The Gauss-Markov Theorem: *OLS is* ***BLUE***

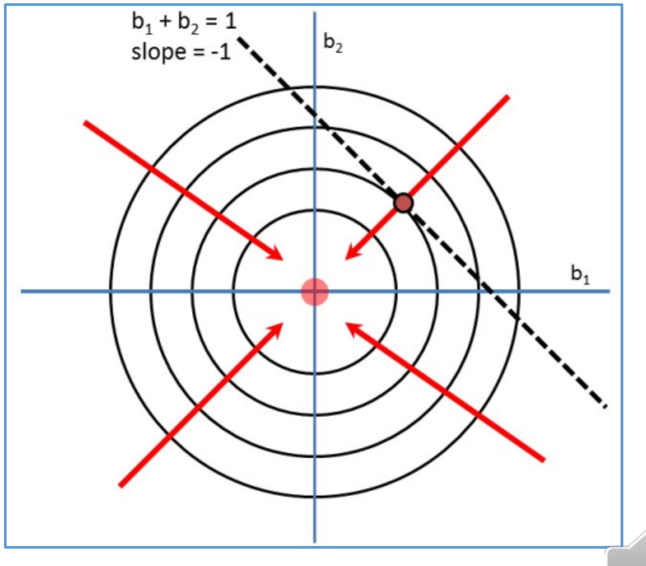
- **The Sample Mean is a *BLUE* estimator**
- **What about OLS? … *Start with those SLR conditions***
- ***We've Seen This Before!***
- ***… Linear Unbiased Estimators (LUEs)***
- ***… Getting to BLUE***
- **The Gauss-Markov Theorem: OLS = *BLUE***
  - **Some Intuition**

The Sample Mean is ***BLUE***

- Recall our analysis of the Sample Mean estimator (of the mean of the distribution):
- We looked at linear unbiased estimators (LUEs):
    - $b_1 Y_1 + b_2 Y_2 + \ldots + b_n Y_n$, where $\sum_{i=1}^{n} b_i = 1$.
- To find the Best Linear Unbiased Estimator (BLUE), we looked for the particular set of coefficients $\{b_i\}$, which minimized the variance within the group/class of LUEs.
- That amounted to solving the optimization problem:

  min $Var(\sum b_i Y_i) = \sigma^2 \sum b_i^{\,2}$ subject to $\sum_{i=1}^{n} b_i = 1$

- This is a constrained optimization problem, with solution $b_i^* = \dfrac{1}{n}$ for all i
- … which is the Sample Mean: $\bar{Y} = \dfrac{1}{n} \sum Y_i$.



What about OLS? *Start with those SLR conditions*

- Now assume SLR.1-SLR.5 and turn to the challenge of finding the BLUE estimator of the parameter β_1 of the linear model: $Y = \beta_0 + \beta_1 X + U$. (We will focus only on estimating the slope parameter here.)
- The analysis will be conditioned on a particular sample of the x_i's, and so each of the randomly determined values of the dependent variable will be defined by:

  SLR.1: $Y_i = \beta_0 + \beta_1 x_i + U_i$, where

  SLR.4: $E(Y_i \mid x_i) = \beta_0 + \beta_1 x_i$ since $E(U_i \mid x_i) = 0$, and

  SLR.5: $Var(Y_i) = Var(U_i) = \sigma^2$ (homoskedasticity).

We've Seen This Before!: … LUEs

- Consider the following general linear estimator (since we are conditioning on the x_i's, the estimator will be linear in the Y_i's): $b_0 + \sum b_i Y_i$
- We require the estimator to be unbiased: $E\left[b_0 + \sum b_i Y\right] = b_0 + E\left[\sum b_i [\beta_0 + \beta_1 x_i + U_i]\right]$, so

  $$E\left[b_0 + \sum b_i Y\right] = b_0 + \beta_0 \sum b_i + \beta_1 \sum b_i x_i + \sum b_i E(U_i \mid x_i)\,.$$

- But by SLR.4, the conditional means of the U_i's are all 0, and so we require that:

  $b_0 + \beta_0 \sum b_i + \beta_1 \sum b_i x_i \equiv \beta_1$, for all parameter values β_0 *and* β_1 .

- This requires that:
  - the intercept must be zero: $b_0 = 0$
  - the coefficients must sum to zero: $\sum b_i = 0$
  - the products of the b's and their respective x's must sum to one: $\sum b_i x_i = 1$.

We've Seen This Before!: … Getting to **BLUE**

Challenge	Objective	Constraints			**BLUE**
		b_0	$\sum b_i$	$\sum b_i x_i$	
Estimate μ	minimize $Var(\sum b_i Y_i) = \sigma^2 \sum b_i^2$	$=0$	$=1$		**Sample Mean**
Estimate β_1	minimize $Var(\sum b_i Y_i) = \sigma^2 \sum b_i^2$	$=0$	$=0$	$=1$	**OLS Estimator**

OLS: To find the BLUE estimator of β_1, we want to solve the following constrained optimization problem:

$$\min \ Var\left[\sum b_i Y_i\right] = \sigma^2 \sum b_i^2 \ \text{ subject to } \sum b_i = 0 \ \text{ and } \sum b_i x_i = 1.$$

Note the similarity to the Sample Mean BLUE constrained optimization problem:

$$\min \ Var(\sum b_i Y_i) = \sigma^2 \sum b_i^2 \ \text{ subject to } \sum_{i=1}^{n} b_i = 1.$$

(The objective functions are the same; the constraints differ in number and are slightly different.)

The Gauss-Markov Theorem: *OLS is* ***BLUE***

- I'll skip the details (see the handouts)… but:

  ***The envelope, please… And the winner is*:**

- Given SLR.1-5, and conditional on the x's, the ***BLUE*** estimator of β_1 is:

$$B_1 = \sum w_i \frac{(Y_i - \bar{Y})}{(x_i - \bar{x})} \text{, where } w_i = \frac{(x_i - \bar{x})^2}{\sum (x_j - \bar{x})^2} = \frac{(x_i - \bar{x})^2}{(n-1)S_{xx}} \geq 0 \text{ and } \sum w_i = 1.$$

- The winner is: ***The OLS estimator!***

  Who knew that minimizing SSRs could turn out so well?

- For the given sample, the estimate will be: $\hat{\beta}_i = \sum \frac{(x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$.

[OLS is BLUE!]

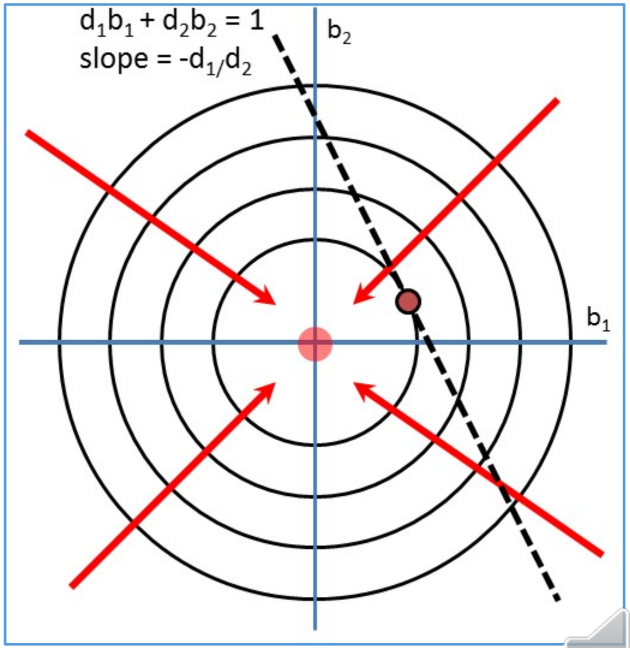
*OLS is **BLUE**: Some Intuition*

- The BLUE optimization problem is: $min\ \ \sigma^2 \sum b_i^2 \ \ s.t. \sum b_i = 0$ and $\sum b_i x_i = 1$.
- We can collapse the two constraints into one (see the handout) :

  $\sum b_i \left(x_i - \bar{x}\right) = \sum b_i d_i = 1 \ldots$ where $d_i = \left(x_i - \bar{x}\right),\ i = 1, \ldots, n$
- With this modification, the BLUE optimization problem becomes

  $min\ \ \sigma^2 \sum b_i^2 \ \ s.t. \sum \left(x_i - \bar{x}\right) b_i = \sum d_i b_i = 1.$
- At the optimum: $b_i^* = \dfrac{d_i}{\sum_{j=1}^{n} d_j^2} = \dfrac{\left(x_i - \bar{x}\right)}{\sum_j \left(x_j - \bar{x}\right)^2}$

  And so, we have ***OLS!*** (Note that the Figure looks virtually identical to what you saw with the Sample Mean as BLUE analysis.)



The Gauss-Markov Theorem: *OLS is* ***BLUE***

Onwards to *Inference*!