

BÀI TẬP MÔN THỐNG KÊ NHIỀU CHIỀU

Tuần 7

Thành viên	MSSV	Bài tập
Đình Anh Huy	18110103	7.12, 7.13, 7.20, 7.25
Nguyễn Đức Vũ Duy	18110004	7.9, 7.19, 7.21, 7.26

Bài tập 7.9. Xem xét bộ dữ liệu sau về một biến z_1 và 2 responses Y_1, Y_2 .

z_1	-2	-1	0	1	2
y_1	5	3	4	2	1
y_2	-3	-1	-1	2	3

Tìm ước lượng bình phương tối thiểu của tham số trong model đường thẳng hồi quy 2 chiều

$$Y_{j1} = \beta_{01} + \beta_{11}z_{j1} + \epsilon_{j1}$$

$$Y_{j2} = \beta_{02} + \beta_{12}z_{j1} + \epsilon_{j2}, j = 1, 2, 3, 4, 5$$

Tính ma trận các giá trị fit được $\hat{\mathbf{Y}}$ và sai số $\hat{\epsilon}$ với $\hat{\mathbf{Y}} = [\mathbf{y}_1 \mathbf{y}_2]$. Kiểm chứng tổng bình phương và phân rã tích chéo.

$$\mathbf{Y}'\mathbf{Y} = \hat{\mathbf{Y}}'\hat{\mathbf{Y}} + \hat{\epsilon}'\hat{\epsilon}$$

Lời giải

Từ giả thiết, ta có: $\mathbf{Z} = \begin{bmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$ Và, $\mathbf{Y} = \begin{bmatrix} 5 & -3 \\ 3 & -1 \\ 4 & -1 \\ 2 & 2 \\ 1 & 3 \end{bmatrix}$. Khi đó, ước lượng bình phương nhỏ

nhất của tham số cần tìm là:

$$\hat{\beta} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y} = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.1 \end{bmatrix} \begin{bmatrix} 15 & 0 \\ -9 & 15 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ -0.9 & 1.5 \end{bmatrix}$$

Do đó, giá trị dự đoán là:

$$\hat{\mathbf{y}} = \mathbf{Z}\hat{\beta} = \begin{bmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ -0.9 & 1.5 \end{bmatrix} = \begin{bmatrix} 4.8 & -3 \\ 3.9 & -1.5 \\ 3 & 0 \\ 2.1 & 1.5 \\ 1.2 & 3 \end{bmatrix}$$

Vậy nên, ta tính được sai số:

$$\hat{\epsilon} = \mathbf{Y} - \hat{\mathbf{y}} = \begin{bmatrix} 5 & -3 \\ 3 & -1 \\ 4 & -1 \\ 2 & 2 \\ 1 & 3 \end{bmatrix} - \begin{bmatrix} 4.8 & -3 \\ 3.9 & -1.5 \\ 3 & 0 \\ 2.1 & 1.5 \\ 1.2 & 3 \end{bmatrix} = \begin{bmatrix} 0.2 & 0 \\ -0.9 & 0.5 \\ 1 & -1 \\ -0.1 & 0.5 \\ -0.2 & 0 \end{bmatrix}$$

Ta kiểm chứng giả thuyết trên như sau, ta có:

$$\hat{\mathbf{Y}}'\hat{\mathbf{Y}} + \hat{\epsilon}'\hat{\epsilon} = \begin{bmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}^T \begin{bmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} + \begin{bmatrix} 0.2 & 0 \\ -0.9 & 0.5 \\ 1 & -1 \\ -0.1 & 0.5 \\ -0.2 & 0 \end{bmatrix}^T \begin{bmatrix} 0.2 & 0 \\ -0.9 & 0.5 \\ 1 & -1 \\ -0.1 & 0.5 \\ -0.2 & 0 \end{bmatrix} = \begin{bmatrix} 55 & -15 \\ -15 & 24 \end{bmatrix}$$

Trong khi đó,

$$\mathbf{Y}'\mathbf{Y} = \begin{bmatrix} 5 & -3 \\ 3 & -1 \\ 4 & -1 \\ 2 & 2 \\ 1 & 3 \end{bmatrix}^T \begin{bmatrix} 5 & -3 \\ 3 & -1 \\ 4 & -1 \\ 2 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 55 & -15 \\ -15 & 24 \end{bmatrix}$$

Do đó, $\mathbf{Y}'\mathbf{Y} = \hat{\mathbf{Y}}'\hat{\mathbf{Y}} + \hat{\epsilon}'\hat{\epsilon}$.

Bài tập 7.12. Cho vectơ trung bình và ma trận hiệp phương sai của Y, Z_1 và Z_2 ,

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_Y \\ \boldsymbol{\mu}_Z \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \\ -2 \end{bmatrix} \quad \text{và} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{YY} & \boldsymbol{\sigma}'_{ZY} \\ \boldsymbol{\sigma}_{ZY} & \boldsymbol{\Sigma}_{ZZ} \end{bmatrix} = \begin{bmatrix} 9 & 3 & 1 \\ 3 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

xác định từng ý sau đây

- (a) Linear predictor $\beta_0 + \beta_1 Z_1 + \beta_2 Z_2$ tốt nhất của Y .
- (b) Sai số bình phương trung bình của linear predictor tốt nhất.
- (c) Hệ số tương quan bội của tổng thể.
- (d) Hệ số tương quan từng phần $\rho_{Y Z_1 \cdot Z_2}$.

Lời giải

(a) Ta có

$$\begin{aligned}\beta &= \Sigma_{ZZ}^{-1} \sigma_{ZY} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \end{bmatrix} \\ \beta_0 &= \mu_Y - \beta' \mu_Z = 4 - \begin{bmatrix} 2 & -1 \end{bmatrix} \begin{bmatrix} 3 \\ -2 \end{bmatrix} = -4\end{aligned}$$

Khi đó, theo kết quả 7.12, ta có linear predictor tốt nhất của Y là

$$\beta_0 + \beta' Z = -4 + \begin{bmatrix} 2 & -1 \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = -4 + 2Z_1 - Z_2.$$

(b) Trung bình sai số bình phương (mean square error) là

$$\sigma_{YY} - \sigma'_{ZY} \Sigma_{ZZ}^{-1} \sigma_{ZY} = 9 - \begin{bmatrix} 3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = 4$$

(c) Ta có

$$\sigma'_{ZY} \Sigma_{ZZ}^{-1} \sigma_{ZY} = \begin{bmatrix} 3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = 5$$

Theo (7-48), ta có population multiple correlation coefficient là

$$\rho_{Y(Z)} = +\sqrt{\frac{\sigma'_{ZY} \Sigma_{ZZ}^{-1} \sigma_{ZY}}{\sigma_{YY}}} = \sqrt{\frac{5}{9}} \approx 0.745$$

(d) Xét phân hoạch mới của Σ như sau

$$\Sigma = \left[\begin{array}{c|c} \Sigma_{YZ_1} & \Sigma'_{ZY} \\ \hline \Sigma_{ZY} & \Sigma_{Z_2Z_2} \end{array} \right] = \left[\begin{array}{cc|c} 9 & 3 & 1 \\ 3 & 2 & 1 \\ \hline 1 & 1 & 1 \end{array} \right]$$

Theo kết quả 7.14, ta có ma trận hiệp phương sai của $\begin{bmatrix} Y \\ Z_1 \end{bmatrix}$ khi biết Z_2 là

$$\Sigma_{YZ_1 \cdot Z_2} = \Sigma_{YZ_1} - \Sigma_{YZ} \Sigma_{Z_2Z_2}^{-1} \Sigma_{ZY} = \begin{bmatrix} 9 & 3 \\ 3 & 2 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} (1)^{-1} \begin{bmatrix} 1 & 1 \end{bmatrix} = \begin{bmatrix} 8 & 2 \\ 2 & 1 \end{bmatrix}$$

Khi đó, theo (7-56), ta có partial correlation coefficient giữa Y và Z_1 , loại bỏ Z_2 là

$$\rho_{YZ_1 \cdot Z_2} = \frac{\sigma_{YZ_1 \cdot Z_2}}{\sqrt{\sigma_{YY \cdot Z_2}} \sqrt{\sigma_{Z_1Z_1 \cdot Z_2}}} = \frac{2}{\sqrt{8}\sqrt{1}} = \frac{\sqrt{2}}{2} \approx 0.707$$

Bài tập 7.13. Điểm thi của các sinh viên đại học được mô tả trong ví dụ 5.5 có

$$\bar{\mathbf{z}} = \begin{bmatrix} \bar{z}_1 \\ \bar{z}_2 \\ \bar{z}_3 \end{bmatrix} = \begin{bmatrix} 527.74 \\ 54.69 \\ 25.13 \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} 5691.34 & & \\ 600.51 & 126.05 & \\ 217.25 & 23.37 & 23.11 \end{bmatrix}$$

Giả sử các biến có cùng phân phối chuẩn đồng thời.

- Thu thập ước lượng hợp lý cực đại của các tham số cho predicting Z_1 từ Z_2 và Z_3 .
- Tính hệ số tương quan bội ước tính $R_{Z_1|Z_{(2)}}$.
- Xác định hệ số tương quan từng phần ước tính R_{Z_1, Z_2, Z_3} .

Lời giải

- Xét phân hoạch của vectơ trung bình và ma trận hiệp phương sai của Z_1, Z_2 và Z_3 như sau

$$\bar{\mathbf{z}} = \begin{bmatrix} \bar{Z}_1 \\ \bar{\mathbf{Z}}_{(2)} \end{bmatrix} = \begin{bmatrix} 527.74 \\ 54.69 \\ 25.13 \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} s_{Z_1 Z_1} & \mathbf{s}_{Z_1 \mathbf{Z}_{(2)}} \\ \mathbf{s}_{\mathbf{Z}_{(2)} Z_1} & \mathbf{S}_{\mathbf{Z}_{(2)} \mathbf{Z}_{(2)}} \end{bmatrix} = \begin{bmatrix} 5691.34 & 600.51 & 217.25 \\ 600.51 & 126.05 & 23.37 \\ 217.25 & 23.37 & 23.11 \end{bmatrix}$$

Do Z_1, Z_2, Z_3 được giả sử rằng có cùng phân phối chuẩn đồng thời nên áp dụng kết quả 7.13 ta có ước lượng hợp lý cực đại của các hệ số là

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \mathbf{S}_{\mathbf{Z}_{(2)} \mathbf{Z}_{(2)}}^{-1} \mathbf{s}_{\mathbf{Z}_{(2)} Z_1} = \begin{bmatrix} 126.05 & 23.37 \\ 23.37 & 23.11 \end{bmatrix}^{-1} \begin{bmatrix} 600.51 \\ 217.25 \end{bmatrix} \approx \begin{bmatrix} 3.72 \\ 5.64 \end{bmatrix} \\ \hat{\beta}_0 &= \bar{Z}_1 - \hat{\boldsymbol{\beta}}' \bar{\mathbf{Z}}_{(2)} = 527.74 - \begin{bmatrix} 3.72 & 5.64 \end{bmatrix} \begin{bmatrix} 54.69 \\ 25.13 \end{bmatrix} = 182.56 \end{aligned}$$

- Ta có

$$\mathbf{s}_{\mathbf{Z}_{(2)} Z_1}' \mathbf{S}_{\mathbf{Z}_{(2)} \mathbf{Z}_{(2)}}^{-1} \mathbf{s}_{\mathbf{Z}_{(2)} Z_1} = \begin{bmatrix} 600.51 & 217.25 \end{bmatrix} \begin{bmatrix} 126.05 & 23.37 \\ 23.37 & 23.11 \end{bmatrix}^{-1} \begin{bmatrix} 600.51 \\ 217.25 \end{bmatrix} \approx 3458.28$$

Theo (7-28), ta có multiple correlation coefficient là

$$R_{Z_1(Z_2, Z_3)} = +\sqrt{\frac{\mathbf{s}_{\mathbf{Z}_{(2)} Z_1}' \mathbf{S}_{\mathbf{Z}_{(2)} \mathbf{Z}_{(2)}}^{-1} \mathbf{s}_{\mathbf{Z}_{(2)} Z_1}}{s_{Z_1 Z_1}}} = \sqrt{\frac{3458.28}{5691.34}} \approx 0.78$$

- (c) Xét một phân hoạch khác của vectơ trung bình và ma trận hiệp phương sai của Z_1, Z_2 và Z_3 như sau

$$\bar{\mathbf{z}} = \begin{bmatrix} \bar{\mathbf{Z}}_{(1)} \\ \bar{\mathbf{Z}}_3 \end{bmatrix} = \begin{bmatrix} 527.74 \\ 54.69 \\ 25.13 \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} \Sigma_{\mathbf{Z}_{(1)}\mathbf{Z}_{(1)}} & \Sigma_{\mathbf{Z}_{(1)}\mathbf{Z}_3} \\ \Sigma_{\mathbf{Z}_3\mathbf{Z}_{(1)}} & \Sigma_{\mathbf{Z}_3\mathbf{Z}_3} \end{bmatrix} = \begin{bmatrix} 5691.34 & 600.51 & 217.25 \\ 600.51 & 126.05 & 23.37 \\ 217.25 & 23.37 & 23.11 \end{bmatrix}$$

Theo kết quả 7.14, ta có ma trận hiệp phương sai của $\mathbf{Z}_{(1)} = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}$ khi biết Z_3 là

$$\Sigma_{\mathbf{Z}_{(1)} \cdot \mathbf{Z}_3} = \Sigma_{\mathbf{Z}_{(1)}\mathbf{Z}_{(1)}} - \Sigma_{\mathbf{Z}_{(1)}\mathbf{Z}_3} \Sigma_{\mathbf{Z}_3\mathbf{Z}_3}^{-1} \Sigma_{\mathbf{Z}_3\mathbf{Z}_{(1)}} = \begin{bmatrix} 3649.04 & 380.82 \\ 380.82 & 102.42 \end{bmatrix}$$

Khi đó, theo (7-56), ta có partial correlation coefficient giữa Z_1 và Z_2 , loại bỏ Z_3 là

$$R_{Z_1, Z_2 \cdot Z_3} = \frac{\sigma_{Z_1 Z_2 \cdot Z_3}}{\sqrt{\sigma_{Z_1 Z_1 \cdot Z_3}} \sqrt{\sigma_{Z_2 Z_2 \cdot Z_3}}} = \frac{380.82}{\sqrt{3649.04} \sqrt{102.42}} \approx 0.62$$