Linear Regression

Newton's three sisters

February 16, 2025

Department of Statistics
Sungshin Women's University

Outline

Hat matrix

- Hat matrix defined by $\hat{y} = Hy$

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

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

- Some properties

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

Residual Sum of Square

- RSS defined

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

- Using hat matrix

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

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

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 $\text{rank}(X) = p + 1$

$$\text{rank}(H) \leq \min\{\text{rank}(X(X^T X)^{-1}), \text{rank}(X)\} \leq \text{rank}(X) = p + 1$$
$$\text{rank}(H) \geq \text{rank}(HX) = \text{rank}(X) = p + 1$$

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

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

Residual Sum of Squares (RSS) and Eigenvalue

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

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

Statistical Properties of the Residual Sum of Squares (RSS)

- Let $w \in \mathbb{R}^{N-p-1}$ be
  - Average
    $E[v] = E[P\varepsilon] = 0$
    $E[w] = 0$
  - Covariance
    $E[vv^t] = E[P\varepsilon(P\varepsilon)^T] = PE[\varepsilon\varepsilon^t]P = P\sigma^2 IP^T = \sigma^2 I$
    $E[ww^T] = \sigma^2 I$
- We have RSS

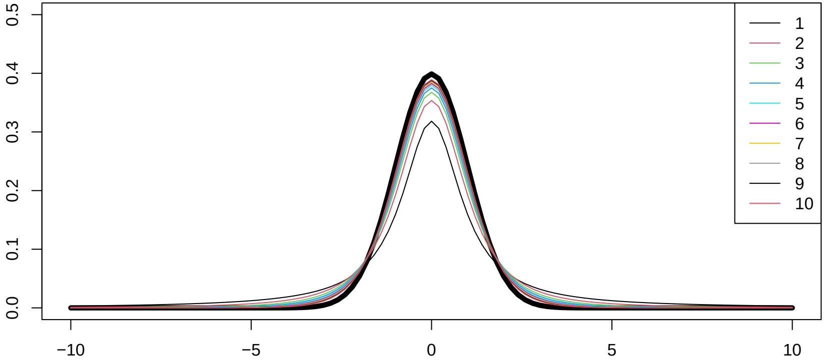
$$\frac{RSS}{\sigma^2} \sim \chi^2_{N-p-1}$$

Outline

Test statistic T

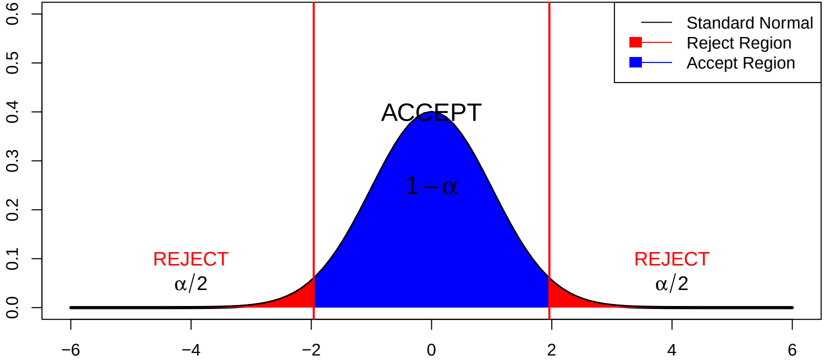
- A t distribution with $N - P - 1$ degrees of freedom
- We decide that hypothesis $\beta_j = 0$ should be rejected.
- $U \sim N(0,1),\ V \sim \chi^2_m,$

$$T \triangleq U/\sqrt{V/m}$$



Significance level

- $\alpha = 0.01,\ 0.05$
- Null hypothesis $\beta_j = 0$



Example 23

- For $p = 1$, since

$$X^T X = \begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_N \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = N \begin{bmatrix} 1 & \bar{x} \\ \bar{x} & \frac{1}{N} \sum_{i=1}^{N} x_i^2 \end{bmatrix}$$

- The inverse is

$$(X^T X)^{-1} = \frac{1}{N} \begin{bmatrix} 1 & \bar{x} \\ \bar{x} & \frac{1}{N} \sum_{i=1}^{N} x_i^2 \end{bmatrix}^{-1} = \frac{1}{\sum_{i=1}^{N} (x_i - \bar{x})^2} \begin{bmatrix} \frac{1}{N} \sum_{i=1}^{N} x_i^2 & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix}$$

- Which means that

$$B_0 = \frac{\frac{1}{N} \sum_{i=1}^{N} x_i^2}{\sum_{i=1}^{N} (x_i - \bar{x})^2} \quad \text{and} \quad B_1 = \frac{1}{\sum_{i=1}^{N} (x_i - \bar{x})^2}$$

Example 23 (contd.)

- For $B = (X^T X)^{-1}, B\sigma^2$ is covariance matrix of $\hat{\beta}$
- $B_j\sigma^2$ is the variance of $\hat{\beta}_j$
- Because $\bar{x}$ is positive, the correlation between $\hat{\beta}_0$ and $\hat{\beta}_1$ is negative

$$t = \frac{\hat{\beta}_j - \beta_j}{\mathrm{SE}(\hat{\beta}_j)} \sim t_{N-p-1}$$

Statistical Independence in Regression

- It remains to be shown that U and V are independent

$$U \triangleq \frac{\hat{\beta}_j - \beta_j}{\sqrt{B_j}\sigma} \sim N(0,1) \quad \text{and} \quad V \triangleq \chi^2_{N-p-1}$$

- Sufficient to show tha $y - \hat{y}$ and $\hat{\beta} - \beta$ are independent

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

- From $E\varepsilon\varepsilon^T = \sigma^2 I$ and $HX = X$,

$$E(\hat{\beta} - \beta)(y - \hat{y})^T = 0$$

Outline

RSS, ESS, and TSS

- We define a matrix $W \in \mathbb{R}^{N \times N}$ such that all the elements are $1/N$ $Wy \in \mathbb{R}^N$ are $\bar{y} = Wy = \sum_{i=1}^{N} y_i$ for $y_1, \cdots, y_N \in \mathbb{R}$
- Residual sum of squares RSS

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

- Explained sum of squres ESS

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

- Total sum of squres TSS

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

Sample - based correlation

- Coefficient of determination

$$R^2 = \frac{\text{ESS}}{\text{TSS}} = 1 - \frac{\text{RSS}}{\text{TSS}}$$

- Correlation between the covariates and response

$$\hat{\rho} \triangleq \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2 \sum_{i=1}^{N}(y_i - \bar{y})^2}}$$

$$\frac{\text{ESS}}{\text{TSS}} = \frac{\hat{\beta}_1^2 ||x - \bar{x}||^2}{||y - \bar{y}||^2} = \left\{ \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \right\}^2 \frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$$

$$= \frac{\left\{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})\right\}^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2] \sum_{i=1}^{N}(y_i - \bar{y})^2} = \hat{\rho}^2$$

Variance Inflation Factors

- Variance inflation factors

$$\mathrm{VIF} \triangleq \frac{1}{1 - R^2_{X_j | X_{-j}}}$$

- The minimum value of VIF is one, and we say that the collinearity of covariate is strong when its VIF value is large

Outline

Confidence Interval

- We have showed how to obtain the estimate $\hat{\beta}$ of $\beta \in \mathbb{R}^{p+1}$, confidence interval of $\hat{\beta}$ as follows

$$\beta_i = \hat{\beta}_i \pm t_{N-p-1}(\alpha/2)\mathrm{SE}(\hat{\beta}_i), \quad \text{for} \ \ i = 0, 1, \cdots, p$$

- Confidence interval of $x_*\hat{\beta}$ for another point $x_* \in \mathbb{R}^{p+1}$
  - The average
  $$E[x_*\hat{\beta}] = x_* E[\hat{\beta}]$$
  - The variance
  $$V[x_*\hat{\beta}] = x_* V(\hat{\beta}) x_*^T = \sigma^2 x_* (X^T X)^{-1} x_*^T$$

- We define

$$\hat{\sigma} \triangleq \sqrt{\mathrm{RSS}/(N-p-1)}, \quad \mathrm{SE}(x_*\hat{\beta}) \triangleq \hat{\sigma}\sqrt{x_*(X^T X)^{-1} x_*^T}$$

- $\mathrm{C} \sim t_{N-p-1}$

- variance in the difference between $x_*\hat{\beta}$ and $y_* \triangleq x_*\beta + \varepsilon$

$$V[x_*\hat{\beta} - (x_*\beta + \varepsilon)] = V[x_*(\hat{\beta} - \beta)] + V[\varepsilon] = \sigma^2 x_*(X^TX)^{-1}x_*^T + \sigma^2$$

- Similarly, we can derive the following

$$P \triangleq \frac{x_*\hat{\beta} - y_*}{\mathrm{SE}(x_*\hat{\beta} - y_*)} = \frac{x_*\hat{\beta} - y_*}{\sigma(1 + \sqrt{x_*(X^TX)^{-1}x_*^T})} \Big/ \sqrt{\frac{\mathrm{RSS}}{\sigma^2} \Big/ (N-p-1)} \sim t_{N-p-1}$$

- The confidence and prediction intervals

$$\begin{aligned} x_*\beta &= x_*\hat{\beta} \pm t_{N-p-1}(\alpha/2)\hat{\sigma}\sqrt{x_*(X^TX)^{-1}x_*^T} \\ y_* &= x_*\hat{\beta} \pm t_{N-p-1}(\alpha/2)\hat{\sigma}\sqrt{1 + x_*(X^TX)^{-1}x_*^T} \end{aligned}$$

Q & A

Thank you :)