

OLS Assumptions

- What assumptions do we need for OLS regression to work?
 - This depends on what we mean by "work."
 - Different guarantees require different sets of assumptions.
- Our population model may meet only a weak set of assumptions.
 - These assumptions are realistic for almost any real dataset.
 - With OLS regression, limited ability to draw meaning from results.
- There is a famous set of fairly strict assumptions: Gauss-Markov.
 - These assumptions are often unrealistic for real datasets.
 - If they hold, we get much stronger guarantees about OLS estimates.
 - Gauss-Markov theorem: under certain assumptions, OLS is **BLUE**.

BLUE

- Best Linear Unbiased Estimator: BLUE
 - This is what is often meant by OLS "working."
 - OLS coefficients are estimators of the population parameters.

BLUE: Terms Explained

- **Best:** relative efficiency
 - OLS coefficients are random variables, and we want them to be as precise as possible.
 - OLS coefficients have smallest variance of all linear unbiased estimators.
- **Linear:** OLS estimates are a linear function of the y 's.
 - You can see in the matrix representation that the vector y is multiplied by a matrix, $(X'X)^{-1}X'$, and matrix multiplication is a linear operation.
- **Unbiased:** each $\hat{\beta}_j$ is an unbiased estimator for the true parameter β_j .
 $E(\hat{\beta}_j) = \beta_j$

BLUE is the most well-known benchmark for OLS performance.

Next up: a look at the assumptions that underlie the theorem