Jacob Miller - Final Exam

(3) $y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad , \quad i = 1 \dots n$

failure: $y_i = \beta_0 + \beta_1 \cdot 0 + \varepsilon_i = \beta_0 + \varepsilon_i$

success: $y_i = \beta_0 + \beta_1 \cdot 1 + \varepsilon_i = \beta_0 + \beta_1 + \varepsilon_i$

$$X = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \qquad y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

1) $$X^T X = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$$

2) $$X^T Y = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} y_1 + y_2 \\ y_2 \end{bmatrix}$$

3) $$\hat{\beta} = (X^T X)^{-1} X^T y = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} y_1 + y_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} y_1 + \cancel{y_2 - y_2} \\ -y_1 - y_2 + 2y_2 \end{bmatrix}$$

$$\hat{\beta} = \begin{bmatrix} y_1 \\ -y_1 + y_2 \end{bmatrix}$$

Jacob Miller - Final Exam

4 $H = X(X^TX)^{-1}X^T$

a) show hat matrix is symmetric

$$H^T = (X(X^TX)^{-1}X^T)^T \longrightarrow [(X^TX)^{-1}]^T = (X^TX)^{-1}$$
$$= X[(X^TX)^{-1}]^T \cdot X^T$$
$$= X(X^TX)^{-1}X^T$$
$\therefore$ $\underline{H^T = H}$, symmetric

b) show hat matrix is idempotent

$$HH = (X(X^TX)^{-1}X^T)(X(X^TX)^{-1}X^T)$$
$$= X(X^TX)^{-1} \cdot [X^TX] \cdot [X^TX]^{-1} \cdot X^T$$
$$= X(X^TX)^{-1}X^T$$
$\therefore$ $\underline{HH = H}$, idempotent

c) show
- $0 < h_{ii} < 1$ for all i
- $\sum h_{ii} = p$

$$\sum h_{ii} = \text{trace}(H)$$
$$= tr(X(X^TX)^{-1}X^T) \longrightarrow \text{trace}(AB) = \text{trace}(BA)$$
$$= tr(X^TX \cdot (X^TX)^{-1})$$
$\hookrightarrow$ X is $(n \times p)$, $\therefore$ X^TX is dimension $(p \times p)$

$\therefore$ $(X^TX) \cdot (X^TX)^{-1} = I \cdot p$

$$= tr(I_p)$$
$\therefore$ $\underline{\sum h_{ii} = p}$

also, H is positive semi definite, $\therefore$ eigenvalues are positive

$\therefore$ $0 < h_{ii} < 1$

d) continuing from part c:

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

if X is dimensions $(p \times p)$, then $(X^TX) = (p \times p)$

$\therefore (X^TX) \cdot (X^TX)^{-1} = I_p$

$$Y = HY + (I - H)Y$$
$$= \hat{Y} + \text{residual}$$
$\therefore \hat{Y} = Y - \text{residual}$

&

h_{ii} are leverage points

Jacob Miller - final Exam

(5) 1) CI: $\hat{\beta}_0 - t_{\alpha/2,n-2} \cdot se(\hat{\beta}_0) \le \beta_0 \le \hat{\beta}_0 + t_{\alpha/2,n-2} \cdot se(\hat{\beta}_0)$

$\beta_0 = 40.40$, $se = 1.92$

$t_{\alpha/2,n-2} = t_{0.05/2,\,15-2} = 2.160$

$\therefore \beta_0 = 40.40 \pm 2.160(1.92) = \underline{36.25 \text{ to } 44.55}$ ⇒ 90% confidence interval

↳ in this case, β_0 tells us the mean clarity of lakes in ecoregion A

2) $H_0: \beta_0 = \beta_1 = \beta_2$
H_A: at least one: $\beta_0 \neq \beta_1 \neq \beta_2$

Regression P-Value of 0.000 causes rejection of H_0 & therefore conclude expected lake clarity is NOT the same for all 3 regions

3) $H_0: \beta_0 = \beta_2$ (because indicator variable X_2 associated w/ ecoregion B)
$H_A: \beta_0 \neq \beta_2$

$$F_0 = \frac{SS_R(\beta_0, \beta_2 | \beta_1)/2}{MS_{res}} \rightarrow SS_R(\beta_0,\beta_2|\beta_1) = 2273.7 - 240.1 = 2033.6$$

$$= \frac{2033.6/2}{18.43}$$

$f_0 = 55.17 \quad \therefore p \ll 0.05$

⇒ Reject H_0, conclude expected lake clarity NOT same in A & B