10.1. Best Linear Unbiased Estimates

Definition: The *Best Linear Unbiased Estimate (BLUE)* of a parameter θ based on data $\mathbf{Y}$ is

1. a linear function of $\mathbf{Y}$. That is, the estimator can be written as $\mathbf{b}'\mathbf{Y}$,
2. unbiased ($E[\mathbf{b}'\mathbf{Y}] = \theta$), and
3. has the smallest variance among all unbiased linear estimators.

Theorem 10.1.1: For any linear combination $\mathbf{c}'\theta$, $\mathbf{c}'\hat{\mathbf{Y}}$ is the BLUE of $\mathbf{c}'\theta$, where $\hat{\mathbf{Y}}$ is the least-squares orthogonal projection of $\mathbf{Y}$ onto $\mathcal{R}(\mathbf{X})$. *Proof:* See lecture notes # 8

Corollary 10.1.2: If $\text{rank}(\mathbf{X}_{n\times p}) = p$, then, for any $\mathbf{a}$, $\mathbf{a}'\hat{\boldsymbol{\beta}}$ is the BLUE of $\mathbf{a}'\boldsymbol{\beta}$.

Note: The Gauss-Markov theorem generalizes this result to the less than full rank case, for *certain* linear combinations $\mathbf{a}'\boldsymbol{\beta}$ (the *estimable functions*).

Proof of Corollary 10.1.2:

$$\begin{aligned}
\theta &= \mathbf{X}\boldsymbol{\beta} \\
\mathbf{X}'\theta &= \mathbf{X}'\mathbf{X}\boldsymbol{\beta} \\
(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\theta &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta} \\
\Rightarrow \mathbf{a}'\boldsymbol{\beta} &= \underbrace{\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'}_{\mathbf{c}'}\theta
\end{aligned}$$

So $\mathbf{a}'\boldsymbol{\beta} = \mathbf{c}'\theta$ where $\mathbf{c}' = \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

Now, $\mathbf{a}'\hat{\boldsymbol{\beta}} = \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ and

$$\begin{aligned}
\mathbf{c}'\hat{\mathbf{Y}} &= \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{Y}} \\
&= \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\
&= \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}
\end{aligned}$$

Therefore, since $\mathbf{a}'\hat{\boldsymbol{\beta}} = \mathbf{c}'\hat{\mathbf{Y}}$, it is the BLUE of $\mathbf{a}'\beta = \mathbf{c}'\theta$.

10.2. Estimable Functions

In the less than full rank case, only certain linear combinations of the components of $\boldsymbol{\beta}$ can be unbiasedly estimated.

Definition: A linear combination $\mathbf{a}'\boldsymbol{\beta}$ is *estimable* if it has a linear unbiased estimate, i.e., $E[\mathbf{b}'\mathbf{Y}] = \mathbf{a}'\boldsymbol{\beta}$ for some $\mathbf{b}$ for all $\boldsymbol{\beta}$.

Lemma 10.2.1:

(i) $\mathbf{a}'\boldsymbol{\beta}$ is estimable if and only if $\mathbf{a} \in \mathcal{R}(\mathbf{X}')$.

Proof: $E[\mathbf{b}'\mathbf{Y}] = \mathbf{b}'\mathbf{X}\boldsymbol{\beta}$, which equals $\mathbf{a}'\boldsymbol{\beta}$ for all $\boldsymbol{\beta}$ if and only if $\mathbf{a} = \mathbf{X}'\mathbf{b}$.

(ii) If $\mathbf{a}'\boldsymbol{\beta}$ is estimable, there is a unique $\mathbf{b}_* \in \mathcal{R}(\mathbf{X})$ such that $\mathbf{a} = \mathbf{X}'\mathbf{b}_*$.

Proof: $\mathbf{a}'\boldsymbol{\beta}$ is estimable so using (i) $\mathbf{a} = \mathbf{X}'\mathbf{b}$. Any $\mathbf{b} \in \Re^n$ can be uniquely decomposed as $\mathbf{b} = \mathbf{b}_* + \tilde{\mathbf{b}}$, where $\mathbf{b}_* \in \mathcal{R}(\mathbf{X})$, and $\tilde{\mathbf{b}} \in \mathcal{R}(\mathbf{X})^\perp$. Then

$$\mathbf{a} = \mathbf{X}'\mathbf{b} = \mathbf{X}'\mathbf{b}_* + \mathbf{X}'\tilde{\mathbf{b}} = \mathbf{X}'\mathbf{b}_*.$$

Comment: Part (i) of the lemma may be a little bit surprising since all of a sudden we are talking about the row space of $\mathbf{X}$, not the column space. However, the idea behind the result need not be mysterious. Every observation we have is an unbiased estimate of its expected value; the expected value of an observation is some linear combination of parameters. Such linear combinations of parameters is therefore estimable. These correspond exactly to the rows of $\mathbf{X}$. Clearly, also, linear combinations of estimable functions should be estimable. These are the vectors that are spanned by the rows of $\mathbf{X}$ – the row space of $\mathbf{X}$.

10.3. Gauss-Markov Theorem

Note: In the full rank case ($r = p$), any $\mathbf{a}'\boldsymbol{\beta}$ is estimable. In particular,

$$\mathbf{a}'\hat{\boldsymbol{\beta}} = \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \equiv \mathbf{b}'\mathbf{Y}$$

is a linear unbiased estimate of $\mathbf{a}'\boldsymbol{\beta}$. In this case we also know that $\mathbf{a}'\hat{\boldsymbol{\beta}}$ is the BLUE (Corollary 10.1.2).

Theorem 10.3.1: (Gauss-Markov). If $\mathbf{a}'\boldsymbol{\beta}$ is estimable, then

(i) $\mathbf{a}'\hat{\boldsymbol{\beta}}$ is unique (i.e., the same for all solutions to the normal equations $\hat{\boldsymbol{\beta}}$).

(ii) $\mathbf{a}'\hat{\boldsymbol{\beta}}$ is the BLUE of $\mathbf{a}'\boldsymbol{\beta}$.

Proof:

(i) By Lemma 10.2.1, $\mathbf{a} = \mathbf{X}'\mathbf{b}_*$ for a unique $\mathbf{b}_* \in \mathcal{R}(\mathbf{X})$. Therefore,

$$\mathbf{a}'\hat{\boldsymbol{\beta}} = \mathbf{b}_*'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{b}_*'\hat{\mathbf{Y}}$$

is unique because $\hat{\mathbf{Y}}$ is unique. (In fact $\mathbf{b}_*'\hat{\mathbf{Y}} = \mathbf{b}_*'\mathbf{Y}$ since $\mathbf{b}_* \in \mathcal{R}(\mathbf{X})$, so that $\mathbf{b}_*'(\mathbf{Y} - \hat{\mathbf{Y}}) = \mathbf{b}_*'\hat{\boldsymbol{\varepsilon}} = 0$.)

(ii) By Theorem 10.1.1, $\mathbf{b}_*'\hat{\mathbf{Y}}$ is the BLUE of $\mathbf{b}_*'\theta$. But, $\mathbf{a}'\hat{\boldsymbol{\beta}} = \mathbf{b}_*'\hat{\mathbf{Y}}$ from part (i) and $\mathbf{a}'\boldsymbol{\beta} = \mathbf{b}_*'\mathbf{X}\boldsymbol{\beta} = \mathbf{b}_*'\theta$.

10.4. The Variance of $\mathbf{a}'\hat{\boldsymbol{\beta}}$

Lemma 10.4.1: If $\mathbf{a}'\boldsymbol{\beta}$ is estimable then

$$\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X} = \mathbf{a}'$$

for any generalized inverse $(\mathbf{X}'\mathbf{X})^{-}$.

Proof: If $\mathbf{a}'\boldsymbol{\beta}$ is estimable, then $\mathbf{a} = \mathbf{X}'\mathbf{b}_*$, $\mathbf{b}_* \in \mathcal{R}(\mathbf{X})$ by Lemma 10.2.1. Then

$$\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X} = \mathbf{b}_*'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X} = \mathbf{b}_*'\mathbf{P}\mathbf{X} = \mathbf{b}_*'\mathbf{X} = \mathbf{a}',$$

regardless of the generalized inverse used.

Theorem 10.4.2: If $\mathbf{a}'\boldsymbol{\beta}$ is estimable, then

$$\text{var}(\mathbf{a}'\hat{\boldsymbol{\beta}}) = \sigma^2\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{a}.$$

Proof: Using an estimate $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y}$, $\text{var}(\mathbf{a}'\hat{\boldsymbol{\beta}}) =$

$$\begin{aligned} \text{var}(\mathbf{a}'\hat{\boldsymbol{\beta}}) &= \text{var}(\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y}) \\ &= \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'(\sigma^2\mathbf{I})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{a} \\ &= \sigma^2\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{a} \\ \text{(by the Lemma)} &= \sigma^2\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{a}. \end{aligned}$$

Note that

$$\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{a} = \mathbf{b}_*'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{b}_* = \mathbf{b}_*'\mathbf{P}\mathbf{b}_*$$

is unique (same for all generalized inverses $(\mathbf{X}'\mathbf{X})^{-}$).

In-class exercise: One–way ANOVA with K groups. There are K groups with J observations from each group. The model is

$$Y_{kj} = \mu + \alpha_k + \epsilon_{kj}$$

for $k = 1, \ldots, K$ and $j = 1, \ldots, J$. As usual, $E[\epsilon] = \mathbf{0}$ and $\text{var}(\epsilon) = \sigma^2\mathbf{I}$. In this setting we are almost never interested in the μ parameter (why not?). What are the estimable functions of the α parameters?