Regression Analysis 2022-05-25

Q2 (a) No intercept in the model : $\sum_{i=1}^{3m} x_i^2 = 8m$

Hence $$\hat{\beta} = \frac{\sum_i x_i y_i}{\sum_i x_i^2} = \frac{-2\sum_{i=1}^{m} y_i + 2\sum_{i=2m+1}^{3m}}{8m} = \frac{-\bar{y}_{(1)} + \bar{y}_{(3)}}{4}$$

(b) $$\widehat{Cov}(\hat{\beta}) = \hat{\sigma}^2 (X^T X)^{-1} = \frac{\hat{\sigma}^2}{8m}$$ where:

$$\hat{\sigma}^2 = \frac{\sum_i e_i^2}{n-p} = \frac{\sum_i (y_i - \hat{\beta} x_i)^2}{3m-1}$$

(c) $$y = \frac{\bar{y}_{(1)} + \bar{y}_{(3)}}{2} + \frac{-\bar{y}_{(1)} + \bar{y}_{(3)}}{4} x$$

Q3 (a) minimize $(y - X\beta)^T (y - X\beta)$

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

$$Cov(\hat{\beta} \mid X) = (X^T X)^{-1} X^T \Sigma X (X^T X)^{-1}$$

(b) minimize $(y - X\beta)^T \Sigma^{-1} (y - X\beta)$

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

$$Cov(\hat{\beta} \mid X) = (X^T \Sigma^{-1} X)^{-1}$$

(c, d) GLS is BLUE

Q4 (a) $Temp = \beta_0 + \beta_1 P + \beta_2 ATemp + e$

(b) $e \mid$ regressors is normal

Q5 (a) Let $W_i = \begin{cases} y_i, & i \le m \\ z_j & \text{otherwise} \end{cases}$

$V_i = \begin{cases} x_i & i \le m \\ 0 & \text{otherwise} \end{cases}$

$$U_i = \begin{cases} 0 & i \leq m \\ V_i & \text{otherwise.} \end{cases}$$

Model: $W_i = \beta_0 + \beta_1 V_i + \Gamma_1 U_i + \varepsilon_i$

(b) $\beta_1 = \gamma_1$

(d) $$\frac{(RSS_0 - RSS_1)/1}{RSS_1/(m+n-3)} \sim F(1, m+n-3)$$

Q6 (a) $$X = \begin{bmatrix} 1/\sqrt{6} & -\frac{1}{2} \\ 1/\sqrt{6} & -\frac{1}{2} \\ 1/\sqrt{6} & 0 \\ 1/\sqrt{6} & 0 \\ 1/\sqrt{6} & \frac{1}{2} \\ 1/\sqrt{6} & \frac{1}{2} \end{bmatrix} \qquad X^TX = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

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

(b, c) t statistic

Q7 a) $\text{MaxSalary} = \beta_0 + NE + NW + \text{score} + e$.

$e|X \sim N(0, \sigma^2 I)$

(b) t tests: $H_0: \beta_i = 0$. $H_1: \beta_i \neq 0$

F test: $H_0: \beta_1 = \beta_2 = \beta_3 = 0$ $H_1$: some nonzero.

(c) $$\frac{(RSS_0 - RSS_1)/3}{RSS_1/491} = 736.3$$

$$\frac{(RSS_0 - RSS_3)/2}{RSS_3/492} = 1096$$

$$\frac{(RSS_0 - RSS_2)/1}{RSS_2/493} = 2194.$$

Then $RSS_0 = \left(\frac{2 \cdot 1036}{492} + 1\right) RSS_3 = \left(\frac{2194}{493} + 1\right) RSS_2.$

Hence $\frac{(RSS_2 - RSS_3)/1}{RSS_3/492} = 0.45 < 3.86 = F_{0.95}(1, 492)$

We can also obtain F value from $(-0.729)^2$. difference is the rounding error.

Q8 (a) $\text{Max Salary} = \beta_0 + \beta_1 \frac{NW}{NE} + \beta_2 \text{Score} + e$

Normal distribution

(b) Non-normal data