Gauss Markov & Predictive Distributions
Merlise Clyde

STA721 Linear Models
Duke University

September 14, 2017
Outline

Topics
  ▶ Gauss-Markov Theorem
  ▶ Estimability and Prediction

Readings: Christensen Chapter 2, Chapter 6.3, (Appendix A, and Appendix B as needed)
Gauss-Markov Theorem

Theorem

*Under the assumptions:*

\[ E[Y] = \mu \]
Gauss-Markov Theorem

Theorem

*Under the assumptions:*

\[
E[Y] = \mu \\
\text{Cov}(Y) = \sigma^2 I_n
\]
Gauss-Markov Theorem

Theorem

Under the assumptions:

\[ E[Y] = \mu \]
\[ \text{Cov}(Y) = \sigma^2 I_n \]

every estimable function \( \psi = \lambda^T \beta \) has a unique unbiased linear estimator \( \hat{\psi} \) which has minimum variance in the class of all unbiased linear estimators.
Gauss-Markov Theorem

Theorem

Under the assumptions:

\[ E[Y] = \mu \]
\[ \text{Cov}(Y) = \sigma^2 I_n \]

every estimable function \( \psi = \lambda^T \beta \) has a unique unbiased linear estimator \( \hat{\psi} \) which has minimum variance in the class of all unbiased linear estimators. \( \hat{\psi} = \lambda^T \hat{\beta} \) where \( \hat{\beta} \) is any set of ordinary least squares estimators.
Lemma

- If $\psi = \lambda^T \beta$ is estimable, there exists a unique linear unbiased estimator of $\psi = a^*^T Y$ with $a^* \in C(X)$. 
Unique Unbiased Estimator

Lemma

- If $\psi = \lambda^T \beta$ is estimable, there exists a unique linear unbiased estimator of $\psi = a^*^T Y$ with $a^* \in C(X)$.
- If $a^T Y$ is any unbiased linear estimator of $\psi$ then $a^*$ is the projection of $a$ onto $C(X)$, i.e. $a^* = P_x a$. 
Unique Unbiased Estimator

Proof

- Since $\psi$ is estimable, there exists an $a \in \mathbb{R}^n$ for which
  $E[a^T Y] = \lambda^T \beta = \psi$ with $\lambda^T = a^T X$
Unique Unbiased Estimator

Proof

Since \( \psi \) is estimable, there exists an \( a \in \mathbb{R}^n \) for which
\[
E[a^T Y] = \lambda^T \beta = \psi \text{ with } \lambda^T = a^T X
\]

Let \( a = a^* + u \) where \( a^* \in C(X) \) and \( u \in C(X)^\perp \)
Unique Unbiased Estimator

Proof

- Since $\psi$ is estimable, there exists an $a \in \mathbb{R}^n$ for which $E[a^T Y] = \lambda^T \beta = \psi$ with $\lambda^T = a^T X$
- Let $a = a^* + u$ where $a^* \in C(X)$ and $u \in C(X)^\perp$
- Then

$$\psi = E[a^T Y] = E[a^*^T Y] + E[u^T Y]$$
Unique Unbiased Estimator

Proof

Since $\psi$ is estimable, there exists an $a \in \mathbb{R}^n$ for which

$E[a^T Y] = \lambda^T \beta = \psi$ with $\lambda^T = a^T X$

Let $a = a^* + u$ where $a^* \in C(X)$ and $u \in C(X)^\perp$

Then

\[
\psi = E[a^T Y] = E[a^*^T Y] + E[u^T Y] = E[a^*^T Y] + 0
\]

$E[u^T Y] = u^T X \beta$

since $u \perp C(X)$ (i.e. $u \in C(X)^\perp$) $E[u^T Y] = 0$
Unique Unbiased Estimator

Proof

Since $\psi$ is estimable, there exists an $a \in \mathbb{R}^n$ for which $E[a^T Y] = \lambda^T \beta = \psi$ with $\lambda^T = a^T X$

Let $a = a^* + u$ where $a^* \in C(X)$ and $u \in C(X)^\perp$

Then $\psi = E[a^T Y] = E[a^*^T Y] + E[u^T Y]$

$$= E[a^*^T Y] + 0$$

$$E[u^T Y] = u^T X \beta$$

since $u \perp C(X)$ (i.e. $u \in C(X)^\perp$) $E[u^T Y] = 0$

Thus $a^*^T Y$ is also an unbiased linear estimator of $\psi$ with $a^* \in C(X)$
Uniqueness

Proof.
Suppose that there is another \( v \in C(X) \) such that \( E[v^T Y] = \psi \).
Then for all \( \beta \)

\[
(a^* - v)^T X \beta = 0
\]

implies \( (a^* - v) \in C(X) \perp \), but by assumption \( (a^* - v) \in C(X) \) \( (C(X) \text{ is a vector space}) \)
the only vector in BOTH is 0, so \( a^* = v \)

Therefore \( a^* T Y \) is the unique linear unbiased estimator of \( \psi \) with \( a^* \in C(X) \).
Proof.
Suppose that there is another \( v \in C(X) \) such that \( E[v^T Y] = \psi \).
Then for all \( \beta \)

\[
0 = E[a^*^T Y] - E[v^T Y]
\]

Therefore \( a^*^T Y \) is the unique linear unbiased estimator of \( \psi \) with \( a^* \in C(X) \).
Uniqueness

Proof.
Suppose that there is another \( v \in C(X) \) such that \( E[v^T Y] = \psi \).
Then for all \( \beta \)

\[
0 = E[a^*^T Y] - E[v^T Y] = (a^* - v)^T X \beta
\]

Therefore \( a^* \in C(X) \) is the unique linear unbiased estimator of \( \psi \) with \( a^* \in C(X) \).
Uniqueness

Proof.
Suppose that there is another \( \mathbf{v} \in C(\mathbf{X}) \) such that \( E[\mathbf{v}^T \mathbf{Y}] = \psi \).
Then for all \( \beta \)

\[
0 = E[a^*^T \mathbf{Y}] - E[\mathbf{v}^T \mathbf{Y}]
= (a^* - \mathbf{v})^T \mathbf{X} \beta
\]

So \( (a^* - \mathbf{v})^T \mathbf{X} = 0 \) for all \( \beta \).
Uniqueness

Proof.
Suppose that there is another \( \mathbf{v} \in C(\mathbf{X}) \) such that \( E[\mathbf{v}^T \mathbf{Y}] = \psi \).
Then for all \( \beta \)

\[
0 = E[\mathbf{a}^*^T \mathbf{Y}] - E[\mathbf{v}^T \mathbf{Y}]
= (\mathbf{a}^* - \mathbf{v})^T \mathbf{X} \beta
\]

So \( (\mathbf{a}^* - \mathbf{v})^T \mathbf{X} = 0 \) for all \( \beta \)

- Implies \( (\mathbf{a}^* - \mathbf{v}) \in C(\mathbf{X})^\perp \)
Uniqueness

Proof.
Suppose that there is another $v \in C(X)$ such that $E[v^T Y] = \psi$. Then for all $\beta$

$$0 = E[a^*^T Y] - E[v^T Y]$$
$$= (a^* - v)^T X \beta$$
So $(a^* - v)^T X = 0$ for all $\beta$

- Implies $(a^* - v) \in C(X)^\perp$
- but by assumption $(a^* - v) \in C(X)$
Uniqueness

Proof.
Suppose that there is another $v \in C(X)$ such that $E[v^T Y] = \psi$. Then for all $\beta$

$$0 = E[a^*T Y] - E[v^T Y] = (a^* - v)^T X \beta$$

So $(a^* - v)^T X = 0$ for all $\beta$

- Implies $(a^* - v) \in C(X)^\perp$
- but by assumption $(a^* - v) \in C(X)$ ($C(X)$ is a vector space)
Uniqueness

Proof.
Suppose that there is another \( v \in C(X) \) such that \( E[v^T Y] = \psi \).
Then for all \( \beta \)

\[
0 = E[a^*^T Y] - E[v^T Y] = (a^* - v)^T X \beta
\]

So \( (a^* - v)^T X = 0 \) for all \( \beta \)

- Implies \( (a^* - v) \in C(X) \perp \)
- but by assumption \( (a^* - v) \in C(X) \) (\( C(X) \) is a vector space)
- the only vector in BOTH is \( 0 \), so \( a^* = v \)
Uniqueness

Proof.
Suppose that there is another \( v \in C(X) \) such that \( E[v^T Y] = \psi \).
Then for all \( \beta \)

\[
0 = E[a^*^T Y] - E[v^T Y] = (a^* - v)^T X \beta
\]

So \( (a^* - v)^T X = 0 \) for all \( \beta \)

- Implies \( (a^* - v) \in C(X)^\perp \)
- but by assumption \( (a^* - v) \in C(X) \) (\( C(X) \) is a vector space)
- the only vector in BOTH is \( 0 \), so \( a^* = v \)

Therefore \( a^*^T Y \) is the unique linear unbiased estimator of \( \psi \) with \( a^* \in C(X) \). \( \square \)
Proof of Minimum Variance (G-M)

Let \( a^* \) be the unique unbiased linear estimator of \( \psi \) with \( a^* \in C(X) \).

Let \( a^T Y \) be any unbiased estimate of \( \psi \); \( a^* = a + u \) with \( a^* \in C(X) \) and \( u \in C(X) \perp \).

\[
\text{Var}(a^T Y) = a^T \text{Cov}(Y) a = \sigma^2 \|a\|^2 = \sigma^2 (\|a^*\|^2 + \|u\|^2 + 2a^* u^T) = \sigma^2 (\|a^*\|^2 + \|u\|^2) + 0 = \text{Var}(a^T Y) + \sigma^2 \|u\|^2 \geq \text{Var}(a^T Y)
\]

with equality if and only if \( a = a^* \).

Hence \( a^T Y \) is the unique linear unbiased estimator of \( \psi \) with minimum variance. "BLUE" = Best Linear Unbiased Estimator.
Proof of Minimum Variance (G-M)

- Let \( a^*^T Y \) be the unique unbiased linear estimator of \( \psi \) with \( a^* \in C(X) \).
- Let \( a^T Y \) be any unbiased estimate of \( \psi \); \( a = a^* + u \) with \( a^* \in C(X) \) and \( u \in C(X)\perp \).

\[
\text{Var}(a^T Y) = a^T \text{Cov}(Y) a = \sigma^2 \|a\|^2 = \sigma^2 (\|a^*\|^2 + \|u\|^2 + 2a^*^T u) = \sigma^2 (\|a^*\|^2 + \|u\|^2) + 0 = \text{Var}(a^*^T Y) + \sigma^2 \|u\|^2 \geq \text{Var}(a^*^T Y) \text{ with equality if and only if } a = a^* \text{.}
\]

Hence \( a^*^T Y \) is the unique linear unbiased estimator of \( \psi \) with minimum variance. 

"BLUE" = Best Linear Unbiased Estimator
Proof of Minimum Variance (G-M)

Let $a^* Y$ be the unique unbiased linear estimator of $\psi$ with $a^* \in C(X)$.

Let $a^T Y$ be any unbiased estimate of $\psi$; $a = a^* + u$ with $a^* \in C(X)$ and $u \in C(X)^\perp$

$$\text{Var}(a^T Y) = a^T \text{Cov}(Y)a$$
Proof of Minimum Variance (G-M)

- Let $a^*^T Y$ be the unique unbiased linear estimator of $\psi$ with $a^* \in C(X)$.

- Let $a^T Y$ be any unbiased estimate of $\psi$; $a = a^* + u$ with $a^* \in C(X)$ and $u \in C(X)^\perp$

$$\text{Var}(a^T Y) = a^T \text{Cov}(Y) a$$

$$= \sigma^2 \|a\|^2$$
Proof of Minimum Variance (G-M)

- Let $\mathbf{a}^* \mathbf{Y}$ be the unique unbiased linear estimator of $\psi$ with $\mathbf{a}^* \in C(\mathbf{X})$.

- Let $\mathbf{a}^T \mathbf{Y}$ be any unbiased estimate of $\psi$; $\mathbf{a} = \mathbf{a}^* + \mathbf{u}$ with $\mathbf{a}^* \in C(\mathbf{X})$ and $\mathbf{u} \in C(\mathbf{X})^\perp$

\[
\text{Var}(\mathbf{a}^T \mathbf{Y}) = \mathbf{a}^T \text{Cov}(\mathbf{Y}) \mathbf{a} \\
= \sigma^2 \|\mathbf{a}\|^2 \\
= \sigma^2 (\|\mathbf{a}^*\|^2 + \|\mathbf{u}\|^2 + 2\mathbf{a}^T \mathbf{u})
\]
Proof of Minimum Variance (G-M)

- Let \( a^*^T Y \) be the unique unbiased linear estimator of \( \psi \) with \( a^* \in C(X) \).
- Let \( a^T Y \) be any unbiased estimate of \( \psi \); \( a = a^* + u \) with \( a^* \in C(X) \) and \( u \in C(X) \perp \)

\[
\text{Var}(a^T Y) = a^T \text{Cov}(Y) a = \sigma^2 \|a\|^2 = \sigma^2 (\|a^*\|^2 + \|u\|^2 + 2a^T u) = \sigma^2 (\|a^*\|^2 + \|u\|^2) + 0
\]

Hence \( a^*^T Y \) is the unique linear unbiased estimator of \( \psi \) with minimum variance "BLUE" = Best Linear Unbiased Estimator
Proof of Minimum Variance (G-M)

- Let $a^*^T Y$ be the unique unbiased linear estimator of $\psi$ with $a^* \in C(X)$.
- Let $a^T Y$ be any unbiased estimate of $\psi$; $a = a^* + u$ with $a^* \in C(X)$ and $u \in C(X) \perp$

\[
\text{Var}(a^T Y) = a^T \text{Cov}(Y) a \\
= \sigma^2 \|a\|^2 \\
= \sigma^2 (\|a^*\|^2 + \|u\|^2 + 2a^T u) \\
= \sigma^2 (\|a^*\|^2 + \|u\|^2) + 0 \\
= \text{Var}(a^*^T Y) + \sigma^2 \|u\|^2
\]
Proof of Minimum Variance (G-M)

- Let $a^*^T Y$ be the unique unbiased linear estimator of $\psi$ with $a^* \in C(X)$.

- Let $a^T Y$ be any unbiased estimate of $\psi$; $a = a^* + u$ with $a^* \in C(X)$ and $u \in C(X) \perp$

$$
\text{Var}(a^T Y) = a^T \text{Cov}(Y) a \\
= \sigma^2 \|a\|^2 \\
= \sigma^2 (\|a^*\|^2 + \|u\|^2 + 2a^T u) \\
= \sigma^2 (\|a^*\|^2 + \|u\|^2) + 0 \\
= \text{Var}(a^*^T Y) + \sigma^2 \|u\|^2 \\
\geq \text{Var}(a^*^T Y)
$$
Proof of Minimum Variance (G-M)

Let $a^*^TY$ be the unique unbiased linear estimator of $\psi$ with $a^* \in C(X)$.

Let $a^TY$ be any unbiased estimate of $\psi$; $a = a^* + u$ with $a^* \in C(X)$ and $u \in C(X)^\perp$

$$\begin{align*}
  \text{Var}(a^TY) & = a^T\text{Cov}(Y)a \\
  & = \sigma^2\|a\|^2 \\
  & = \sigma^2(\|a^*\|^2 + \|u\|^2 + 2a^T u) \\
  & = \sigma^2(\|a^*\|^2 + \|u\|^2) + 0 \\
  & = \text{Var}(a^*^TY) + \sigma^2\|u\|^2 \\
  & \geq \text{Var}(a^*^TY)
\end{align*}$$

with equality if and only if $a = a^*$

Hence $a^*^TY$ is the unique linear unbiased estimator of $\psi$ with minimum variance

"BLUE" = Best Linear Unbiased Estimator
Proof of Minimum Variance (G-M)

Let $\mathbf{a}^* \mathbf{Y}$ be the unique unbiased linear estimator of $\psi$ with $\mathbf{a}^* \in C(\mathbf{X})$.

Let $\mathbf{a}^T \mathbf{Y}$ be any unbiased estimate of $\psi$; $\mathbf{a} = \mathbf{a}^* + \mathbf{u}$ with $\mathbf{a}^* \in C(\mathbf{X})$ and $\mathbf{u} \in C(\mathbf{X})^\perp$

\[
\text{Var}(\mathbf{a}^T \mathbf{Y}) = \mathbf{a}^T \text{Cov}(\mathbf{Y}) \mathbf{a} \\
= \sigma^2 \| \mathbf{a} \|^2 \\
= \sigma^2 (\| \mathbf{a}^* \|^2 + \| \mathbf{u} \|^2 + 2 \mathbf{a}^T \mathbf{u}) \\
= \sigma^2 (\| \mathbf{a}^* \|^2 + \| \mathbf{u} \|^2) + 0 \\
= \text{Var}(\mathbf{a}^* \mathbf{Y}) + \sigma^2 \| \mathbf{u} \|^2 \\
\geq \text{Var}(\mathbf{a}^* \mathbf{Y})
\]

with equality if and only if $\mathbf{a} = \mathbf{a}^*$

Hence $\mathbf{a}^* \mathbf{Y}$ is the unique linear unbiased estimator of $\psi$ with minimum variance
Proof of Minimum Variance (G-M)

- Let $a^*^T Y$ be the unique unbiased linear estimator of $\psi$ with $a^* \in C(X)$.

- Let $a^T Y$ be any unbiased estimate of $\psi$; $a = a^* + u$ with $a^* \in C(X)$ and $u \in C(X) \perp$

\[
\text{Var}(a^T Y) = a^T \text{Cov}(Y)a \\
= \sigma^2 \|a\|^2 \\
= \sigma^2 (\|a^*\|^2 + \|u\|^2 + 2a^T u) \\
= \sigma^2 (\|a^*\|^2 + \|u\|^2) + 0 \\
= \text{Var}(a^*^T Y) + \sigma^2 \|u\|^2 \\
\geq \text{Var}(a^*^T Y)
\]

with equality if and only if $a = a^*$

Hence $a^*^T Y$ is the unique linear unbiased estimator of $\psi$ with minimum variance "BLUE" = Best Linear Unbiased Estimator.
Proof.
Show that $\hat{\psi} = a^*^T Y = \lambda^T \hat{\beta}$
Proof.
Show that $\hat{\psi} = a^*^T Y = \lambda^T \hat{\beta}$
Since $a^* \in C(X)$ we have $a^* = P_X a^*$
Proof.
Show that \( \hat{\psi} = a^* \, T \, Y = \lambda^T \hat{\beta} \)
Since \( a^* \in C(X) \) we have \( a^* = P_X a^* \)

\[
a^* \, T \, Y = a^* \, T \, P_X \, T \, Y
\]
Proof.
Show that $\hat{\psi} = a^*^T Y = \lambda^T \hat{\beta}$
Since $a^* \in C(X)$ we have $a^* = P_X a^*$

\[
\begin{align*}
a^*^T Y &= a^*^T P_X^T Y \\
&= a^*^T P_X Y
\end{align*}
\]
Continued

Proof.
Show that $\hat{\psi} = a^*^T Y = \lambda^T \hat{\beta}$
Since $a^* \in C(X)$ we have $a^* = P_X a^*$

$$a^*^T Y = a^*^T P_Y^T Y = a^*^T P_X Y = a^*^T X \hat{\beta}$$
Proof.
Show that \( \hat{\psi} = a^*^T Y = \lambda^T \hat{\beta} \)
Since \( a^* \in C(X) \) we have \( a^* = P_X a^* \)

\[
a^*^T Y = a^*^T P_X^T Y = a^*^T P_X Y = a^*^T X \hat{\beta} = \lambda^T \hat{\beta}
\]
Proof.

Show that \( \hat{\psi} = a^*^T Y = \lambda^T \hat{\beta} \)

Since \( a^* \in C(X) \) we have \( a^* = P_X a^* \)

\[
a^*^T Y = a^*^T P_X^T Y \\
= a^*^T P_X Y \\
= a^*^T X \hat{\beta} \\
= \lambda^T \hat{\beta}
\]

for \( \lambda^T = a^{*^T} X \) or \( \lambda = X^T a \)
Gauss-Markov Theorem says that OLS has minimum variance in the class of all Linear Unbiased estimators.
Gauss-Markov Theorem says that OLS has minimum variance in the class of all Linear Unbiased estimators.

- Requires just first and second moments.
Gauss-Markov Theorem says that OLS has minimum variance in the class of all Linear Unbiased estimators

- Requires just first and second moments
- Additional assumption of normality, OLS = MLEs have minimum variance out of ALL unbiased estimators (MVUE); not just linear estimators
MVUE

- Gauss-Markov Theorem says that OLS has minimum variance in the class of all Linear Unbiased estimators
- Requires just first and second moments
- Additional assumption of normality, OLS = MLEs have minimum variance out of ALL unbiased estimators (MVUE); not just linear estimators (requires Completeness and Rao-Blackwell Theorem - next semester)
Gauss-Markov Theorem says that OLS has minimum variance in the class of all Linear Unbiased estimators.

Requires just first and second moments.

Additional assumption of normality, OLS = MLEs have minimum variance out of ALL unbiased estimators (MVUE); not just linear estimators (requires Completeness and Rao-Blackwell Theorem - next semester)
Prediction

For predicting at new $x_*$ is there always a unique unbiased estimator of $E[Y \mid x_*]$?
Prediction

- For predicting at new $x_*$ is there always a unique unbiased estimator of $E[Y \mid x_*]$?
- If one does exist, how do we know that if we are given $\lambda$?
Existence

- $x^* \beta$ has a unique unbiased estimator if $x^* \equiv \lambda = X^T a$

What about out of sample prediction?
Existence

- $x_\beta$ has a unique unbiased estimator if $x_\star \equiv \lambda = X^T a$
- Clearly if $x_\star = x_i$ (ith row of observed data) then it is estimable with $a$ equal to the vector with a 1 in the $i$th position even if $X$ is not full rank!
Existence

- $\mathbf{x}_* \beta$ has a unique unbiased estimator if $\mathbf{x}_* \equiv \lambda = \mathbf{X}^T \mathbf{a}$
- Clearly if $\mathbf{x}_* = \mathbf{x}_i$ (ith row of observed data) then it is estimable with $\mathbf{a}$ equal to the vector with a 1 in the ith position even if $\mathbf{X}$ is not full rank!
- What about out of sample prediction?
Existence

- \( \mathbf{x}_* \beta \) has a unique unbiased estimator if \( \mathbf{x}_* \equiv \lambda = \mathbf{X}^T \mathbf{a} \)
- Clearly if \( \mathbf{x}_* = \mathbf{x}_i \) (ith row of observed data) then it is estimable with \( \mathbf{a} \) equal to the vector with a 1 in the ith position even if \( \mathbf{X} \) is not full rank!
- What about out of sample prediction?
Example

\[
x_1 = -4:4 \\
x_2 = c(-2, 1, -1, 2, 0, 2, -1, 1, -2) \\
x_3 = 3x_1 - 2x_2 \\
x_4 = x_2 - x_1 + 4 \\
Y = 1 + x_1 + x_2 + x_3 + x_4 + \begin{bmatrix} -.5 & .5 & .5 & -.5 & 0 & .5 & -.5 & -.5 & .5 \end{bmatrix} \\
\text{dev.set} = \text{data.frame}(Y, x_1, x_2, x_3, x_4) \\
\text{lm1234} = \text{lm}(Y \sim x_1 + x_2 + x_3 + x_4, \text{data}=\text{dev.set}) \\
\text{round}(\text{coefficients} (\text{lm1234}), 4) \\
\begin{array}{cccccc}
\text{(Intercept)} & x_1 & x_2 & x_3 & x_4 \\
5 & 3 & 0 & \text{NA} & \text{NA} \\
\end{array}
\]

\[
\text{lm3412} = \text{lm}(Y \sim x_3 + x_4 + x_1 + x_2, \text{data} = \text{dev.set}) \\
\text{round}(\text{coefficients} (\text{lm3412}), 4) \\
\begin{array}{cccccc}
\text{(Intercept)} & x_3 & x_4 & x_1 & x_2 \\
-19 & 3 & 6 & \text{NA} & \text{NA} \\
\end{array}
\]
In Sample Predictions

cbind(dev.set, predict(lm1234), predict(lm3412))

## Y x1 x2 x3 x4 predict(lm1234) predict(lm3412)
## 1 -7.5 -4 -2 -8 6  -7  -7
## 2 -3.5 -3  1 -11 8  -4  -4
## 3 -0.5 -2 -1 -4 5  -1  -1
## 4  1.5 -1  2 -7 7   2   2
## 5  5.0  0  0  0 4   5   5
## 6  8.5  1  2 -1 5   8   8
## 7 10.5  2 -1  8 1  11  11
## 8 13.5  3  1  7 2  14  14
## 9 17.5  4 -2 16 -2  17  17

Both models agree for estimating the mean at the observed X points!
### Out of Sample

```r
out = data.frame(test.set,
                 Y1234=predict(lm1234, new=test.set),
                 Y3412=predict(lm3412, new=test.set))
out
```

```
##         x1 x2 x3 x4  Y1234  Y3412
## 1  3  1  7  2   14     14
## 2  6  2 14  4   23    47
## 3  6  2 14  0   23     23
## 4  0  0  0  4    5      5
## 5  0  0  0  0    5    -19
## 6  1  2  3  4    8    14
```

Agreement for cases 1, 3, and 4 only! Can we determine that without finding the predictions and comparing?
Out of Sample

```r
out = data.frame(test.set,
                 Y1234=predict(lm1234, new=test.set),
                 Y3412=predict(lm3412, new=test.set))
out
```

```
##   x1 x2 x3 x4  Y1234  Y3412
## 1  1  3  1  7   14     14
## 2  2  6  2 14   23    47
## 3  6  2 14  0   23     23
## 4  0  0  0  4    5     5
## 5  0  0  0  0    5   -19
## 6  1  2  3  4    8    14
```

Agreement for cases 1, 3, and 4 only! Can we determine that without finding the predictions and comparing?
Determining Estimable $\lambda$

- Estimable means that $\lambda^T = a^T X$ for $a \in C(X)$

$\lambda \perp C(X)^T$ is the null space of $Xv \perp C(X)^T$: $Xv = 0 \iff v \in N(X)$

If $P$ is a projection onto $C(X)^T$ then $I - P$ is a projection onto $N(X)$ and therefore $(I - P)\lambda = 0$ if $\lambda$ is estimable.
Determining Estimable $\lambda$

- Estimable means that $\lambda^T = a^T X$ for $a \in C(X)$
- Transpose: $\lambda = X^T a$ for $a \in C(X)$
Determining Estimable $\lambda$

- Estimable means that $\lambda^T = a^T X$ for $a \in C(X)$
- Transpose: $\lambda = X^T a$ for $a \in C(X)$
- $\lambda \in C(X^T)$ ($\lambda \in R(X)$)

$\lambda \perp C(X^T)$, $\lambda \perp N(X)$

If $P$ is a projection onto $C(X^T)$ then $I - P$ is a projection onto $N(X)$ and therefore $(I - P)\lambda = 0$ if $\lambda$ is estimable.
Determining Estimable $\lambda$

- Estimable means that $\lambda^T = a^T X$ for $a \in C(X)$
- Transpose: $\lambda = X^T a$ for $a \in C(X)$
- $\lambda \in C(X^T)$ ($\lambda \in R(X)$)
- $\lambda \perp C(X^T)\perp$
Determining Estimable $\lambda$

- Estimable means that $\lambda^T = a^T X$ for $a \in C(X)$
- Transpose: $\lambda = X^T a$ for $a \in C(X)$
- $\lambda \in C(X^T) \ (\lambda \in R(X))$
- $\lambda \perp C(X^T)^\perp$
- $C(X^T)^\perp$ is the null space of $X$
Determining Estimable $\lambda$

- Estimable means that $\lambda^T = a^T X$ for $a \in C(X)$
- Transpose: $\lambda = X^T a$ for $a \in C(X)$
- $\lambda \in C(X^T)$ ($\lambda \in R(X)$)
- $\lambda \perp C(X^T)^\perp$
- $C(X^T)^\perp$ is the null space of $X$

\[ v \perp C(X^T) : Xv = 0 \iff v \in N(X) \]
Determining Estimable $\lambda$

- Estimable means that $\lambda^T = a^T X$ for $a \in C(X)$
- Transpose: $\lambda = X^T a$ for $a \in C(X)$
- $\lambda \in C(X^T)$ ($\lambda \in R(X)$)
- $\lambda \perp C(X^T)^\perp$
- $C(X^T)^\perp$ is the null space of $X$

\[
v \perp C(X^T) : Xv = 0 \iff v \in N(X)
\]

- $\lambda \perp N(X)$
Determining Estimable \( \lambda \)

- Estimable means that \( \lambda^T = a^T X \) for \( a \in C(X) \)
- Transpose: \( \lambda = X^T a \) for \( a \in C(X) \)
- \( \lambda \in C(X^T) \) (\( \lambda \in R(X) \))
- \( \lambda \perp C(X^T)^\perp \)
- \( C(X^T)^\perp \) is the null space of \( X \)

\[
\mathbf{v} \perp C(X^T) : X\mathbf{v} = 0 \iff \mathbf{v} \in N(X)
\]

- \( \lambda \perp N(X) \)
- if \( P \) is a projection onto \( C(X^T) \) then \( I - P \) is a projection onto \( N(X) \) and therefore \( (I - P)\lambda = 0 \) if \( \lambda \) is estimable
Determining Estimable $\lambda$

- Estimable means that $\lambda^T = a^T X$ for $a \in C(X)$
- Transpose: $\lambda = X^T a$ for $a \in C(X)$
- $\lambda \in C(X^T)$ ($\lambda \in R(X)$)
- $\lambda \perp C(X^T)^\perp$
- $C(X^T)^\perp$ is the null space of $X$

\[ v \perp C(X^T) : Xv = 0 \iff v \in N(X) \]

- $\lambda \perp N(X)$
- if $P$ is a projection onto $C(X^T)$ then $I - P$ is a projection onto $N(X)$ and therefore $(I - P)\lambda = 0$ if $\lambda$ is estimable

Take $P_{X^T} = (X^T X)(X^T X)^{-}$ as a projection onto $C(X^T)$ and show $(I - P_{X^T})\lambda = 0_p$
Example

```r
library("estimability")
cbind(epredict(lm1234, test.set), epredict(lm3412, test.set))
```

```r
## [,1] [,2]
## 1   14  14
## 2 NA  NA
## 3   23  23
## 4   5   5
## 5 NA  NA
## 6 NA  NA
```

Rows 2, 5, and 6 are not estimable! No linear unbiased estimator
Summary

- When BLUEs exist, under normality they are MVUE (ditto for prediction - BLUP)
Summary

- When BLUEs exist, under normality they are MVUE (ditto for prediction - BLUP)
- BLUE/BLUP do not always for estimation/prediction if $\mathbf{X}$ is not full rank
Summary

- When BLUEs exist, under normality they are MVUE (ditto for prediction - BLUP)
- BLUE/BLUP do not always for estimation/prediction if $\mathbf{X}$ is not full rank
- May occur with redundancies for modest $p < n$ and of course $p > n$
Summary

- When BLUEs exist, under normality they are MVUE (ditto for prediction - BLUP)
- BLUE/BLUP do not always for estimation/prediction if $\mathbf{X}$ is not full rank
- may occur with redundancies for modest $p < n$ and of course $p > n$
- Eliminate redundancies by removing variables (variable selection)
Summary

- When BLUEs exist, under normality they are MVUE (ditto for prediction - BLUP)
- BLUE/BLUP do not always for estimation/prediction if $\mathbf{X}$ is not full rank
- may occur with redundancies for modest $p < n$ and of course $p > n$
- Eliminate redundancies by removing variables (variable selection)
- Consider alternative estimators (Bayes and related)
What about some estimator $g(Y)$ that is not unbiased?
Other Estimators

What about some estimator \( g(Y) \) that is not unbiased?

- Mean Squared Error for estimator \( g(Y) \) of \( \lambda^T \beta \) is

\[
E[g(Y) - \lambda^T \beta]^2 = \text{Var}(g(Y)) + \text{Bias}^2(g(Y))
\]

where \( \text{Bias} = E[g(Y)] - \lambda^T \beta \).
Other Estimators

What about some estimator $g(Y)$ that is not unbiased?

- Mean Squared Error for estimator $g(Y)$ of $\lambda^T \beta$ is
  \[
  E[g(Y) - \lambda^T \beta]^2 = \text{Var}(g(Y)) + \text{Bias}^2(g(Y))
  \]
  where $\text{Bias} = E[g(Y)] - \lambda^T \beta$

- Bias vs Variance tradeoff
Other Estimators

What about some estimator $g(Y)$ that is not unbiased?

- Mean Squared Error for estimator $g(Y)$ of $\lambda^T \beta$ is

$$E[g(Y) - \lambda^T \beta]^2 = \text{Var}(g(Y)) + \text{Bias}^2(g(Y))$$

where $\text{Bias} = E[g(Y)] - \lambda^T \beta$

- Bias vs Variance tradeoff

- Can have smaller MSE if we allow some Bias!
Bayes

- Next Class Bayes Theorem & Conjugate Normal-Gamma Prior/Posterior distributions
- Read Chapter 2 in Christensen or Wakefield 5.7
- Review Multivariate Normal and Gamma distributions