10.1. Best Linear Unbiased Estimates

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

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

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

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

**Note:** The Gauss-Markov theorem generalizes this result to the less than full rank case, for *certain* linear combinations $a'\beta$ (the *estimable functions*).
Proof of Corollary 10.1.2:

\[ \theta = X\beta \]
\[ X'\theta = X'X\beta \]
\[ (X'X)^{-1}X'\theta = (X'X)^{-1}X'X\beta = \beta \]
\[ \Rightarrow a'\beta = \underbrace{a'(X'X)^{-1}X'}_{c'} \theta \]

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

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

\[ c'\hat{Y} = a'(X'X)^{-1}X'\hat{Y} \]
\[ = a'(X'X)^{-1}X'X(X'X)^{-1}X'Y \]
\[ = a'(X'X)^{-1}X'Y \]

Therefore, since \( a'\hat{\beta} = c'\hat{Y} \), it is the BLUE of \( a'\beta = c'\theta \).
10. ESTIMABLE FUNCTIONS AND GAUSS-MARKOV THEOREM

10.2. Estimable Functions

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

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

**Lemma 10.2.1:**

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

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

(ii) If $a'\beta$ is estimable, there is a unique $b_\ast \in \mathcal{R}(X)$ such that $a = X'b_\ast$.

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

$$a = X'b = X'b_\ast + X'\tilde{b} = X'b_\ast.$$  

**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 $X$, 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 $X$. Clearly, also, linear combinations of estimable functions should be estimable. These are the vectors that are spanned by the rows of $X$ – the row space of $X$. 
10.3. Gauss-Markov Theorem

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

\[
a'\hat{\beta} = a'(X'X)^{-1}X'Y \equiv b'Y
\]

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

**Theorem 10.3.1:** (Gauss-Markov). If \( a'\beta \) is estimable, then

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

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

**Proof:**

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

\[
a'\hat{\beta} = b_*'X\hat{\beta} = b_*'\hat{Y}
\]

is unique because \( \hat{Y} \) is unique. (In fact \( b_*'\hat{Y} = b_*'Y \) since \( b_* \in \mathcal{R}(X) \), so that \( b_*'(Y - \hat{Y}) = b_*'\hat{e} = 0 \).)

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

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

$$\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = \mathbf{a}'$$

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

**Proof:** If $\mathbf{a}' \mathbf{\beta}$ 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 $\mathbf{a}' \mathbf{\beta}$ is estimable, then

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

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

$$\text{var}(\mathbf{a}' \hat{\mathbf{\beta}}) = \text{var}(\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y})$$

$$= \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'(\sigma^{2}\mathbf{I})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{a}$$

$$= \sigma^{2}\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{a}$$

(by the Lemma) $$= \sigma^{2}\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{a}.$$ 

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, $E[\epsilon] = 0$ and $\text{var}(\epsilon) = \sigma^2 I$. In this setting we are almost never interested in the $\mu$ parameter (why not?). What are the estimable functions of the $\alpha$ parameters?