Last lecture: OLS $\hat{\beta}_{(p+1)\times 1} = (X^T X)^{-1} X^T Y$

for linear regression model $Y_{n\times 1} = X_{n\times(p+1)} \beta_{(p+1)\times 1} + e_{n\times 1}$

This lecture

1. Estimation of σ^2
2. Uncertainty Quantification of $\hat{\beta}$.

1. Est. of σ^2. (Recall $\sigma^2 = Var(e_i \mid x)$ for any $i = 1, \cdots, n$)

Some useful quantities:

- $\underset{\in \mathbb{R}}{\hat{y}_i} \triangleq x_i^T \hat{\beta}$ is called the fitted value

  $x_i = \begin{pmatrix} x_{i,1} \\ \vdots \\ x_{i,p} \end{pmatrix}_{(p+1)\times 1}$

  ($\hat{Y}_i$ can be viewed as an estimator of $E[Y_i \mid x_i] = x_i^T \underset{\text{"unknown"}}{\beta}$)

- $\hat{e}_i \triangleq Y_i - \hat{Y}_i$ is called the residual

  $= Y_i - x_i^T \hat{\beta}$ (can be viewed as "estimator" of e_i)

  statistical error $e_i = Y_i - E[Y_i \mid x_i]$

★ To estimate σ^2, let's first consider an ideal case — e_i's are observed.

$\boxed{\{e_1, \cdots, e_n\} \quad E(e_i) = 0,\ Var(e_i) = \sigma^2 \ \forall i = 1, \cdots, n}$. then we can estimate σ^2 by the sample variance of e_i's.

$$\tilde{\sigma}^2 = \frac{\sum_{i=1}^{n} (e_i - \bar{e})^2}{n-1} \quad \text{where } \bar{e} = \frac{\sum_{i=1}^{n} e_i}{n}$$

We can not use $\tilde{\sigma}^2$ in practice since e_i's are not observable.

★ In practice, we use $\hat{e}_i$'s to estimate σ^2:

we estimate σ^2 by $$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (\hat{e}_i - \bar{\hat{e}})^2}{n-(p+1)}$$

where $\bar{\hat{e}} = \dfrac{\sum_{i=1}^{n} \hat{e}_i}{n}$.

Note that $\bar{\hat{e}} = 0 \Leftarrow$ Why? This follows from the normal equations. particularly.

$$\Downarrow$$

$$X^T (\underbrace{Y_{n\times 1} - X\hat{\beta}}) = 0$$

$$X^T = \begin{bmatrix} 1 & x_{11} \cdots x_{1p} \\ \vdots & \vdots \quad\quad \vdots \\ 1 & x_{n1} \cdots x_{np} \end{bmatrix}^T = \begin{bmatrix} 1 & \cdots & 1 \\ x_{11} & \cdots & x_{n1} \\ \vdots & & \vdots \\ x_{1p} & \cdots & x_{np} \end{bmatrix}_{(p+1)\times n}$$

$$Y - X\hat{\beta} = \begin{pmatrix} Y_1 - x_1^T\hat{\beta} \\ \vdots \\ Y_n - x_n^T\hat{\beta} \end{pmatrix} = \begin{pmatrix} \hat{e}_1 \\ \vdots \\ \hat{e}_n \end{pmatrix} = \hat{e}_{n\times 1} \text{ residual vector.}$$

$x_1^T\hat{\beta} = \hat{Y}_1$, $x_n^T\hat{\beta} = \hat{Y}_n$ from last lecture

first row of $X^T \cdot \hat{e}_{n\times 1} = 0$
$(1, \cdots, 1)$

$$\Rightarrow \sum_{i=1}^{n} \hat{e}_i = 0 \quad \& \quad \bar{\hat{e}} = 0.$$

So. our estimator for σ^2 is

$$\boxed{\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \hat{e}_i^2}{n-(p+1)}}.$$

where $\sum_{i=1}^{n} \hat{e}_i^2$ is called the Residual Sum of Squares. (RSS).

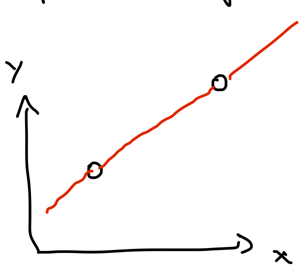
n : sample size.

$p+1$: ~~~~ # β parameters in the mean function

$n-(p+1)$: degrees of freedom. (d.f.)

E.g. $p=1$ Simple linear reg.

$n=2$



$\left.\begin{matrix}\hat{e}_1 = 0 \\ \hat{e}_2 = 0\end{matrix}\right\} \Rightarrow \hat{\sigma}^2 = 0$

d.f $= 0$.

☆ $\hat{\sigma}^2 = \dfrac{\sum_{i=1}^{n} \hat{e}_i^2 \leftarrow \text{RSS}}{n-(p+1) \leftarrow \text{d.f.}}$ is an $\underbrace{\text{unbiased estimator}}$ of σ^2.

that is, $E\hat{\sigma}^2 = \sigma^2$.

(this will be proved in a later lecture).

§2. Properties of $\hat{\beta}$.

We know $\hat{\beta} = (X^T X)^{-1} X^T Y$ (from last lecture)

from model assumption $Y = X\beta + e$

$$= (X^T X)^{-1} X^T (X\beta + e)$$

$$= (X^T X)^{-1} X^T X \beta + (X^T X)^{-1} X^T e.$$

$$\boxed{\hat{\beta} = \beta + (X^T X)^{-1} X^T e}$$

→ helpful to study theoretical properties of $\hat{\beta}$.

Next, we derive $E(\hat{\beta} \mid X)$ and $\underbrace{Var}_{\text{Covariance matrix}}(\hat{\beta} \mid X)$.

① $E(\hat{\beta} \mid X) = E[\beta + (X^T X)^{-1} X^T e \mid X]$

$E(\beta \mid X) = \beta$

$$= \beta + E((X^T X)^{-1} X^T e \mid X)$$

(for any const. matrix A. $E[Ae] = A E[e]$)

by $E(e \mid X) = 0$

$$= \beta.$$

This means that OLS estimator $\hat{\beta}$ is an unbiased estimator of β.

$$\boxed{E[\hat{\beta} \mid X] = \beta.}$$

② $Var(\hat{\beta}_{(p+1)\times 1} \mid X)$ covariance matrix of $\hat{\beta}$

$(p+1) \times (p+1)$.

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{pmatrix}$$

with (k row, j column) entry being

$Cov(\hat{\beta}_{k-1}, \hat{\beta}_{j-1})$.

$$Var(\hat{\beta} \mid X) = Var\left(\beta + (X^TX)^{-1}X^Te \,\middle|\, X\right)$$

for a random var. Z

$Var(Z) = Var(Z + \text{const.})$

$$= Var\left(\underbrace{(X^TX)^{-1}X^T}e \mid X\right)$$

$$\boxed{\begin{array}{l}\text{For any } A \\ Var(Ae) = A\, Var(e)\, A^T\end{array}}$$

$$= (X^TX)^{-1}X^T \underbrace{Var(e \mid X)}_{= \sigma^2 I_n} \underbrace{\left[(X^TX)^{-1}X^T\right]^T}_{= X(X^TX)^{-1}}$$

from our ~~model~~ variance function assumption

because $\left((X^TX)^{-1}\right)^T = (X^TX)^{-1}$ due to X^TX is symmetric.

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

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

That is

$$\boxed{Var(\hat{\beta} \mid X) = \sigma^2 (X^TX)^{-1}}$$

This implies that for any $k, j \in \{1, \cdots, p+1\}$

$$\underbrace{Cov(\hat{\beta}_{k-1}, \hat{\beta}_{j-1})}_{\text{kth row \& jth column of the $(p+1)\times(p+1)$ covariance matrix of } \hat{\beta}_{p+1} = \begin{pmatrix}\hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p\end{pmatrix}} = \text{the k^{th} row \& j^{th} column entry of the matrix } \sigma^2 (X^TX)^{-1}.$$

Particularly, when $j = k \in \{1, \cdots, p+1\}$, we have

$$Var(\hat{\beta}_{j-1} | X) = \text{the } j\text{th diagnal entry of the matrix } \sigma^2 (X^T X)^{-1}.$$

e.g. $j=1$. $Var(\hat{\beta}_0 | X)$ = the first diagnal entry of $\sigma^2 (X^T X)^{-1}$.

$j=2$ $Var(\hat{\beta}_1 | X)$ = the 2nd ..... of $\sigma^2 (X^T X)^{-1}$.

$\vdots$ $\vdots$

In practice, σ^2 is unknown, so to compute $Var(\hat{\beta} | X) = \sigma^2 (X^T X)^{-1}$

we replace σ^2 by $\hat{\sigma}^2 = \dfrac{\sum_{i=1}^{n} \hat{e}_i^2}{n-(p+1)}$.

$\Rightarrow$ We have an <u>estimated</u> $Var(\hat{\beta} | X)$

$$\widehat{Var}(\hat{\beta} | X) = \hat{\sigma}^2 (X^T X)^{-1}$$

☆ (estimated) standard error (s.e.) of $\hat{\beta}_{j-1} = \sqrt{\widehat{Var}(\hat{\beta}_{j-1} | X)}$ for $j = 1, \cdots, p+1$.

e.g $j=1$. $s.e.(\hat{\beta}_0) = \sqrt{\widehat{Var}(\hat{\beta}_0 | X)} \leftarrow$ 1st diagnal entry of $\hat{\sigma}^2 (X^T X)^{-1}$.

$j=2$ $\vdots$

$\vdots$ $\vdots$

☆ If we further assume $e_{n\times 1} \sim N(0, \sigma^2 I_n)$

then

$$\hat{\beta}_{(p+1)\times 1} | X \sim N(\underset{E(\hat{\beta}|X)}{\beta}, \underset{Var(\hat{\beta}|X)}{\sigma^2 (X^T X)^{-1}}).$$