统计线性模型 Assignment 4

12111603 谭致恒

November 2023

1 Question 1

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 2 & -1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \theta \\ \phi \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix} \triangleq X\beta + \vec{\varepsilon}$$

where $X \triangleq \begin{bmatrix} 1 & 0 \\ 2 & -1 \\ 1 & 2 \end{bmatrix}$, $\beta \triangleq \begin{bmatrix} \theta \\ \phi \end{bmatrix}$ and $\vec{\varepsilon} \triangleq \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix}$.

We perform OLS to solve the optimization problem:

$$\min_{\beta} (Y - X\beta)^T (Y - X\beta)$$

Then we can derive the solution

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

which indicates

$$\begin{aligned} \hat{\beta} = \begin{bmatrix} \hat{\theta} \\ \hat{\phi} \end{bmatrix} &= \left(\begin{bmatrix} 1 & 2 & 1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 2 & -1 \\ 1 & 2 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 2 & 1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{6} & 0 \\ 0 & \frac{1}{5} \end{bmatrix} \begin{bmatrix} Y_1 + 2Y_2 + Y_3 \\ -Y_2 + 2Y_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{6}Y_1 + \frac{1}{3}Y_2 + \frac{1}{6}Y_3 \\ -\frac{1}{5}Y_2 + \frac{2}{5}Y_3 \end{bmatrix} \end{aligned}$$

Therefore, the OLS estimators are

$$\hat{\theta}_{OLS} = \tfrac{1}{6}Y_1 + \tfrac{1}{3}Y_2 + \tfrac{1}{6}Y_3 \text{ and } \hat{\phi}_{OLS} = -\tfrac{1}{5}Y_2 + \tfrac{2}{5}Y_3.$$

2 Question 2

$$Y = \begin{bmatrix} Y^a_{m\times 1} \\ Y^b_{m\times 1} \\ Y^c_{n\times 1} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_{m\times 1} & \mathbf{0}_{m\times 1} \\ \mathbf{1}_{m\times 1} & \mathbf{1}_{m\times 1} \\ \mathbf{1}_{n\times 1} & -\mathbf{2}_{n\times 1} \end{bmatrix} \begin{bmatrix} \theta \\ \phi \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix} \triangleq X\beta + \varepsilon$$

By performing OLS then we can derive

$$\begin{aligned} \hat{\beta} &= (X^TX)^{-1}X^TY \\ &= \left(\begin{bmatrix} 2m+n & m-2n \\ m-2n & m+4n \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{1}^T_{m\times 1}Y^a + \mathbf{1}^T_{m\times 1}Y^b + \mathbf{1}^T_{n\times 1}Y^c \\ \mathbf{1}^T_{m\times 1}Y^b - \mathbf{2}^T_{n\times 1}Y^c \end{bmatrix} \\ &= \frac{1}{m^2+13mn} \begin{bmatrix} m+4n & -m+2n \\ -m+2n & 2m+n \end{bmatrix} \begin{bmatrix} \mathbf{1}^T_{m\times 1}Y^a + \mathbf{1}^T_{m\times 1}Y^b + \mathbf{1}^T_{n\times 1}Y^c \\ \mathbf{1}^T_{m\times 1}Y^b - \mathbf{2}^T_{n\times 1}Y^c \end{bmatrix} \end{aligned}$$

Therefore, the OLS estimators are

$$\hat{\theta}_{OLS} = \frac{m+4n}{m^2+13mn}(\mathbf{1}^T_{m\times 1}Y^a + \mathbf{1}^T_{m\times 1}Y^b + \mathbf{1}^T_{n\times 1}Y^c) + \frac{-m+2n}{m^2+13mn}(\mathbf{1}^T_{m\times 1}Y^b - \mathbf{2}^T_{n\times 1}Y^c)$$

and

$$\hat{\phi}_{OLS} = \frac{-m+2n}{m^2+13mn}(\mathbf{1}^T_{m\times 1}Y^a + \mathbf{1}^T_{m\times 1}Y^b + \mathbf{1}^T_{n\times 1}Y^c) + \frac{2m+n}{m^2+13mn}(\mathbf{1}^T_{m\times 1}Y^b - \mathbf{2}^T_{n\times 1}Y^c).$$

Since

$$\begin{aligned} Cov(\hat{\beta}) &= Cov((X^TX)^{-1}X^TY) = (X^TX)^{-1}X^TCov(Y)X(X^TX)^{-1} \\ &= \sigma^2(X^TX)^{-1} = \frac{\sigma^2}{m^2+13mn} \begin{bmatrix} m+4n & -m+2n \\ -m+2n & 2m+n \end{bmatrix} \end{aligned}$$

Thus, if $m = 2n$, then from the covariance matrix we can figure out that $Cov(\hat{\theta}, \hat{\phi}) = 0$, which means they are uncorrelated.

3 Question 3

3.1 (a)

Since $Cov(\hat{\beta}) = Cov((X^TX)^{-1}X^TY) = (X^TX)^{-1}X^TCov(Y)X(X^TX)^{-1} = \sigma^2(X^TX)^{-1}$ and $\mathbb{E}\hat{\beta} = \beta$, then we have

$$\begin{aligned}
MSE(\tilde{\beta}) &= \mathbb{E}(c\hat{\beta} - \beta)^T(c\hat{\beta} - \beta) \\
&= \mathbb{E}(c^2\hat{\beta}^T\hat{\beta} - 2c\beta^T\hat{\beta} + \beta^T\beta) \\
&= c^2\mathbb{E}(\hat{\beta}^T\hat{\beta}) + (1-2c)\beta^T\beta \\
&= c^2[trace(\sigma^2(X^TX)^{-1}) + \beta^T\beta] + (1-2c)\beta^T\beta \\
&= c^2\sigma^2 trace((X^TX)^{-1}) + (c-1)^2\beta^T\beta.
\end{aligned}$$

3.2 (b)

Since $X^TX \succeq \mathbf{0}$, then $(X^TX)^{-1} \succeq \mathbf{0}$, which means

$$trace((X^TX)^{-1}) \geq 0.$$

Thus, $MSE(\tilde{\beta}) = c^2\sigma^2 trace((X^TX)^{-1}) + (c-1)^2\beta^T\beta$ is a convex quadratic function openning up for variable c, which has a minimizer.

Let its derivative $\frac{\mathrm{d}(MSE)}{\mathrm{d}c} = 0$, we can solve that the minimizer

$$c^* = \frac{\beta^T\beta}{\beta^T\beta + \sigma^2 trace((X^TX)^{-1})}$$

and obviously $c^* \leq 1$ is feasible.

Therefore,

$$c^* = \frac{\beta^T\beta}{\beta^T\beta + \sigma^2 trace((X^TX)^{-1})}.$$

3.3 (c)

Since the eigenvalues of X^TX are $1, 2, 3, 4, 5$, then the eigenvalues of $(X^TX)^{-1}$ are $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}$. Hence,

$$trace((X^TX)^{-1}) = \sum_{i=1}^{5} \frac{1}{\lambda_i} = \frac{137}{60}.$$

Therefore,

$$c^* = \frac{\beta^T\beta}{\beta^T\beta + \sigma^2 trace((X^TX)^{-1})} = \frac{55}{55 + \frac{137}{60}} = \frac{3300}{3437}.$$

4 Question 4

4.1 (a)

Since X_1 has full column rank, then $(X_1^T X_1)^{-1}$ exists. Then comes

$$\begin{aligned}
&\hat{\beta} = (X^T X)^{-1} X^T Y \\
\Rightarrow\ & X^T X \hat{\beta} = X^T Y \\
\Rightarrow\ & (X_1^T X_1 + X_2^T X_2)\hat{\beta} = X^T Y = X_1^T Y_1 + X_2^T Y_2 \\
\Rightarrow\ & X_1^T X_1 \hat{\beta} = X_1^T Y_1 + X_2^T Y_2 - X_2^T X_2 \hat{\beta} = X_1^T Y_1 + X_2^T e_2 \\
\Rightarrow\ & \hat{\beta} = (X_1^T X_1)^{-1} X_1^T Y_1 + (X_1^T X_1)^{-1} X_2^T e_2 \\
\Rightarrow\ & \hat{\beta} = \hat{\beta}^* + (X_1^T X_1)^{-1} X_2^T e_2.
\end{aligned}$$

Thus, we have already verified that

$$\hat{\beta} - \hat{\beta}^* = (X_1^T X_1)^{-1} X_2^T e_2 = M_1^{-1} X_2^T e_2.$$

4.2 (b)

Consider the difference between e_2 and e_2^*, we have

$$e_2 - e_2^* = Y_2^* - Y_2 = X_2(\hat{\beta}^* - \hat{\beta}).$$

Thus, we can obtain that

$$e_2 = e_2^* - X_2(\hat{\beta} - \hat{\beta}^*) = e_2^* - X_2 (X_1^T X_1)^{-1} X_2^T e_2.$$

Hence,

$$e_2 = (I + X_2 (X_1^T X_1)^{-1} X_2^T)^{-1} e_2^*.$$

(Since $I + X_2(X_1^T X_1)^{-1} X_2^T \succ X_2(X_1^T X_1)^{-1} X_2^T) \succeq \mathbf{0}$, then we can say that the inverse $(I + X_2(X_1^T X_1)^{-1} X_2^T)^{-1}$ exists.)

Now $\hat{\beta} - \hat{\beta}^*$ becomes

$$\hat{\beta} - \hat{\beta}^* = (X_1^T X_1)^{-1} X_2^T (I + X_2 (X_1^T X_1)^{-1} X_2^T)^{-1} e_2^*.$$

After discussion with my friend, she found a brief form of $\hat{\beta} - \hat{\beta}^*$, now we derive it as follows.

Since $e_2 = e_2^* - X_2(\hat{\beta} - \hat{\beta}^*)$, then we can rewrite (a) as

$$\hat{\beta} - \hat{\beta}^* = M_1^{-1} X_2^T (e_2^* - X_2(\hat{\beta} - \hat{\beta}^*)),$$

then follows

$$\hat{\beta} - \hat{\beta}^* = (M_1 + X_2^T X_2)^{-1} X_2^T e_2^* = (X^T X)^{-1} X_2^T e_2^*.$$

4.3 (c)

We have

$$X_1 = \left(\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -3 & -2 & -1 & 0 & 1 & 2 & 3 \end{bmatrix}^T \right)_{7\times 2}$$

and

$$X_2 = \begin{bmatrix} 1 & 4 \end{bmatrix}_{1\times 2}$$

We can calculate that $e_2^* = 4 - X_2\hat{\beta}^* = 4 - (-2) = 6$.

Thus,

$$\begin{aligned} \hat{\beta} &= \begin{bmatrix} 6 \\ -2 \end{bmatrix} + \begin{bmatrix} \frac{1}{7} & 0 \\ 0 & \frac{1}{28} \end{bmatrix} \begin{bmatrix} 1 \\ 4 \end{bmatrix} \left(1 + \begin{bmatrix} 1 & 4 \end{bmatrix} \begin{bmatrix} \frac{1}{7} & 0 \\ 0 & \frac{1}{28} \end{bmatrix} \begin{bmatrix} 1 \\ 4 \end{bmatrix} \right)^{-1} \cdot 6 \\ &= \begin{bmatrix} 6 \\ -2 \end{bmatrix} + \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix} \\ &= \begin{bmatrix} 6.5 \\ -1.5 \end{bmatrix} \end{aligned}$$

Therefore, the new parameter estimate $\hat{\beta}$ is $\begin{bmatrix} 6.5 \\ -1.5 \end{bmatrix}$.

Now, we verify the brief form

$$\hat{\beta} - \hat{\beta}^* = (M_1 + X_2^T X_2)^{-1} X_2^T e_2^* = (X^T X)^{-1} X_2^T e_2^*.$$

Calculate $X^T X = \begin{bmatrix} 8 & 4 \\ 4 & 44 \end{bmatrix}$, thus

$$\begin{aligned}
\hat{\beta} &= \hat{\beta}^* + (X^T X)^{-1} X_2^T e_2^* \\
&= \begin{bmatrix} 6 \\ -2 \end{bmatrix} + \begin{bmatrix} \frac{11}{84} & -\frac{1}{84} \\ -\frac{1}{84} & \frac{1}{42} \end{bmatrix} \begin{bmatrix} 6 \\ 24 \end{bmatrix} = \begin{bmatrix} 6 \\ -2 \end{bmatrix} + \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix} \\
&= \begin{bmatrix} 6.5 \\ -1.5 \end{bmatrix}
\end{aligned}$$

The results are consistent.