Best Linear Unbiased Estimators

STA 721: Lecture 4

Merlise Clyde (clyde@duke.edu)
Duke University

Outline

- Characterizing Linear Unbiased Estimators
- Gauss-Markov Theorem
- Best Linear Unbiased Estimators

Readings: - Christensen Chapter 1-2 and Appendix B - Seber & Lee Chapter 3

Full Rank Case

- Model: $\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$
- Minimal Assumptions:
  - Mean $\boldsymbol{\mu} \in C(\mathbf{X})$ for $\mathbf{X} \in \mathbb{R}^{n \times p}$
  - Errors $\textsf{E}[\boldsymbol{\epsilon}] = \mathbf{0}_n$

▼ **Definition:** Linear Unbiased Estimators (LUEs)

An estimator $\tilde{\boldsymbol{\beta}}$ is a **Linear Unbiased Estimator** (LUE) of $\boldsymbol{\beta}$ if

1. linearity: $\tilde{\boldsymbol{\beta}} = \mathbf{A}\mathbf{Y}$ for $\mathbf{A} \in \mathbb{R}^{p \times n}$
2. unbiasedness: $\textsf{E}[\tilde{\boldsymbol{\beta}}] = \boldsymbol{\beta}$ for all $\boldsymbol{\beta} \in \mathbb{R}^p$

The class of linear unbiased estimators is the same for every model with parameter space $\boldsymbol{\beta} \in \mathbb{R}^p$ and $P \in \mathcal{P}$, for any collection $\mathcal{P}$ of mean-zero distributions over $\mathbb{R}^n$.

Linear Unbiased Estimators (LUEs)

- Let N be an ONB for $\boldsymbol{\mathcal{N}} = \boldsymbol{\mathcal{M}}^{\perp} = N(\mathbf{X}^T)$:
  - $\mathsf{N}^T \mathbf{m} = \mathsf{N}^T \mathbf{X}\mathbf{b} = \mathbf{0} \quad \forall \mathbf{m} = \mathbf{X}\mathbf{b} \in \boldsymbol{\mathcal{M}}$
  - $\mathsf{N}^T \mathsf{N} = \mathbf{I}_{n-p}$

Consider another linear estimator $\tilde{\boldsymbol{\beta}} = \mathbf{A}\mathbf{Y}$

LUEs continued

Since each column of $\mathbf{H}$ is in $\boldsymbol{\mathcal{N}}$ there exists a $\mathbf{G} \in \mathbb{R}^{p \times (n-p)} \ni \mathbf{H} = \mathsf{N}\mathbf{G}^T$

Rewriting $\delta = \tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}$:

$$\begin{aligned}
\tilde{\boldsymbol{\beta}} &= \hat{\boldsymbol{\beta}} + \delta \\
&= \hat{\boldsymbol{\beta}} + \mathbf{H}^T\mathbf{Y} \\
&= \hat{\boldsymbol{\beta}} + \mathbf{G}\mathsf{N}^T\mathbf{Y}
\end{aligned}$$

- therefore $\tilde{\boldsymbol{\beta}}$ is linear and unbiased:

$$\begin{aligned}
\mathsf{E}[\tilde{\boldsymbol{\beta}}] &= \mathsf{E}[\hat{\boldsymbol{\beta}} + \mathbf{G}\mathsf{N}^T\mathbf{Y}] \\
&= \boldsymbol{\beta} + \mathsf{E}[\mathbf{G}\mathsf{N}^T\mathbf{X}\boldsymbol{\beta}] \\
&= \boldsymbol{\beta}
\end{aligned}$$

Characterization of LUEs

Summary of previous results:

▼ Theorem

An estimator $\tilde{\boldsymbol{\beta}}$ is a linear unbiased estimator of $\boldsymbol{\beta}$ in a linear statistical model if and only if

$$\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} + \mathbf{H}^T\mathbf{Y}$$

for some $\mathbf{H} \in \mathbb{R}^{n \times p}$ such that $\mathbf{X}^T\mathbf{H} = \mathbf{0}$ or equivalently for some $\mathbf{G} \in \mathbb{R}^{p \times (n-p)}$

$$\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} + \mathbf{G}\mathrm{N}^T\mathbf{Y}$$

Numerical

```r
# X is model matrix; Y is response
  p = ncol(X)
  n = nrow(X)
  G = matrix(rnorm(p*(n-p)), nrow=p, ncol=n-p)
  H = MASS::Null(X) %*% t(G)
  btilde = bhat + t(H) %*% Y
```

infinite number of LUEs!

LUEs via Generalized Inverses

Let $\tilde{\boldsymbol{\beta}} = \mathbf{A}\mathbf{Y}$ be a LUE in the statistical linear model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ with $\mathbf{X}$ full column rank p

$$\begin{aligned}\textsf{E}[\tilde{\boldsymbol{\beta}}] &= \textsf{E}[\mathbf{A}\mathbf{Y}] \\ &= \mathbf{A}\textsf{E}[\mathbf{Y}] \\ &= \mathbf{A}\mathbf{X}\boldsymbol{\beta} \quad \forall \boldsymbol{\beta} \in \mathbb{R}^p\end{aligned}$$

- Must have $\mathbf{A}\mathbf{X} = \mathbf{I}_p$ ($\mathbf{A}$ is a generalized inverse of $\mathbf{X}$)
- $\mathbf{X}\mathbf{X}^{-}\mathbf{X} = \mathbf{X}$
- one generalized inverse is $\mathbf{X}_{MP}^{-} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$
- $\mathbf{X}_{MP}^{-} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T = \mathbf{V}\boldsymbol{\Delta}^{-1}\mathbf{U}^T$ (using SVD of $\mathbf{X} = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}^T$)
- $\mathbf{A}$ is a generalized inverse of $\mathbf{X}$ iff $\mathbf{A} = \mathbf{X}_{MP}^{-} + \mathbf{H}^T$ for $\mathbf{H} \in \mathbb{R}^{n \times p} \ni \mathbf{H}^T\mathbf{U} = \mathbf{0}$
- $\mathbf{A}\mathbf{Y} = (\mathbf{X}_{MP}^{-} + \mathbf{H}^T)\mathbf{Y} = \hat{\boldsymbol{\beta}} + \mathbf{H}^T\mathbf{Y}$

Best Linear Unbiased Estimators

- the distribution of values of any unbiased estimator is *centered* around $\boldsymbol{\beta}$
- out of the infinite number of LUEs is there one that is more *concentrated* around $\boldsymbol{\beta}$?
- is there an unbiased estimator that has a lower variance than all other unbiased estimators?
- Recall variance-covariance matrix of a random vector $\mathbf{Z}$ with mean $\boldsymbol{\theta}$

$$\textsf{Cov}[\mathbf{Z}] \equiv \textsf{E}[(\mathbf{Z} - \boldsymbol{\theta})(\mathbf{Z} - \boldsymbol{\theta})^T]$$
$$\textsf{Cov}[\mathbf{Z}]_{ij} = \textsf{E}[(z_i - \theta_i)(z_j - \theta_j)]$$

Lemma

Let $\mathbf{A} \in \mathbb{R}^{q \times p}$ and $\mathbf{b} \in \mathbb{R}^q$ with $\mathbf{Z}$ a random vector in $\mathbb{R}^p$ then

$$\textsf{Cov}[\mathbf{A}\mathbf{Z} + \mathbf{b}] = \mathbf{A}\textsf{Cov}[\mathbf{Z}]\mathbf{A}^T \geq 0$$

Variance of Linear Unbiased Estimators

Let's look at the variance of any LUE under assumption $\textsf{Cov}[\boldsymbol{\epsilon}] = \sigma^2 \mathbf{I}_n$

- for $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} = \boldsymbol{\beta} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\epsilon}$

$$\begin{aligned}
\textsf{Cov}[\hat{\boldsymbol{\beta}}] &= \textsf{Cov}[\boldsymbol{\beta} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\epsilon}] \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\textsf{Cov}[\boldsymbol{\epsilon}]\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}
\end{aligned}$$

- Covariance is increasing in σ^2 and generally decreasing in n
- Rewrite $\mathbf{X}^T\mathbf{X}$ as $\mathbf{X}^T\mathbf{X} = \sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T$ (a sum of n outer-products)

Variance of Arbitrary LUE

- for $\tilde{\boldsymbol{\beta}} = \left((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T + \mathbf{H}^T\right)\mathbf{Y} = \boldsymbol{\beta} + \left((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T + \mathbf{H}^T\right)\boldsymbol{\epsilon}$
- recall $\mathbf{X}_{MP}^{-} \equiv (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$

$$\begin{aligned}
\textsf{Cov}[\tilde{\boldsymbol{\beta}}] &= \textsf{Cov}[\left(\mathbf{X}_{MP}^{-} + \mathbf{H}^T\right)\boldsymbol{\epsilon}] \\
&= \sigma^2 \left(\mathbf{X}_{MP}^{-} + \mathbf{H}^T\right)\left(\mathbf{X}_{MP}^{-} + \mathbf{H}^T\right)^T \\
&= \sigma^2 \left(\mathbf{X}_{MP}^{-}(\mathbf{X}_{MP}^{-})^T + \mathbf{X}_{MP}^{-}\mathbf{H} + \mathbf{H}^T(\mathbf{X}_{MP}^{-})^T + \mathbf{H}^T\mathbf{H}\right) \\
&= \sigma^2 \left((\mathbf{X}^T\mathbf{X})^{-1} + \mathbf{H}^T\mathbf{H}\right)
\end{aligned}$$

- Cross-product term $\mathbf{H}^T(\mathbf{X}_{MP}^{-})^T = \mathbf{H}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} = \mathbf{0}$
- Therefor the $\textsf{Cov}[\tilde{\boldsymbol{\beta}}] = \textsf{Cov}[\hat{\boldsymbol{\beta}}] + \mathbf{H}^T\mathbf{H}$
- the sum of a positive definite matrix plus a positive semi-definite matrix

Gauss-Markov Theorem

Is Cov$[\tilde{\boldsymbol{\beta}}] \geq$ Cov$[\hat{\boldsymbol{\beta}}]$ in some sense?

Definition: Loewner Ordering

For two positive semi-definite matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$, we say that $\boldsymbol{\Sigma}_1 > \boldsymbol{\Sigma}_2$ if $\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2$ is positive definite, $\mathbf{x}^T(\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2)\mathbf{x}) > 0$, and $\boldsymbol{\Sigma}_1 \geq \boldsymbol{\Sigma}_2$ if $\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2$ is positive semi-definite, $\mathbf{x}^T(\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2)\mathbf{x}) \geq 0$

- Since Cov$[\tilde{\boldsymbol{\beta}}] -$ Cov$[\hat{\boldsymbol{\beta}}] = \mathbf{H}^T\mathbf{H}$, we have that Cov$[\tilde{\boldsymbol{\beta}}] \geq$ Cov$[\hat{\boldsymbol{\beta}}]$

Theorem: Gauss-Markov

Let $\tilde{\boldsymbol{\beta}}$ be a linear unbiased estimator of $\boldsymbol{\beta}$ in a linear model where E$[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta}$, $\boldsymbol{\beta} \in \mathbb{R}^p$, $\mathbf{X}$ rank p, and Cov$[\mathbf{Y}] = \sigma^2\mathbf{I}_n$, $\sigma^2 > 0$. Then Cov$[\tilde{\boldsymbol{\beta}}] \geq$ Cov$[\hat{\boldsymbol{\beta}}]$ where $\hat{\boldsymbol{\beta}}$ is the OLS estimator and is the **Best Linear Unbiased Estimator** (BLUE) of $\boldsymbol{\beta}$.

▼ **Theorem:** Gauss-Markov Theorem (Classic)

For $\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$, with $\boldsymbol{\mu} \in \boldsymbol{\mathcal{M}}$, $\textsf{E}[\boldsymbol{\epsilon}] = \mathbf{0}_n$ and $\textsf{Cov}[\boldsymbol{\epsilon}] = \sigma^2 \mathbf{I}_n$ and $\mathbf{P}$ the orthogonal projection onto $\boldsymbol{\mathcal{M}}$, $\mathbf{P}\mathbf{Y} = \hat{\boldsymbol{\mu}}$ is the BLUE of $\boldsymbol{\mu}$ out of the class of LUEs $\mathbf{A}\mathbf{Y}$ where $\textsf{E}[\mathbf{A}\mathbf{Y}] = \boldsymbol{\mu}$, $\mathbf{A} \in \mathbb{R}^{n \times n}$ equality iff $\mathbf{A} = \mathbf{P}$

▶ **Proof**

Estimation of Linear Functionals of $\boldsymbol{\mu}$

If $\mathbf{PY} = \hat{\boldsymbol{\mu}}$ is the BLUE of $\boldsymbol{\mu}$, is $\mathbf{BPY} = \mathbf{B}\hat{\boldsymbol{\mu}}$ the BLUE of $\mathbf{B}\boldsymbol{\mu}$?

Yes! Similar proof as above to show that out of the class of LUEs $\mathbf{AY}$ of $\mathbf{B}\boldsymbol{\mu}$ where $\mathbf{A} \in \mathbb{R}^{d \times n}$ that

$$\mathsf{E}[\|\mathbf{AY} - \mathbf{B}\boldsymbol{\mu}\|^2] \geq \mathsf{E}[\|\mathbf{BPY} - \mathbf{B}\boldsymbol{\mu}\|^2]$$

with equality iff $\mathbf{A} = \mathbf{BP}$.

What about linear functionals of $\boldsymbol{\beta}$, $\boldsymbol{\Lambda}^T\boldsymbol{\beta}$, for $\mathbf{X}$ rank $r \leq p$?

- $\hat{\boldsymbol{\beta}}$ is not unique if $r < p$ even though $\hat{\boldsymbol{\mu}}$ is unique ($\hat{\boldsymbol{\beta}}$ is not BLUE)
- Since $\mathbf{B}\boldsymbol{\mu} = \mathbf{BX}\boldsymbol{\beta}$ is always identifiable, the only linear functions of $\boldsymbol{\beta}$ that are identifiable and can be estimated uniquely are functions of $\mathbf{X}\boldsymbol{\beta}$, i.e. estimates in the form $\boldsymbol{\Lambda}^T\boldsymbol{\beta} = \mathbf{BX}\boldsymbol{\beta}$ or $\boldsymbol{\Lambda} = \mathbf{X}^T\mathbf{B}^T$.
- columns of $\boldsymbol{\Lambda}$ must be in the $C(\mathbf{X}^T)$
- detailed discussion and proof in Christensen Ch. 2 for scalar functionals $\lambda^T\beta$.

BLUE of $\boldsymbol{\Lambda\beta}$

If $\mathbf{\Lambda}^T = \mathbf{B}\mathbf{X}$ for some matrix $\mathbf{B}$ then

- $\textsf{E}[\mathbf{B}\mathbf{P}\mathbf{Y}] = \textsf{E}[\mathbf{\Lambda}\hat{\boldsymbol{\beta}}] = \mathbf{\Lambda}^T\boldsymbol{\beta}$
- The unique OLS estimate of $\mathbf{\Lambda}^T\boldsymbol{\beta}$ is $\mathbf{\Lambda}^T\hat{\boldsymbol{\beta}}$
- $\mathbf{B}\mathbf{P}\mathbf{Y} = \mathbf{\Lambda}^T\hat{\boldsymbol{\beta}}$ is the BLUE of $\mathbf{\Lambda}^T\boldsymbol{\beta}$

$$\textsf{E}[\|\mathbf{B}\mathbf{P}\mathbf{Y} - \mathbf{B}\boldsymbol{\mu}\|^2] \le \textsf{E}[\|\mathbf{A}\mathbf{Y} - \mathbf{B}\boldsymbol{\mu}\|^2]$$
$$\Leftrightarrow$$
$$\textsf{E}[\|\mathbf{\Lambda}^T\hat{\boldsymbol{\beta}} - \mathbf{\Lambda}^T\boldsymbol{\beta})\|^2] \le \textsf{E}[\|\mathbf{L}^T\hat{\boldsymbol{\beta}} - \mathbf{\Lambda}^T\boldsymbol{\beta}\|^2]$$

for LUE $\mathbf{A}\mathbf{Y}$ and $\mathbf{L}^T\hat{\boldsymbol{\beta}}$ of $\mathbf{\Lambda}^T\boldsymbol{\beta}$

Proof of Cross-Product

Let $\mathbf{D} = \mathbf{HP}$ and write

$$\begin{aligned}\mathsf{E}[(\mathbf{H}^T(\mathbf{Y} - \mu))^T\mathbf{P}(\mathbf{Y} - \boldsymbol{\mu})] &= \mathsf{E}[(\mathbf{Y} - \mu))^T\mathbf{HP}(\mathbf{Y} - \boldsymbol{\mu})] \\ &= \mathsf{E}[(\mathbf{Y} - \mu))^T\mathbf{D}(\mathbf{Y} - \boldsymbol{\mu})]\end{aligned}$$

$$\begin{aligned}\mathsf{E}[(\mathbf{Y} - \mu))^T\mathbf{D}(\mathbf{Y} - \boldsymbol{\mu})] =& \mathsf{E}[\mathsf{tr}(\mathbf{Y} - \mu))^T\mathbf{D}(\mathbf{Y} - \boldsymbol{\mu}))] \\ =& \mathsf{E}[\mathsf{tr}(\mathbf{D}(\mathbf{Y} - \boldsymbol{\mu})(\mathbf{Y} - \mu)^T)] \\ =& \mathsf{tr}(\mathsf{E}[\mathbf{D}(\mathbf{Y} - \boldsymbol{\mu})(\mathbf{Y} - \mu)^T]) \\ =& \mathsf{tr}(\mathbf{D}\mathsf{E}[(\mathbf{Y} - \boldsymbol{\mu})(\mathbf{Y} - \mu)^T]) \\ =& \sigma^2\mathsf{tr}(\mathbf{D}\mathbf{I}_n)\end{aligned}$$

Since $\mathsf{tr}(\mathbf{D}) = \mathsf{tr}(\mathbf{HP}) = \mathsf{tr}(\mathbf{PH})$ we can conclude that the cross-product term is zero.