STA 502. Problemset 2

Q4. The multiple regression model

$y = X\beta + \varepsilon$ with K explanatory variables.

Q4-1.

Answer>

(In Scalar format) $y = \beta_0 + \beta_1 X_1 + \cdots + \beta_K X_K + \varepsilon$

$\Downarrow$

For example, matrix $X_1 = \begin{pmatrix} 2 \\ 1 \\ 3 \\ 4 \\ \vdots \end{pmatrix} \cdot \beta_1$ ← unit: 10,000kg

matrix $X_1' = \begin{pmatrix} 20 \\ 10 \\ 30 \\ 40 \\ \vdots \end{pmatrix} \cdot \beta_1$ ← unit: 1000kg

In that case,

matrix $X_1' = 10 \cdot X_1$

$\hookrightarrow y = \beta_0 + \underline{10 \cdot X_1 \cdot \beta_1'} + \beta_2 X_2 + \cdots + \varepsilon$

$\uparrow$

$\beta_1' = \frac{1}{10}\beta_1$

$\therefore$ To correct this error,

we must multiply $\frac{1}{10}\left(\frac{1}{\lambda}\right)$ by

the explanatory variable.

Q4-2.

Answer>

There are no such things that affect coefficients except X, Y.

$\Downarrow$

For example, $y = \alpha + \beta \log x + \varepsilon$

(differentiate both sides)

$$\frac{dy}{dx} = \beta \cdot \frac{1}{x}$$

$$dy = \beta \cdot \frac{1}{x} dx$$

when $x = 10 \cdots y = \frac{1}{10}\beta$

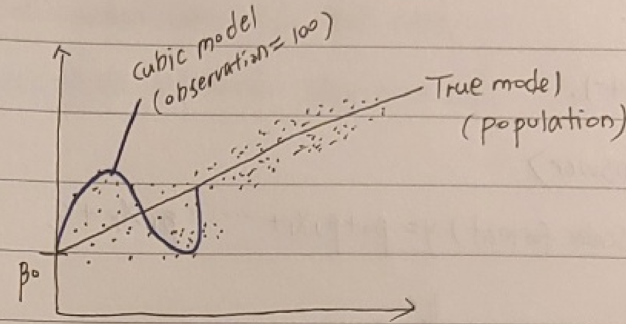
(β doesn't change).

Q5-1.

fitting model : $Y = \beta_0 + \beta_1 X^1 + \beta_2 X^2 + \beta_3 X^3 + \varepsilon$. (cubic model)

True model : $Y = \beta_0 + \beta_1 X + \varepsilon$.

Answer) If true model is linear, RSS for the linear regression is lower than others. The linear part of cubic could fit well, but other parts might not fit well.



Q5-2.

Answer) If the true model is not linear, the traing RSS for the cubic regression is lower than RSS for the linear regression. That's because RSS of the cubic regression has more explanatory variables than others. We could fit the training observation better.