Thống kê nhiều chiều: Bài tập nhóm tuần 07

Lê Hoàng Đức - 18110075
Nguyễn Phú Thành - 18110014

11/5/2021

Bài tập	Thành viên
6.5, 6.8, 6.19, 6.23, 6.28, 6.33	Nguyễn Phú Thành
6.6, 6.9, 6.22, 6.24, 6.30	Lê Hoàng Đức

Bài 6.5 Một nhà nghiên cứu xét 3 chỉ tiêu được dùng để đánh giá mức nguy hiểm của một cơn đau tim. Các giá trị của các chỉ tiêu này được đo trên $n = 40$ bệnh nhân và được các giá trị thống kê sau:

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

(a) Bằng (6-16), Kiểm định giả thuyết các trung bình của các chỉ tiêu là bằng nhau với mức ý nghĩa $\alpha = 0.05$

Bài giải:

Cho $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n$ là mẫu ngẫu nhiên chỉ số đo 3 chỉ tiêu trên n bệnh nhân, được lấy từ tổng thể có phân phối $\mathcal{N}_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ và ma trận:

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

- Giả thuyết thống kê:

$$\begin{aligned} H_0 &: \ \boldsymbol{C\mu} = \boldsymbol{0} \\ H_1 &: \ \boldsymbol{C\mu} \neq \boldsymbol{0} \end{aligned}$$

- Mức ý nghĩa $\alpha = 0.05$
- Khi H_0 đúng ta có thống kê:

$$T^2 = n(\boldsymbol{C}\bar{\boldsymbol{X}})^T(\boldsymbol{CSC}^T)^{-1}(\boldsymbol{C}\bar{\boldsymbol{X}}) \sim \frac{2(n-1)}{n-2} F_{2,n-2}$$

Với mẫu thực nghiệm, ta có:

- $n = 40, \alpha = 0.05$
- $F_{2,n-2}(1-\alpha) = F_{2,38}(0.95) \approx 3.245$
  Khi đó: $\dfrac{2(n-1)}{n-2} F_{2,n-2}(1-\alpha) \approx 6.661$

- $\bar{\boldsymbol{x}} = \begin{bmatrix} 46.1 \\ 57.3 \\ 50.4 \end{bmatrix}$
- $\boldsymbol{S} = \begin{bmatrix} 101.3 & 63.0 & 71.0 \\ 63.0 & 80.2 & 55.6 \\ 71.0 & 55.6 & 97.4 \end{bmatrix}$
- $t^2 = n(\boldsymbol{C}\bar{\boldsymbol{x}})^T(\boldsymbol{CSC}^T)^{-1}(\boldsymbol{C}\bar{\boldsymbol{x}}) \approx 90.494$

Vì $t^2 > \dfrac{2(n-1)}{n-2}F_{2,n-2}(1-\alpha)$ nên ta bác bỏ H_0 với mức ý nghĩa $\alpha = 0.05$
Như vậy, với mức ý nghĩa $\alpha = 0.05$, ta kết luận trung bình của 3 chỉ tiêu đánh giá mức nguy hiểm của cơn đau tim là khác nhau
(b) Đánh giá sự sai khác giữa các trung bình bằng cách sử dụng khoảng tin cậy đồng thời 95% ở (6-18)
Bài giải:
Theo (6 - 18), với vectơ $\boldsymbol{c}$ bất kì, ta có khoảng tin cậy đồng thời $(1-\alpha)$ cho $\boldsymbol{c}^T\boldsymbol{\mu}$ là:

$$\boldsymbol{c}^T\bar{\boldsymbol{x}} \pm \sqrt{\frac{2(n-1)}{n-2}F_{2,n-2}(1-\alpha)}\sqrt{\frac{\boldsymbol{c}^T\boldsymbol{S}\boldsymbol{c}}{n}}$$

Với mẫu thực nghiệm, ta có:

- $n = 40, \alpha = 0.05$
- $F_{2,n-2}(1-\alpha) = F_{2,38}(0.95) \approx 3.245$
  Khi đó $\sqrt{\dfrac{2(n-1)}{n-2}F_{2,n-2}(1-\alpha)} \approx 2.581$
- $\bar{\boldsymbol{x}} = \begin{bmatrix} 46.1 \\ 57.3 \\ 50.4 \end{bmatrix}$
- $\boldsymbol{S} = \begin{bmatrix} 101.3 & 63.0 & 71.0 \\ 63.0 & 80.2 & 55.6 \\ 71.0 & 55.6 & 97.4 \end{bmatrix}$

Cho $\boldsymbol{c}$ lần lượt là $[1, 0, -1]^T, [0, 1, -1]^T$ ta có các khoảng tin cậy đồng thời 95% của:

$$\begin{aligned} \mu_1 - \mu_3 : &\quad (-7.373, -1.227) \\ \mu_2 - \mu_3 : &\quad (3.575, 10.225) \end{aligned}$$

Nhận xét:

1. Trung bình của chỉ tiêu 1 thấp hơn trung bình của chỉ tiêu 3
2. Trung bình của chỉ tiêu 2 lớn hơn trung bình của chỉ tiêu 3
3. Trong hai khoảng tin cậy của $\mu_1 - \mu_3$ và $\mu_2 - \mu_3$ đều không chứa 0

Bài 6.6 Sử dụng dữ liệu cho "treatments" 2 và 3 ở bài 6.8:
(a). Tính $\boldsymbol{S}_{pooled}$.
Bài giải Ma trận $\boldsymbol{S}_{pooled}$ được định nghĩa:

$$\begin{aligned}\boldsymbol{S}_{pooled} &= \frac{\left(\sum_{j=1}^{n_2}\boldsymbol{x}_{2j}-\bar{\boldsymbol{x}}_2\right)\left(\sum_{j=1}^{n_2}\boldsymbol{x}_{2j}-\bar{\boldsymbol{x}}_2\right)^T+\left(\sum_{j=1}^{n_3}\boldsymbol{x}_{3j}-\bar{\boldsymbol{x}}_3\right)\left(\sum_{j=1}^{n_3}\boldsymbol{x}_{3j}-\bar{\boldsymbol{x}}_3\right)^T}{n_2+n_3-2}\\ &= \frac{n_2-1}{n_2+n_3-2}\boldsymbol{S}_2+\frac{n_3-1}{n_2+n_3-2}\boldsymbol{S}_3\end{aligned}$$

Ta có: $\boldsymbol{S}_2=\begin{bmatrix}1 & -1.5\\ -1.5 & 3\end{bmatrix}$, $\boldsymbol{S}_3=\begin{bmatrix}2 & -1.33\\ -1.33 & 1.33\end{bmatrix}$, từ đó ta có $\boldsymbol{S}_{pooled}=\frac{2}{5}\boldsymbol{S}_2+\frac{3}{5}\boldsymbol{S}_3=\begin{bmatrix}1.6 & -1.4\\ -1.4 & 2\end{bmatrix}$
(b) Kiểm định giả thuyết $H_0:\boldsymbol{\mu}_2-\boldsymbol{\mu}_3=\boldsymbol{0}$ với $\alpha=.01$.
Bài giải Ta có:

$$\bar{\boldsymbol{x}_2}-\bar{\boldsymbol{x}_3}=\begin{bmatrix}-1\\ 2\end{bmatrix}$$

Thống kê T^2 được cho bởi:

$$\begin{aligned}T^2 &= (\bar{\boldsymbol{x}_2}-\bar{\boldsymbol{x}_3})^T\left[\left(\frac{1}{n_2}+\frac{1}{n_3}\right)\boldsymbol{S}_{pooled}\right]^{-1}(\bar{\boldsymbol{x}_2}-\bar{\boldsymbol{x}_3})\\ &= \begin{bmatrix}-1\\ 2\end{bmatrix}^T\left[(1/3+1/4)\begin{bmatrix}1.6 & -1.4\\ -1.4 & 2\end{bmatrix}\right]^{-1}\begin{bmatrix}-1\\ 2\end{bmatrix}\\ &= 3.8709\end{aligned}$$

và

$$\begin{aligned}c^2 &= \frac{(n_2+n_3-2)p}{(n_2+n_3-p-1)}F_{p,n_2+n_3-p-1}(\alpha)\\ &= \frac{10}{4}F_{2,4}(.01)\\ &= \frac{5}{2}\times 18.00\\ &= 45.0\end{aligned}$$

Vì $T^2=3.8709<c^2=45.0$, nên ta không thể bác bỏ giả thuyết H_0 với mức ý nghĩa $\alpha=.01$. Như vậy kết luận của kiểm định là hai trung bình $\boldsymbol{\mu}_2$ và $\boldsymbol{\mu}_3$ là như nhau.

(c) Xây dựng các khoảng tin cậy đồng thời 99% cho các giá trị $\mu_{2i}-\mu_{3i}, i=1,2$
. **Bài giải** Từ kết quả 6.3: Cho $c^2=[(n_1+n_2-2)p/(n_1+n_2-p-1)]F_{p,n_1+n_2-p-1}(\alpha)$. Với mức ý nghĩa $1-\alpha$.

$$\boldsymbol{a}^T(\bar{\boldsymbol{X}}_1-\bar{\boldsymbol{X}}_2)\pm c\sqrt{\boldsymbol{a}^T\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\boldsymbol{S}_{pooled}\boldsymbol{a}}$$

sẽ chứa $\boldsymbol{a}^T(\mu_1-\mu_2)$ với mọi $\boldsymbol{a}$. Từ đó, $\mu_{1i}-\mu_{2i}$ có khoảng tin cậy:

$$(\bar{X}_{1i}-\bar{X}_{2i})\pm c\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)s_{ii,pooled}}, i=1,2,...,p$$

Ta có,

$$\begin{aligned} c^2 &= \frac{(n_2+n_3-2)p}{(n_2+n_3-p-1)}F_{p,n_2+n_3-p-1}(\alpha) \\ &= \frac{10}{4}F_{2,4}(.01) \\ &= \frac{5}{2}\times 18.00 \\ &= 45.0 \end{aligned}$$

Vậy, khoảng tin cậy đồng thời 99% cho từng thành phần của các trung bình mẫu khác nhau là:

$$\begin{aligned} \mu_{21}-\mu_{31} &: (2.0-3.0) \pm \sqrt{45.0}\sqrt{(1/3+1/4)\times 1.6} \\ &\quad -7.48 \le \mu_{21}-\mu_{31} \le 5.48 \\ \mu_{22}-\mu_{32} &: (4.0-2.0) \pm \sqrt{45.0}\sqrt{(1/3+1/4)\times 2.0} \\ &\quad -5.24 \le \mu_{22}-\mu_{32} \le 9.24 \end{aligned}$$

Bài 6.8 Các quan sát của 2 "responses" với 3 "treatments" được thu thập và cho bởi:

$$\begin{aligned} &\text{Treatment 1:} && \begin{bmatrix}6\\7\end{bmatrix}, \begin{bmatrix}5\\9\end{bmatrix}, \begin{bmatrix}8\\6\end{bmatrix}, \begin{bmatrix}4\\9\end{bmatrix}, \begin{bmatrix}7\\9\end{bmatrix} \\ &\text{Treatment 2:} && \begin{bmatrix}3\\3\end{bmatrix}, \begin{bmatrix}1\\6\end{bmatrix}, \begin{bmatrix}2\\3\end{bmatrix} \\ &\text{Treatment 3:} && \begin{bmatrix}2\\3\end{bmatrix}, \begin{bmatrix}5\\1\end{bmatrix}, \begin{bmatrix}3\\1\end{bmatrix}, \begin{bmatrix}2\\3\end{bmatrix} \end{aligned}$$

(a) Tách từng quan sát thành trung bình, "treatment" và phần dư như (6-39), Xây dựng các mảng tương ứng cho từng biến (như ví dụ 6.9)

Bài giải:

Với biến thứ nhất ta có:

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

$$\text{Quan sát} = (\text{Trung bình tổng thể}) + (\text{Ảnh hương của treatment}) + (\text{Phần dư sai số})$$

Tương tự, với biến thứ hai ta có:

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

$$\text{Quan sát} = (\text{Trung bình tổng thể}) + (\text{Ảnh hương của treatment}) + (\text{Phần dư sai số})$$

(b) Sử dụng kết quả đã tính ở câu (a), xây dựng bảng MANOVA một yếu tố

Ta có:

$$\begin{aligned}
\boldsymbol{B} &= \sum_{l=1}^{3} n_l(\bar{\boldsymbol{x}}_l - \bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_l - \bar{\boldsymbol{x}})^T \\
&= 5\begin{bmatrix}2\\3\end{bmatrix}\begin{bmatrix}2 & 3\end{bmatrix} + 3\begin{bmatrix}-2\\-1\end{bmatrix}\begin{bmatrix}-2 & -1\end{bmatrix} + 4\begin{bmatrix}-1\\-3\end{bmatrix}\begin{bmatrix}-1 & -3\end{bmatrix} \\
&= \begin{bmatrix}36 & 48\\48 & 84\end{bmatrix} \\
\boldsymbol{W} &= \sum_{l=1}^{3}\sum_{j=1}^{n_l}(\boldsymbol{x}_{lj} - \bar{\boldsymbol{x}}_l)(\boldsymbol{x}_{lj} - \bar{\boldsymbol{x}}_l)^T \\
&= \begin{bmatrix}0\\-1\end{bmatrix}\begin{bmatrix}0 & -1\end{bmatrix} + \begin{bmatrix}-1\\1\end{bmatrix}\begin{bmatrix}-1 & 1\end{bmatrix} + \begin{bmatrix}2\\-2\end{bmatrix}\begin{bmatrix}2 & -2\end{bmatrix} + \cdots + \begin{bmatrix}-1\\1\end{bmatrix}\begin{bmatrix}-1 & 1\end{bmatrix} \\
&= \begin{bmatrix}18 & -13\\-13 & 18\end{bmatrix}
\end{aligned}$$

Khi đó:

$$\boldsymbol{B} + \boldsymbol{W} = \begin{bmatrix}54 & 35\\35 & 102\end{bmatrix}$$

Như vậy ta có bảng MANOVA:

Source of variation	SSP	df
Treatment	$\boldsymbol{B} = \begin{bmatrix}36 & 48\\48 & 84\end{bmatrix}$	2
Residual	$\boldsymbol{W} = \begin{bmatrix}18 & -13\\-13 & 18\end{bmatrix}$	9
Total	$\boldsymbol{B} + \boldsymbol{W} = \begin{bmatrix}54 & 35\\35 & 102\end{bmatrix}$	11

(c) Tính thống kê "Wilk's lambda" Λ^* và sử dụng bảng 6.3 để kiểm định giả thuyết H_0 : "Response" không bị ảnh hưởng bởi "treatment" với mức ý nghĩa $\alpha = 0.01$. Lặp lại phép kiểm định này với thống kê xấp xỉ có hiệu chỉnh của Bartlett. So sánh hai kết luận

Bài giải:

- Giả thuyết thống kê:

$$\begin{aligned}
H_0 &: \text{ Các "response" không bị ảnh hưởng bởi các "treatment"} \\
H_1 &: \text{ Các "response" bị ảnh hưởng bởi các "treatment"}
\end{aligned}$$

- Mức ý nghĩa $\alpha = 0.01$
- $p = 2, g = 3, \sum_{l=1}^{g} n_l = 12$
- Khi H_0 đúng, ta có thống kê:

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

trong đó $\Lambda^* = \dfrac{|\boldsymbol{W}|}{|\boldsymbol{B}+\boldsymbol{W}|}$

Với mẫu thực nghiệm ta có:

- $F_{2(g-1),2(\sum n_l - g-1)}(1-\alpha) = F_{4,16}(0.99) \approx 4.773$
- $\boldsymbol{W} = \begin{bmatrix} 18 & -13 \\ -13 & 18 \end{bmatrix}$
- $\boldsymbol{B}+\boldsymbol{W} = \begin{bmatrix} 54 & 35 \\ 35 & 102 \end{bmatrix}$
- $\Lambda^* = \dfrac{|\boldsymbol{W}|}{|\boldsymbol{B}+\boldsymbol{W}|} \approx 0.036$
  Khi đó: $\left(\dfrac{\sum n_l - g - 1}{g-1}\right)\left(\dfrac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right) \approx 17.082$

Vì $17.082 > F_{4,16}(0.99)$ nên với mức ý nghĩa $\alpha = 0.01$ ta bác bỏ H_0, tức các "response" bị ảnh hưởng bởi các "treatment"
Tương tự, với thống kê xấp xỉ có hiệu chỉnh của Bartlett thì khi H_0 đúng, ta có thống kê:

$-\left(n-1-\dfrac{p+g}{2}\right)\ln\Lambda^*$ có xấp xỉ phân phối chi bình phương với bậc tự do $p(g-1)$

trong đó $n = \sum n_l$
Với mẫu thực nghiệm ta có:

- $\Lambda^* \approx 0.036$
  Khi đó: $-\left(n-1-\dfrac{p+g}{2}\right)\ln\Lambda^* \approx 28.256$
- $\chi^2_{p(g-1)}(\alpha) = \chi^2_4(0.01) \approx 13.277$

Vì $28.256 > \chi^2_4(0.01)$ nên với mức ý nghĩa $\alpha = 0.01$, ta bác bỏ giả thuyết H_0, tức các "response" bị ảnh hưởng bởi các "treatment"
Như vậy kết luận của hai phép kiểm định là như nhau

Bài 6.9 Sử dụng contrast matrix $\boldsymbol{C}$ ở (6-13), kiểm tra mối quan hệ $\boldsymbol{d_j} = \boldsymbol{C}\boldsymbol{x_j}$, $\bar{\boldsymbol{d}} = \boldsymbol{C}\bar{\boldsymbol{x}}$ và $\boldsymbol{S}_d = \boldsymbol{C}\boldsymbol{S}\boldsymbol{C}^T$.

Bài giải:
Cho $j \in \{1,2,...,n\}$ bất kỳ, khi đó:

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

$$= \begin{bmatrix} x_{1j1} - x_{2j1} \\ \vdots \\ x_{1jp} - x_{2jp} \end{bmatrix} = \begin{bmatrix} d_{j1} \\ \vdots \\ d_{jp} \end{bmatrix} = \boldsymbol{d}_j$$

$$\Rightarrow \boldsymbol{C}\boldsymbol{x}_j = \boldsymbol{d}_j \quad (\forall j \in \{1,2,...,n\})$$

Ta có: $\boldsymbol{C}\bar{\boldsymbol{x}} = \boldsymbol{C}\left(\frac{1}{n}\sum_{j=1}^{n}\boldsymbol{x}_j\right) = \frac{1}{n}\sum_{j=1}^{n}\boldsymbol{C}\boldsymbol{x}_j = \frac{1}{n}\sum_{j=1}^{n}\boldsymbol{d}_j = \bar{\boldsymbol{d}}$. Từ đó, với mọi $j \in \{1, 2, ..., n\}$:

$$\begin{aligned}\boldsymbol{d}_j - \bar{\boldsymbol{d}} &= \boldsymbol{C}\boldsymbol{x}_j - \boldsymbol{C}\bar{\boldsymbol{x}} \\ &= \boldsymbol{C}(\boldsymbol{x}_j - \bar{\boldsymbol{x}})\end{aligned}$$

Ta được,

$$\begin{aligned}\boldsymbol{S}_d &= \frac{1}{n-1}\sum_{j=1}^{n}(\boldsymbol{d}_j - \bar{\boldsymbol{d}})(\boldsymbol{d}_j - \bar{\boldsymbol{d}})^T \\ &= \frac{1}{n-1}\sum_{j=1}^{n}[\boldsymbol{C}(\boldsymbol{x}_j - \bar{\boldsymbol{x}})][\boldsymbol{C}(\boldsymbol{x}_j - \bar{\boldsymbol{x}})]^T \\ &= \boldsymbol{C}(\boldsymbol{x}_j - \bar{\boldsymbol{x}})(\boldsymbol{x}_j - \bar{\boldsymbol{x}})^T\boldsymbol{C}^T \\ &= \boldsymbol{C}\left(\frac{1}{n-1}\sum_{j=1}^{n}(\boldsymbol{x}_j - \bar{\boldsymbol{x}})(\boldsymbol{x}_j - \bar{\boldsymbol{x}})^T\right)\boldsymbol{C}^T \\ &= \boldsymbol{C}\boldsymbol{S}\boldsymbol{C}^T\end{aligned}$$

Các mối quan hệ được thành lập.

Bài 6.19 Trong giai đoạn đầu của việc nghiên cứu chi phí vận chuyển sữa từ doanh trại đến công ty, một cuộc khảo sát được thực hiện với các công ty vận chuyển sữa. Dữ liệu về chi phí bao gồm X_1 = chi phí nhiên liệu, X_2 = chi phí sửa chữa và X_3 = chi phí vốn, tất cả được tính trên mỗi mile, và cho trong bảng 6.10 với $n_1 = 36$ xe tải dùng gasoline và $n_2 = 23$ xe tải dùng diesel

(a) Kiểm định sự sai khác giữa các vectơ trung bình với mức ý nghĩa $\alpha = 0.01$

Bài giải:

Trước hết ta sẽ kiểm định xem phương sai của 2 mẫu có bằng nhau hay không

Gọi $\boldsymbol{X}_{11}, \boldsymbol{X}_{12}, \cdots, \boldsymbol{X}_{1n_1}$ là mẫu ngẫu nhiên chỉ chi phí của các xe sử dụng gasoline và $\boldsymbol{X}_{21}, \boldsymbol{X}_{22}, \cdots, \boldsymbol{X}_{2n_2}$ là mẫu ngẫu nhiên chỉ chi phí của các xe sử dụng diesel

Cả hai mẫu ngẫu nhiên lấy từ 2 tổng thể có vectơ trung bình lần lượt là $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ và ma trận hiệp phương sai lần lượt là $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$

- Giả thuyết thống kê:

$$\begin{aligned}H_0 :&\quad \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 \\ H_1 :&\quad \boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2\end{aligned}$$

- Mức ý nghĩa $\alpha = 0.01$

Khi H_0 đúng, phép kiểm định Box cho thống kê:

$$C = (1-u)\left\{\left[\sum_l (n_l - 1)\ln|\boldsymbol{S}_{\text{pooled}}|\right] - \sum_l [(n_l - 1)\ln|\boldsymbol{S}_l|]\right\}$$

xấp xỉ phân phối Chi bình phương với bậc tự do:

$$v = \frac{1}{2}p(p+1)(g-1)$$

và

$$u = \left[\sum_l \frac{1}{(n_l - 1)} - \frac{1}{\sum_l (n_l - 1)}\right] \frac{2p^2 + 3p - 1}{6(p+1)(g-1)}$$

trong đó p là số biến và g là số tổng thể
Với mẫu thực nghiệm ta có:

- Số biến $p = 3$, số tổng thể $g = 2$
- Số lượng tổng thể thứ nhất $n_1 = 36$
- Số lượng tổng thể thứ hai $n_2 = 23$
- $c \approx 30.544$
- Điểm tới hạn $\chi^2_{p(p+1)(g-1)/2}(\alpha) = \chi^2_6(0.01) \approx 16.812$

Vì $c > \chi^2_6(0.01)$ nên với mức ý nghĩa $\alpha = 0.01$ ta bác bỏ H_0. Như vậy với mức ý nghĩa 1% ta kết luận $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$
Ta tiến hành kiểm định sự sai khác về vectơ trung bình của 2 tổng thể với giả định $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$

- Giả thuyết thống kê:

$$\begin{aligned} H_0 &: \quad \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \\ H_1 &: \quad \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2 \end{aligned}$$

- Mức ý nghĩa $\alpha = 0.01$
- $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$ và $n_1 - p = 33$, $n_2 - p = 20$ đủ lớn

Theo kết quả 6.4, khi H_0 đúng, ta có thống kê:

$$T^2 = \left(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2\right)^T \left[\frac{1}{n_1}\boldsymbol{S}_1 + \frac{1}{n_2}\boldsymbol{S}_2\right]^{-1} \left(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2\right)$$

xấp xỉ phân phối Chi bình phương với bậc tự do p
Với mẫu thực nghiệm ta có:

- $n_1 = 36, n_2 = 23$
- $\bar{\boldsymbol{x}}_1 \approx \begin{bmatrix} 12.219 \\ 8.113 \\ 9.590 \end{bmatrix}$, $\bar{\boldsymbol{x}}_2 \approx \begin{bmatrix} 10.106 \\ 10.762 \\ 18.168 \end{bmatrix}$
- $\boldsymbol{S}_1 \approx \begin{bmatrix} 23.013 & 12.366 & 2.907 \\ 12.366 & 17.544 & 4.773 \\ 2.907 & 4.773 & 13.963 \end{bmatrix}$, $\boldsymbol{S}_2 \approx \begin{bmatrix} 4.362 & 0.760 & 2.362 \\ 0.760 & 25.851 & 7.686 \\ 2.362 & 7.686 & 46.654 \end{bmatrix}$
- $\chi^2_p(\alpha) = \chi^2_3(0.01) \approx 11.345$

- $t^2 = (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)^T \left[\frac{1}{n_1}\boldsymbol{S}_1 + \frac{1}{n_2}\boldsymbol{S}_2\right]^{-1} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) \approx 43.176$

Vì $t^2 > \chi_3^2(0.01)$ nên với mức ý nghĩa $\alpha = 0.01$ ta bác bỏ H_0
Như vậy với mức ý nghĩa $\alpha = 0.01$ ta kết luận chi phí trung bình ở hai loại xe tải gasoline và diesel là khác nhau
(b) Nếu giả thuyết các vectơ trung bình bằng nhau bị bác bỏ ở câu a, tìm tổ hợp tuyến tính của các phần tử của vectơ trung bình chịu trách nhiệm cho sự bác bỏ này
Bài giải:
Theo **Ví dụ 6.5,** tổ hợp tuyến tính của các phần tử của vectơ trung bình chịu trách nhiệm cho sự bác bỏ này có hệ số:

$$\hat{\boldsymbol{a}} \propto \left(\frac{1}{n_1}\boldsymbol{S}_1 + \frac{1}{n_2}\boldsymbol{S}_2\right)^{-1} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) = \begin{bmatrix} 4.044 \\ -1.560 \\ -3.556 \end{bmatrix}$$

(c) Xây dựng khoảng tin cậy 99% cho hiệu số giữa từng cặp các phần tử của vectơ trung bình
Bài giải:
Theo **Kết quả 6.4**, với mẫu thực nghiệm cho trước và vectơ $\boldsymbol{a}$ bất kì, ta có khoảng tin cậy đồng thời $(1-\alpha)\%$ cho $\boldsymbol{a}^T(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ là:

$$\boldsymbol{a}^T(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) \pm \sqrt{\chi_p^2(\alpha)}\sqrt{\boldsymbol{a}^T\left(\frac{1}{n_1}\boldsymbol{S}_1 + \frac{1}{n_2}\boldsymbol{S}_2\right)\boldsymbol{a}}$$

Với mẫu thực nghiệm ta có:

- $n_1 = 36, n_2 = 23$
- $\bar{\boldsymbol{x}}_1 \approx \begin{bmatrix} 12.219 \\ 8.113 \\ 9.590 \end{bmatrix}$, $\bar{\boldsymbol{x}}_2 \approx \begin{bmatrix} 10.106 \\ 10.762 \\ 18.168 \end{bmatrix}$
- $\boldsymbol{S}_1 \approx \begin{bmatrix} 23.013 & 12.366 & 2.907 \\ 12.366 & 17.544 & 4.773 \\ 2.907 & 4.773 & 13.963 \end{bmatrix}$, $\boldsymbol{S}_2 \approx \begin{bmatrix} 4.362 & 0.760 & 2.362 \\ 0.760 & 25.851 & 7.686 \\ 2.362 & 7.686 & 46.654 \end{bmatrix}$
- $\chi_p^2(\alpha) = \chi_3^2(0.01) \approx 11.345$

Lần lượt cho $\boldsymbol{a}^T = [1, 0, 0]$, $\boldsymbol{a}^T = [0, 1, 0]$ và $\boldsymbol{a}^T = [0, 0, 1]$ ta được:

Khoảng tin cậy đồng thời 99% cho $\mu_{11} - \mu_{21}$ là: $(-0.954, 5.180)$
Khoảng tin cậy đồng thời 99% cho $\mu_{12} - \mu_{22}$ là: $(-6.925, 1.626)$
Khoảng tin cậy đồng thời 99% cho $\mu_{13} - \mu_{23}$ là: $(-13.813, -3.342)$

Nhận xét: Trong 3 khoảng tin cậy chỉ có khoảng tin cậy 99% cho $\mu_{13} - \mu_{23}$ không chứa 0 và chặn dưới cũng như chặn trên đều âm nên ta có thể nói, với độ tin cậy 99%, $\mu_{13} < \mu_{23}$
(d) Nhận xét về các giả định trong các phép kiểm định đã thực hiện. Chú ý rằng, các quan sát thứ 9 và thứ 21 cho dữ liệu của xe tải sử dụng gasoline được cho là các giá trị ngoại lai. Thực hiện lại câu (a) sau khi xóa các giá trị ngoại lai này. Nhận xét kết quả cho được
Bài giải:
Một trong những giả định khi thực hiện phép kiểm định trong câu (a) là $n_1 - p = 33$ và $n_2 - p = 20$

là đủ lớn cho thống kê T^2 xấp xỉ phân phối Chi bình phương.
Ta có thể kiểm tra nhận định này bằng cách sử dụng phân phối xấp xỉ cho T^2 khi cỡ mẫu hai tổng thể nhỏ ở (6-28) và (6-29)
Bằng cách sử dụng (6-29) ta có điểm tới hạn là:

$$\frac{vp}{v-p+1}F_{p,v-p+1}(1-\alpha) = \frac{45\times 3}{45-3+1}F_{3,45-3+1}(0.99) \approx 13.414$$

và thống kê:

$$t^2 \approx 43.176$$

Như vậy nếu xem cỡ mẫu hai tổng thể là nhỏ thì với mức ý nghĩa $\alpha = 0.01$ ta vẫn bác bỏ giả thuyết $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ như ở câu (a)
Sau khi loại bỏ ngoại lai, thì các kết luận của các phép kiểm định thực hiện ở câu (a) vẫn không thay đổi
Bài 6.22 Các nhà khoa học thực hiện khảo sát đánh giá chức năng phổi thông qua những người không có bệnh lý bằng việc sử dụng máy chạy bộ cho đến khi không còn thể lực. Các kết quả của 4 đại lượng đo sự hấp thụ Oxy của 25 người nam và 25 người nữ được cho trong Table 6.12 (trang 348). Cụ thể là các đặc trưng sau:

$$\begin{aligned}
X_1 &= \text{Lượng ô xy tiêu thụ (L/min)}\\
X_2 &= \text{Lượng ô xy tiêu thụ (mL/kg/min)}\\
X_3 &= \text{Lượng ô xy tiêu thụ tối đa (L/min)}\\
X_4 &= \text{Lượng ô xy tiêu thụ tối đa (mL/kg/min)}
\end{aligned}$$

(a) Thông qua dữ liệu của 2 giới tính kiểm định tính bằng của trung bình hai nhóm. Với $\alpha = .05$. Nếu bác bỏ $\boldsymbol{H}_0 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \mathbf{0}$, tìm tổ hợp tuyến tính chịu trách nhiệm cho sự bác bỏ này.
Bài giải
Ta đi kiểm định tính bằng nhau của phương sai từ hai mẫu. Gọi $\boldsymbol{X}_{11}, \boldsymbol{X}_{12}, \ldots, \boldsymbol{X}_{1n_1}$ là mẫu ngẫu nhiên chỉ lượng oxy tiêu thụ từ nhóm nam và $\boldsymbol{X}_{21}, \boldsymbol{X}_{22}, \ldots, \boldsymbol{X}_{2n_2}$ là mẫu ngẫu nhiên chỉ lượng oxy tiêu thụ từ nhóm nữ. Cả hai mẫu có vector trung bình lần lượt là $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ và ma trận hiệp phương sai lần lượt là $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$.

- Gỉa thuyết thống kê:

$$\begin{aligned}
H_0 &: \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2\\
H_1 &: \boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2
\end{aligned}$$

- Mức ý nghĩa $\alpha = 0.05$

Khi H_0 đúng, phép kiểm định Box cho thống kê:

$$C = (1-u)\left\{\left[\sum_l (n_l - 1)\right]\ln|\boldsymbol{S}_{pooled}| - \sum_l \left[(n_l-1)\ln|\boldsymbol{S}_l|\right]\right\}$$

xấp xỉ phân phối Chi bình phương với bậc tự do:

$$v = \frac{1}{2}p(p+1)(g-1)$$

và

$$u = \left[\sum_l \frac{1}{(n_l - 1)} - \frac{1}{\sum_l (n_l - 1)}\right] \frac{2p^2 + 3p - 1}{6(p+1)(g-1)}$$

Từ mẫu thực nghiệm, ta có:

- p = 4, g = 2
- $n_1 = n_2 = 25$
- C $\approx$ 58.002
- Điểm tới hạn $\chi^2_{p(p+1)(g-1)/2}(\alpha) = \chi^2_{10}(0.05) \approx 18.307$

Ta có, C $> \chi^2_{10}(0.05)$ nên giải thuyết H_0 bị bác bỏ. Kết luận rằng hai mẫu có phương sai khác nhau.
Tiếp đến, đi kiểm định giả thuyết:

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$$
$$H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$$

Ta có trung bình mẫu:

$$\bar{\boldsymbol{x}_1} = \begin{bmatrix} 0.397 \\ 5.329 \\ 3.687 \\ 49.420 \end{bmatrix}$$
$$\bar{\boldsymbol{x}_2} = \begin{bmatrix} 0.313 \\ 5.178 \\ 2.315 \\ 38.154 \end{bmatrix}$$

Mức ý nghĩa: $\alpha = 0.05$. Theo (6-28), khi H_0 đúng, phép kiểm định T^2 cho thống kê:

$$T^2 = (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)^T \left[\frac{1}{n_1}\boldsymbol{S}_1 + \frac{1}{n_2}\boldsymbol{S}_2\right]^{-1} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) \tag{6-27}$$
$$\approx 96.373$$

và điểm tới hạn: $\dfrac{vp}{v-p+1}F_{p,v-p+1}(\alpha)$, trong đó:

$$v = \frac{p + p^2}{\sum_i \frac{1}{n_i}\left\{tr\left[\left(\frac{1}{n_i}\boldsymbol{S}_i\left(\frac{1}{n_1}\boldsymbol{S}_1 + \frac{1}{n_2}\boldsymbol{S}_2\right)^{-1}\right)^2\right] + \left(tr\left[\frac{1}{n_i}\boldsymbol{S}_i\left(\frac{1}{n_1}\boldsymbol{S}_1 + \frac{1}{n_2}\boldsymbol{S}_2\right)^{-1}\right]\right)^2\right\}}$$

Từ mẫu thực nghiệm:

- $p = 4$
- $n_1 = n_2 = 25$
- $v \approx 45.253$
- $\dfrac{vp}{v-p+1}F_{p,v-p+1}(\alpha) = \dfrac{45.253 \times 4}{45.253-4+1}F_{4,45.253-4+1}(0.05) \approx 11.11$

Như vậy, $T^2 > \dfrac{vp}{v-p+1}F_{p,v-p+1}(\alpha)$ nên ta chấp nhận giả thuyết H_0, với mức ý nghĩa $\alpha = 0.05$.
Tổ hợp tuyến tính của các phần tử của vectơ trung bình chịu trách nhiệm cho sự bác bỏ này có hệ số:

$$\hat{\boldsymbol{a}} \propto \left(\frac{1}{n_1}\boldsymbol{S}_1 + \frac{1}{n_2}\boldsymbol{S}_2\right)^{-1}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) = \begin{bmatrix} 1242.48 \\ -79.7 \\ -77.851 \\ 9.885 \end{bmatrix}$$

(b) Xây dựng khoảng tin cậy đồng thời 95% cho mỗi $\mu_{1i} - \mu_{2i}, i = 1, 2, 3, 4$. So sánh với khoảng tin cậy Bonferroni.
Bài giải
Từ kết quả 6.3: Cho $c^2 = [(n_1 + n_2 - 2)p/(n_1 + n_2 - p - 1)]F_{p,n_1+n_2-p-1}(\alpha)$. Với độ tin cậy $1 - \alpha$.

$$\boldsymbol{a}^T(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2) \pm c\sqrt{\boldsymbol{a}^T\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\boldsymbol{S}_{pooled}\boldsymbol{a}}$$

sẽ chứa $\boldsymbol{a}^T(\mu_1 - \mu_2)$ với mọi $\boldsymbol{a}$. Từ đó, $\mu_{1i} - \mu_{2i}$ có khoảng tin cậy:

$$(\bar{X}_{1i} - \bar{X}_{2i}) \pm c\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)s_{ii,pooled}}, i = 1, 2, ..., p$$

Ta có,

$$\begin{aligned} c^2 &= [(n_1 + n_2 - 2)p/(n_1 + n_2 - p - 1)]F_{p,n_1+n_2-p-1}(\alpha) \\ &\approx 11.0 \end{aligned}$$

Vậy, khoảng tin cậy đồng thời 95% cho từng thành phần của các trung bình mẫu khác nhau là:

$$\begin{aligned} \mu_{11} - \mu_{21} &: (0.397 - 0.313) \pm \sqrt{11.0}\sqrt{(1/25 + 1/25) \times 8.42 \times 10^{-3}} \\ &\quad - 2.5 \times 10^{-}3 \leq \mu_{11} - \mu_{21} \leq 1.69 \times 10^{-1} \\ \mu_{12} - \mu_{22} &: (5.329 - 5.178) \pm \sqrt{11.0}\sqrt{(1/25 + 1/25) \times 1.96} \\ &\quad [-1.163 \leq \mu_{12} - \mu_{22} \leq 1.465 \\ \mu_{13} - \mu_{23} &: (3.687 - 2.315) \pm \sqrt{11.0}\sqrt{(1/25 + 1/25) \times 0.288} \\ &\quad 0.87 \leq \mu_{13} - \mu_{23} \leq 1.875 \\ \mu_{14} - \mu_{24} &: (49.420 - 38.154) \pm \sqrt{11.0}\sqrt{(1/25 + 1/25) \times 3.925} \\ &\quad 5.38 \leq \mu_{14} - \mu_{24} \leq 17.14 \end{aligned}$$

So sánh với khoảng tin cậy đồng thời Bonferroni:

$$\begin{aligned}
\mu_{11}-\mu_{21} &: (0.397-0.313) \pm t_{24}(0.05/2\times 4)\sqrt{s_{11,pooled}} \\
0.033 &\le \mu_{11}-\mu_{21} \le 0.134 \\
\mu_{12}-\mu_{22} &: (5.329-5.178) \pm t_{24}(0.05/2\times 4)\sqrt{s_{22,pooled}} \\
-0.621 &\le \mu_{12}-\mu_{22} \le 0.922 \\
\mu_{13}-\mu_{23} &: (3.687-2.315) \pm t_{24}(0.05/2\times 4)\sqrt{s_{33,pooled}} \\
1.076 &\le \mu_{13}-\mu_{23} \le 1.668 \\
\mu_{14}-\mu_{24} &: (49.420-38.154) \pm t_{24}(0.05/2\times 4)\sqrt{s_{44,pooled}} \\
7.812 &\le \mu_{14}-\mu_{24} \le 14.719
\end{aligned}$$

(c) Dữ liệu trên được thu thập từ các sinh viên tình nguyện, vì vậy không thể đại diện cho mẫu ngẫu nhiên. Nhận xét hàm ý của thông tin này.

Bài giải

Nghiên cứu trên dựa trên các số liệu thu lập từ các sinh viên tình nguyện, do độ tuổi còn trẻ nên các chỉ số thống kê từ dữ liệu này không thể đại diện cho mẫu ngẫu nhiên. Nên, nếu trường hợp là mẫu ngẫu nhiên với các độ tuổi thu thập dữ liệu khác nhau, các số liệu thống kê trên sẽ có thay đổi và các khoảng tin cậy (miền tin cậy) đồng thời sẽ cho kết quả đáng tin hơn.

Bài 6.23 Xây dựng bảng MANOVA một nhân tố với dữ liệu đo chiều rộng đài hoa và chiều rộng cánh hoa của loài hoa diên vĩ ở bảng 11.5. Xây dựng khoảng tin cậy đồng thời 95% cho sự sai khác trung bình của 2 "response" cho từng cặp của tổng thể. Nhận định giả sử $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \mathbf{\Sigma}_3$

Bài giải:

• Bộ dữ liệu có $g = 3$ tổng thể và có $n_1 = n_2 = n_3 = 50$ quan sát trong từng tổng thể

• Gọi $\boldsymbol{\mu}_{k1}$ là chiều rộng trung bình của đài hoa của tổng thể k $(k = 1, 2, 3)$, $\boldsymbol{\mu}_{k2}$ là chiều rộng trung bình của cánh hoa của tổng thể k $(k = 1, 2, 3)$

• **Xây dựng bảng MANOVA**

Từ bộ dữ liệu, ta xây dựng được bảng MANOVA sau:

Source of variation	SSP	df
Treatment	$\boldsymbol{B} \approx \begin{bmatrix} 11.345 & -22.933 \\ -22.933 & 80.413 \end{bmatrix}$	2
Residual	$\boldsymbol{W} \approx \begin{bmatrix} 16.962 & 4.808 \\ 4.808 & 6.157 \end{bmatrix}$	147
Total	$\boldsymbol{B}+\boldsymbol{W} \approx \begin{bmatrix} 28.307 & -18.124 \\ -18.124 & 86.57 \end{bmatrix}$	149

• **Xây dựng khoảng tin cậy đồng thời** 95% **cho sự sai khác trung bình của 2 "response" cho từng cặp của tổng thể**

Theo **Kết quả 6.5** ta có khoảng tin cậy đồng thời $(1-\alpha)$ cho $\mu_{ki}-\mu_{li}$ là:

$$\bar{x}_{ki}-\bar{x}_{li} \pm t_{n-g}\left(\frac{\alpha}{pg(g-1)}\right)\sqrt{\frac{w_{ii}}{n-g}\left(\frac{1}{n_k}+\frac{1}{n_l}\right)}$$

trong đó $n = \sum_{k=1}^{g} n_k$ và w_{ii} là phần tử thứ i trên đường chéo chính của ma trận $\boldsymbol{W}$

Với $\alpha = 0.05, n = 150, g = 3, p = 2$ thì: $t_{n-g}\left(\frac{\alpha}{pg(g-1)}\right) = t_{147}\left(\frac{1}{240}\right) \approx 2.674$

Khi đó, khoảng tin cậy đồng thời 95% cho độ chênh lệch chiều rộng đài hoa giữa các tổng thể là:

$$\begin{aligned}&\text{Giữa tổng thể 1 và 2 là:} \quad (0.476, 0.840)\\ &\text{Giữa tổng thể 2 và 3 là:} \quad (-0.386, -0.022)\\ &\text{Giữa tổng thể 1 và 3 là:} \quad (0.272, 0.636)\end{aligned}$$

Tương tự, khoảng tin cậy đồng thời 95% cho độ chênh lệch chiều rộng cánh hoa giữa các tổng thể là:

$$\begin{aligned}&\text{Giữa tổng thể 1 và 2 là:} \quad (-1.189, -0.971)\\ &\text{Giữa tổng thể 2 và 3 là:} \quad (-0.809, -0.591)\\ &\text{Giữa tổng thể 1 và 3 là:} \quad (-1.889, -1.671)\end{aligned}$$

- Kiểm định $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3$
  - Giả thuyết thống kê:

$$\begin{aligned}H_0: &\quad \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3\\ H_1: &\quad \exists i, j \in \{1, 2, 3\}, i \neq j \text{ sao cho } \boldsymbol{\Sigma}_i \neq \boldsymbol{\Sigma}_j\end{aligned}$$

  - Mức ý nghĩa $\alpha = 0.05$

Khi H_0 đúng, phép kiểm định Box cho thống kê:

$$C = (1-u)\left\{\left[\sum_l (n_l - 1)\ln|\boldsymbol{S}_{\text{pooled}}|\right] - \sum_l [(n_l - 1)\ln|\boldsymbol{S}_l|]\right\}$$

xấp xỉ phân phối Chi bình phương với bậc tự do:

$$v = \frac{1}{2}p(p+1)(g-1)$$

và

$$u = \left[\sum_l \frac{1}{(n_l - 1)} - \frac{1}{\sum_l (n_l - 1)}\right]\frac{2p^2 + 3p - 1}{6(p+1)(g-1)}$$

trong đó p là số biến và g là số tổng thể
Với mẫu thực nghiệm ta có:

- $p = 2, g = 3, n_1 = n_2 = n_3 = 50$. Khi đó $v = \frac{1}{2}p(p+1)(g-1) = 6$ và $u \approx 0.0197$
- $c \approx 51.794$
- $\chi^2_v(\alpha) = \chi^2_6(0.01) \approx 12.592$

Vì $c > \chi^2_6(0.01)$ nên với mức ý nghĩa $\alpha = 0.05$, ta bác bỏ H_0. Như vậy, với mức ý nghĩa $\alpha = 0.05$ ta kết luận ma trận phương sai của 3 tổng thể không bằng nhau

Bài 6.24 Các nhà nghiên cứu đã đề cập đến sự thay đổi của kích cỡ hộp sọ người qua thời gian là một bằng chứng của sự giao phối giữa người dân bản địa và dân nhập cư. 4 đặc trưng được đo thông qua các hộp sọ của người Ai Cập với 3 thời kỳ khác nhau: thời kỳ 1 là 4000 năm (TCN), thời kỳ 2 là 3300 năm (TCN), thời kỳ 3 là 1850 năm (TCN). Dữ liệu được cho ở bảng 6.13. Các đặc trưng như sau:

X_1 = maximum breadth of skull (mm)
X_2 = basibregmatic height of skull (mm)
X_3 = basialveolar length of skull (mm)
X_4 = nasal height of skull (mm)

Xây dựng bảng MANOVA một chiều cho dữ liệu trên. Với $\alpha = 0.05$. Xây dựng khoảng tin cậy đồng thời 95% để xác định sự sai khác giữa các thành phần trung bình của 3 nhóm mẫu qua 3 thời kỳ. Các giả thuyết MANOVA có phù hợp với dữ liệu trên ? Giải thích.

Bài giải

• Xây dựng bảng MANOVA

Từ bộ dữ liệu, ta xây dựng được bảng MANOVA sau:

Source of variation	SSP	df
Treatment	$\boldsymbol{B} \approx \begin{bmatrix} 150.2 & 20.3 & -161.83 & 5.03 \\ 20.3 & 20.6 & -38.73 & 6.43 \\ -161.83 & -38.73 & 190.28 & -10.85 \\ 5.03 & 6.43 & -10.85 & 2.02 \end{bmatrix}$	2
Residual	$\boldsymbol{W} \approx \begin{bmatrix} 1.78 \times 10^3 & 1.725 \times 10^2 & 1.289 \times 10^2 & 2.89 \times 10^2 \\ 1.725 \times 10^2 & 1.925 \times 10^3 & 1.78 \times 10^2 & 1.72 \times 10^2 \\ 1.289 \times 10^2 & 1.78 \times 10^2 & 2.15 \times 10^3 & -1.7 \\ 2.89 \times 10^2 & 1.72 \times 10^2 & -1.7 & 8.4 \times 10^2 \end{bmatrix}$	87
Total	$\boldsymbol{B} + \boldsymbol{W} \approx \begin{bmatrix} 1935.6 & 192.8 & -32.86 & 294.6 \\ 192.8 & 1944.9 & 140.06 & 178.3 \\ -32.86 & 140.06 & 2343.2 & -12.55 \\ 294.66 & 178.33 & -12.55 & 842.22 \end{bmatrix}$	89

• Xây dựng khoảng tin cậy đồng thời 95% **cho sự sai khác trung bình cho từng cặp của tổng thể**

Theo **Kết quả 6.5** ta có khoảng tin cậy đồng thời $(1 - \alpha)$ cho $\mu_{ki} - \mu_{li}$ là:

$$\bar{x}_{ki} - \bar{x}_{li} \pm t_{n-g}\left(\frac{\alpha}{pg(g-1)}\right)\sqrt{\frac{w_{ii}}{n-g}\left(\frac{1}{n_k} + \frac{1}{n_l}\right)}$$

trong đó $n = \sum_{k=1}^{g} n_k$ và w_{ii} là phần tử thứ i trên đường chéo chính của ma trận $\boldsymbol{W}$

Với $\alpha = 0.05, n = 90, g = 3, p = 4$ thì: $t_{n-g}\left(\frac{\alpha}{pg(g-1)}\right) = t_{87}\left(\frac{1}{480}\right) \approx 2.943$

Khi đó, khoảng tin cậy đồng thời 95% cho độ chênh lệch giữa trung bình của đặc trưng X_1 giữa các tổng thể là:

Giữa tổng thể 1 và 2 là: $(-4.44, 2.44)$
Giữa tổng thể 2 và 3 là: $(-5.54, -1.34)$
Giữa tổng thể 1 và 3 là: $(-6.54, 0.34)$

Tương tự, khoảng tin cậy đồng thời 95% cho độ chênh lệch trung bình của đặc trưng X_2 giữa các tổng thể là:

Giữa tổng thể 1 và 2 là: $(-2.67, 4.47)$
Giữa tổng thể 2 và 3 là: $(-4.67, 2.47)$
Giữa tổng thể 1 và 3 là: $(-3.77, 3.37)$

Tương tự, khoảng tin cậy đồng thời 95% cho độ chênh lệch trung bình của đặc trưng X_3 giữa các tổng thể là:

$$\begin{aligned}&\text{Giữa tổng thể 1 và 2 là:} \quad (-3.68, 3.88)\\ &\text{Giữa tổng thể 2 và 3 là:} \quad (-0.74, 6.81)\\ &\text{Giữa tổng thể 1 và 3 là:} \quad (-0.64, 6.91)\end{aligned}$$

Tương tự, khoảng tin cậy đồng thời 95% cho độ chênh lệch trung bình của đặc trưng X_4 giữa các tổng thể là:

$$\begin{aligned}&\text{Giữa tổng thể 1 và 2 là:} \quad (-2.06, 2.66)\\ &\text{Giữa tổng thể 2 và 3 là:} \quad (-2.69, 2.02)\\ &\text{Giữa tổng thể 1 và 3 là:} \quad (-2.39, 2.32)\end{aligned}$$

Bài 6.28 Kiểm định vectơ trung bình của các đặc trưng của 2 giống ruồi cho ở bảng 6.15 có bằng nhau hay không với mức ý nghĩa $\alpha = 0.05$. Nếu giả thuyết này bị bác bỏ, thành phần (đặc trưng) nào khiến cho giả thuyết này bị bác bỏ ?

Bài giải:

Gọi $\boldsymbol{X}_{11}, \boldsymbol{X}_{12}, \cdots, \boldsymbol{X}_{1n_1}, \boldsymbol{X}_{21}, \boldsymbol{X}_{22}, \cdots, \boldsymbol{X}_{2n_2}$ lần lượt là hai mẫu ngẫu nhiên lấy từ 2 giống ruồi có kì vọng lần lượt là $\boldsymbol{\mu}_1$ và $\boldsymbol{\mu}_2$, có ma trận hiệp phương sai lần lượt là $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$

• Ta thực hiện kiểm định $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$

- Giả thuyết thống kê:

$$\begin{aligned}H_0: &\quad \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2\\ H_1: &\quad \boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2\end{aligned}$$

- Mức ý nghĩa $\alpha = 0.01$

Khi H_0 đúng, phép kiểm định Box cho thống kê:

$$C = (1-u)\left\{\left[\sum_l (n_l - 1)\ln|\boldsymbol{S}_{\text{pooled}}|\right] - \sum_l [(n_l - 1)\ln|\boldsymbol{S}_l|]\right\}$$

xấp xỉ phân phối Chi bình phương với bậc tự do:

$$v = \frac{1}{2}p(p+1)(g-1)$$

và

$$u = \left[\sum_l \frac{1}{(n_l - 1)} - \frac{1}{\sum_l (n_l - 1)}\right]\frac{2p^2 + 3p - 1}{6(p+1)(g-1)}$$

trong đó p là số biến và g là số tổng thể

Với mẫu thực nghiệm ta có:

- $p = 7, g = 2, n_1 = 35, n_2 = 35$. Khi đó $v = 28, u \approx 0.109$
- $c \approx 64.539$

- $\chi^2_v(\alpha) = \chi^2_{28}(0.01) \approx 48.278$

Vì $c > \chi^2_{28}(0.01)$ nên với mức ý nghĩa $\alpha = 0.01$ ta bác bỏ H_0. Như vậy, với mức ý nghĩa $\alpha = 0.01$, ta kết luận $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$

• Ta thực hiện kiểm định $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$

- Giả thuyết thống kê:

$$\begin{aligned} H_0: &\quad \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \\ H_1: &\quad \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2 \end{aligned}$$

- Mức ý nghĩa $\alpha = 0.05$
- $n_1 = 35, n_2 = 35, p = 7$. Khi đó $n_1 - p = 28 < 30$ và $n_2 - p = 28 < 30$

Khi H_0 đúng, ta có thống kê:

$$T^2 = (\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)^T \left[\frac{1}{n_1}\boldsymbol{S}_1 + \frac{1}{n_2}\boldsymbol{S}_2\right]^{-1} (\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)$$

xấp xỉ $\dfrac{vp}{v-p+1}F_{p,v-p+1}$ trong đó v được định nghĩa bởi $(6-29)$ trong sách

Với mẫu thực nghiệm, ta có:

- $p = 7, v = 66$
- $t^2 \approx 106.135$
- $\dfrac{vp}{v-p+1}F_{p,v-p+1}(\alpha) = \dfrac{462}{60}F_{7,60}(0.01) \approx 16.682$

Vì $t^2 > \dfrac{462}{60}F_{7,60}(0.01)$ nên với mức ý nghĩa $\alpha = 0.01$ ta bác bỏ H_0

Như vậy, với mức ý nghĩa $\alpha = 0.01$ ta kết luận có sự khác nhau về các đặc trưng của 2 giống ruồi này

• Nếu giả thuyết này bị bác bỏ, thành phần (đặc trưng) nào khiến cho giả thuyết này bị bác bỏ ?

Bằng cách sử dụng (5-23) và (6-30), với vectơ $\boldsymbol{a}$ cho trước, ta xây dựng được khoảng tin cậy đồng thời $1-\alpha$ cho $\boldsymbol{a}^T(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ là:

$$\left(\boldsymbol{a}^T(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) - c\sqrt{\boldsymbol{a}^T\left(\frac{1}{n_1}\boldsymbol{S}_1 + \frac{1}{n_2}\boldsymbol{S}_2\right)\boldsymbol{a}}, \boldsymbol{a}^T(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) + c\sqrt{\boldsymbol{a}^T\left(\frac{1}{n_1}\boldsymbol{S}_1 + \frac{1}{n_2}\boldsymbol{S}_2\right)\boldsymbol{a}}\right)$$

trong đó $c^2 = \dfrac{vp}{v-p+1}(\alpha)$

Lần lượt cho $\boldsymbol{a}^T = [0, \cdots, 0, 1, 0, \cdots, 0]$ với số 1 nằm ở vị trí thứ $i = 1, 2, \cdots, 7$. Ta được các khoảng tin cậy đồng thời 95% cho $\mu_{1i} - \mu_{2i}$:

$$\begin{aligned} i = 1: &\quad (-8.745, 2.973) \\ i = 2: &\quad (-4.811, 3.154) \\ i = 3: &\quad (-6.420, -1.466) \\ i = 4: &\quad (-1.845, 1.559) \\ i = 5: &\quad (-7.994, -0.749) \\ i = 6: &\quad (-1.161, 0.989) \\ i = 7: &\quad (-0.629, 1.314) \end{aligned}$$

Ta thấy khoảng tin cậy đồng thời 95% cho $\mu_{13} - \mu_{23}$ và $\mu_{15} - \mu_{25}$ không chứa 0 nên biến thứ 3 và 5 là nguyên nhân khiến cho giả thuyết bị bác bỏ

Bài 6.30

(a) Kiểm định sự bằng nhau giữa hai trung bình, với $\alpha = 0.05$

Bài giải Gọi $\boldsymbol{X}_{11}, \boldsymbol{X}_{12}, \cdots, \boldsymbol{X}_{1n_1}, \boldsymbol{X}_{21}, \boldsymbol{X}_{22}, \cdots, \boldsymbol{X}_{2n_2}$ lần lượt là hai mẫu ngẫu nhiên lấy từ những người tham gia trước và sau chương trình 1 năm có kì vọng lần lượt là $\boldsymbol{\mu}_1$ và $\boldsymbol{\mu}_2$, có ma trận hiệp phương sai lần lượt là $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$

• Ta thực hiện kiểm định $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$

- Giả thuyết thống kê:

$$\begin{aligned} H_0: &\quad \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 \\ H_1: &\quad \boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2 \end{aligned}$$

- Mức ý nghĩa $\alpha = 0.05$

Khi H_0 đúng, phép kiểm định Box cho thống kê:

$$C = (1-u)\left\{\left[\sum_l (n_l - 1)\ln|\boldsymbol{S}_{\text{pooled}}|\right] - \sum_l [(n_l - 1)\ln|\boldsymbol{S}_l|]\right\}$$

xấp xỉ phân phối Chi bình phương với bậc tự do:

$$v = \frac{1}{2}p(p+1)(g-1)$$

và

$$u = \left[\sum_l \frac{1}{(n_l - 1)} - \frac{1}{\sum_l (n_l - 1)}\right]\frac{2p^2 + 3p - 1}{6(p+1)(g-1)}$$

trong đó p là số biến và g là số tổng thể

Với mẫu thực nghiệm ta có:

- $p = 6, g = 2, n_1 = 24, n_2 = 24$. Khi đó $v = 21, u \approx 0.138$
- $C \approx 7.686$
- $\chi^2_v(\alpha) = \chi^2_{21}(0.05) \approx 32.67$

Vì $C < \chi^2_{21}(0.05)$ nên với mức ý nghĩa $\alpha = 0.05$ ta chấp nhận H_0. Như vậy, với mức ý nghĩa $\alpha = 0.05$, ta kết luận $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$

Tiếp đến, đi kiểm định giả thuyết:

$$\begin{aligned} H_0 &: \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \\ H_1 &: \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2 \end{aligned}$$

Ta có trung bình mẫu:

$$\bar{\boldsymbol{x}_1} = \begin{bmatrix} 0.840 \\ 0.813 \\ 1.785 \\ 1.729 \\ 0.697 \\ 0.686 \end{bmatrix}$$

$$\bar{\boldsymbol{x}_2} = \begin{bmatrix} 0.840 \\ 0.810 \\ 1.778 \\ 1.716 \\ 0.712 \\ 0.686 \end{bmatrix}$$

Mức ý nghĩa: $\alpha = 0.05$. Theo (6-28), khi H_0 đúng, phép kiểm định T^2 cho thống kê:

$$\begin{aligned} T^2 &= (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)^T \left[\frac{1}{n_1}\boldsymbol{S}_1 + \frac{1}{n_2}\boldsymbol{S}_2\right]^{-1} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) \\ &\approx 0.829 \end{aligned} \tag{6-27}$$

và điểm tới hạn: $\dfrac{vp}{v-p+1}F_{p,v-p+1}(\alpha)$, trong đó:

$$v = \frac{p + p^2}{\sum_i \frac{1}{n_i}\left\{tr\left[\left(\frac{1}{n_i}\boldsymbol{S}_i\left(\frac{1}{n_1}\boldsymbol{S}_1 + \frac{1}{n_2}\boldsymbol{S}_2\right)^{-1}\right)^2\right] + \left(tr\left[\frac{1}{n_i}\boldsymbol{S}_i\left(\frac{1}{n_1}\boldsymbol{S}_1 + \frac{1}{n_2}\boldsymbol{S}_2\right)^{-1}\right]\right)^2\right\}}$$

Từ mẫu thực nghiệm:

- $p = 6$
- $n_1 = n_2 = 24$
- $v \approx 47.069$
- $\dfrac{vp}{v-p+1}F_{p,v-p+1} = \dfrac{47.069 \times 6}{47.069 - 6 + 1}F_{6,47.069-6+1}(0.05) \approx 15.598$

Như vậy, $T^2 \leq \dfrac{vp}{v-p+1}F_{p,v-p+1}$ nên ta chấp nhận giả thuyết H_0, với mức ý nghĩa $\alpha = 0.05$.
Như vậy, với mức ý nghĩa $\alpha = 0.05$ ta kết luận $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$
(b) Xây dựng khoảng tin cậy đồng thời 95 % cho các trung bình khác nhau.
Bài giải
Từ kết quả 6.3: Cho $c^2 = [(n_1 + n_2 - 2)p/(n_1 + n_2 - p - 1)]F_{p,n_1+n_2-p-1}(\alpha)$. Với mức ý nghĩa $1 - \alpha$.

$$\boldsymbol{a}^T(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2) \pm c\sqrt{\boldsymbol{a}^T\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\boldsymbol{S}_{pooled}\boldsymbol{a}}$$

sẽ chứa $\boldsymbol{a}^T(\mu_1 - \mu_2)$ với mọi $\boldsymbol{a}$. Từ đó, $\mu_{1i} - \mu_{2i}$ có khoảng tin cậy:

$$(\bar{X}_{1i} - \bar{X}_{2i}) \pm c\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)s_{ii,pooled}}, i = 1, 2, ..., p$$

Ta có,

$$\begin{aligned} c^2 &= [(n_1+n_2-2)p/(n_1+n_2-p-1)]F_{p,n_1+n_2-p-1}(\alpha) \\ &\approx 15.68 \end{aligned}$$

Vậy, khoảng tin cậy đồng thời 95% cho từng thành phần của các trung bình mẫu khác nhau là:

$$\begin{aligned} i=1: &\quad (-0.137, 0.137) \\ i=2: &\quad (-0.119, 0.125) \\ i=3: &\quad (-0.342, 0.357) \\ i=4: &\quad (-0.306, 0.331) \\ i=5: &\quad (-0.135, 0.105) \\ i=6: &\quad (-0.120, 0.120) \end{aligned}$$

(c) So sánh khoảng tin cậy ở câu (b) với khoảng tin cậy Bonferroni
Bài làm
Khoảng tin cậy đồng thời 95% Bonferroni:

$$\begin{aligned} i=1: &\quad (-0.070, 0.070) \\ i=2: &\quad (-0.059, 0.066) \\ i=3: &\quad (-0.172, 0.186) \\ i=4: &\quad (-0.151, 0.176) \\ i=5: &\quad (-0.077, 0.046) \\ i=6: &\quad (-0.062, 0.061) \end{aligned}$$

Bài 6.33
(a) Xây dựng bảng MANOVA hai nhân tố. Kiểm định ảnh hưởng bởi giống, ảnh hưởng bởi thời gian và ảnh hưởng bởi tương tác giữa giống và thời gian với mức ý nghĩa $\alpha = 0.05$
Bài giải:
Ta gọi nhân tố thứ nhất là ảnh hưởng bởi giống, nhân tố thứ hai là ảnh hưởng bởi thời gian
Với mẫu thực nghiệm ta có bảng MANOVA sau:

Source of Variation	SSP	df
Factor 1	$\begin{pmatrix} 965.181 & 1377.602 \\ 1377.602 & 2026.856 \end{pmatrix}$	2
Factor 2	$\begin{pmatrix} 1275.248 & 2644.927 \\ 2644.927 & 5573.806 \end{pmatrix}$	2
Interaction	$\begin{pmatrix} 795.808 & 375.963 \\ 375.963 & 193.549 \end{pmatrix}$	4
Residual	$\begin{pmatrix} 76.659 & 37.930 \\ 37.930 & 1769.642 \end{pmatrix}$	27
Total	$\begin{pmatrix} 3112.896 & 4436.422 \\ 4436.422 & 9563.853 \end{pmatrix}$	35

• Kiểm định ảnh hưởng bởi giống loài trên các quan sát với mức ý nghĩa $\alpha = 0.05$

- Giả thuyết thống kê:

$$\begin{aligned} H_0: &\quad \text{Các quan sát không ảnh hưởng bởi giống loài} \\ H_1: &\quad \text{Các quan sát bị ảnh hưởng bởi giống loài} \end{aligned}$$

- Mức ý nghĩa $\alpha = 0.05$

Với mẫu thực nghiệm ta có:

- $\Lambda^* = \dfrac{|\text{SSP}_{\text{res}}|}{|\text{SSP}_{\text{fac1}} + \text{SSP}_{\text{res}}|} \approx 0.069$
- $n = 4, g = 3, b = 3, p = 2$
- $c = -\left[gb(n-1) - \dfrac{p+1-(g-1)}{2}\right] \ln \Lambda^* \approx 70.939$
- $\chi^2_{(g-1)p}(\alpha) = \chi^2_4(0.05) \approx 9.488$

Vì $c > \chi^2_4(0.05)$ nên với mức ý nghĩa $\alpha = 0.05$ ta bác bỏ H_0
Như vậy, với mức ý nghĩa $\alpha = 0.05$, ta kết luận các quan sát bị ảnh hưởng bởi yếu tố giống loài
• Kiểm định ảnh hưởng bởi thời gian trên các quan sát với mức ý nghĩa $\alpha = 0.05$

- Giả thuyết thống kê:

$$\begin{aligned} H_0: &\quad \text{Các quan sát không ảnh hưởng bởi yếu tố thời gian} \\ H_1: &\quad \text{Các quan sát bị ảnh hưởng bởi yếu tố thời gian} \end{aligned}$$

- Mức ý nghĩa $\alpha = 0.05$

Với mẫu thực nghiệm ta có:

- $\Lambda^* = \dfrac{|\text{SSP}_{\text{res}}|}{|\text{SSP}_{\text{fac2}} + \text{SSP}_{\text{res}}|} \approx 0.049$
- $n = 4, g = 3, b = 3, p = 2$
- $c = -\left[gb(n-1) - \dfrac{p+1-(b-1)}{2}\right] \ln \Lambda^* \approx 79.833$
- $\chi^2_{(b-1)p}(\alpha) = \chi^2_4(0.05) \approx 9.488$

Vì $c > \chi^2_4(0.05)$ nên với mức ý nghĩa $\alpha = 0.05$ ta bác bỏ H_0
Như vậy, với mức ý nghĩa $\alpha = 0.05$, ta kết luận các quan sát bị ảnh hưởng bởi yếu tố thời gian
• Kiểm định ảnh hưởng bởi sự tương tác giữa thời gian và giống loài trên các quan sát với mức ý nghĩa $\alpha = 0.05$

- Giả thuyết thống kê:

$$\begin{aligned} H_0: &\quad \text{Các quan sát không ảnh hưởng bởi tương tác giữa thời gian và giống loài} \\ H_1: &\quad \text{Các quan sát bị ảnh hưởng bởi tương tác giữa thời gian và giống loài} \end{aligned}$$

- Mức ý nghĩa $\alpha = 0.05$

Với mẫu thực nghiệm ta có:

- $\Lambda^* = \dfrac{|\text{SSP}_{\text{res}}|}{|\text{SSP}_{\text{int}} + \text{SSP}_{\text{res}}|} \approx 0.087$
- $n = 4, g = 3, b = 3, p = 2$
- $c = -\left[gb(n-1) - \dfrac{p+1-(g-1)(b-1)}{2}\right] \ln \Lambda^* \approx 67.129$
- $\chi^2_{(g-1)(b-1)p}(\alpha) = \chi^2_8(0.05) \approx 15.51$

Vì $c > \chi^2_8(0.05)$ nên với mức ý nghĩa $\alpha = 0.05$ ta bác bỏ H_0
Như vậy, với mức ý nghĩa $\alpha = 0.05$, ta kết luận các quan sát bị ảnh hưởng bởi sự tương tác giữa yếu tố thời gian và yếu tố giống loài
(c) Lập bảng ANOVA 2 nhân tố cho từng biến và rút ra kết luận
Bài giải:
• Thực hiện bảng ANOVA 2 nhân tố cho biến X_1 ta được:

	Sum of Squares	df	F	$\mathbb{P}(>F)$
Species	965.181	2.0	169.973	$5.027e-16$
Time	1275.248	2.0	224.578	$1.492e-17$
Species : Time	795.808	4.0	70.073	$7.341e-14$
Residual	76.659	27.0		

Nhận xét: Các p-giá trị của bảng ANOVA 2 nhân tố cho X_1 đều nhỏ hơn 0.05 nên với mức ý nghĩa $\alpha = 0.05$ ta kết luận biến X_1 bị ảnh hưởng bởi nhân tố giống loài, nhân tố thời gian cũng như bị ảnh hưởng bởi sự tương tác của giống loài qua thời gian
• Tương tự, thực hiện bảng ANOVA 2 nhân tố cho biến X_2 ta được:

	Sum of Squares	df	F	$\mathbb{P}(>F)$
Species	2026.856	2.0	15.462	$3.348e-05$
Time	5573.806	2.0	42.521	$4.537e-09$
Species:Time	193.549	4.0	0.738	0.574
Residual	1769.642	27.0		

Nhận xét: Các p-giá trị của bảng ANOVA 2 nhân tố cho X_2 ở hai dòng đầu đều nhỏ hơn 0.05, nhưng dòng thứ 3 có p-giá trị lớn hơn 0.05 nên với mức ý nghĩa $\alpha = 0.05$ ta kết luận biến X_2 bị ảnh hưởng bởi nhân tố giống loài, nhân tố thời gian nhưng không bị ảnh hưởng bởi sự tương tác của giống loài qua thời gian