Linear Regression

Newton's three sisters

February 9, 2025

Department of Statistics Sungshin Women's University

- 1 Least Squares Method
- 2 Multiple Regression
- 3 Distribution $\hat{\beta}$
- 4 Distribution of RSS Values
- ${\color{red} {\bf 5}}{\color{black} {\bf I}}$ Hypothesis Testing for $\hat{\beta}_j \neq 0$
- 6 Coefficient of Determination and the Detection of Collinearity
- 7 Confidence and Prediction Intervals

- 1 Least Squares Method
- 2 Multiple Regression
- 3 Distribution (
- 4 Distribution of RSS Values
- ${\color{red} {\bf 5}}$ Hypothesis Testing for $\hat{\beta}_j \neq 0$
- 6 Coefficient of Determination and the Detection of Collinearity
- 7 Confidence and Prediction Intervals

- 1 Least Squares Method
- 2 Multiple Regression
- 3 Distribution $\hat{\beta}$
- 4 Distribution of RSS Values
- ${\color{red} {\bf 5}}$ Hypothesis Testing for $\hat{\beta}_j \neq 0$
- 6 Coefficient of Determination and the Detection of Collinearity
- 7 Confidence and Prediction Interval

- 1 Least Squares Method
- 2 Multiple Regression
- 3 Distribution $\hat{\beta}$
- 4 Distribution of RSS Values
- ${\color{red} {\bf 5}}$ Hypothesis Testing for $\hat{\beta}_j \neq 0$
- 6 Coefficient of Determination and the Detection of Collinearity
- 7 Confidence and Prediction Intervals

Hat matrix

• Hat matrix defined by

$$\hat{y} = X \hat{\beta} = X (X^T X)^{-1} X^T y = H y$$

$$H \triangleq X (X^T X)^{-1} X^T$$

• some properties

$$\begin{split} H^2 &= X(X^TX)^{-1}X^T \cdot X(X^TX)^{-1}X^T = X(X^TX)^{-1}X^T = H \\ (I-H)^2 &= I-2H+H^2 = I-H \\ HX &= X(X^TX)^{-1}X^T \cdot X = X \end{split}$$

Residual sum of square

• RSS defined

$$\mathrm{RSS} \triangleq ||y - \hat{y}||^2$$

• Using hat matrix

$$\begin{split} y - \hat{y} &= y - Hy = (I - H)y = (I - H)(X\beta + \varepsilon) \\ &= (X - HX)\beta + (I - H)\varepsilon = (I - H)\varepsilon \end{split}$$

$$RSS \triangleq ||y - \hat{y}||^2 = \{(I - H)\varepsilon\}^T (I - H)\varepsilon = \varepsilon^T (I - H)^2 \varepsilon = \varepsilon^T (I - H)\varepsilon$$

Eigenvalues of H and Null space of (I - H)

• Proof by contrapositive

$$Hx = x \Rightarrow (I - H)x = 0$$

 $(I - H)x = 0 \Rightarrow Hx = x$

• Dimensions of the eigenspaces of H is p+1

Proof using
$$rank(X) = p + 1$$

$$\begin{aligned} & \operatorname{rank}(H) \leq \min\{\operatorname{rank}(X(X^TX)^{-1}), \operatorname{rank}(X)\} \leq \operatorname{rank}(X) = p+1 \\ & \operatorname{rank}(H) \geq \operatorname{rank}(HX) = \operatorname{rank}(X) = p+1 \end{aligned}$$

 \bullet Dimensions of the null space of I-H is N-(p+1)

$$P(I-H)P^T = \mathrm{diag}(\underbrace{1,\dots,1}_{N-p-1},\underbrace{0,\dots,0}_{p+1})$$

제목 뭐라고 하지..

• We define $v = P\varepsilon \in \mathbb{R}^N$, then from $\varepsilon = P^T v$

$$\begin{split} \text{RSS} &= \varepsilon^T (I - H) \varepsilon = (P^T v)^T (I - H) P^T v = v^T P (I - H) P^T v \\ &= [v_1, \cdots, v_{N-p-1}, v_{N-p}, \cdots, v_n] \begin{bmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & 0 & \cdots & \cdots & \vdots \\ \vdots & 0 & 1 & \cdots & \cdots & 0 \\ 0 & 0 & 0 & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & \cdots & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_{N-p-1} \\ v_{N-p} \\ \vdots \\ v_N \end{bmatrix} \\ &= \sum_{i=1}^{N-p-1} v_i^2 \end{split}$$

• Let
$$w \in \mathbb{R}^{N-p-1}$$

Average

$$E[v] = E[P\varepsilon] = 0$$

$$E[w] = 0$$

Covariance

$$\begin{split} E[vv^t] &= E[P\varepsilon(P\varepsilon)^T] = PE[\varepsilon\varepsilon^t]P = P\sigma^2IP^T = \sigma^2I\\ E[ww^T] &= \sigma^2I \end{split}$$

• We have RSS

$$\frac{RSS}{\sigma^2} \sim \chi^2_{N-p-1}$$

- 1 Least Squares Method
- 2 Multiple Regression
- 3 Distribution $\hat{\beta}$
- 4 Distribution of RSS Values
- **5** Hypothesis Testing for $\hat{\beta}_j \neq 0$
- 6 Coefficient of Determination and the Detection of Collinearity
- 7 Confidence and Prediction Intervals

- 1 Least Squares Method
- 2 Multiple Regression
- 3 Distribution $\hat{\beta}$
- 4 Distribution of RSS Values
- ${\color{red} {\bf 5}}$ Hypothesis Testing for $\hat{\beta}_j \neq 0$
- 6 Coefficient of Determination and the Detection of Collinearity
- 7 Confidence and Prediction Intervals

- \bullet We define a matrix $W \in \mathbb{R}^{N \times N}$ such that all the elements are 1/N
- Explained sum of squres ESS

ESS
$$\triangleq ||\hat{y} - \bar{y}||^2 = ||\hat{y} - Wy||^2 = ||(H - W)y||^2$$

• Total sum of squres TSS

$$\mathrm{TSS} \triangleq ||y - \bar{y}||^2 = ||(I - W)y||^2$$

- 1 Least Squares Method
- 2 Multiple Regression
- 3 Distribution $\hat{\beta}$
- 4 Distribution of RSS Values
- ${\color{red} {\bf 5}}$ Hypothesis Testing for $\hat{\beta}_j \neq 0$
- 6 Coefficient of Determination and the Detection of Collinearity
- 7 Confidence and Prediction Intervals

Q & A

Thank you:)