BÀI TẬP MÔN THỐNG KÊ NHIỀU CHIỀU

Tuần 6

Đinh Anh Huy - 18110103

Nguyễn Đức Vũ Duy - 18110004

Bài tập 6.5. Một nghiên cứu sinh xem xét ba chỉ số đo mức độ nghiêm trọng của các cơn đau tim. Giá trị của các chỉ số này cho $n = 40$ bệnh nhân đau tim được tiếp nhận ở phòng cấp cứu tại bệnh viện cho các thống kê tóm tắt như sau

$$\overline{\mathbf{x}} = \begin{bmatrix} 46.1 \\ 57.3 \\ 50.4 \end{bmatrix} \text{ và } \mathbf{S} = \begin{bmatrix} 101.3 & 63.0 & 71.0 \\ 63.0 & 80.2 & 55.6 \\ 71.0 & 55.6 & 97.4 \end{bmatrix}$$

(a) Cả 3 chỉ số được tính cho mỗi bệnh nhân. Kiểm định trung bình của các chỉ số bằng cách sử dụng công thức (6-16) với mức ý nghĩa $\alpha = 0.05$.

(b) Đánh giá sự sai khác giữa các cặp chỉ số trung bình bằng cách sử dụng khoảng tin cậy đồng thời 95%.

Lời giải

(a) Với $\boldsymbol{\mu'} = [\mu_1, \mu_2, \mu_3, \mu_4]$, thì ta có ma trận contrast $\mathbf{C}$ như sau

$$\mathbf{C} = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$$

Ta cần kiểm định giả thuyết $H_0 : \mathbf{C}\boldsymbol{\mu} = \mathbf{0}$ với giả thuyết $H_1 : \mathbf{C}\boldsymbol{\mu} \neq \mathbf{0}$.

Giả sử H_0 đúng, ta xét thống kê T^2 cho kiểm định giả thuyết $H_0 : \mathbf{C}\boldsymbol{\mu} = \mathbf{0}$.

Theo đề, ta có các thống kê được tóm tắt như sau

$$\overline{\mathbf{x}} = \begin{bmatrix} 46.1 \\ 57.3 \\ 50.4 \end{bmatrix} \text{ và } \mathbf{S} = \begin{bmatrix} 101.3 & 63.0 & 71.0 \\ 63.0 & 80.2 & 55.6 \\ 71.0 & 55.6 & 97.4 \end{bmatrix}$$

Khi đó ta có

$$\mathbf{C}\overline{\mathbf{x}} = \begin{bmatrix} 11.2 \\ -6.9 \end{bmatrix}; \quad \mathbf{CSC'} = \begin{bmatrix} 55.5 & -32.6 \\ -32.6 & 66.4 \end{bmatrix}$$

Trên mẫu thực nghiệm, ta có giá trị thống kê

$$T^2 = n(\mathbf{C}\overline{\mathbf{x}})'(\mathbf{CSC}')^{-1}(\mathbf{C}\overline{\mathbf{x}}) = 40 \begin{bmatrix} 11.2 & -6.9 \end{bmatrix} \begin{bmatrix} 55.5 & -32.6 \\ -32.6 & 66.4 \end{bmatrix} \begin{bmatrix} 11.2 \\ -6.9 \end{bmatrix} \approx 90.49457$$

Mặt khác, với mức ý nghĩa $\alpha = 0.05$, ta có

$$\frac{(n-1)(q-1)}{(n-q+1)}F_{q-1,n-q+1}(\alpha) = \frac{39\times 2}{38}F_{2,38}(0.05) \approx \frac{39\times 2}{38}3.2448 \approx 6.66037$$

Vì $T^2 = 90.49457 > 6.66037$, nên ta bác bỏ giả thuyết H_0 với mức ý nghĩa $\alpha = 0.05$. Do đó với mức ý nghĩa $\alpha = 0.05$ thì các chỉ số trung bình không bằng nhau.

(b) Với

$$\mathbf{c}_1' = [-1, 1, 0] \quad \text{và} \quad \mathbf{c}_2' = [0, -1, 1]$$

thì ta có khoảng tin cậy đồng thời 95% của

* $\mathbf{c}_1'\boldsymbol{\mu} = \mu_2 - \mu_1$:

$$\overline{x}_2 - \overline{x}_1 \pm \sqrt{\frac{39\times 2}{38}F_{2,38}(0.05)}\sqrt{\frac{\mathbf{c}_1'\mathbf{S}\mathbf{c}_1}{40}} = (57.3-46.1)\pm\sqrt{6.66037}\sqrt{\frac{55.5}{40}} = 11.2\pm 3.0399$$

* $\mathbf{c}_2'\boldsymbol{\mu} = \mu_3 - \mu_2$:

$$\overline{x}_3 - \overline{x}_2 \pm \sqrt{\frac{39\times 2}{38}F_{2,38}(0.05)}\sqrt{\frac{\mathbf{c}_2'\mathbf{S}\mathbf{c}_2}{40}} = (50.4-57.3)\pm\sqrt{6.66037}\sqrt{\frac{66.4}{40}} = -6.9\pm 3.3251$$

Từ 2 khoảng tin cậy đồng thời 95% ở trên, ta có thể thấy rằng không có khoảng tin cậy nào chứa giá trị 0. Điều đó cho thấy rằng các giá trị trung bình khác nhau từng đôi một.

Bài tập 6.6. Sử dụng dữ liệu cho treatments 2 và 3 từ bài 6.8

(a) Tính $\boldsymbol{S_{pooled}}$

(b) Kiểm định $H_0 : \boldsymbol{\mu_2} - \boldsymbol{\mu_3} = 0$ sử dụng một phương pháp 2 mẫu với $\alpha = .01$

c) Xây dựng khoảng tin cậy đồng thời 99% cho sự chênh lệch $\mu_{2i} - \mu_{3i}$, $i = 1, 2$

Lời giải

(a) Theo đề bài, ta có bộ dữ liệu như sau:

treatment 2: $\begin{bmatrix}3\\3\end{bmatrix}, \begin{bmatrix}1\\6\end{bmatrix}, \begin{bmatrix}2\\3\end{bmatrix}$

treatment 3: $\begin{bmatrix}2\\3\end{bmatrix}, \begin{bmatrix}5\\1\end{bmatrix}, \begin{bmatrix}3\\1\end{bmatrix}, \begin{bmatrix}2\\3\end{bmatrix}$

Từ đó, ta tính được:

$$\overline{\boldsymbol{X_2}} = \begin{bmatrix}2\\4\end{bmatrix}, \boldsymbol{S_2} = \begin{bmatrix}1 & -1.5\\-1.5 & 3\end{bmatrix}$$

Và

$$\overline{\boldsymbol{X_3}} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \boldsymbol{S_3} = \begin{bmatrix} 2 & -1.333 \\ -1.333 & 1.333 \end{bmatrix}$$

Khi đó,

$$\boldsymbol{S_{pooled}} = \frac{n_2 - 1}{n_2 + n_3 - 2}\boldsymbol{S_2} + \frac{n_2 - 1}{n_1 + n_2 - 2}\boldsymbol{S_3} = \begin{bmatrix} 1.6 & -1.4 \\ -1.4 & 2 \end{bmatrix}$$

(b) Xét kiểm định $H_0 : \boldsymbol{\mu_2} - \boldsymbol{\mu_3} = 0$ với $\alpha = .01$: Áp dụng kết quả 6.2, ta tính:

$$T^2 = [\overline{\boldsymbol{X_2}} - \overline{\boldsymbol{X_3}}]'[(\frac{1}{n_2} + \frac{1}{n_3})\boldsymbol{S_{pooled}}]^{-1}[\overline{\boldsymbol{X_2}} - \overline{\boldsymbol{X_3}}] = 3.87096$$

Với $p = 2$,

$$c^2 = \frac{(n_2 + n_3 - 2).p}{(n_2 + n_3 - p - 1)}F_{p,n_2+n_3-p-1}(0.01) = 45$$

Ta có, $T^2 < c^2$ nên ta không có đủ dữ kiện để bác bỏ H_0 với mức ý nghĩa $\alpha = 0.01$.

c) Áp dụng kết quả 6.3, ta có khoảng tin cậy đồng thời 99% cho $\mu_{21} - \mu_{31}$ là:

$$(\overline{X_{21}} - \overline{X_{31}}) \pm c\sqrt{(\frac{1}{n_2} + \frac{1}{n_3})s_{11,pooled}} = (2 - 3) \pm \sqrt{45}\sqrt{(\frac{1}{3} + \frac{1}{4})1.6} = [-7.48, 5.48]$$

Tương tự, ta có khoảng tin cậy đồng thời 99% cho $\mu_{22} - \mu_{32}$ là:

$$(\overline{X_{22}} - \overline{X_{32}}) \pm c\sqrt{(\frac{1}{n_2} + \frac{1}{n_3})s_{22,pooled}} = (4 - 2) \pm \sqrt{45}\sqrt{(\frac{1}{3} + \frac{1}{4})2} = [-5.24, 9.24]$$

Bài tập 6.8. Các quan trắc trên 2 phản hồi được thu thập từ 3 treatments. Các vectơ quan trắc $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ được cho như sau

$$\text{Treatment 1: } \begin{bmatrix} 6 \\ 7 \end{bmatrix}, \begin{bmatrix} 5 \\ 9 \end{bmatrix}, \begin{bmatrix} 8 \\ 6 \end{bmatrix}, \begin{bmatrix} 4 \\ 9 \end{bmatrix}, \begin{bmatrix} 7 \\ 9 \end{bmatrix}$$

$$\text{Treatment 2: } \begin{bmatrix} 3 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 6 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

$$\text{Treatment 3: } \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 5 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

(a) Chia nhỏ các quan trắc thành các thành phần mean, treatment và residual như trong (6-39). Xây dựng các ma trận tương ứng cho mỗi biến.

(b) Sử dụng thông tin từ câu (a), xây dựng bảng one-way MANOVA.

(c) Ước lượng Wilk's lambda, Λ^*, và sử dụng bảng 6.3 cho kiểm định treatment effect. Với mức ý nghĩa $\alpha = 0.01$, lặp lại kiểm định bằng cách sử dụng xấp xỉ chi bình phương với Bartlett's correction. So sánh các kết luận.

Lời giải

Ta có mẫu sau

$$
\begin{aligned}
&\text{Treatment 1:}\quad \mathbf{x}_{11} = \begin{bmatrix} 6 \\ 7 \end{bmatrix}, \quad \mathbf{x}_{12} = \begin{bmatrix} 5 \\ 9 \end{bmatrix}, \quad \mathbf{x}_{13} = \begin{bmatrix} 8 \\ 6 \end{bmatrix}, \quad \mathbf{x}_{14} = \begin{bmatrix} 4 \\ 9 \end{bmatrix}, \quad \mathbf{x}_{15} = \begin{bmatrix} 7 \\ 9 \end{bmatrix}, \quad (n_1 = 5) \\
&\text{Treatment 2:}\quad \mathbf{x}_{21} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}, \quad \mathbf{x}_{22} = \begin{bmatrix} 1 \\ 6 \end{bmatrix}, \quad \mathbf{x}_{23} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \quad (n_2 = 3) \\
&\text{Treatment 3:}\quad \mathbf{x}_{31} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \quad \mathbf{x}_{32} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}, \quad \mathbf{x}_{33} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \quad \mathbf{x}_{34} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \quad (n_3 = 4)
\end{aligned}
$$

trong đó, số lượng biến là $p = 2$, số lượng nhóm là $g = 3$ và tổng số lượng quan trắc là $n = n_1 + n_2 + n_3 = 12$. Khi đó, ta có

$$
\begin{aligned}
\overline{\mathbf{x}}_1 &= \frac{1}{n_1}\sum_{j=1}^{n_1} \mathbf{x}_{1j} = \frac{1}{5}\left(\begin{bmatrix} 6 \\ 7 \end{bmatrix} + \begin{bmatrix} 5 \\ 9 \end{bmatrix} + \begin{bmatrix} 8 \\ 6 \end{bmatrix} + \begin{bmatrix} 4 \\ 9 \end{bmatrix} + \begin{bmatrix} 7 \\ 9 \end{bmatrix}\right) = \begin{bmatrix} 6 \\ 8 \end{bmatrix} \\
\overline{\mathbf{x}}_2 &= \frac{1}{n_2}\sum_{j=1}^{n_2} \mathbf{x}_{2j} = \frac{1}{3}\left(\begin{bmatrix} 3 \\ 3 \end{bmatrix} + \begin{bmatrix} 1 \\ 6 \end{bmatrix} + \begin{bmatrix} 2 \\ 3 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 4 \end{bmatrix} \\
\overline{\mathbf{x}}_3 &= \frac{1}{n_3}\sum_{j=1}^{n_3} \mathbf{x}_{3j} = \frac{1}{4}\left(\begin{bmatrix} 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 5 \\ 1 \end{bmatrix} + \begin{bmatrix} 3 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 \\ 3 \end{bmatrix}\right) = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \\
\overline{\mathbf{x}} &= \frac{1}{n}\sum_{i=1}^{3}\sum_{j=1}^{n_i} \mathbf{x}_{ij} = \begin{bmatrix} 4 \\ 5 \end{bmatrix}
\end{aligned}
$$

(a) Xét phản hồi đầu tiên thu được từ các quan trắc. Ta muốn biểu diễn các quan trắc thành tổng của các thành phần overall mean, treatment effect và residual với từng phần tử có dạng như sau

$$
\begin{array}{ccccccc}
\text{observation} & = & \text{mean} & + & \text{treatment effect} & + & \text{residual} \\
(x_{\ell j}) & = & (\overline{x}) & + & (\overline{x}_\ell - \overline{x}) & + & (x_{\ell j} - \overline{x}_\ell)
\end{array}
$$

trong đó $\ell = 1, 2, 3; j = 1, ..., n_\ell$. Ta có

$$\begin{aligned}
6 = x_{11}^{(1)} &= \overline{x}^{(1)} + (\overline{x}_1^{(1)} - \overline{x}^{(1)}) + (x_{11}^{(1)} - \overline{x}_1^{(1)}) \\
&= 4 + (6-4) + (6-6) \\
&= 4 + 2 + 0 \\
&\vdots \\
3 = x_{21}^{(1)} &= \overline{x}^{(1)} + (\overline{x}_2^{(1)} - \overline{x}^{(1)}) + (x_{21}^{(1)} - \overline{x}_2^{(1)}) \\
&= 4 + (2-4) + (3-2) \\
&= 4 + (-2) + 1 \\
&\vdots \\
2 = x_{31}^{(1)} &= \overline{x}^{(1)} + (\overline{x}_3^{(1)} - \overline{x}^{(1)}) + (x_{31}^{(1)} - \overline{x}_3^{(1)}) \\
&= 4 + (3-4) + (2-3) \\
&= 4 + (-1) + (-1) \\
&\vdots
\end{aligned}$$

Khi đó ta thu được

$$\begin{bmatrix} 6 & 5 & 8 & 4 & 7 \\ 3 & 1 & 2 & & \\ 2 & 5 & 3 & 2 & \end{bmatrix} = \begin{bmatrix} 4 & 4 & 4 & 4 & 4 \\ 4 & 4 & 4 & & \\ 4 & 4 & 4 & 4 & \end{bmatrix} + \begin{bmatrix} 2 & 2 & 2 & 2 & 2 \\ -2 & -2 & -2 & & \\ -1 & -1 & -1 & -1 & \end{bmatrix} + \begin{bmatrix} 0 & -1 & 2 & -2 & 1 \\ 1 & -1 & 0 & & \\ -1 & 2 & 0 & -1 & \end{bmatrix}$$

Hơn nữa, ta có các bình phương độ dài sau

$$\begin{aligned}
SS_{obs} &= 6^2 + 5^2 + 8^2 + 4^2 + 7^2 + 3^2 + 1^2 + 2^2 + 2^2 + 5^2 + 3^2 + 2^2 = 246 \\
SS_{mean} &= 12 \times 4^2 = 192 \\
SS_{tr} &= 2^2 \times 5 + (-2)^2 \times 3 + (-1)^2 \times 4 = 36 \\
SS_{res} &= 0^2 + (-1)^2 + 2^2 + (-2)^2 + 1^2 + 1^2 + (-1)^2 + 0^2 + (-1)^2 + 2^2 + 0^2 + (-1)^2 = 18
\end{aligned}$$

Khi đó

$$\begin{array}{ccccccc} SS_{obs} & = & SS_{mean} & + & SS_{tr} & + & SS_{res} \\ 246 & = & 192 & + & 36 & + & 18 \end{array}$$

và

$$\text{Total SS (corrected)} = SS_{obs} - SS_{mean} = 246 - 192 = 54$$

Tương tự khi xét phản hồi thứ hai thu được từ các quan trắc. Ta thu được

$$\begin{bmatrix} 7 & 9 & 6 & 9 & 9 \\ 3 & 6 & 3 & & \\ 3 & 1 & 1 & 3 & \end{bmatrix} = \begin{bmatrix} 5 & 5 & 5 & 5 & 5 \\ 5 & 5 & 5 & & \\ 5 & 5 & 5 & 5 & \end{bmatrix} + \begin{bmatrix} 3 & 3 & 3 & 3 & 3 \\ -1 & -1 & -1 & & \\ -3 & -3 & -3 & -3 & \end{bmatrix} + \begin{bmatrix} -1 & 1 & -2 & 2 & 1 \\ -1 & 2 & -1 & & \\ 1 & -1 & -1 & 1 & \end{bmatrix}$$

Hơn nữa, ta có các bình phương độ dài sau

$$\begin{aligned}
SS_{obs} &= 7^2+9^2+6^2+9^2+9^2+3^2+6^2+3^2+3^2+1^2+1^2+3^2 = 402\\
SS_{mean} &= 12\times 5^2 = 300\\
SS_{tr} &= 3^2\times 5+(-1)^2\times 3+(-3)^2\times 4 = 84\\
SS_{res} &= (-1)^2+1^2+(-2)^2+1^2+1^2+(-1)^2+2^2+(-1)^2+1^2+(-1)^2+(-1)^2+1^2 = 18
\end{aligned}$$

Khi đó

$$\begin{array}{ccccccc}
SS_{obs} & = & SS_{mean} & + & SS_{tr} & + & SS_{res}\\
402 & = & 300 & + & 84 & + & 18
\end{array}$$

và

$$\text{Total SS (corrected)} = SS_{obs} - SS_{mean} = 402-300 = 102$$

(b) Bằng cách nhân từng dòng của các ma trận tương ứng cho cả hai biến, ta thu được cross product contributions như sau

$$\begin{aligned}
\text{Mean:}\quad & 4\times 5+4\times 5+\ldots+4\times 5 = 12\times 20 = 240\\
\text{Treatment:}\quad & 5\times 2\times 3+3\times(-2)\times(-1)+4\times(-1)\times(-3) = 48\\
\text{Residual:}\quad & 0\times(-1)+(-1)\times 1+2\times(-2)+(-2)\times 1+\ldots+0\times(-1)+(-1)\times 1 = -13
\end{aligned}$$

và khi đó

$$\begin{aligned}
\text{Total (corrected) cross product} &= \text{total cross product} - \text{mean cross product}\\
&= (240+48+(-13))-240 = 35
\end{aligned}$$

Kết hợp các kết quả thu được từ câu (a), ta có được bảng one-way MANOVA như sau

Source of variance	Matrix of sum of squares and cross products	Degrees of freedom (d.f.)
Treatment	$\mathbf{B} = \begin{bmatrix} 36 & 48\\ 48 & 84\end{bmatrix}$	$3-1=2$
Residual	$\mathbf{W} = \begin{bmatrix} 18 & -13\\ -13 & 18\end{bmatrix}$	$5+3+4-3=9$
Total (corrected)	$\mathbf{B}+\mathbf{W} = \begin{bmatrix} 54 & 35\\ 35 & 102\end{bmatrix}$	$5+3+4-1=11$

(c) Phát biểu giả thuyết

$$\begin{cases} H_0: \tau_1=\tau_2=\tau_3=0\\ H_1: \exists i\in\{1,2,3\}: \tau_i\neq 0\end{cases}$$

Giả sử giả thuyết H_0 đúng, ta sử dụng thống kê Wilk's lambda để kiểm định giả thuyết trên. Khi đó ta có giá trị thống kê

$$\Lambda^* = \frac{|\mathbf{W}|}{|\mathbf{B}+\mathbf{W}|} = \frac{18^2-(-13)^2}{54\times 102-35^2} \approx 0.0362$$

Do $p=2$ và $g=3\geq 2$ nên ta xét thống kê

$$\left(\frac{\sum n_\ell - g - 1}{g-1}\right)\left(\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right) \sim F_{2(g-1),2(\sum n_\ell - g-1)}$$

Khi đó ta có giá trị thống kê là

$$\left(\frac{\sum\limits_{\ell=1}^{3} n_\ell - g - 1}{g-1}\right)\left(\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right) = \left(\frac{12-3-1}{3-1}\right)\frac{1-\sqrt{0.0362}}{\sqrt{0.0362}} \approx 17.0235$$

Mặt khác, với mức ý nghĩa $\alpha = 0.01$, ta có

$$F_{2(g-1),2(\sum n_\ell - g-1)}(\alpha) = F_{4,16}(0.01) \approx 4.7726$$

Vì $17.0235 > 4.7726$ nên ta bác bỏ giả thuyết H_0 với mức ý nghĩa $\alpha = 0.01$. Do đó ta kết luận rằng có sự chênh lệch giữa các treatment.

Một cách tiếp cận khác, ta cần kiểm định giả thuyết

$$H_0: \tau_1 = \tau_2 = \tau_3 = 0$$

Xét thống kê

$$-\left(n-1-\frac{p+g}{2}\right)\ln\Lambda^* \sim \chi^2_{p(g-1)}$$

Ta có giá trị thống kê

$$-\left(n-1-\frac{p+g}{2}\right)\ln\Lambda^* = -\left(12-1-\frac{2+3}{2}\right)\ln(0.0362) \approx 28.2089$$

Mặt khác, với mức ý nghĩa $\alpha = 0.01$, ta có

$$\chi^2_{p(g-1)}(\alpha) = \chi^2_4(0.01) \approx 13.2767$$

Vì $28.2089 > 13.2767$ nên ta bác bỏ giả thuyết H_0 với mức ý nghĩa $\alpha = 0.01$. Như vậy, tương tự với cách tiếp cận kiểm định giả thuyết H_0 bằng thống kê Wilk's Lambda, ta có thể kết luận rằng tồn tại sự chênh lệch giữa các treatment.

Bài tập 6.9. Sử dụng ma trận nghịch $\boldsymbol{C}$, kiểm tra $\boldsymbol{d_j} = \boldsymbol{C}\boldsymbol{x_j}, \overline{\boldsymbol{d}} = \boldsymbol{C}\overline{\boldsymbol{x}}, \boldsymbol{S_d} = \boldsymbol{CSC'}$

Lời giải

Ta có:

$$\boldsymbol{C}\boldsymbol{x_j} = \begin{bmatrix} 1 & 0 & \cdots & 0 & -1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & -1 \end{bmatrix} \begin{bmatrix} x_{1j1} \\ \vdots \\ x_{1jp} \\ x_{2j1} \\ \vdots \\ x_{2jp} \end{bmatrix} = \begin{bmatrix} x_{1j1} - x_{2j1} \\ \vdots \\ x_{1jp} - x_{2jp} \end{bmatrix} = \begin{bmatrix} d_{j1} \\ \vdots \\ d_{jp} \end{bmatrix} = \boldsymbol{d_j}$$

Khi đó,

$$\overline{\boldsymbol{d}} = \frac{1}{n}\sum_{j=1}^{n} \boldsymbol{d_j} = \frac{1}{n}\sum_{j=1}^{n} \boldsymbol{C}\boldsymbol{x_j} = \boldsymbol{C}(\sum_{j=1}^{n} \boldsymbol{x_j}) = \boldsymbol{C}\overline{\boldsymbol{x}}$$

Vậy nên,

$$\begin{aligned} \boldsymbol{S_d} &= \frac{1}{n-1}\sum_{j=1}^{n}(\boldsymbol{d_j} - \overline{\boldsymbol{d}})(\boldsymbol{d_j} - \overline{\boldsymbol{d}})' \\ &= \frac{1}{n-1}\sum_{j=1}^{n}(\boldsymbol{C}(\boldsymbol{x_j} - \overline{\boldsymbol{x}}))(\boldsymbol{C}(\boldsymbol{x_j} - \overline{\boldsymbol{x}}))' \\ &= \frac{1}{n-1}\sum_{j=1}^{n}\boldsymbol{C}(\boldsymbol{x_j} - \overline{\boldsymbol{x}})(\boldsymbol{x_j} - \overline{\boldsymbol{x}})'\boldsymbol{C}' \\ &= \boldsymbol{C}(\boldsymbol{\frac{1}{n-1}\sum_{j=1}^{n}}(\boldsymbol{x_j} - \overline{\boldsymbol{x}})(\boldsymbol{x_j} - \overline{\boldsymbol{x}})')\boldsymbol{C}' \\ &= \boldsymbol{CSC'} \end{aligned}$$