

Regression Analysis 2023-05-29

Q1 (a) $\hat{\beta} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} = -1.66$

$\hat{\alpha} = \bar{y} - \hat{\beta} \bar{x} = 73$

(b) We use t -test. The error variance is estimated by $\hat{\sigma}^2 = 0.15$.

We then get $\text{Cov} \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \hat{\sigma}^2 (X^T X)^{-1} = \begin{bmatrix} 0.225 & -0.015 \\ -0.015 & 0.0012 \end{bmatrix}$

Then the t value is $t = \frac{-1.66}{\sqrt{0.0012}} = -47.92$.

The critical value is 4.30. $t_{0.975}(2)$

We then reject $H_0 \beta = 0$.

(c) $73 - 1.66 \cdot 25 = 31.5$

(d) $31.5 \pm t_{0.975}(2) \sqrt{\hat{\sigma}^2 (1 + [1 \ 25] (X^T X)^{-1} \begin{bmatrix} 1 \\ 25 \end{bmatrix})}$

Q2 (a) $\hat{\beta}_1 = \frac{-2/1 + 0 + 2/3}{8}$

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \bar{y}$

(b) $\text{Cov} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \sigma^2 (X^T X)^{-1} = \sigma^2 \left(\begin{bmatrix} 1 & -2 \\ 1 & 0 \\ 1 & 2 \end{bmatrix}^T \begin{bmatrix} 1 & -2 \\ 1 & 0 \\ 1 & 2 \end{bmatrix} \right)^{-1}$

Q3. (a) GLS minimizes $(y - X\beta)^T \Omega^{-1} (y - X\beta)$

The minimizer is $\hat{\beta} = (X^T \Omega^{-1} X)^{-1} X^T \Omega^{-1} y$

(b) $E(\hat{\beta} | X) = (X^T \Omega^{-1} X)^{-1} X^T \Omega^{-1} E(y | X) = \beta$ unbiased

$V(\hat{\beta} | X) = (X^T \Omega^{-1} X)^{-1} X^T \Omega^{-1} V(y | X) \Omega^{-1} X (X^T \Omega^{-1} X)^{-1}$
 $= \sigma^2 (X^T \Omega^{-1} X)^{-1}$

Q4 a) $A^{**} = 475.5 / 3.451$

$B^{**} = -353 / 476.3$

$C^{**} = 332.9 \cdot (1 - 0.394)$

$D^{**} = -1035 / (1 - 2.162)$

b) age, bmi, children, smokeyes, region southeast
region southwest

c) $256.9 \cdot 32 + 339.2 \cdot 26 + 475.5 \cdot 2 - 1035$

d) $C^{**} < 0$ so the charge for female is higher
charge for northwest is higher than other regions

Q5 a) $A^{**} = 5.31821 / 2.94460$

$B^{**} = -0.09260 / 0.04587$

b) Yes, P value < 0.05

c) $\log \mu = 5.31821 - 0.09260 \cdot \text{Temp}$

d) $\exp(5.31821 - 0.09260 \cdot 31)$

e) Hard to say because Temp is a continuous regressor