Agenda:

1. Recap classical linear model
2. Unbiasedness of OLS
3. distribution of t-statistic
4. OLS is BLUE

Recall classical linear model:

linearity: $y = X\beta^* + \epsilon$

exogeneity: $E[\epsilon | X] = 0$

spherical error variance: $E[\epsilon\epsilon^T | X] = \sigma^2 I$ $\left\{ E[\epsilon_i^2 | X] = \sigma^2 \quad i=1,\dots,n \right.$

Unbiasedness of OLS:

Recall under linear model w/ Gaussian errors:

$$\hat\beta - \beta^* | X \sim N(0, \sigma^2 (X^TX)^{-1})$$

equivalently

$$\hat\beta | X \sim N(\beta^*, \sigma^2 (X^TX)^{-1})$$

$$E[\hat\beta | X] \overset{\text{claim}}{=} \beta^*$$

$$= E[\underbrace{(X^TX)^{-1}X^Ty}_{\text{expression for } \hat\beta} | X] \qquad \left(\begin{array}{l}\text{definition of} \\ \text{OLS estimator } \hat\beta\end{array}\right)$$

$$= E[(X^TX)^{-1}X^T(X\beta^* + \epsilon) | X] \qquad (\text{linearity})$$

$$= E[\cancel{(X^TX)^{-1}}\cancel{X^TX}\beta^* + (X^TX)^{-1}X^T\epsilon | X]$$

$$= E[\beta^* + (X^TX)^{-1}X^T\epsilon | X]$$

$$= \underset{p\times 1}{\beta^*} + \underset{p\times n\ n\times p}{(X^TX)^{-1}}\underset{p\times n}{X^T}\underbrace{\cancel{E[\epsilon|X]}^{0}}_{n\times 1} \qquad (\text{exogeneity})$$

$$E[(X^TX)^{-1}X^T\epsilon | X] = (X^TX)^{-1}X^TE[\epsilon | X]$$

$E[(X^TX)^{-1}X^T E[\epsilon|X]] = (X^TX)^{-1}X^T \, 0 = 0$

$\int (X^TX)^{-1}X^T \epsilon \, p(\epsilon|X)\, d\epsilon = (X^TX)^{-1}X^T \int \epsilon \, p(\epsilon|X)\, d\epsilon$

$Var[\hat\beta | X] \overset{(claim)}{=} \sigma^2 (X^TX)^{-1}$

$\hookrightarrow = Var[\hat\beta - \beta^* | X]$

$= Var[(X^TX)^{-1}X^Ty - \beta^* | X]$ (def of OLS)

$= Var[(X^TX)^{-1}X^T(X\beta^* + \epsilon) - \beta^* | X]$ (linearity)

$= Var[\cancel{(X^TX)^{-1}} \cancel{X^TX}\beta^* + (X^TX)^{-1}X^T\epsilon - \beta^* | X]$

$= Var[(X^TX)^{-1}X^T\epsilon | X]$

$= (X^TX)^{-1}X^T \, Var[\epsilon|X] \, X(X^TX)^{-1}$

$= (X^TX)^{-1}X^T(\sigma^2 I)X(X^TX)^{-1}$

$= \sigma^2 (X^TX)^{-1} \cancel{X^TX(X^TX)^{-1}}$

$= \sigma^2 (X^TX)^{-1}$

Aside:
$z \sim$ random vector in $\mathbb{R}^k$
$A \in \mathbb{R}^{l \times k}$
$Var[Az] = A \, Var[z] \, A^T$

$s^2_{ML} = \frac{1}{n}\sum_{i=1}^n (y_i - x_i^T\hat\beta)^2 \leftarrow SSR(\hat\beta)$

$\overset{(approx)}{\approx} \frac{1}{n}\sum_{i=1}^n (y_i - x_i^T\beta^*)^2$

$= \frac{1}{n}\sum_{i=1}^n \epsilon_i^2$

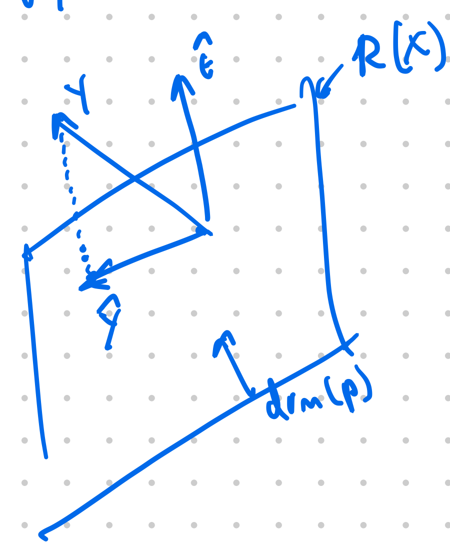
$y_i - \hat y_i = [y - \hat y]_i$

$s^2 = \frac{1}{n-p}\sum_{i=1}^n (y_i - x_i^T\hat\beta)^2$

$= \frac{1}{n-p}\sum_{i=1}^n (y_i - \hat y_i)^2$

$= \frac{1}{n-p}\sum_{i=1}^n \hat\epsilon_i^2$

$= \frac{1}{n-p}\|\hat\epsilon\|_2^2 \qquad \hat\epsilon = \begin{bmatrix} \hat\epsilon_1 \\ \vdots \\ \hat\epsilon_n \end{bmatrix}$



$$= \frac{1}{n-p} \|(I-H)\epsilon\|_2^2$$

$$= \frac{1}{n-p} \epsilon^T (I-H)^2 \epsilon$$

$$= \frac{1}{n-p} \epsilon^T (I-H) \epsilon$$

$$= \frac{1}{n-p} tr(\epsilon^T (I-H) \epsilon)$$

$H = X(X^TX)^{-1}X^T$

$$= \frac{1}{n-p} tr(\epsilon \epsilon^T (I-H))$$

$$E[s^2 | X] = E\left[\frac{1}{n-p} tr(\epsilon \epsilon^T (I-H)) | X\right]$$

(ideally) ||

σ^2

$$= \frac{1}{n-p} E[tr(\epsilon\epsilon^T (I-H)) | X]$$

$$= \frac{1}{n-p} tr(E[\epsilon\epsilon^T (I-H) | X])$$

$$= \frac{1}{n-p} tr(E[\epsilon\epsilon^T | X] (I-H))$$

$$= \frac{1}{n-p} tr(\sigma^2 I (I-H))$$

$$= \frac{\sigma^2}{n-p} tr(I-H)$$

$$= \frac{\sigma^2}{n-p} \cdot (n-p)$$

$$= \sigma^2$$

Claim: let P be projector onto subspace of dim k

then $tr(P) = k$

let U be an orthobasis of subspace

then $P = UU^T$

$$tr(P) = tr(UU^T)$$
$$= tr(U^TU)$$
$$= tr(I_k)$$

---

distribution of t-statistic

1. let $Z \sim N(0,1)$, let K be a χ^2_d independent of Z

   the $Z / \sqrt{\frac{K}{d}} \sim t_d$

2. If $Z \sim N(0, I_k)$, then $\|Z\|_2^2 \sim \chi^2_k$

   If $Z_1, \ldots, Z_k \overset{ind}{\sim} N(0,1)$, then $\sum_{i=1}^k Z_i^2 \sim \chi^2_k$

z-statistic: $\dfrac{\hat{\beta}_j}{SE[\hat{\beta}_j]}$, $\quad SE[\hat{\beta}_j] = \left(\sigma^2 \left[(X^T X)^{-1}\right]_{j,j}\right)^{\frac{1}{2}}$

t-statistic: $\dfrac{\hat{\beta}_j}{\widehat{SE}[\hat{\beta}_j]}$, $\quad \widehat{SE}[\hat{\beta}_j] = \left(s^2 \left[(X^T X)^{-1}\right]_{j,j}\right)^{\frac{1}{2}}$

$$\frac{\hat{\beta}_j}{\widehat{SE}[\hat{\beta}_j]} = \underbrace{\frac{\hat{\beta}_j}{SE[\hat{\beta}_j]}}_{N(0,1)} \cdot \frac{SE[\hat{\beta}_j]}{\widehat{SE}[\hat{\beta}_j]} =$$