Least Squares Estimation of the Expected Value of the Response Vector

Consider the General Linear Model (GLM) previously introduced, where

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon},$$

with $E(\boldsymbol{\varepsilon}) = \mathbf{0}$.

Suppose we want to estimate

$$E(\boldsymbol{y}) = \boldsymbol{X\beta}$$

based on our observed response $\boldsymbol{y}$.

Because we know

$$E(\boldsymbol{y}) = \boldsymbol{X\beta} \in \mathcal{C}(\boldsymbol{X}),$$

a reasonable strategy may be to find the vector in $\mathcal{C}(\boldsymbol{X})$ that is closest to $\boldsymbol{y}$ in terms of Euclidean distance.

For example, suppose

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} [\mu] + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}.$$

Then

$$E\left(\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} [\mu] = \begin{bmatrix} \mu \\ \mu \end{bmatrix}.$$

If we observe $\boldsymbol{y} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$ and have to guess

$$E(\boldsymbol{y}) = \boldsymbol{X\beta} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} [\mu] = \begin{bmatrix} \mu \\ \mu \end{bmatrix},$$

it may be reasonable to find the value of μ that makes $\begin{bmatrix} \mu \\ \mu \end{bmatrix}$ as close to $\begin{bmatrix} 4 \\ 2 \end{bmatrix}$ as possible.

In general, we seek a vector $\hat{\boldsymbol{\beta}} \ni \boldsymbol{X}\hat{\boldsymbol{\beta}}$ is closer to $\boldsymbol{y}$ than any other vector in $\mathcal{C}(\boldsymbol{X})$.

If

$$Q(\boldsymbol{b}) = \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}\|^2$$
$$= (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}),$$

we seek $\hat{\boldsymbol{\beta}} \ni Q(\hat{\boldsymbol{\beta}}) \leq Q(\boldsymbol{b}) \ \forall \ \boldsymbol{b} \in \mathbb{R}^p$.

If

$$Q(\hat{\boldsymbol{\beta}}) \leq Q(\boldsymbol{b}) \ \forall \ \boldsymbol{b} \in \mathbb{R}^p,$$

$\boldsymbol{X}\hat{\boldsymbol{\beta}}$ is called the Least Squares Estimate (LSE) of $E(\boldsymbol{y}) = \boldsymbol{X}\boldsymbol{\beta}$.

Suppose $\boldsymbol{P_X}$ is the orthogonal projection matrix onto $\mathcal{C}(\boldsymbol{X})$.

Show that $\boldsymbol{P_X y}$ is the unique vector in $\mathcal{C}(\boldsymbol{X})$ that is closest to $\boldsymbol{y}$.

Because $\boldsymbol{P_X y} \in \mathcal{C}(\boldsymbol{X})$, we know $\exists$ at least one vector $\boldsymbol{b} \ni \boldsymbol{P_X y} = \boldsymbol{Xb}$.

We can take $\hat{\boldsymbol{\beta}}$ to be any such vector $\boldsymbol{b}$. Then we have shown that $\hat{\boldsymbol{\beta}}$ will minimize $Q(\boldsymbol{b})$ over $\boldsymbol{b} \in \mathbb{R}^p$.

Now we know that the least square estimator of $E(\boldsymbol{y}) = \boldsymbol{X}\beta$ is $\boldsymbol{X}\hat{\beta} = \boldsymbol{P_X}\boldsymbol{y}$.

How do we find $\boldsymbol{P_X}$?

We need to find the symmetric and idempotent matrix that projects onto $\mathcal{C}(\boldsymbol{X})$.

In the next set of notes we will show

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