EE559

HOMEWORK-2

1. To proove Gauss-Markov Theorem,

Assume a linear model;

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

$y \in R^{n\times 1}$
$X \in R^{n\times p}$
$\beta \in R^{p\times 1}$
$\varepsilon \in n\times 1$ vectors of errors

$E[\varepsilon] = 0$
$var(\varepsilon) = \sigma^2$

↳ least squares estimator $\hat{\beta}$ is given by,

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

Unbiasedness of $\hat{\beta}$ $\quad [E[\hat{\beta}]] = \beta$.

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

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

$\Rightarrow \hat{\beta}$ is unbiased.

↳ Any other linear Unbiased Estimator.

if $c^T y$ is any other linear Unbiased estimator, of $a^T\beta$

$$E[c^T y] = a^T\beta$$

$y = X\beta + \varepsilon$

$\therefore E[c^T y] = c^T E[X\beta + \varepsilon] = c^T X\beta$

Given it is unbiased, comparing $c^T X\beta = a^T\beta$

we get, $c^T X = a^T$

Since $c^T X = a^T$ & any other estimator, it could potentially have additional components & c^T can be written as, $c^T = a^T \hat{\beta} + b$

$$c^T y = a^T \hat{\beta} + b^T y$$

i.e c^T includes a component that projects y onto X to capture $a^T \beta \to (a^T (X^T X)^{-1} X^T)$ which multiplied by y gives estimate $a^T \hat{\beta}$, where $\hat{\beta} = (X^T X^{-1}) X^T y$.

plus another component b^T.

where, $b^T X = 0$ {null space of X}

$$\therefore E[c^T y] = a^T E[\hat{\beta}] + b^T E[y] = a^T \beta$$

Variance of $\hat{\beta}$

$$\text{Var}(\hat{\beta}) = \text{Var}((X^T X)^{-1} X^T y) = (X^T X)^{-1} X^T \underbrace{\text{Var}(y)}_{= \sigma^2} X (X^T X)^{-1}$$

$$\because \text{Var}[bX] = b \,\text{Var}[X]\, b^T$$

$\because \text{Var}(X\beta + \varepsilon) = \text{Var}(\varepsilon) = \sigma$

$$\therefore \text{var}(\hat{\beta}) = \sigma^2 (X^T X)^{-1} \underbrace{X^T X (X^T X)^{-1}}_{I}$$

$$\Rightarrow \text{var}(\hat{\beta}) = \sigma^2 (X^T X)^{-1} \Rightarrow \text{var}(a^T \hat{\beta}) = \sigma^2 a^T (X^T X^{-1}) a$$

Variance of $c^T y$

$$\begin{aligned}
\text{Var}(c^T y) &= \text{Var}(a^T \hat{\beta} + b^T y) \\
&= \text{Var}(a^T \hat{\beta}) + \text{Var}(b^T y) \\
&= a^T \text{var}(\hat{\beta}) a + \sigma^2 b^T b \\
&= \underline{\sigma^2 a^T (X^T X)^{-1} a} + \sigma^2 b^T b
\end{aligned}$$

also, $b^T b \geq 0$

$\therefore \operatorname{Var}(c^T y) \geq \sigma^2 a^T (X^T X)^{-1} a$.

$$\boxed{\text{i.e. } \operatorname{Var}(c^T y) \geq \operatorname{Var}(a^T \hat{\beta})}$$

Another method. [From discussion class reference]

Let $\tilde{\beta} = Cy$ be another estimator β

where, $C = (X^T X)^{-1} X^T + D$

where, $D =$ non-zero matrix.

$$\begin{aligned}
\therefore E[\tilde{\beta}] &= E[Cy] \\
&= E[((X^T X)^{-1} X^T + D)(X\beta + \varepsilon)] \\
&= ((X^T X)^{-1} X^T + D) X\beta + ((X^T X)^{-1} X^T + D) \underbrace{E[\varepsilon]}_{0} \\
&= ((X^T X)^{-1} X^T + D) X\beta \\
&= \underbrace{(X^T X)^{-1} X^T X}_{I} \beta + DX\beta \\
&= (I + DX)\beta
\end{aligned}$$

$\Rightarrow E[\tilde{\beta}] = \beta$ when $DX = 0$

so it is unbiased.

$$\begin{aligned}
\operatorname{Var}(\tilde{\beta}) &= \operatorname{Var}(Cy) \\
&= C \operatorname{Var}(y) C^T \\
&= \sigma^2 C C^T \\
&\sigma^2 ((X^T X)^{-1} X^T + D)(X (X^T X)^{-1} + D^T) \\
&= \sigma^2 ((X^T X^{-1}) X^T X (X^T X)^{-1} + (X^T X)^{-1} X^T D^T + DX (X^T X)^{-1} + DD^T \\
&= \sigma^2 (X^T X)^{-1} + \sigma^2 (X^T X)^{-1} (DX)^T + \sigma^2 DX (X^T X)^{-1} + \sigma^2 DD^T \\
&= \underbrace{\sigma^2 (X^T X)^{-1}} + \sigma^2 DD^T \\
&= \operatorname{Var}(\hat{\beta}) + \sigma^2 DD^T
\end{aligned}$$

$$\boxed{\Rightarrow \operatorname{Var}(\tilde{\beta}) \geq \operatorname{Var}(\hat{\beta})}$$

2. Linear Regression with Orthogonal Design.

Soln:

$x_0, ----x_p$ of X are orthogonal.

$\Rightarrow$ dot product $= 0$

i.e for any two cols, $x_0 ---- x_p$; x_i & x_j $(i \neq j)$
we have

$$x_i^T x_j = 0$$

$\therefore$ In X matrix, all the elements of X will be zero, <u>except diagonal</u> elements

Estimating $\hat{\beta}$.

we have,

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

now since X is orthogonal matrix,

$$\begin{bmatrix} x_0 & x_1 & --- & x_p \\ | & | & & | \\ | & | & & | \end{bmatrix}^T \begin{bmatrix} x_0 & x_1 & --- & x_p \\ | & | & & | \\ | & | & & | \end{bmatrix} = X^T X$$

$$X^T X = \begin{bmatrix} x_0^T x_0 & 0 & 0 & --- \\ 0 & x_1^T x_1 & & \\ | & & \ddots & \\ | & & & x_p^T x_p \end{bmatrix} = \mathrm{diag}(x_0^T x_0, x_1^T x_1, --- x_p^T x_p) = D$$

$$\therefore (X^T X)^{-1} = D^{-1} = \begin{bmatrix} 1/x_0^T x_0 & & 0 \\ & 1/x_1^T x_1 & \\ & & \ddots \\ 0 & & x_p^T x_p \end{bmatrix}$$

$$\therefore \hat{\beta} = D^{-1} X^T y$$

$$\Rightarrow \hat{\beta} = \begin{bmatrix} \frac{1}{x_0^T x_0} & \cdots & 0 \\ \vdots & \frac{1}{x_1^T x_0} & \vdots \\ 0 & \cdots & \frac{1}{x_p^T x_0} \end{bmatrix} \begin{bmatrix} x_0 & x_1 & \cdots & x_p \\ \vdots & \vdots & & \end{bmatrix}^T y$$

$$\Rightarrow \hat{\beta} = \begin{bmatrix} \frac{x_0^T}{x_0^T x_0} y \\ \frac{x_1^T}{x_1^T x_1} y \\ \vdots \\ \frac{x_p^T}{x_p^T x_p} y \end{bmatrix} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{bmatrix}$$

$$\boxed{\Rightarrow \hat{\beta}_j = \frac{x_j^T}{x_j^T x_j} y}$$

3. Minimum Norm Solution:

when $X^T X$ not invertible $\Rightarrow$ no unique solution.

$X \in R^{n \times (p+1)}$ $\quad r =$ rank of matrix

SVD of $X = U \Sigma V^T$ $\quad U \in R^{n \times (p+1)}$ $\quad U^T U = I_r$

$V \in R^{(p+1) \times r}$ $\quad V^T V = I_r$

$\Sigma \in R^{n + (p+1)}$ $\quad \Sigma = \text{diag}(\sigma_1 \cdots \sigma_r)$

(a) To show $\beta_{mins} = V \Sigma^{-1} U^T y$ is solution to normal equat

normal equation is given by,

$$X^T X \beta = X^T y$$

$\because X^T X$ is not invertible, we need to make use of SVD to show, $\beta_{mns} = V\Sigma^{-1}U^T y$ satisfies normal eq^n, let's substitute β_{mns} in normal equation,

$$X^T X \beta_{mns} = X^T y$$

we need to proove after substitution, we get L.H.S = R.H.S

$$X^T X \underbrace{V\Sigma^{-1}U^T y}_{\beta_{mns}} = X^T y$$

now substituting $X = U\Sigma V^T$,

$$(U\Sigma V^T)^T (U\Sigma V^T) V\Sigma^{-1}U^T y = (U\Sigma V^T)^T y$$

$$= (V\Sigma \underline{U^T})(\underline{U}\Sigma V^T) V\Sigma^{-1}U^T y = V\Sigma U^T y$$

we have $U^T U = I_r$

$$\therefore V\Sigma\Sigma \underline{V^T V}\Sigma^{-1}U^T y = V\Sigma U^T y \quad ; V^T V = I_r$$

$$V\Sigma \underbrace{\Sigma\Sigma^{-1}}_{I} U^T y = V\Sigma U^T y$$

$$\Rightarrow V\Sigma U^T y = V\Sigma U^T y = X^T y$$

$\therefore$ β_{mns} satisfies normal eq^n.

(b) To show: for any other solution β to normal eq^n
$\|\beta\| \geq \|\beta_{mns}\|$

Proof:

normal equation is given by,

$$X^T X\beta = X^T y \quad \text{where } \beta = \beta_{mns}$$

when $X = U\Sigma V^T$,

minimum norm solution is given by

$$\beta_{mns} = V\Sigma^{-1}U^T y$$

Suppose, β is any solution to normal equation.

$$\Rightarrow X^T X\beta = X^T y$$

taking difference between β_{mns} & β,

Let,

$$\beta - \beta_{mns} = b \quad \text{for some vector } b.$$

Using $\beta_{mns} = \beta + b$ in normal eqn,

$$X^T X(\beta_{mns} + b) = X^T y$$

$$X^T X \beta_{mns} + X^T X b = X^T y$$

$\because$ previously proved, $X^T X \beta_{mns} = X^T y$

$$X^T y + X^T X b = X^T y$$

$$\Rightarrow X^T X b = 0$$

$\Rightarrow$ <u>b</u> lies in <u>null space</u> of $X^T X$

$$\Rightarrow \beta = \beta_{mns} + b \quad \text{——(i)}$$

$\therefore$ any general solution to normal equation can be written as sum of minimum norm solution and any vector in null space of $X^T X$

$\therefore$ to show, $\|\beta\| \geq \|\beta_{mns}\|$

norm of β can be written as:

$$\|\beta\| = \|\beta_{mns} + b\|$$

$\because X^T X b = 0$ i.e b is $\perp$ β_{mns},

$$\|\beta\|^2 = \|\beta_{mns} + b\|^2 = \|\beta_{mns}\|^2 + \|b\|^2$$ by Pythagorean theorem

$$\Rightarrow \|\beta\|^2 \geq \|\beta_{mns}\|^2.$$

$$\boxed{\Rightarrow \|\beta\| \geq \|\beta_{mns}\|}$$

(c) $V\Sigma^{-1}U^T$ → to check if this the Pseudo-inverse of X.

Penrose properties for pseudo-inverse is given by:

1. $XX^{+}X = X$
2. $X^{+}XX^{+} = X^{+}$
3. $(XX^{+})^T = XX^{+}$
4. $(X^{+}X)^T = X^{+}X$

$X^{+} = V\Sigma^{-1}U^T$ {to check}.

Property 1: $XX^{+}X = X$

$$XX^{+}X = (U\Sigma V^T)(V\Sigma^{-1}U^T)(U\Sigma V^T)$$

$$= U\Sigma \underbrace{V^T V}_{I} \Sigma^{-1} \underbrace{U^T U}_{I} \Sigma V^T$$

$$= U\Sigma\Sigma^{-1}\Sigma V^T$$

$$= U\Sigma V^T = X$$

$+ \quad + \quad +$

Property 2:

$$X^{+}XX^{+} = X^{+}$$

$$\underbrace{(V\Sigma^{-1}U^{T})(U\Sigma V^{T})}_{I}(V\Sigma^{-1}U^{T})$$

$$= V\underbrace{\Sigma^{-1}\Sigma}_{I}\underbrace{V^{T}V}_{I}\Sigma^{-1}U^{T}$$

$$V\Sigma^{-1}U^{T}$$

$$= X^{+}$$

Property 3: $(XX^{+})^{T} = XX^{+}$

$$(U\Sigma\underbrace{\underbrace{V^{T}V}_{I}\Sigma^{-1}}_{I}U^{T})^{T}$$

$$=$$

$$= (UU^{T})^{T}$$

$$= UU^{T} \quad \text{—(i)}$$

$$XX^{+} = U\underbrace{\Sigma\underbrace{V^{T}V}_{I}\Sigma^{-1}}_{I}U^{T}$$

$$= UU^{T} \quad \text{—(ii)}$$

$$(i) = (ii)$$

Property 4: $(X^{+}X)^{T} = X^{+}X$

$$(X^{+}X)^{T} = ((V\underbrace{\Sigma^{-1}U^{T})(U\Sigma}_{I}V^{T}))^{T}$$

$$= (VV^{T})^{T} = VV^{T} \quad \text{—(i)}$$

$$X^{+}X = U\Sigma^{-1}U^{T}U\Sigma V^{T} = VV^{T} \quad \text{—(ii)}$$

$$\Rightarrow (i) = (ii)$$

Since all four Penrose properties are satisfied, $V\Sigma^{-1}U^{T}$ is pseudo-inverse of X.