STK3100 Exercises, Week 2

Vinnie Ko

September 7, 2018

Exercise 2.11

i)

$$\left(\boldsymbol{I} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}\right)^{\mathrm{T}} = \boldsymbol{I}^{\mathrm{T}} - \frac{1}{n}\left(\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}\right)^{\mathrm{T}} = \boldsymbol{I} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}$$

So, $\boldsymbol{I} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}$ is symmetric.

ii)

$$\begin{aligned}\left(\boldsymbol{I} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}\right)\left(\boldsymbol{I} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}\right) &= \boldsymbol{I} - \frac{2}{n}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}} + \frac{1}{n^2}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}} \\ &= \boldsymbol{I} - \frac{2}{n}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}} + \frac{n}{n^2}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}} \\ &= \boldsymbol{I} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}\end{aligned}$$

So, $\boldsymbol{I} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}$ is idempotent.

iii)

$$\left(\boldsymbol{I} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}\right)\boldsymbol{y} = \boldsymbol{y} - \mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{y} - \overline{y}\mathbf{1}_n$$

Exercise 2.12

Let $\boldsymbol{X}$ be a $n \times p$ matrix of full rank, then $\operatorname{rank}(\boldsymbol{X}) = p$. Since $\boldsymbol{H}$ is a projection matrix, $\operatorname{rank}(\boldsymbol{H}) = \operatorname{tr}(\boldsymbol{H})$. So, we have

$$\operatorname{rank}(\boldsymbol{H}) = \operatorname{tr}(\boldsymbol{H}) = \operatorname{tr}(\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}) = \operatorname{tr}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}) = \operatorname{tr}(\boldsymbol{I}_p) = p = \operatorname{rank}(\boldsymbol{X}).$$

Exercise 1.17

Recall the definition of *model space*:

$$C(\boldsymbol{X}) = \{\boldsymbol{\eta} : \text{there is a } \boldsymbol{\beta} \text{ such that } \boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta}\}.$$

Let $\boldsymbol{A}$ be a non-singular matrix, then by definition, there exists $\boldsymbol{A}^{-1}$ that satisfies $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$. This allows us to write $\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}\boldsymbol{I}\boldsymbol{\beta} = \boldsymbol{X}\boldsymbol{A}\boldsymbol{A}^{-1}\boldsymbol{\beta}$ which yields

$$C(\boldsymbol{X}\boldsymbol{A}) = \{\boldsymbol{\eta} : \text{there is a } \boldsymbol{A}^{-1}\boldsymbol{\beta} \text{ such that } \boldsymbol{\eta} = (\boldsymbol{X}\boldsymbol{A})(\boldsymbol{A}^{-1}\boldsymbol{\beta})\}.$$

If we can show that vector space of $\boldsymbol{\beta}$ and vector space $\boldsymbol{A}^{-1}\boldsymbol{\beta}$ are same, namely $\mathbb{R}^p$, we have $C(\boldsymbol{X}) = C(\boldsymbol{X}\boldsymbol{A})$.

Proof.
Let $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k\}$ be a set of vectors that (linear) span $\mathbb{R}^p$. i.e. We can write any element of vector space $\boldsymbol{\beta}$ as a linear combination of $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k\}$.
Since $\boldsymbol{A}\boldsymbol{v}_i \in \mathbb{R}^p$ for each $1 \leq i \leq k$, we can write

$$\boldsymbol{A}\boldsymbol{v}_i = a_1\boldsymbol{v}_1 + \cdots + a_k\boldsymbol{v}_k$$

where a_i's are constants. By multiplying $\boldsymbol{A}^{-1}$ on the both sides, we get

$$\boldsymbol{v}_i = a_1\boldsymbol{A}^{-1}\boldsymbol{v}_1 + \cdots + a_k\boldsymbol{A}^{-1}\boldsymbol{v}_k$$

So, any element in the spanning set $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k\}$ can be written as a linear combination of $\{\boldsymbol{A}^{-1}\boldsymbol{v}_1, \cdots, \boldsymbol{A}^{-1}\boldsymbol{v}_k\}$. In other words, $\{\boldsymbol{A}^{-1}\boldsymbol{v}_1, \cdots, \boldsymbol{A}^{-1}\boldsymbol{v}_k\}$ is also a spanning set of $\mathbb{R}^p$. Thus, vector space of $\boldsymbol{\beta}$ and vector space of $\boldsymbol{A}^{-1}\boldsymbol{\beta}$ are same. $\square$

Exercise 2.13

Let $\boldsymbol{X}$ be a $n \times p$ model matrix of full rank and $\boldsymbol{A}$ be an arbitrary $p \times p$ non-singular matrix. Further, let $\boldsymbol{P}_X$ denote the projection matrix onto model space $C(\boldsymbol{X})$ and $\boldsymbol{P}_{XA}$ denote the projection matrix onto model space $C(\boldsymbol{X}\boldsymbol{A})$. Those projections matrices are unique (see p.34 of the book) and in the context of linear model, they are also called as *hat matrix*.

By using the least square principle, we can obtain the expression for the hat matrix $\boldsymbol{P}_{XA}$ and we can write it further to show that $\boldsymbol{P}_{XA} = \boldsymbol{P}_X$:

$$\begin{aligned}
\boldsymbol{P}_{XA} &= \boldsymbol{X}\boldsymbol{A}\left((\boldsymbol{X}\boldsymbol{A})^{\mathrm{T}}\boldsymbol{X}\boldsymbol{A}\right)^{-1}(\boldsymbol{X}\boldsymbol{A})^{\mathrm{T}} \\
&= \boldsymbol{X}\boldsymbol{A}\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}} \\
&= \boldsymbol{X}\boldsymbol{A}\boldsymbol{A}^{-1}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}(\boldsymbol{A}^{\mathrm{T}})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}} \\
&= \boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}} \\
&= \boldsymbol{P}_X
\end{aligned}$$

And since $\boldsymbol{P}_{XA} = \boldsymbol{P}_X$, we also have $\boldsymbol{\eta}_{XA} = \boldsymbol{P}_{XA} \cdot \boldsymbol{y} = \boldsymbol{P}_X \cdot \boldsymbol{y} = \boldsymbol{\eta}_X$.

Exercise 2.14

i)
In linear model, projection matrix is hat matrix. So,

$$\boldsymbol{P}_X\boldsymbol{X} = \boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} = \boldsymbol{X}.$$

ii)
By definition of model space, we have

$$C(\boldsymbol{X}) = \{\boldsymbol{\eta} : \text{there exists } \boldsymbol{\beta} \in \mathbb{R}^p \text{ such that } \boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta}\}$$

and

$$C(\boldsymbol{P}_X) = \{\boldsymbol{\tau} : \text{there exists } \boldsymbol{a} \in \mathbb{R}^n \text{ such that } \boldsymbol{\tau} = \boldsymbol{P}_X \boldsymbol{a}\}.$$

Take an arbitrary $\boldsymbol{\tau} \in C(\boldsymbol{P}_X)$. Then, $\boldsymbol{\tau} = \boldsymbol{P}_X \boldsymbol{a}$ for an $\boldsymbol{a} \in \mathbb{R}^n$. Note that

$$\boldsymbol{\tau} = \boldsymbol{P}_X \boldsymbol{a} = \boldsymbol{X} \underbrace{(\boldsymbol{X}^{\mathrm{T}} \boldsymbol{X})^{-1} \boldsymbol{X}^{\mathrm{T}} \boldsymbol{a}}_{\boldsymbol{c}} = \boldsymbol{X} \boldsymbol{c} \in \mathbb{R}^p \quad \text{and} \quad \boldsymbol{c} = (\boldsymbol{X}^{\mathrm{T}} \boldsymbol{X})^{-1} \boldsymbol{X}^{\mathrm{T}} \boldsymbol{a} \in \mathbb{R}^p.$$

Hence, $\boldsymbol{\tau} \in C(\boldsymbol{X})$.

Similarly, we can do the other direction. Then we have $C(\boldsymbol{X}) = C(\boldsymbol{P}_X)$.

Exercise 2.15

i)
When model a is a special case of model b, we have $\boldsymbol{P}_a \boldsymbol{P}_b = \boldsymbol{P}_b \boldsymbol{P}_a = \boldsymbol{P}_a$ (p.36 of the book). Since null model is nested within every possible linear model with an intercept, we can directly use this result for between $\boldsymbol{P}_0$ (the hat matrix of null model) and arbitrary hat matrix $\boldsymbol{H}$:

$$\boldsymbol{P}_0 \boldsymbol{H} = \boldsymbol{H} \boldsymbol{P}_0 = \boldsymbol{P}_0.$$

ii)

$$\begin{aligned}
\boldsymbol{P}_0 &= \mathbf{1}_n (\mathbf{1}_n^{\mathrm{T}} \mathbf{1}_n)^{-1} \mathbf{1}_n^{\mathrm{T}} \\
&= \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^{\mathrm{T}} \\
&= \frac{1}{n} \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}
\end{aligned}$$

So,

$$\begin{aligned}
\boldsymbol{P}_0 \boldsymbol{H} &= \frac{1}{n} \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix} \cdot \begin{bmatrix} h_{1,1} & \cdots & h_{1,n} \\ \vdots & \ddots & \vdots \\ h_{n,1} & \cdots & h_{n,n} \end{bmatrix} \\
&= \frac{1}{n} \begin{bmatrix} \sum_{i=1}^{n} h_{i,1} & \cdots & \sum_{i=1}^{n} h_{i,n} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} h_{i,1} & \cdots & \sum_{i=1}^{n} h_{i,n} \end{bmatrix}
\end{aligned}$$

From i), we have $\boldsymbol{P}_0 \boldsymbol{H} = \boldsymbol{P}_0$. So,

$$\begin{bmatrix} \sum_{i=1}^{n} h_{i,1} & \cdots & \sum_{i=1}^{n} h_{i,n} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} h_{i,1} & \cdots & \sum_{i=1}^{n} h_{i,n} \end{bmatrix} = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}$$

and this implies $\sum_{i=1}^{n} h_{i,j} = 1$, for $j = 1, \cdots, n$. Thus, all columns sums of $\boldsymbol{H}$ are 1.

By repeating the same procedure for $\boldsymbol{H}\boldsymbol{P}_0 = \boldsymbol{P}_0$ gives $\sum_{j=1}^{n} h_{i,j} = 1$, for $i = 1, \cdots, n$. Thus, all row sums of $\boldsymbol{H}$ are 1.

Exercise 2.16

(a)

Note that $\boldsymbol{v}_2 \in C(\boldsymbol{X})^{\perp}$. So, $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{v}_2 = \boldsymbol{0}$ and $\boldsymbol{v}_2^{\mathrm{T}}\boldsymbol{X} = \boldsymbol{0}^{\mathrm{T}}$ (p.32 of the book).

$$
\begin{aligned}
\boldsymbol{v}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{G}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} &= (\boldsymbol{v}_1 + \boldsymbol{v}_2)^{\mathrm{T}}\boldsymbol{X}\boldsymbol{G}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} \\
&= \boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{X}\boldsymbol{G}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} + \boldsymbol{v}_2^{\mathrm{T}}\boldsymbol{X}\boldsymbol{G}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} \\
&= \boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{X}\boldsymbol{G}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} \\
&= (\boldsymbol{X}\boldsymbol{b})^{\mathrm{T}}\boldsymbol{X}\boldsymbol{G}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} \\
&= \boldsymbol{b}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{G}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} \\
&= \boldsymbol{b}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} && \text{(by the definition of } \textit{generalized inverse}\text{)} \\
&= (\boldsymbol{X}\boldsymbol{b})^{\mathrm{T}}\boldsymbol{X} \\
&= \boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{X} \\
&= \boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{X} + \boldsymbol{v}_2^{\mathrm{T}}\boldsymbol{X} \\
&= (\boldsymbol{v}_1 + \boldsymbol{v}_2)^{\mathrm{T}}\boldsymbol{X} \\
&= \boldsymbol{v}^{\mathrm{T}}\boldsymbol{X}
\end{aligned}
$$

Since $\boldsymbol{v}$ is arbitrary, $\boldsymbol{X}\boldsymbol{G}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} = \boldsymbol{X}$.

(b)

From (a), we have $\boldsymbol{X}\boldsymbol{G}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} = \boldsymbol{X}$ and this results holds for any generalized inverse since we only used the definition of generalized inverse to derive this result.

Imagine that we repeated (a) with another generalized inverse $\boldsymbol{H}$, then we have $\boldsymbol{X}\boldsymbol{H}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} = \boldsymbol{X}\boldsymbol{G}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$.

$$
\begin{aligned}
\boldsymbol{X}\boldsymbol{H}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} &= \boldsymbol{X}\boldsymbol{G}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} \\
\boldsymbol{X}\boldsymbol{H}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} \cdot \boldsymbol{b} &= \boldsymbol{X}\boldsymbol{G}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} \cdot \boldsymbol{b} \\
\boldsymbol{X}\boldsymbol{H}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{v}_1 &= \boldsymbol{X}\boldsymbol{G}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{v}_1 \\
\boldsymbol{X}\boldsymbol{H}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{v}_1 + \boldsymbol{X}^{\mathrm{T}}\boldsymbol{v}_2\right) &= \boldsymbol{X}\boldsymbol{G}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{v}_1 + \boldsymbol{X}^{\mathrm{T}}\boldsymbol{v}_2\right) \\
\boldsymbol{X}\boldsymbol{H}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{v} &= \boldsymbol{X}\boldsymbol{G}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{v}
\end{aligned}
$$

Since $\boldsymbol{b}$ is arbitrary, this result holds for all $\boldsymbol{v}$. Thus, $\boldsymbol{X}\boldsymbol{H}\boldsymbol{X}^{\mathrm{T}} = \boldsymbol{X}\boldsymbol{G}\boldsymbol{X}^{\mathrm{T}}$.