Giang Vu - 4445745
STAT 2131
Sep 24, 2020

Homework 5

1) a) $f(B) = (Y - XB)^T (Y - XB)$

$= Y^T Y - \beta^T X^T Y - Y^T X B + B^T X^T X B$

$= Y^T Y - 2\beta^T X^T Y + B^T X^T X B$

(because $Y^T X B = (Y^T X B)^T = B^T X^T Y$ is a ~~constant~~ scalar)

To find $\hat{\beta}$ that minimizes $f(B)$, we take the derivative of $f(B)$ with respect to B and set to 0 & solve for $\hat{\beta}$

$\nabla_B f(B) = 0 - 2X^T Y + (X^T X + X^T X) B$

$= 0 - 2X^T Y + 2X^T X B$

Set $\nabla_B f(B) = 0$ and solve for $\hat{\beta}$

$\Rightarrow -2X^T Y + 2X^T X \hat{\beta} = 0$

$\Rightarrow 2X^T X \hat{\beta} = 2X^T Y$

$\Rightarrow \underbrace{(X^T X)^{-1} (X^T X)}_{I_{(p \times p)}} \hat{\beta} = (X^T X)^{-1} X^T Y$

$\Rightarrow I_{(p \times p)} \cdot \hat{\beta} = (X^T X)^{-1} X^T Y$

$\Rightarrow \hat{\beta} = \underset{B \in \mathbb{R}^p}{\operatorname{argmin}} f(B) = (X^T X)^{-1} X^T Y$

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

$\Rightarrow \hat{Y} = HY$ (using definition of H and result in a)

$\hat{\varepsilon} = Y - \hat{Y} = Y - HY$ (using result above)

$= (I - H) Y$

$H = (\underset{n \times p}{X} \underset{p \times n}{(X^T} \underset{n \times p}{X)^{-1}} \underset{p \times n}{X^T})$ has dimension $n \times n$

$\Rightarrow \hat{\varepsilon} = (I_n - H) Y = QY$

2) a) To prove $Im(H) = Im(X)$, we prove $Im(H) \subset Im(X)$ and $Im(X) \subset Im(H)$

- Proving $Im(H) \subset Im(X)$

Let $\vec{a}$ be a random vector in $Im(H)$ ($\vec{a} \in \mathbb{R}^n$)

$\Rightarrow \vec{a}$ has the form $\underset{n\times n}{H}\underset{n\times 1}{\vec{v}}$ for $\vec{v} \in \mathbb{R}^n$

$\Rightarrow \vec{a} = X(X^TX)^{-1}X^T\vec{v} = \underset{n\times p}{X}\underbrace{\left[\underset{p\times n\ n\times p}{(X^TX)^{-1}}\underset{p\times n\ n\times 1}{X^T\vec{v}}\right]}$ a (px1) vector

$\Rightarrow \vec{a} \in Im(X)$ as well

$\Rightarrow Im(H) \subset Im(X)$

- Proving $Im(X) \subset Im(H)$

Let $\vec{b}$ be a random vecter in $Im(X)$ ($\vec{b} \in \mathbb{R}^n$)

$\Rightarrow \vec{b} = \underset{n\times p}{X}\underset{p\times 1}{\vec{u}}$ for any $\vec{u} \in \mathbb{R}^p$

We have

$(X^TX)^{-1}X^T\vec{v}$ is a vector in $\mathbb{R}^p$ ($\vec{v} \in \mathbb{R}^n$)

$\Rightarrow (X^TX)^{-1}X^T\vec{v}$ can be one possible value for $\vec{u}$

$\Rightarrow \vec{b}$ then will become $X(X^TX)^{-1}X^T\vec{v}$

$\Rightarrow \vec{b} \in Im(H)$ as well $\Rightarrow Im(X) \subset Im(H)$

$\Rightarrow Im(X) = Im(H)$

b) i) $H^T = (X(X^TX)^{-1}X^T)^T = (X^T)^T((X^TX)^{-1})^TX^T$

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

$Q^T = (I_n - H)^T = I_n^T - H^T = I_n - H = Q$

(because I_n is symmetric and we have $H^T = H$ above)

$\Rightarrow$ H and Q are symmetric

ii) $H^2 = (X(X^TX)^{-1}X^T)(X(X^TX)^{-1}X^T)$

$= X(X^TX)^{-1}\left[(X^TX)(X^TX)^{-1}\right]X^T$

$= X(X^TX)^{-1} \cdot I_p \cdot X^T$

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

$Q^2 = (I_n - H)^2 = I_n^2 - 2I_nH + H^2$

$= I_n - 2H + H = I_n - H$ (using $H^2 = H$)

$\Rightarrow$ H and Q are idempotent

iii) $HQ = H(I_n - H) = HI_n - H^2 = H - H = 0$

(using $H^2 = H$ in ii))

$\Rightarrow$ H and Q are orthogonal to one another

c) i) $v \in \mathbb{R}^n \Rightarrow \underset{n\times n}{H}\underset{n\times 1}{v} = X(X^TX)^{-1}X^Tv = \underset{n\times p}{X}\underbrace{\left((X^TX)^{-1}X^Tv\right)}_{p\times 1 \text{ vector}}$

$\Rightarrow Hv \in Im(X)$ for $v \in \mathbb{R}^n$

If $u \in Im(X) \Rightarrow \underset{n \times p}{u = X} \underset{p \times 1}{a}$ for $u \in \mathbb{R}^n$

$\Rightarrow Hu = (X(X^T X)^{-1} X^T)(Xa) = X[(X^T X)^{-1}(X^T X)]a = X I_p a$

$\Rightarrow Hu = Xa = u$ (for $u \in Im(X)$)

2) c) (continued)

ii) $\|v-u\|_2^2 = \|v - Hv + Hv - u\|_2^2$

$= \|(v-Hv) + (Hv-u)\|_2^2$

$= [(v-Hv)+(Hv-u)]^T[(v-Hv)+(Hv-u)]$

$= \underset{(1)}{(v-Hv)^T(v-Hv)} + \underset{(2)}{(v-Hv)^T(Hv-u)}$

$+ \underset{(3)}{(Hv-u)^T(v-Hv)} + \underset{(4)}{(Hv-u)^T(Hv-u)}$

Let's look at $(1) = (v-Hv)^T(v-Hv) = \|v-Hv\|_2^2$

$(4) = (Hv-u)^T(Hv-u) = \|Hv-u\|_2^2$

$(2) = (v-Hv)^T(Hv-u)$

$= v^THv - v^Tu - v^TH^THv + v^TH^Tu$

$= v^THv - v^Tu - v^THv + v^Tu$

(because $H^TH = HH = H$

and $H^Tu = Hu = u$)

$\Rightarrow (2) = (v^THv - v^THv) + (v^Tu - v^Tu) = 0$

$(3) = (Hv-u)^T(v-Hv)$

$= v^TH^Tv - v^THHv - u^Tv + u^THv$

$= v^THv - v^THv - u^Tv + u^Tv$

(because $H = H^T = HH$

and $u^TH = (Hu)^T = u^T$)

$= 0$

$\Rightarrow \|v-u\|_2^2 = (1) + (2) + (3) + (4)$

$= \|v-Hv\|_2^2 + 0 + 0 + \|Hv-u\|_2^2$

$= \|v-Hv\|_2^2 + \|Hv-u\|_2^2$

which is $\geq \|v-Hv\|_2^2$

Therefore, $u = Hv = \underset{u\in Im(X)}{\text{argmin}} \|v-u\|_2^2$

Scanned with CamScanner

3) a) For an eigenvector $\vec{v_i}$ of H, we have

$$H\vec{v_i} = \lambda_i \vec{v_i} \quad (\lambda_i \text{ is an eigenvalue of } H)$$

$$\Rightarrow (H \cdot H\vec{v_i} \Rightarrow) \; H^2\vec{v_i} = H \cdot \lambda_i \vec{v_i}$$

$$\Rightarrow \quad H^2\vec{v_i} = \lambda_i H\vec{v_i}$$

$$\Rightarrow \quad H^2\vec{v_i} = \lambda_i \lambda_i \vec{v_i}$$

$$\Rightarrow \quad H^2\vec{v_i} = \lambda_i^2 \vec{v_i} \quad (1)$$

But also because H is idempotent, $H^2 = H$

$$\Rightarrow \quad H^2\vec{v_i} = H\vec{v_i} = \lambda_i \vec{v_i} \quad (2)$$

$$\xrightarrow{(1),(2)} \quad \lambda_i^2 \vec{v_i} = \lambda_i \vec{v_i} \quad \Rightarrow \quad \lambda_i^2 = \lambda_i$$

$$\Rightarrow \quad \lambda_i = 0 \text{ or } \lambda_i = 1$$

When $\lambda_i = 1 \Rightarrow H\vec{v_i} = \vec{v_i}$, meaning applying orthogonal projection H on $\vec{v_i}$ doesn't change $\vec{v_i}$, that happens when $\vec{v_i}$ lies in the space we are projecting onto with H, which is Im(X) (proved in 2c)

Im(X) is column space of X, which is also a full rank matrix with rank(X) = p $\Rightarrow$ Im(X) contains p linearly independent columns of X $\Rightarrow$ there are p situations for $\vec{v_i}$ to be ~~equ~~ on the Im(X) $\Rightarrow$ there are p eigenvalues $\lambda_i = 1$

H is a $(n \times n)$ matrix that has n eigenvalues $\Rightarrow$ the remaining $(n-p)$ eigenvalues will be zeros.

b) Let λ_{Hi} be an i^{th} eigenvalue of H

λ_{Qi} be an i^{th} eigenvalue of Q

We have $Q\vec{v_i} = \lambda_{Qi}\vec{v_i}$

$$\Rightarrow (I - H)\vec{v_i} = \lambda_{Qi}\vec{v_i}$$

$$\Rightarrow \vec{v_i} - H\vec{v_i} = \lambda_{Qi}\vec{v_i}$$

$$\Rightarrow \vec{v_i} - \lambda_{Hi}\vec{v_i} = \lambda_{Qi}\vec{v_i}$$

$$\Rightarrow (1 - \lambda_{Hi})\vec{v_i} = \lambda_{Qi}\vec{v_i}$$

$\Rightarrow \lambda_{Qi} = 1 - \lambda_{Hi} \Rightarrow$ Q has p eigenvalues of 0 and $(n-p)$ eigenvalues of 1

c) $Tr(H) = \sum_{i=1}^{n} \lambda_i = \lambda_1 + \lambda_2 + \ldots + \lambda_n = \underbrace{p\cdot(1)}_{\text{p eigenvalues of 1}} + \underbrace{(n-p)\cdot(0)}_{\text{the rest is 0}} = p$

$Tr(Q) = \sum_{i=1}^{n} \lambda_i = (n-p)\cdot(1) + p\cdot(0) = n - p$

4) a) $\hat{\beta} = (X^TX)^{-1}X^TY = (X^TX)^{-1}X^T(X\beta+\epsilon)$

$= \underbrace{(X^TX)^{-1}(X^TX)}_{I}\beta + (X^TX)^{-1}X^T\epsilon$

$= \beta + (X^TX)^{-1}X^T\epsilon$

$\Rightarrow E(\hat{\beta}) = E\left(\beta + (X^TX)^{-1}X^T\epsilon\right)$

$= E(\beta) + E\left((X^TX)^{-1}X^T\epsilon\right)$

$= \beta + (X^TX)^{-1}X^T E(\epsilon)$

We have $E(\epsilon) = 0$ by assumption

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

b) We have $\hat{y} = HY$ and $\hat{\epsilon} = Y - \hat{y} = Y - HY = (I-H)Y$

$\Rightarrow Cov(\hat{y}, \hat{\epsilon}) = Cov(HY, (I-H)Y)$

$= E\{[HY - E(HY)][(I-H)Y - E((I-H)Y)]^T\}$

$= E\{H(Y-E(Y))[(I-H)(Y-E(Y))]^T\}$

$= E[H(Y-E(Y))(Y-E(Y))^T(I-H)^T]$

$= E\{H[(Y-E(Y))(Y-E(Y))^T]Q^T\}$

But $Q^T = Q$ (from 2b) and $(Y-E(Y))(Y-E(Y))^T = Cov(Y,Y)$

$\Rightarrow Cov(\hat{y}, \hat{\epsilon}) = E(H\,Cov(Y,Y)Q)$

$= E(H\,Var(Y)Q)$

We have $Var(Y) = Var(X\beta+\epsilon) = Var(\epsilon) = \sigma^2 I_n$

(becase X are not random, X_i are scalars)

$\Rightarrow Cov(\hat{y}, \hat{\epsilon}) = E(H\sigma^2 I_n Q) = \sigma^2 E(HQ)$ (σ^2 is constant)

Also from 2b, $HQ = 0$

$\Rightarrow Cov(\hat{y}, \hat{\epsilon}) = \sigma^2(0) = 0$

c) If $Var(\epsilon)$ is not a multiple of identity matrix I_n

$\Rightarrow Cov(\hat{y}, \hat{\epsilon}) = E(H\,Var(\epsilon)Q) \neq 0$

$\Rightarrow$ 4b) doesn't hold if $Var(\epsilon) \neq \sigma^2 I_n$

d) (i) We have $E(\epsilon) = 0$ and $HX = X\underbrace{[(X^TX)^{-1}X^TX]}_{I} = X$

- Prove $\hat{Y} \perp\!\!\!\perp \hat{\epsilon}$ by proving $E(\hat{y}\hat{\epsilon}^T) = 0$

$E(\hat{Y}\hat{\epsilon}^T) = E(\underset{n\times n}{H}\underset{n\times 1}{Y}\underset{1\times n}{\hat{\epsilon}^T}) = E(H(X\beta+\epsilon)\hat{\epsilon}^T)$

$= E(HX\beta\hat{\epsilon}^T + H\epsilon\hat{\epsilon}^T)$

Also $\hat{\epsilon} = (I-H)Y \Rightarrow \hat{\epsilon}^T = Y^T(I-H)^T$

$\Rightarrow E(\hat{y}\hat{\epsilon}^T) = E[HX\beta Y^T(I-H)^T + H\epsilon Y^T(I-H)^T]$

$= \underbrace{HX\beta\, E(Y^T)(I-H)^T}_{(1)} + \underbrace{\cancel{X}\, H\, E(\epsilon Y^T)(I-H)^T}_{(2)}$

We have $E(Y^T) = E(\epsilon^T + \beta^T X^T) = E(\epsilon^T) = 0$

$\Rightarrow (1) = 0$

We have $E(\epsilon Y^T) = E(\epsilon(\epsilon^T + \beta^T X^T)) = E(\epsilon\epsilon^T) + \underbrace{E(\epsilon \overbrace{\beta^T X^T}^{\text{constant}})}_{=0}$

$$= E(\epsilon\epsilon^T) = E\begin{bmatrix} \epsilon_1|X \\ \epsilon_2|X \\ \vdots \\ \epsilon_n|X \end{bmatrix}[\epsilon_1|X\ \epsilon_2|X \ldots \epsilon_n|X]$$

$$\Rightarrow E(\epsilon Y^T) = E\begin{bmatrix} \epsilon_1^2|X & \epsilon_1\epsilon_2|X & \ldots & \epsilon_1\epsilon_n|X \\ \epsilon_2\epsilon_1|X & \epsilon_2^2|X & \ldots & \epsilon_2\epsilon_n|X \\ \vdots & & & \\ \epsilon_n\epsilon_1|X & \epsilon_n\epsilon_2|X & \ldots & \epsilon_n\epsilon_n|X \end{bmatrix} = \begin{bmatrix} E(\epsilon_1^2|X) & \ldots & E(\epsilon_1\epsilon_n|X) \\ \vdots & & \\ E(\epsilon_n\epsilon_1|X) & \ldots & E(\epsilon_n^2|X) \end{bmatrix}$$

$$= \begin{bmatrix} \sigma^2 & 0 & \ldots & 0 \\ 0 & \sigma^2 & & \\ \vdots & & \ddots & \\ 0 & & & \sigma^2 \end{bmatrix} = \sigma^2 I_n \quad (= Var(\epsilon))$$

$\Rightarrow$ (2) becomes $H\sigma^2 I_n(I-H)^T = \sigma^2 H Q^T = \sigma^2 \cdot 0 = 0$

$\Rightarrow E(\hat{Y}\hat{\epsilon}^T) = (1) + (2) = 0$

$\Rightarrow$ $\hat{Y}$ and $\hat{\epsilon}$ are independent

- Proving $\hat{\beta} \perp\!\!\!\perp \hat{\epsilon}$ by proving $E(\hat{\epsilon}\hat{\beta}^T) = 0$

$E(\hat{\epsilon}\hat{\beta}^T) = E(\underbrace{(I-H)\epsilon}_{n\times 1}(\underbrace{(X^TX)^{-1}X^TY}_{1\times p})^T)$

$= E((I-H)\epsilon Y^T X(X^TX)^{-1})$ (because $(X^TX)^{-1}$ is symmetric

$= (I-H)E(\epsilon Y^T)X(X^TX)^{-1}$ $\Rightarrow ((X^TX)^{-1})^T = (X^TX)^{-1}$)

We show above that $E(\epsilon Y^T) = \sigma^2 I_n$

$\Rightarrow E(\hat{\epsilon}\hat{\beta}^T) = (I-H)\sigma^2 I_n X(X^TX)^{-1}$

$= \sigma^2(I-H)X(X^TX)^{-1}$

$= \sigma^2(X - HX)(X^TX)^{-1}$

$= \sigma^2(X-X)(X^TX)^{-1}$ (using $X = HX$)

$= \sigma^2(0)(X^TX)^{-1} = 0$

$\Rightarrow$ $\hat{\beta}$ is independent of $\hat{\epsilon}$

ii) Because $E(\hat{\epsilon}\hat{\beta}^T)$ and $E(\hat{y}\hat{\epsilon}^T)$ have been shown above to be written in some form that involves $E(\epsilon\epsilon^T) = Var(\epsilon)$ and they can only be zero when $Var(\epsilon) = \sigma^2 I_n$

$\Rightarrow$ i) does not hold for when $Var(\epsilon) \neq \sigma^2 I_n$

iii) If $\epsilon \sim N(0_n, \sigma^2 I_n)$, then using 4b) $Cov(\hat{y}, \hat{\epsilon}) = 0$

$\Rightarrow$ We can expect correlation of $\hat{y}$ and $\hat{\epsilon}$ to be equal to 0

$\Rightarrow$ When we plot $\hat{Y}$ vs $\hat{\epsilon}$, we would expect a plot with no pattern / relationship at all.