The Orthogonal Projection Matrix onto $\mathcal{C}(\boldsymbol{X})$ and the Normal Equations

Theorem 2.1:

$\boldsymbol{P_X}$, the orthogonal projection matrix onto $\mathcal{C}(\boldsymbol{X})$, is equal to $\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'$.

The matrix $\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'$ satisfies

(a) $[\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'][\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'] = [\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}']$

(b) $\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y} \in \mathcal{C}(\boldsymbol{X}) \ \forall \ \boldsymbol{y} \in \mathbb{R}^n$

(c) $\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{x} = \boldsymbol{x} \ \forall \ \boldsymbol{x} \in \mathcal{C}(\boldsymbol{X})$

(d) $\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}' = [\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}']'$

(e) $\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'$ is the same $\forall$ GI of $\boldsymbol{X}'\boldsymbol{X}$.

It is useful to establish a few results before proving Theorem 2.1.

Show that if $\underset{n\times n}{\boldsymbol{A}}$ is a symmetric matrix and $\underset{n\times n}{\boldsymbol{G}}$ is a GI of $\underset{n\times n}{\boldsymbol{A}}$, then $\underset{n\times n}{\boldsymbol{G}'}$ is also GI of $\underset{n\times n}{\boldsymbol{A}}$.

Result 2.4:

$$X'XA = X'XB \Longleftrightarrow XA = XB.$$

One direction is obvious ($\Longleftarrow$).

Can you prove the other ($\Longrightarrow$)?

Result 2.5:

Suppose $(\boldsymbol{X}'\boldsymbol{X})^-$ is any GI of $\boldsymbol{X}'\boldsymbol{X}$. Then $(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'$ is a GI of $\boldsymbol{X}$, i.e.,

$$\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{X}.$$

Corollary to Result 2.5:

For $(\boldsymbol{X}'\boldsymbol{X})^-$ any GI of $\boldsymbol{X}'\boldsymbol{X}$,

$$\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}' = \boldsymbol{X}'.$$

Now we can prove Theorem 2.1:

First show $X(X'X)^{-}X'$ is idempotent.

Now show that

$$\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y} \in \mathcal{C}(\boldsymbol{X}) \ \forall \ \boldsymbol{y} \in \mathbb{R}^n.$$

Now show that

$$\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{z} = \boldsymbol{z} \; \forall \; \boldsymbol{z} \in \mathcal{C}(\boldsymbol{X}).$$

Now we know $\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'$ is a projection matrix onto $\mathcal{C}(\boldsymbol{X})$.

To be the unique projection matrix onto $\mathcal{C}(\boldsymbol{X})$, $\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'$ must be symmetric.

To show symmetry of $\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'$, it will help to first show that

$$\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})_1^{-}\boldsymbol{X}' = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})_2^{-}\boldsymbol{X}'$$

for any two GIs of $\boldsymbol{X}'\boldsymbol{X}$ denoted $(\boldsymbol{X}'\boldsymbol{X})_1^{-}$ and $(\boldsymbol{X}'\boldsymbol{X})_2^{-}$.

Now show that $\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'$ is symmetric.

We have shown that $\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'$ is the orthogonal projection matrix onto $\mathcal{C}(\boldsymbol{X})$.

We know that it is the only symmetric projection matrix onto $\mathcal{C}(\boldsymbol{X})$ by Result A.16.

Example:

Find the orthogonal projection matrix onto $\mathcal{C}(\underset{n\times 1}{\mathbf{1}})$.

Also, simplify $\boldsymbol{P_X y}$ in this case.

By Corollary A.4 to Result A.16, we know that

$$\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'$$

is the orthogonal projection matrix onto $\mathcal{C}(\boldsymbol{X})^{\perp} = \mathcal{N}(\boldsymbol{X}')$.

This is Result 2.6.

By Result A.4, any $\boldsymbol{y} \in \mathbb{R}^n$ may be written as

$$\boldsymbol{y} = \boldsymbol{s} + \boldsymbol{t},$$

where $\boldsymbol{s} \in \mathcal{C}(\boldsymbol{X})$ and $\boldsymbol{t} \in \mathcal{C}(\boldsymbol{X})^\perp = \mathcal{N}(\boldsymbol{X}')$.

Furthermore, the vectors $\boldsymbol{s} \in \mathcal{C}(\boldsymbol{X})$ and $\boldsymbol{t} \in \mathcal{C}(\boldsymbol{X})^\perp$ are unique.

Because

$$\boldsymbol{y} = \boldsymbol{P_X y} + (\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y}$$

$$\boldsymbol{P_X y} \in \mathcal{C}(\boldsymbol{X}) \text{ and } (\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y} \in \mathcal{C}(\boldsymbol{X})^{\perp},$$

we get $\boldsymbol{s} = \boldsymbol{P_X y}$ and $\boldsymbol{t} = (\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y}$.

Moreover,

$$\hat{\boldsymbol{y}} \equiv \boldsymbol{P_X y} = \text{ the vector of fitted values}$$

$$\hat{\boldsymbol{\varepsilon}} \equiv (\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y} = \boldsymbol{y} - \hat{\boldsymbol{y}} = \text{ the vector of residuals.}$$

Let $\boldsymbol{P_W}$ and $\boldsymbol{P_X}$ denote the orthogonal projection matrices onto $\mathcal{C}(\boldsymbol{W})$ and $\mathcal{C}(\boldsymbol{X})$, respectively.

Suppose $\mathcal{C}(\boldsymbol{W}) \subseteq \mathcal{C}(\boldsymbol{X})$. Show that

$$\boldsymbol{P_W P_X} = \boldsymbol{P_X P_W} = \boldsymbol{P_W}.$$

Theorem 2.2:

If $\mathcal{C}(\boldsymbol{W}) \subseteq \mathcal{C}(\boldsymbol{X})$, then $\boldsymbol{P_X} - \boldsymbol{P_W}$ is the orthogonal projection matrix onto $\mathcal{C}((\boldsymbol{I} - \boldsymbol{P_W})\boldsymbol{X})$.

We have previously seen that

$$\begin{aligned} Q(\boldsymbol{b}) &= (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}) \\ &= \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}\|^2 \geq \|\boldsymbol{y} - \boldsymbol{P}_X\boldsymbol{y}\|^2 \end{aligned}$$

$\forall\ \boldsymbol{b} \in \mathbb{R}^p$ with equality iff $\boldsymbol{X}\boldsymbol{b} = \boldsymbol{P}_X\boldsymbol{y}$.

Now we know that $\boldsymbol{P_X} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'$.

Thus, $\hat{\boldsymbol{\beta}}$ minimizes $Q(\boldsymbol{b})$ iff $\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y}$.

By Result 2.4, this equation is equivalent to

$$\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y}.$$

Because $X'X(X'X)^{-}X' = X'$, $X'X\hat{\beta} = X'X(X'X)^{-}X'y$ is equivalent to

$$X'X\hat{\beta} = X'y.$$

This system of linear equations is known as the Normal Equations (NE).

We have established <u>Result 2.3:</u>

$\hat{\beta}$ is a solution to the NE $(\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} = \boldsymbol{X}'\boldsymbol{y})$ iff $\hat{\beta}$ minimizes $Q(\boldsymbol{b})$.

Corollary 2.1:

The NE are consistent.

By Result A.13, $\hat{\boldsymbol{\beta}}$ is a solution to $\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} = \boldsymbol{X}'\boldsymbol{y}$ iff

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y} + [\boldsymbol{I} - (\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{X}]\boldsymbol{z}$$

for some $\boldsymbol{z} \in \mathbb{R}^p$.

Corollary 2.3:

$\boldsymbol{X}\hat{\boldsymbol{\beta}}$ is invariant to the choice of a solution $\hat{\boldsymbol{\beta}}$ to the NE, i.e., if $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_2$ are any two solutions to the NE, then $\boldsymbol{X}\hat{\boldsymbol{\beta}}_1 = \boldsymbol{X}\hat{\boldsymbol{\beta}}_2$.

We will finish this set of notes with some other results from Section 2.2 of the text.

Lemma 2.1: $\mathcal{N}(\boldsymbol{X}'\boldsymbol{X}) = \mathcal{N}(\boldsymbol{X})$.

Result 2.2: $\mathcal{C}(\boldsymbol{X}'\boldsymbol{X}) = \mathcal{C}(\boldsymbol{X}')$.

Corollary 2.2: $rank(\boldsymbol{X}'\boldsymbol{X}) = rank(\boldsymbol{X})$.