Agenda

1. Recap properties of OLS under exogeneity
2. OLS is BLUE
3. distribution of t-statistic

Classical linear model:

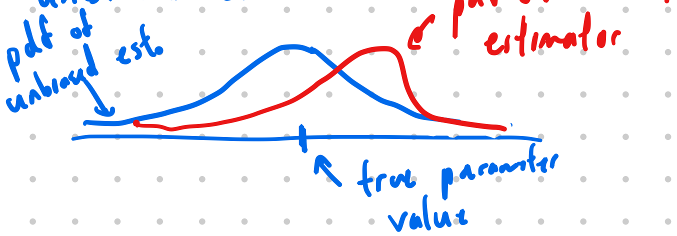
1. linearity : $y = X\beta^* + \epsilon$
2. exogeneity : $E[\epsilon | X] = 0$
3. spherical error variance : $E[\epsilon\epsilon^T | X] = \sigma^2 I_n$

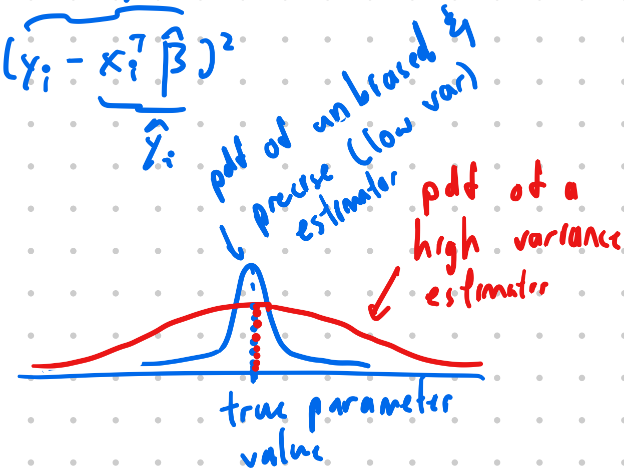
Under classical linear model:

1. OLS estimator is unbiased: $E[\hat{\beta} | X] = \beta^*$
2. $Var[\hat{\beta} | X] = \sigma^2 (X^T X)^{-1}$
3. unbiasedness of $s^2 \triangleq \frac{1}{n-p} \sum_{i=1}^n (\overbrace{y_i - \underbrace{x_i^T \hat{\beta}}_{\hat{y}_i}}^{\hat{\epsilon}_i})^2$

   $E[s^2 | X] = \sigma^2$

unbiasedness:





OLS is BLUE.

$\underbrace{\text{Best Linear Unbiased Estimator}}$

Let $\tilde{\beta}$ be another estimator of β^*.

$\tilde{\beta}$ is a linear estimator if $\tilde{\beta} = Ay$

Note OLS estimator is linear : $\hat{\beta} = (X^T X)^{-1} X^T y$

Best here means min variance.

$\tilde{\beta}$ is unbiased if $E[\tilde{\beta} | X] = \beta^*$

OLS is BLUE means $\underline{Var[\tilde{\beta} | X]} - \underline{Var[\hat{\beta} | X]}$ is positive semi-definite. (PSD)

Recall a square matrix $\underset{}{\overset{p\times p \text{ matrix}}{A \in \mathbb{R}^{d\times d}}}$ is $\overset{p\times p}{\text{PSD}}$ iff

$$\underset{1\times d}{x^T}\,\underset{d\times d}{A}\,\underset{d\times 1}{x} \geq 0 \quad \text{for all } x \in \mathbb{R}^d$$

Suppose $Var[\tilde{\beta}|X] - Var[\hat{\beta}|X]$ is PSD

By def of PSD, $\underbrace{a^T(Var[\tilde{\beta}|X] - Var[\hat{\beta}|X])a}$ ≥ 0 for any $a \in \mathbb{R}^p$

$$a^T\left(E[(\tilde{\beta}-\beta^*)(\tilde{\beta}-\beta^*)^T|X] - E[(\hat{\beta}-\beta^*)(\hat{\beta}-\beta^*)^T|X]\right)a \geq 0$$

$$E[a^T(\tilde{\beta}-\beta^*)\underbrace{(\tilde{\beta}-\beta^*)^T}_{1\times p}\underset{p\times 1}{a}|X] - E[a^T(\hat{\beta}-\beta^*)(\hat{\beta}-\beta^*)^T a|X] \geq 0$$

$$= E[(a^T(\tilde{\beta}-\beta^*))^2|X] - E[(a^T(\hat{\beta}-\beta^*))^2|X]$$

Pick $a = e_j = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$ ← j-th position

$$= E[(e_j^T(\tilde{\beta}-\beta^*))^2|X] - E[(e_j^T(\hat{\beta}-\beta^*))^2|X]$$

$$= E[(\tilde{\beta}_j-\beta_j^*)^2|X] - E[(\hat{\beta}_j-\beta_j^*)^2|X]$$

<u>Thm (Gauss-Markov)</u>: For any linear & unbiased estimator $\tilde{\beta}$, then $Var[\tilde{\beta}|X] - Var[\hat{\beta}|X]$ is PSD, where $\hat{\beta}$ is OLS estimator

<u>Pf</u>: $\tilde{\beta}$ is linear, so it can be written as $\tilde{\beta} = Ay$ for some matrix $A \in \mathbb{R}^{p\times n}$

Recall OLS estimator is $\hat{\beta} = (X^TX)^{-1}X^Ty$

$\tilde{\beta}$ is unbiased means $0 = E[\tilde{\beta} - \beta^*|X]$

$$= E[\tilde{\beta}|X] - \beta^*$$

$$= E[\tilde{\beta}|X] - E[\hat{\beta}|X] \quad \left(\text{OLS estimator is unbiased}\right)$$

$$= E[Ay|X] - E[(X^TX)^{-1}X^Ty|X]$$

$= E[\underbrace{(A - (X^TX)^{-1}X^T)}_{\Delta} y \mid X]$ $\quad$ Δy is just the difference b/w $\tilde{\beta}$ & $\hat{\beta}$

$= E[\Delta(X\beta^* + \epsilon) \mid X]$ (linearity)

$= E[\Delta X \beta^* \mid X] + E[\Delta \epsilon \mid X]$

$= \Delta X \beta^* + \Delta \cancel{E[\epsilon \mid X]}^{0}$ (exogeneity)

$= \Delta X \beta^*$

The only way $\Delta X \beta^*$ is always zero regardless of the choice of β^* is $\underset{p\times n}{\Delta}\underset{n\times p}{X} = 0_{p\times p}$

$$Var[\tilde{\beta} \mid X] = Var[\tilde{\beta} - \beta^* \mid X]$$
$$= Var[Ay - \beta^* \mid X] \quad (\text{def of } \tilde{\beta})$$
$$= Var[A(X\beta^* + \epsilon) - \beta^* \mid X] \quad (\text{linearity})$$
$$= Var[AX\beta^* + A\epsilon - \beta^* \mid X]$$
$$= Var[(\overbrace{(X^TX)^{-1}X^T + \Delta}^{A})X\beta^* + A\epsilon - \beta^* \mid X] \quad \left(\text{recall } \Delta: A - (X^TX)^{-1}X^T\right)$$
$$= Var[\cancel{(X^TX)^{-1}X^TX\beta^*} + \underbrace{\Delta X}_{=0}\beta^* + A\epsilon - \cancel{\beta^*} \mid X]$$
$$= Var[A\epsilon \mid X]$$
$$= A\underbrace{Var[\epsilon \mid X]}_{E[\epsilon\epsilon^T \mid X]}A^T$$
$$= \sigma^2 A A^T$$
$$= \sigma^2((X^TX)^{-1}X^T + \Delta)((X^TX)^{-1}X^T + \Delta)^T$$
$$= \sigma^2[(X^TX)^{-1}\cancel{X^TX(X^TX)^{-1}} + (X^TX)^{-1}\cancel{X^T\Delta^T}^{0} + \cancel{\Delta X}(X^TX)^{-1} + \Delta\Delta^T]$$
$$= \sigma^2[(X^TX)^{-1} + \Delta\Delta^T]$$

$$Var[\tilde{\beta} \mid X] = Var[\hat{\beta} \mid X] + \sigma^2 \Delta\Delta^T$$

Claim: $\Delta\Delta^T$ is PSD. $\quad$ $(Var[\tilde{\beta} \mid X] - Var[\hat{\beta} \mid X]$

Pick any vector $a \in \mathbb{R}^{p \times p}$

$$a^T \Delta \Delta^T a = (\Delta^T a)^T \Delta a = \|\Delta a\|_2^2 \geq 0$$

$= \sigma^2 \Delta \Delta^T$