Chapter 2: Linear Regression

Distribution of the RSS Values and Hypothesis Testing for $\hat{\beta}_j \neq 0$

JuHyun Kang

February 23, 2025

The Three Sisters of Newton
School of Mathematics, Statistics and Data Science
Sungshin Women's University

Outline

1. Distribution of the RSS Values

2. Hypothesis Testing for $\hat{\beta}_j \neq 0$

Hat Matrix

- Hat matrix defined by

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

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

- Some properties

  1. $H^2 = X(X^TX)^{-1}X^T \cdot X(X^TX)^{-1}X^T = X(X^TX)^{-1}(X^TX)(X^TX)^{-1}X^T$
     $= X(X^TX)^{-1}X^T = H$
  2. $(I-H)^2 = I - 2H + H^2 = I - 2H + H = I - H$
  3. $HX = X(X^TX)^{-1}X^T \cdot X = X$
  4. $H^T = \{X(X^TX)^{-1}X^T\}^T = X(X^TX)^{-1}X^T = H$

Residual Sum of Squares

- RSS defined

$$\text{RSS} := ||y - \hat{y}||^2$$

- Using hat matrix

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

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

- They are only zeros and ones
- Dimensions of the eigenspaces of H and $I - H$ are both $p + 1$

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

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

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

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

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

RSS Matrix

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

$$\text{RSS} = \varepsilon^T(I-H)\varepsilon = (P^Tv)^T(I-H)P^Tv = v^TP(I-H)P^Tv$$

$$= [v_1, \cdots, v_{N-p-1}, v_{N-p}, \cdots, v_n] \begin{bmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & 0 & \cdots & \cdots & \vdots \\ \vdots & 0 & 1 & \cdots & \cdots & 0 \\ 0 & 0 & 0 & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & \cdots & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_{N-p-1} \\ v_{N-p} \\ \vdots \\ v_N \end{bmatrix} = \sum_{i=1}^{N-p-1} v_i^2$$

Distribution of RSS

- Let $w \in \mathbb{R}^{N-p-1}$

- Average

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

  $E[w] = 0$

- Covariance

  $E[vv^t] = E[P\varepsilon(P\varepsilon)^T] = PE[\varepsilon\varepsilon^t]P = P\sigma^2 IP^T = \sigma^2 I$

  $E[ww^T] = \sigma^2 I$

- $w \sim N(0, \sigma^2 I)$, $\frac{w^T w}{\sigma^2}$ is the sum of squares of $N-p-1$ independent standard normal variables

- We have RSS

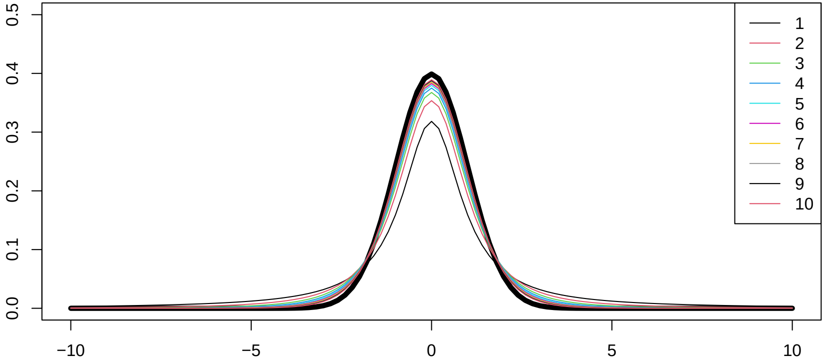
$$\frac{RSS}{\sigma^2} \sim \chi^2_{N-p-1}$$

Outline

Test Statistic T

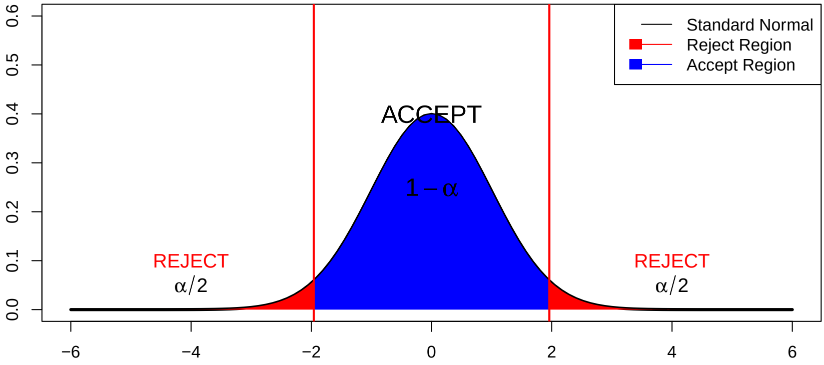
- T distribution with $N - p - 1$ degrees of freedom
- We decide that hypothesis $\beta_j = 0$ should be rejected.
- $U \sim N(0,1),\ V \sim \chi^2_m,$

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



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

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

- 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 that $y - \hat{y}$ 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 $HX = X$,

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

Q & A

Thank you :)