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 **a'y** for some vector 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 X is full rank. Then:

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

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:

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

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

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

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

Gauss-Markov Theorem

- Then:

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

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

Gauss-Markov Theorem

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

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

+

Gauss-Markov Theorem

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

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

- But since **BB’** is positive semidefinite, **c’BB’c** ≥ 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?