Nguyen Xuan Binh 887799 Theory Exercise Week 2

Exercise 2.3: Considering the data set with 3 observations

$$y_1 = (y_{11}, y_{12}) = (1,2) \quad y_2 = (y_{21}, y_{22}) = (3,4) \quad y_3 = (y_{31}, y_{33}) = (5,6)$$

a) Since second variables are permuted, there are $3! = 6$ distinct permutations

b) The distinct permutations are

$$\begin{Bmatrix}(1,2)\\(3,4)\\(5,6)\end{Bmatrix} \begin{Bmatrix}(1,2)\\(3,6)\\(5,4)\end{Bmatrix} \begin{Bmatrix}(1,4)\\(3,2)\\(5,6)\end{Bmatrix} \begin{Bmatrix}(1,4\\(3,6\\(5,2\end{Bmatrix} \begin{Bmatrix}(1,6)\\(3,2)\\(5,4)\end{Bmatrix} \begin{Bmatrix}(1,6)\\(3,4)\\(5,2)\end{Bmatrix}$$

c) Form 5 bootstrap samples: We have $n = 3$ is the size of original data $\Rightarrow n_{boot} = 3$

Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
$\begin{Bmatrix}(1,2)\\(1,2)\\(5,6)\end{Bmatrix}$	$\begin{Bmatrix}(3,4)\\(5,6)\\(5,6)\end{Bmatrix}$	$\begin{Bmatrix}(3,4)\\(1,2)\\(5,6)\end{Bmatrix}$	$\begin{Bmatrix}(5,6)\\(1,2)\\(1,2)\end{Bmatrix}$	$\begin{Bmatrix}(3,4)\\(3,4)\\(3,4)\end{Bmatrix}$

d) Bootstrap samples should contain only the members in the permutation sets in (b)

$\Rightarrow$ Possible bootstrap samples are 1, 2, 4, 6 and 7

Exercise 2.4: Consider the following models

$y = \alpha_0 + \alpha_1 x + \varepsilon \quad (4)$

$y = \beta_0 + \beta_1 x + \beta_2 z + v \quad (5)$

where we have n observations for variables x, y, z

a) $\sum_{i=1}^{n} \hat{\varepsilon}_i^2 \geq \sum_{i=1}^{n} \hat{v}_i^2$ ($\hat{\varepsilon}$ and $\hat{v}$ are estimated residuals)

We have $y = \alpha_0 + \alpha_1 x + \varepsilon = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} [\alpha_0 \ \alpha_1] + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} = X\alpha + \varepsilon \quad (4)$

$y = \beta_0 + \beta_1 x + \beta_2 z + v = \begin{bmatrix} 1 & x_1 & z_1 \\ \vdots & \vdots & \vdots \\ 1 & x_n & z_n \end{bmatrix} [\beta_0 \ \beta_1 \ \beta_2] + \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = X\beta + v \quad (5)$

We have $SSE_4 \geq SSE_5 \Rightarrow$ Model 5 is more accurate than model 4

$\Rightarrow$ The claim is true in the case statement (5) is accurate model where z parameter has non-zero correlation with y

b) $\hat{\alpha}_1$ is statistically significant (5% significant level) but $\hat{\beta}_1$ is not

For this claim to be true, the null hypothesis $\beta_1 = 0$ and the alternate hypothesis $\alpha_1 \neq 0$ must hold true. Rewrite the models:

$y = \alpha_0 + \alpha_1 x + \varepsilon \quad (4)$

$y = \beta_0 + \beta_2 z + v \quad (5)$

For both models to be true $\Rightarrow \alpha_1$ and $\beta_2 \neq 0 \Rightarrow x$ and z are linearly dependent

Indeed, in $y = \beta_0 + \beta_1 x + \beta_2 z + v$ (5), since x and z are linearly dependent, the x-component isn't necessary $\Rightarrow$ This claim is true in the case x and z parameters are linearly dependent

c) $\hat{\alpha}_1$ is not statistically significant (5% significant level), but $\hat{\beta}_1$ is

Null hypothesis: $\alpha_1 = 0$ and alternate hypothesis: $\beta_1 \neq 0$. Rewrite the models

$y = \alpha_0 + \varepsilon \quad (4)$

$y = \beta_0 + \beta_1 x + \beta_2 z + v \quad (5)$

From (4), we see that y can be modeled without effects from x and z. However in (5), x and z explanatory variables are significant in modeling $y \Rightarrow$ contrary: this claim is not true in any situation

d) R^2 of model (4) > R^2 of model (5)

We have $R^2_{(4)} = 1 - \frac{\sum_{i=1}^{n} \hat{\varepsilon}_i^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \qquad R^2_{(5)} = 1 - \frac{\sum_{i=1}^{n} \hat{v}_i^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$

$R^2_{(4)} > R^2_{(5)} \Rightarrow \sum_{i=1}^{n} \hat{\varepsilon}_i^2 < \sum_{i=1}^{n} \hat{v}_i^2 \Rightarrow SSE_4 < SSE_5$

$\Rightarrow$ This claim is true when the sole variable x can predict y better than linear combination of x and z $\Rightarrow$ There may be multicollinearity in (5) of x and z that makes the model less accurate. However, if that's the case, then the null hypothesis $\beta_2 = 0$ is true for model (5)

$\Rightarrow$ If z has non-zero correlation with y, then this statement is not true in any situation