Regression Analysis: 2023-08-18

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

$\hat{\sigma}^2 = \frac{\sum_i e_i^2}{n-p} = 0.15$

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

The t value is $t = \frac{-1.66}{\sqrt{0.0012}} = -47.92$

The quantile of $t(n-p)$ is 4.30. Hence we reject $\beta = 0$

The effect is significant

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

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

Q2 (a) $\hat{\beta}_1 = \frac{-5y_1 + 0 + 5y_3}{50}$

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

(b) $Cov\begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\end{bmatrix} = \sigma^2 (X^T X)^{-1} = \sigma^2 \left(\begin{bmatrix}1 & -5\\ 1 & 0\\ 1 & 5\end{bmatrix}^T \begin{bmatrix}1 & -5\\ 1 & 0\\ 1 & 5\end{bmatrix}\right)^{-1}$

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

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

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

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

(b) age, bmi, children, smokeryes, regionsoutheast, regionsouthwest

(c) $256.9 \cdot 32 + 339.1 \cdot 26 + 475.5 \cdot 2 - 1035$.

(d) $C^{**} < 0$ so charge for female is higher

Charge for northeast is higher than other regions.

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

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

(b) Yes when $\alpha = 0.05 > 0.0435 = P\text{-Value}$.

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

(d) Hard to say because Temp is a continuous regressor.