L4. Multiple Linear Regression.

- Model set up and estimate in matrix notation.
- Gauss - Markow for MLR.
- Fitted value and Residuals./ Estimate of σ^2

1. Model set-up for MLR

In real life, of course we would suspect the response y is related to more than one predictor; then the

$\boxed{\text{Classical Multiple Linear Regression (MLR)}}$

comes to the play:

$$y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2}, + \cdots + \beta_{p-1}, X_{i,p-1} + \varepsilon_i$$

- $X_{i,1}, \cdots, X_{i,p-1}$ are the observed values of $X_1, \cdots, X_{p-1}$ respectively. - Fixed Effects.
- $\varepsilon_i \overset{i.i.d.}{\sim} N(0, \sigma^2)$

All the n observations are assumed to follow the same model:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & X_{1,1} & X_{1,2}, & \cdots, & X_{1,p-1} \\ 1 & X_{2,1} & X_{2,2} & \cdots & X_{2,p-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n,1} & X_{n,2} & \cdots & X_{n,p-1} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

or

$$\vec{y}_{n\times 1} = \vec{X}_{n\times p} \vec{\beta}_{p\times 1} + \vec{\varepsilon}_{n\times 1}$$

- $\vec{y}$: the vector of values in response variable.
- $\vec{X}$: the matrix of intercept and predictors
- $\vec{\beta}$: the parameter vector
- $\vec{\varepsilon}$: the error term vector.

Assumptions: $E(\vec{\varepsilon}) = \vec{0}$

$$Var(\vec{\varepsilon}) = \begin{pmatrix} \sigma^2 & & 0 \\ & \ddots & \\ 0 & & \sigma^2 \end{pmatrix} = \sigma^2 I_n$$

$\vec{\varepsilon} \sim MN(\vec{0}, \sigma^2 I_n)$ *Multivariate Normal Dist.

2. MLR Parameter Estimation.

Now $e_i = y_i - \hat{y}_i = y_i - (\beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \cdots + \beta_{p-1} X_{i,p-1})$

so $\vec{e} = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} - \begin{pmatrix} 1 & X_{1,1} & X_{1,2} & \cdots & X_{1,p-1} \\ 1 & X_{2,1} & X_{2,2} & \cdots & X_{2,p-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n,1} & X_{n,2} & \cdots & X_{n,p-1} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix}$

$= \vec{y} - \vec{X}\vec{\beta}$

recall $S(\beta_0, \beta_1) = \Sigma e_i^2 = \|\vec{e}\|^2 = \vec{e}^T \vec{e}$ in SLR.

so we can define $\boxed{S(\vec{\beta}) = \vec{e}^T \vec{e} = (\vec{y} - \vec{X}\vec{\beta})^T (\vec{y} - \vec{X}\vec{\beta})}$

and find $\vec{\beta} = \arg\min S(\vec{\beta})$

To solve this.

$S(\vec{\beta}) = \vec{y}^T \vec{y} - \underbrace{\vec{y}^T \vec{X} \vec{\beta}}_{1\times 1} - \underbrace{(\vec{X}\vec{\beta})^T \vec{y}}_{1\times 1} + (\vec{X}\vec{\beta})^T (\vec{X}\vec{\beta})$

$= \vec{y}^T \vec{y} - 2\vec{\beta}^T \vec{X}^T \vec{y} + \vec{\beta}^T \vec{X}^T \vec{X} \vec{\beta}$

$\frac{\partial S(\vec{\beta})}{\partial \vec{\beta}} = -2\vec{X}^T \vec{y} + 2\vec{X}^T \vec{X} \vec{\beta} = 0$

Hint: matrix differentiation.

$$\Rightarrow \quad (\vec{X}^T \vec{X}) \vec{\beta} = \vec{X}^T \vec{y}$$

$$\Rightarrow \quad \boxed{\vec{\beta} = (\vec{X}^T \vec{X})^{-1} (\vec{X}^T \vec{y})}$$

check for concavity: $\dfrac{\partial^2 S(\vec{\beta})}{\partial \vec{\beta}^2} = 2 \vec{X}^T \vec{X}$ positive semi-definite.

To avoid $\hat{\vec{\beta}}$ ... we call the LSE for MLR:

$$\vec{b} = (\vec{X}^T \vec{X})^{-1} (\vec{X}^T \vec{y}) = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_{p-1} \end{pmatrix}$$

3. Gauss-Markov Thm for $\vec{b}$ (Goal: understand the proof.)

- The OLSE $\vec{b}$ is the "BLUE" of $\vec{\beta}$.

(i) Linear

(ii) Unbiased

(iii) Covariance Matrix and Variance of $\vec{b}$

(iv) "Smallest"

(i) Linear: $\vec{b} = \underbrace{(\vec{X}^T\vec{X})^{-1}\vec{X}^T}\ \underbrace{\vec{y}}$

linear function of $\vec{y}$.

(ii) Unbiased: $\vec{b} - \vec{\beta} = (\vec{X}^T\vec{X})^{-1}\vec{X}^T\vec{y} - \vec{\beta}$

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

$$= (\vec{X}^T\vec{X})^{-1}\vec{X}^T\vec{\varepsilon}$$

$$E(\vec{b} - \vec{\beta}) = (\vec{X}^T\vec{X})^{-1}\vec{X}^T\ E(\vec{\varepsilon}) = 0$$

so $E(\vec{b}) = E(\vec{\beta})$

(iii) Variance-Covariance Matrix of $\vec{b}$ is

$$\mathrm{Var}\cancel{\mathrm{cov}}(\vec{b}) = E(\vec{b} - \vec{\beta})(\vec{b} - \vec{\beta})^T$$

$$= E\left[(\vec{X}^T\vec{X})^{-1}\vec{X}^T\vec{\varepsilon}\,\vec{\varepsilon}^T\vec{X}(\vec{X}^T\vec{X})^{-1}\right]$$

$$= (\vec{X}^T\vec{X})^{-1}\vec{X}^T \cdot \underbrace{E(\vec{\varepsilon}\,\vec{\varepsilon}^T)}_{=\sigma^2 I_n} \cdot \vec{X}(\vec{X}^T\vec{X})^{-1}$$

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

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

(iv) "Smallest" Variance in the matrix sense.

If we have another unbiased linear estimator $\vec{b}^* = \vec{C}\vec{y}$

recall $\vec{b} = \underbrace{(\vec{X}^T\vec{X})^{-1}\vec{X}^T}_{p\times n}\vec{y}$. let $\vec{C} = (\vec{X}^T\vec{X})^{-1}\vec{X}^T + \underbrace{\vec{D}}_{\substack{p\times n \text{ non-zero} \\ \text{matrix}}}$

Then $E(\vec{b}^*) = E(\vec{C}\vec{Y})$

$$= E[\vec{b} + \vec{D}\vec{Y}]$$

$$= \vec{\beta} + E(\vec{D}(\vec{X}\vec{\beta} + \vec{\varepsilon}))$$

$$= \vec{\beta} + \vec{D}\vec{X}\vec{\beta} = (I_p + \vec{D}\vec{X})\vec{\beta} = \vec{\beta}$$

so if $\vec{b}^*$ is unbiased, then $\vec{D}\vec{X} = \vec{O}_{p\times p}$

The variance - covariance matrix of $\vec{b}^*$ is given by

$$VAR(\vec{b}^*) = VAR(\vec{C}\vec{Y}) = \vec{C}\,Var(\vec{Y})\,\vec{C}^T$$

Hint: $Var(\vec{Y}) = Var(\vec{X}\vec{\beta} + \vec{\varepsilon}) = Var(\vec{\varepsilon}) = \sigma^2 I_n$

$$= \sigma^2 \vec{C}\vec{C}^T$$

$$= \sigma^2 [(\vec{X}^T\vec{X})^{-1}\vec{X}^T + \vec{D}][(\vec{X}^T\vec{X})^{-1}\vec{X}^T + \vec{D}]^T$$

$$= \sigma^2((\vec{X}^T\vec{X})^{-1}\vec{X}^T + \vec{D})(\vec{X}(\vec{X}^T\vec{X})^{-1} + \vec{D}^T)$$

$$= \sigma^2((\vec{X}^T X)^{-1} + (\vec{X}^T\vec{X})^{-1}\vec{X}^T\vec{D}^T + \vec{D}\vec{X}(\vec{X}^T\vec{X})^{-1} + \vec{D}\vec{D}^T)$$

Recall: $\vec{D}\vec{X} = \vec{0}_{p\times p}$.

so $\vec{D}\vec{X}(\vec{X}^T\vec{X})^{-1} = 0$, $(\vec{D}\vec{X}(\vec{X}^T\vec{X})^{-1})^T = 0$

Therefore. $$VAR(\vec{b}^*) = \sigma^2 (X^TX)^{-1} + \sigma^2 \vec{D}\vec{D}^T$$

$$= Var(\vec{b}) + \underbrace{\sigma^2 \vec{D}\vec{D}^T}_{\text{positive semi-definite}}$$

Therefore. $VAR(\vec{b}^*)$ is "larger" than $VAR(\vec{b})$.

4. Fitted Value and Residuals

- The Fitted Value

$$\hat{\vec{y}} = \vec{X}\cdot\vec{b}$$

- The Residuals

$$\vec{e} = \vec{y} - \hat{\vec{y}} = \vec{y} - \vec{X}\vec{b}$$

$$= \vec{y} - \underbrace{\vec{X}(\vec{X}^T\vec{X})^{-1}\vec{X}^T}_{\text{define as } \vec{H} \text{ : Hat Matrix}}\vec{y}$$

$$= \vec{y} - \vec{H}\vec{y} = (I_n - \vec{H})\vec{y}$$

5. Estimate of σ^2:

$$SSE = \sum_{i=1}^{n} e_i^2 = \vec{e}^T \vec{e} = [(I-H)\vec{y}]^T [(I-H)\vec{y}]$$

$$= \vec{y}^T (I-H)^T (I-H) \vec{y}$$

$$= \vec{y}^T (I-H) \vec{y}$$

Hint: $(I-H)^T(I-H) = I-H$ using the def. of H.

Now $\underline{\vec{y} \sim N(\vec{X}\vec{\beta}, \sigma^2 I_n)}$ can you see this?

then $\frac{SSE}{\sigma^2} \sim \chi^2(n-p)$

Thus $E\left(\frac{SSE}{n-p}\right) = \sigma^2$

Define $\boxed{MSE = \frac{SSE}{n-p}, \text{ then } MSE = \hat{\sigma}^2}$

Summary and Take-aways of this lecture: Matrix Notation

MLR: $\vec{y} = \vec{X}\vec{\beta} + \vec{\varepsilon}$, $\vec{\varepsilon} \sim MN(\vec{0}, \sigma^2 I_n)$

OLSE: $\vec{b} = (\vec{X}^T\vec{X})^{-1}(\vec{X}^T)\vec{y}$ is $\boxed{\text{"BLUE"}}$.

with $E(\vec{b}) = \vec{\beta}$, $Var(\vec{b}) = \sigma^2(\vec{X}^T\vec{X})^{-1}$

Fitted value: $\hat{\vec{y}} = \vec{X}\vec{b}$

Residuals: $\vec{e} = (I-H)\vec{y}$, where $H = \vec{X}(\vec{X}^T\vec{X})^{-1}\vec{X}^T$

Estimate of σ^2: $MSE = \frac{SSE}{n-p} = \frac{\vec{e}^T\vec{e}}{n-p}$