SIAI,

MBA in AI/Big Data

STA502 : Math & Stat for MBA

Yeon Shim, Shim

Assignment # 1, Sep 6, 2021

Problem 3 In the two-variable model

<Basic Conditions>

$$X = [x_1 \ x_2], \quad X' = \begin{bmatrix} x_1' \\ x_2' \end{bmatrix}$$

$$X'X = \begin{bmatrix} x_1' \\ x_2' \end{bmatrix} [x_1, x_2] = \begin{bmatrix} x_1'x_1 & x_1'x_2 \\ x_2'x_1 & x_2'x_2 \end{bmatrix}$$

$$(X'X)^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{(x_1'x_1)(x_2'x_2) - (x_1'x_2)(x_2'x_1)} \begin{bmatrix} x_2'x_2 & -x_1'x_2 \\ -x_2'x_1 & x_1'x_1 \end{bmatrix}$$

$$X'y = \begin{bmatrix} x_1' \\ x_2' \end{bmatrix} y = \begin{bmatrix} x_1'y \\ x_2'y \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

<β> $\beta = \frac{1}{2\times 2 - 1\times 1} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$

<$\hat{\beta}$> $\hat{\beta} = (X'X)^{-1}X'Y = \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 3 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

<Given> $y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \quad (i = 1 \cdots 11) \rightarrow df, 11-2=9.$

$(\beta_1 = 1, \beta_2 = 0) \downarrow$

$y_i = 1 \cdot x_{1i} + 0 \cdot x_{2i} + \varepsilon_i$

obs 12	5	-2
obs 13	3	-7

$\Rightarrow$

$y_{12} = 5 + 0 + \varepsilon_{12} = 5 + \varepsilon_{12} = 5 + \varepsilon_{12} \pm 1.28\sigma$

$y_{13} = 3 + 0 + \varepsilon_{13} = 3 + \varepsilon_{13} = 3 + \varepsilon_{13} \pm 1.28\sigma$

<σ^2> $\sigma^2 = \frac{1}{n-2} \sum \hat{\varepsilon}_i^2 = \frac{1}{n-2} \sum (y_i - x_{1i})^2 = \frac{1}{n-2} \left(\sum y_i^2 - 2\sum y_i x_{1i} + \sum x_{1i}^2 \right)$

$= \frac{1}{11-2} \left(\frac{7}{3} - 2\cdot 2 + 2 \right) = \frac{1}{9} \times \frac{1}{3} = \frac{1}{27}$

<σ> $\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{27}} = 0.19245009$

(1) <when 80% Prediction Interval, = 1.28> $t_{0.1}(9) = 1.383$ ($\because$ unknow σ)

$$\hat{\beta}_0 + \beta_1 x^* \pm t_{\frac{\alpha}{2}} S \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}} + 1}$$

obs 12 : $5 + \varepsilon_{12} \Rightarrow 5 \pm 1.383 \times 0.19245 \sqrt{\frac{1}{15} + \frac{5}{2}^2 + 1} = 5 \pm 0.9803 = [4.020, 5.980]$

obs 13 : $3 + \varepsilon_{13} \Rightarrow 3 \pm 1.383 \times 0.19245 \sqrt{\frac{1}{15} + \frac{5}{2}^2 + 1} = 3 \pm 0.6280 = [2.372, 3.628]$

(2) y_{12}, y_{13} (80% prediction Interval)

$obs_{12} = 5 \pm 1.383 \times 0.19245 \sqrt{\frac{1}{15} + \frac{5^2}{2}} = 5 \pm 0.9435 = [4.0565, 5.9435]$

$obs_{13} = 3 \pm 1.383 \times 0.19245 \sqrt{\frac{1}{15} + \frac{3^2}{2}} = 3 \pm 0.5688 = [2.4312, 3.5688]$

(3) (1), (2) differ reason ?

t distribution depends on the degrees of freedom

That means , a wider interval than the corresponding Gaussian-based confidence interval from before.

If we don't know the standard deviation and we estimate it, we're then less certain about our estimate y