Bài Ôn tập Thống kê nhiều chiều

Ngày 14 tháng 6 năm 2013

Bài 1. Xét véc-tơ ngẫu nhiên $\mathbf{X}' = (X_1, \ldots, X_p)$ với véc-tơ kỳ vọng $\boldsymbol{\mu}' = (\mu_1, \ldots, \mu_p)$ và ma trận hiệp phương sai $\mathbf{\Sigma}$. Đặt $\mathbf{C} = \{c_{ij}\}_{q\times p}$ là ma trận hằng số thực và đặt $\mathbf{Z} = \mathbf{CX}$. Chứng tỏ rằng $\mathbb{C}\mathrm{ov}(\mathbf{Z}) = \mathbf{C\Sigma C}'$.

Bài 2. Xét véc-tơ ngẫu nhiên $\mathbf{X}' = (X_1, X_2)$ có phân phối $\mathcal{N}_2(\boldsymbol{\mu}, \mathbf{\Sigma})$, với véc-tơ kỳ vọng $\boldsymbol{\mu} = (0, 2)$ và ma trận hiệp phương sai

$$\mathbf{\Sigma} = \begin{bmatrix} 2 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$

(a) Xác định tâm và các trục của ellip xác suất 50%.

(b) Thực hiện yêu cầu ở câu (a) với xác suất 90%.

(Có thể kiểm tra lại bằng **R**, để vẽ ellip dùng gói ellipse)

Bài 3. Bài **4.14** sách *Applied Multivariate Statistical Analysis* của Johnson & Wichern.

Bài 4. Giả sử rằng véc-tơ ngẫu nhiên $\mathbf{X}' = (X_1, \ldots, X_p)$ có phân phối $\mathcal{N}_p(\boldsymbol{\mu}, \mathbf{\Sigma})$, với véc-tơ kỳ vọng $\boldsymbol{\mu}' = (\mu_1, \ldots, \mu_p)$ và ma trận hiệp phương sai

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_p^2 \end{pmatrix}$$

Chứng tỏ rằng hàm mật độ đồng thời của $(X_1, \ldots, X_p)$ có dạng

$$f(x_1, \ldots, x_p) = \prod_{i=1}^{p} \frac{1}{\sqrt{2\pi\sigma_i}} \exp\left(-\frac{1}{2\sigma_i^2}(x_i - \mu_i)^2\right)$$

Bài 5. Xét biến ngẫu nhiên chuẩn hai chiều $\mathbf{X} \sim \mathcal{N}_2(\boldsymbol{\mu}, \mathbf{\Sigma})$ với $\boldsymbol{\mu}' = (0, 2)$, $\sigma_{11} = 2$, $\sigma_{22} = 1$ và $\rho_{12} = 0.5$ với các σ_{ij} và ρ_{ij}, $i, j = 1, 2$ là các phần tử của ma trận hiệp phương sai $\mathbf{\Sigma}$ và ma trận hệ số tương quan $\boldsymbol{\rho}$.

(a) Viết hàm mật độ đồng thời cho $\mathbf{X}$.

(b) Tìm phân phối có điều kiện của X_1, cho trước $X_2 = x_2$ từ phân phối đồng thời của $\mathbf{X}' = (X_1, X_2)$.

Bài 6. Nếu $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ là mẫu ngẫu nhiên từ $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ với $\boldsymbol{\Sigma}$ không biết. Với giả thuyết $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ và đối thuyết $\boldsymbol{\mu} \neq \boldsymbol{\mu}_0$. Chứng tỏ rằng tỷ lệ hợp lý Λ của kiểm định là

$$\Lambda^{n/2} = \frac{1}{1 + \frac{T^2}{n-1}}$$

với thống kê Hotelling $T^2 = n(\bar{\mathbf{X}} - \boldsymbol{\mu}_0)\mathbf{S}^{-1}(\bar{\mathbf{X}} - \boldsymbol{\mu}_0)'$ suy ra kiểm định dựa trên thống kê T^2 sẽ tương đương với kiểm định tỷ lệ hợp lý.

Bài 7. Xét $\mathbf{X} \sim \mathcal{N}_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ và $\boldsymbol{\mu}' = (2, -3, 1)$ và

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 2 \end{pmatrix}$$

(a) Tìm phân phối của $3X_1 - 2X_2 + X_3$.

(b) Tìm $\mathbf{a}' = (a_1, a_2)$ sao cho X_2 và $X_2 - \mathbf{a}' \begin{pmatrix} X_1 \\ X_3 \end{pmatrix}$ độc lập.

Bài 8. Cho dữ liệu

$$\mathbf{X} = \begin{pmatrix} 2 & 12 \\ 8 & 9 \\ 6 & 9 \\ 8 & 10 \end{pmatrix}$$

(a) Tính giá trị thống kê Hotelling T^2 cho kiểm định giả thuyết $H_0 : \boldsymbol{\mu}' = (7, 11)$ với $\boldsymbol{\mu}' = (\mu_1, \mu_2)$. Xác định phân phối của T^2. Cho kết luận với $\alpha = 0.05$.

(b) Xác định tỷ lệ hợp lý Λ.

(c) Tìm miền tin cậy đồng thời cho véc-tơ kỳ vọng $\boldsymbol{\mu}$.

Bài 9. Xét $\mathbf{X}$ là một véc-tơ ngẫu nhiên p chiều với véc-tơ kỳ vọng $\boldsymbol{\mu}$ và ma trận hiệp phương sai xác định dương $\boldsymbol{\Sigma}$. Các trị riêng của $\boldsymbol{\Sigma}$ là $\lambda_1 \geq \cdots \geq \lambda_p > 0$ và các véc tơ riêng tương ứng là $\mathbf{e}_1$, $\mathbf{e}_2$, $\ldots$, $\mathbf{e}_p$. Đặt ma trận $\mathbf{P} = [\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_p]$ và $\boldsymbol{\Lambda} = \text{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_p\}$ là ma trận đường chéo với các trị riêng trên đường chéo chính.

(a) Chứng tỏ rằng ta có thể phân tích $\boldsymbol{\Sigma}$ dưới dạng $\boldsymbol{\Sigma} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}'$.

Căn bậc hai của ma trận $\boldsymbol{\Sigma}$ cho bởi $\boldsymbol{\Sigma}^{1/2} = \mathbf{P}\boldsymbol{\Lambda}^{1/2}\mathbf{P}'$.

(b) Chứng tỏ $\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2} = \boldsymbol{\Sigma}$.

(c) Dạng nghịch đảo của ma trận $\boldsymbol{\Sigma}^{1/2}$ ký hiệu là $\boldsymbol{\Sigma}^{1/2}$. Biểu diễn dạng nghịch đảo này theo $\mathbf{P}$ và $\boldsymbol{\Lambda}$ và chứng tỏ rằng $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}^{-1/2} = \boldsymbol{\Sigma}^{-1}$.

Xét véc-tơ $\mathbf{Z}' = (Z_1, \ldots, Z_p)$ cho bởi $\mathbf{Z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$.

(d) Chứng tỏ rằng $\mathbb{E}(\mathbf{Z}) = \mathbf{0}$ và $\mathbb{V}\text{ar}(\mathbf{Z}) = \mathbf{I}$.

Giả sử véc-tơ $\mathbf{X}$ có phân phối $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

(e) Chứng tỏ rằng $Z_1, \ldots, Z_p$ là độc lập và có cùng phân phối chuẩn.

(f) Chứng tỏ rằng $(\mathbf{X}-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})$ có phân phối Chi-bình phương với p bậc tự do.

Bài 10. Chứng minh rằng nếu $\mathbf{x}_{ij}$, $i = 1, \ldots, k$, $j = 1, \ldots, n$ là các quan trắc độc lập từ tổng thể có phân phối chuẩn p chiều $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, thì thống kê tỷ lệ hợp lý cho giả thuyết

$$H_0 : \boldsymbol{\mu}_1 = \ldots = \boldsymbol{\mu}_k$$

là

$$\Lambda = \frac{|\mathbf{W}|}{|\mathbf{W}+\mathbf{B}|}$$

với

$$\mathbf{B} = n\sum_{i=1}^{n}(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})'$$

$$\mathbf{E} = \sum_{i=1}^{k}\sum_{j=1}^{n}(\bar{\mathbf{x}}_{ij} - \bar{\mathbf{x}}_i)(\bar{\mathbf{x}}_{ij} - \bar{\mathbf{x}}_i)'$$

Bài 11. **(ANOVA 1 nhân tố)** Trong nghiên cứu về sức căng của sợi tơ tổng hợp, các kỹ sư trong một nhà máy cho rằng sức căng của sợi tơ tổng hợp phụ thuộc vào phần trăm cotton bên trong sợi. 5 mức độ về phần trăm cotton được sử dụng, các quan trắc được chọn theo thứ tự ngẫu nhiên, kết quả cho bởi bảng sau

Phần trăm cotton	Quan trắc				
	1	**2**	**3**	**4**	**5**
15	7	7	15	11	9
20	12	17	12	18	18
25	14	18	18	19	19
30	19	25	22	19	23
35	7	10	11	15	11

(a) Liệu phần trăm cotton có ảnh hưởng đến sức căng của sợi tơ, thực hiện phân tích phương sai với $\alpha = 0.05$.

(b) Sử dụng phương pháp Fisher's LSD với $\alpha = 0.05$ để so sánh trung bình của 5 mức độ cotton.

Bài 12. **(MANOVA 1 nhân tố)** Các quan trắc trên hai biến ngẫu nhiên được thu thập cho 3 nhóm (ứng với 3 phương thức thí nghiệm khác nhau); các véc tơ quan trắc $[x_1 \;\; x_2]$ là

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

$$\text{Nhóm 2}: \begin{bmatrix}3\\3\end{bmatrix}, \begin{bmatrix}1\\6\end{bmatrix}, \begin{bmatrix}2\\3\end{bmatrix}, \begin{bmatrix}2\\5\end{bmatrix}, \begin{bmatrix}2\\2\end{bmatrix}$$

$$\text{Nhóm 3}: \begin{bmatrix}2\\3\end{bmatrix}, \begin{bmatrix}5\\1\end{bmatrix}, \begin{bmatrix}3\\1\end{bmatrix}, \begin{bmatrix}2\\3\end{bmatrix}, \begin{bmatrix}4\\2\end{bmatrix}$$

(a) Tính các ma trận $\mathbf{W}$, $\mathbf{B}$, $\mathbf{T} = \mathbf{W} + \mathbf{B}$ và lập bảng MANOVA 1 nhân tố.

(b) Tính thống kê Wilk's Lambda Λ và xác định thống kê F tương ứng từ thống kê Λ. Kiểm định ảnh hưởng giữa các phương thức thí nghiệm với $\alpha = 0.01$.

Bài 13. Xét $\mathbf{\Sigma}$ là ma trận hiệp phương sai tương ứng với véc-tơ ngẫu nhiên $\mathbf{X}' = (X_1, X_2, \cdots, X_p)$ và $(\lambda_1, \mathbf{e}_1), \ldots, (\lambda_p, \mathbf{e}_p)$ với $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ lần lượt là các cặp trị riêng và véc-tơ riêng xác định từ $\mathbf{\Sigma}$. Thành phần chính thứ i cho bởi

$$\mathbf{Y}_i = \mathbf{e}_i'\mathbf{X} = e_{i1}X_1 + \cdots + e_{ip}X_p, \quad i = 1, 2, \ldots, p$$

Chứng tỏ rằng

$$\begin{aligned} \mathbb{V}\mathrm{ar}(Y_i) &= \mathbf{e}_i'\mathbf{\Sigma}\mathbf{e}_i = \lambda_i \quad i = 1, 2, \ldots, p \\ \mathbb{C}\mathrm{ov}(Y_i, Y_k) &= \mathbf{e}_i'\mathbf{\Sigma}\mathbf{e}_k = 0 \quad i \neq k \end{aligned}$$

Hơn nữa, nếu có một số λ_i bằng nhau, thì sự lựa chọn các véc-tơ hệ số $\mathbf{e}_i$ tương ứng sẽ không duy nhất.

Bài 14. Cho ma trận hiệp phương sai

$$\mathbf{\Sigma} = \begin{bmatrix} 5 & 2 \\ 2 & 2 \end{bmatrix}$$

(a) Xác định các thành phần chính Y_1, Y_2 từ $\mathbf{\Sigma}$. Thành phần chính thứ nhất giải thích bao nhiêu phần trăm sự biến thiên của dữ liệu.

(b) Xác định ma trận hệ số tương quan $\boldsymbol{\rho}$.

(c) Xác định các thành phần chính $\tilde{Y}_1$, $\tilde{Y}_2$ từ $\boldsymbol{\rho}$. Tìm tỷ lệ sự biến thiên do thành phần chính thứ nhất (xác định từ $\boldsymbol{\rho}$) trên tổng sự biến thiên.

(d) Tính các hệ số tương quan ρ_{Y_1,X_1}, ρ_{Y_1,X_2} và ρ_{Y_2,X_1}

Bài 15. Cho véc-tơ ngẫu nhiên $\mathbf{X}' = (X_1, X_2)$ có ma trận hiệp phương sai

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \quad 0 < \rho < 1$$

(a) Xác định các thành phần chính Y_1, Y_2 từ $\mathbf{\Sigma}$.

(b) Nếu thay đổi tỷ lệ của X_1 bằng cách nhân với một hằng số $c > 1$, tìm ma trận hiệp phương sai $\mathbf{\Sigma}_c$ xác định từ $\mathbf{X}_c' = (cX_1, X_2)$. Xác định các thành phần chính từ $\mathbf{\Sigma}_c$. Các thành phần chính thay đổi như thế nào khi thay đổi tỷ lệ của X_1.