ME731 - Métodos em Análise Multivariada
– Análise de Componentes Principais III –

Prof. Carlos Trucíos
ctrucios@unicamp.br
ctruciosm.github.io

Instituto de Matemática, Estatística e Computação Científica,
Universidade Estadual de Campinas

Aula 12

UNICAMP

Agenda I

Resultados Asintóticos

Resultados Asintóticos

Teorema

Sejam $\Sigma > 0$ com autovalores diferentes e $\mathbf{S} \sim n^{-1}W_p(\Sigma, n-1)$ com decomposição espectral $\Sigma = P\Lambda P'$ e $\mathbf{S} = GLG'$. Então,

1. $\sqrt{n-1}(l - \lambda) \rightarrow^D N_p(0, 2\Lambda^2)$ em que $l = (l_1, \cdots, l_p)'$ e $\lambda = (\lambda_1, \cdots, \lambda_p)'$ são as diagonais de L e Λ.
2. $\sqrt{g_j - p_j} \rightarrow^D N_p\left(0, \lambda_j \sum_{k \neq j} \frac{\lambda_k}{(\lambda_k - \lambda_j)^2} p_k p_k'\right)$.
3. Os elemenos em l são asintóticamente independentes dos elementos em G.

Resultados Asintóticos

Teorema

Sejam $\Sigma > 0$ com autovalores diferentes e $\mathbf{S} \sim n^{-1}W_p(\Sigma, n-1)$ com decomposição espectral $\Sigma = P\Lambda P'$ e $\mathbf{S} = GLG'$. Então,

1. $\sqrt{n-1}(l-\lambda) \to^D N_p(0, 2\Lambda^2)$ em que $l = (l_1, \cdots, l_p)'$ e $\lambda = (\lambda_1, \cdots, \lambda_p)'$ são as diagonais de L e Λ.
2. $\sqrt{g_j - p_j} \to^D N_p\left(0, \lambda_j \sum_{k \neq j} \frac{\lambda_k}{(\lambda_k - \lambda_j)^2} p_k p_k'\right)$.
3. Os elemenos em l são asintóticamente independentes dos elementos em G.

Podemos utilizar esses resultados para construir intervalos de confiança e testes de hipóteses.

Resultados Asintóticos

Sejam $\mathbf{X}_1, \cdots, \mathbf{X}_n \sim N_p(\mu, \Sigma)$. Então

$$\mathbf{S} \sim n^{-1} W_p(\Sigma, n-1) \quad e \quad \sqrt{n-1}(l - \lambda) \to^{D} N_p(0, 2\Lambda^2)$$

Resultados Asintóticos

Sejam $\mathbf{X}_1, \cdots, \mathbf{X}_n \sim N_p(\mu, \Sigma)$. Então

$$\mathbf{S} \sim n^{-1} W_p(\Sigma, n-1) \quad e \quad \sqrt{n-1}(l - \lambda) \rightarrow^D N_p(0, 2\Lambda^2)$$

Consequentemente,

$$\sqrt{n-1}(l_j - \lambda_j) \rightarrow^D N_p(0, 2\lambda_j^2) \quad j = 1, \cdots, p.$$

Resultados Asintóticos

Sejam $\mathbf{X}_1, \cdots, \mathbf{X}_n \sim N_p(\mu, \Sigma)$. Então

$$\mathbf{S} \sim n^{-1} W_p(\Sigma, n-1) \quad e \quad \sqrt{n-1}(l - \lambda) \rightarrow^D N_p(0, 2\Lambda^2)$$

Consequentemente,

$$\sqrt{n-1}(l_j - \lambda_j) \rightarrow^D N_p(0, 2\lambda_j^2) \quad j = 1, \cdots, p.$$

Pelo método Delta,

$$\sqrt{n-1}(\log(l_j) - \log(\lambda_j) \rightarrow^D N_p(0, 2)$$

Resultados Asintóticos

Sejam $\mathbf{X}_1, \cdots, \mathbf{X}_n \sim N_p(\mu, \Sigma)$. Então

$$\mathbf{S} \sim n^{-1} W_p(\Sigma, n-1) \quad e \quad \sqrt{n-1}(l - \lambda) \to^D N_p(0, 2\Lambda^2)$$

Consequentemente,

$$\sqrt{n-1}(l_j - \lambda_j) \to^D N_p(0, 2\lambda_j^2) \quad j = 1, \cdots, p.$$

Pelo método Delta,

$$\sqrt{n-1}(\log(l_j) - \log(\lambda_j) \to^D N_p(0, 2)$$

Assim,

$$\sqrt{\frac{n-1}{2}}(\log(l_j) - \log(\lambda_j) \to^D N_p(0, 1) \quad ou \quad \sqrt{\frac{n-1}{2}}(l_j/\lambda_j - 1) \to^D N_p(0, 1)$$

Resultados Asintóticos

A variância explicada pelas primeiras q componentes principais é dada por

$$\psi = \frac{\lambda_1 + \cdots + \lambda_q}{\sum_{i=1}^{p} \lambda_i}.$$

Na prática, ψ é estimado por

$$\hat{\psi} = \frac{l_1 + \cdots + l_q}{\sum_{i=1}^{p} l_i}.$$

Resultados Asintóticos

A variância explicada pelas primeiras q componentes principais é dada por

$$\psi = \frac{\lambda_1 + \cdots + \lambda_q}{\sum_{i=1}^{p} \lambda_i}.$$

Na prática, ψ é estimado por

$$\hat{\psi} = \frac{l_1 + \cdots + l_q}{\sum_{i=1}^{p} l_i}.$$

Sabemos que $\sqrt{n-1}(l - \lambda) \to^{D} N_p(0, 2\Lambda^2)$, então aplicando o método Delta.

Resultados Asintóticos

$$\sqrt{n-1}(\hat{\psi}-\psi) \rightarrow^{D} N(0, D'2\Lambda D),$$

em que $D = (d_1, \cdots, d_p)'$ e

$$d_j = \frac{\partial \psi}{\partial \lambda_j} = \begin{cases} \frac{1-\psi}{Tr(\Sigma)}, & \text{se } 1 \leq j \leq q, \\ \frac{-\psi}{Tr(\Sigma)}, & \text{se } q+1 \leq j \leq p. \end{cases}$$

Resultados Asintóticos

$$\sqrt{n-1}(\hat{\psi}-\psi) \to^{D} N(0, D'2\Lambda D),$$

em que $D = (d_1, \cdots, d_p)'$ e

$$d_j = \frac{\partial \psi}{\partial \lambda_j} = \begin{cases} \dfrac{1-\psi}{Tr(\Sigma)}, & \text{se } 1 \leq j \leq q, \\ \dfrac{-\psi}{Tr(\Sigma)}, & \text{se } q+1 \leq j \leq p. \end{cases}$$

$$\sqrt{n-1}(\hat{\psi}-\psi) \to^{D} N\Big(0, \frac{2Tr(\Sigma^2)}{(Tr(\Sigma))^2}(\psi - 2\beta\psi + \beta)\Big),$$

em que $\beta = \dfrac{\lambda_1^2 + \cdots + \lambda_q^2}{\lambda_1^2 + \cdots + \lambda_p^2}$

SVD

SVD

SVD (Singular Value Decomposition)

$$X = UDV^t \tag{1}$$

onde U, V são ortogonais .

$$X^tX = VD^2V^t \tag{2}$$

que é a decomposição espectral de X^tX com V sendo a matriz de autovetores de D^2 a matriz de autovalores associados.

Caso de Estudo

Caso de Estudo

- Neste notebook discutiremos passo a passo como realizar uma análise de componentes principais na prática.
- Serão discutidos tópicos referentes à interpretação e outros detalhes encontrados em algumas implementações populares.

Referências

Referências

- Härdle, W. K., & Simar, L. (2019). Applied Multivariate Statistical Analysis. Fifth Editon. Springer Nature. Capítulo 11.
- Johnson, R. A., & Wichern, D. W. (2007). Applied multivariate statistical analysis. Sixth Edition. Pearson Prentice Hall. Capítulo 4.
- Mardia, K. V., Kent, J. T., & Bibby, J, M. (1979). Multivariate Analysis. Academic Press. Capítulo 8.
- Peña, D. (2002). Análisis de Datos Multivariantes. Mc Graw Hill. Capítulo 5.
- Husson, F., Lê, S., & Pagès, J. (2017). Exploratory Multivariate Analysis by Example Using R. Second Edition. CRC Press. Capítulo 1