: H_0 is not true.

$$y = X\beta + \epsilon$$

$$X = \begin{bmatrix} 1 & 1 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 1 & 0 & \cdots & 0 \\ \vdots \\ 1 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \qquad \beta = \begin{bmatrix} \mu \\ \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}$$

$$L\beta = z$$

$$\begin{bmatrix} 0 & 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & 0 & -1 & \cdots & 0 \\ \vdots \\ 0 & 1 & 0 & 0 & \cdots & -1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Assumptions

(i) $L\beta = z$, $z \in \mathcal{C}(L)$

(ii) $\mathcal{R}(L) \subseteq \mathcal{R}(X)$

$$\beta^{0} = \beta^{LS} - (X^T X)^{-} L^T \left(L (X^T X)^{-} L^T \right)^{-} (L\beta^{LS} - z)$$

will minimize $(y - X\beta)^T (y - X\beta)$ under the restriction $L\beta = z$