

Stat 608
BLUE Notes



BLUE: Best Linear Unbiased Estimator



- The Gauss-Markov Theorem says that our parameter estimate vector is BLUE:
 - Best: Minimum Variance
 - Linear: A linear combination of Y 's (we can write $\hat{\beta}$ as $\mathbf{a}'\mathbf{y}$ for some vector \mathbf{a})
 - Unbiased: That is, $E[\hat{\beta}|X] = \beta$
 - Estimator: A statistic
- For estimating the mean of the normal distribution, we like the sample mean better than the sample median because of the smaller variability.

+ Gauss-Markov Theorem

- Consider the linear model: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$
- Assume that the errors have mean and covariance 0, that the variance of the errors is constant, and that the design matrix \mathbf{X} is full rank. Then:

For any $\tilde{\boldsymbol{\beta}}$ a linear combination of \mathbf{y}
such that $E[\tilde{\boldsymbol{\beta}}] = \boldsymbol{\beta}$,
$$\text{Var}(\mathbf{c}'\tilde{\boldsymbol{\beta}}) \geq \text{Var}(\mathbf{c}'\hat{\boldsymbol{\beta}})$$

That is, $\mathbf{c}'\hat{\boldsymbol{\beta}}$ is the Best Linear Unbiased Estimator of $\mathbf{c}'\boldsymbol{\beta}$

Note: we didn't assume independence of the errors or any particular distribution of the errors.

+ Gauss-Markov Theorem

■ Proof:

First note that the expected value and variance of $\mathbf{c}'\hat{\boldsymbol{\beta}}$ are:

$$E[\mathbf{c}'\hat{\boldsymbol{\beta}}] = \mathbf{c}'\boldsymbol{\beta} \quad \text{Var}(\mathbf{c}'\hat{\boldsymbol{\beta}}) = \sigma^2 \mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}$$

Next, since $\tilde{\boldsymbol{\beta}}$ is another linear combination of y's, it can be written as:

$$\tilde{\boldsymbol{\beta}} = [(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{B}] \mathbf{y} + \mathbf{b}_0$$

where \mathbf{B} and \mathbf{b}_0 are a constant $p \times n$ matrix and $p \times 1$ vector, respectively.

+ Gauss-Markov Theorem

■ Then:

$$\begin{aligned} E[\tilde{\beta}] &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{B}] \mathbf{y} + \mathbf{b}_0 \\ &= \end{aligned}$$

- Note: Being the best linear unbiased estimator means we are only concerned with unbiased $\tilde{\beta}$ which implies both $\mathbf{B}\mathbf{X} = \mathbf{0}$ and $\mathbf{b}_0 = \mathbf{0}$.

+ Gauss-Markov Theorem

- The variance of $\tilde{\beta}$ is:

$$\begin{aligned}\text{Var}(\tilde{\beta}) &= \text{Var}\left(\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{B}\right] \mathbf{y}\right) \\ &= \text{Var}\left(\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{B}\right] (\mathbf{X}\beta + \mathbf{e})\right)\end{aligned}$$

+ Gauss-Markov Theorem

- So then, the variance of the new estimator is:

$$\begin{aligned}\text{Var}(\mathbf{c}'\tilde{\boldsymbol{\beta}}) &= \mathbf{c}' \text{Var}(\tilde{\boldsymbol{\beta}}) \mathbf{c} \\ &= \mathbf{c}' \left\{ \sigma^2 [(\mathbf{X}'\mathbf{X})^{-1} + \mathbf{B}\mathbf{B}'] \right\} \mathbf{c} \\ &= \text{Var}(\mathbf{c}'\hat{\boldsymbol{\beta}}) + \mathbf{c}'\mathbf{B}\mathbf{B}'\mathbf{c}\end{aligned}$$

- But since $\mathbf{B}\mathbf{B}'$ is positive semidefinite, $\mathbf{c}'\mathbf{B}\mathbf{B}'\mathbf{c} \geq 0$.



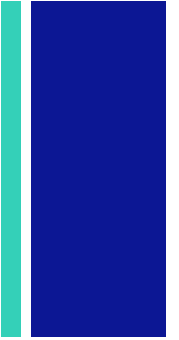
UMVUE: Uniform Minimum Variance Unbiased Estimator



- 610/611: Uniform Minimum Variance Unbiased Estimators
 - Minimum variance for UMVUE is across all unbiased estimators, including non-linear estimators
 - Saying $\hat{\beta}$ is BLUE only means minimum variance across all unbiased linear estimators.
 - To prove an estimator is UMVUE, often the Cramer-Rao Lower Bound is used, which requires some distribution assumptions.
 - The Gauss-Markov Theorem doesn't require any distribution assumptions; $\hat{\beta}$ is BLUE whatever the distribution of the errors.



Example: Median



- Assume that we have iid errors with mean 0.
- For the linear model $y_i = \beta + e_i$ we might use the median as an estimator of our parameter β .
- Is the median the BLUE for parameter β ? Why or why not?



Example: Another Estimator



- Assume that we have iid errors with mean 0. For the model

$$y_i = \beta + e_i$$

consider the estimator for β :

$$\frac{y_1 + y_2}{2}$$

- Is the above estimator the BLUE for estimating β ? Why or why not?