SME0822 Análise Multivariada e Aprendizado Não-Supervisionado

Aula 9a: **Análise de Correlações Canônicas**

Prof. Cibele Russo

cibele@icmc.usp.br

http://www.icmc.usp.br/~cibele

Johnson, R. A., & Wichern, D. W. (2007). Applied Multivariate Statistical Analysis. Prentice Hall.
Seber, G. A. F. (2009) Multivariate observations. John Wiley & Sons.
Anderson, T. W. (2003) An Introduction to Multivariate Statistical Analysis. Wiley. New York. 3rd Edition.

Análise de correlações canônicas

Objetivos:

- **Compreender e quantificar a associação linear** entre dois conjuntos de variáveis.
- **Resumir** a informação de cada conjunto de variáveis nas chamadas variáveis canônicas, para **avaliar a correlação entre elas (correlação canônica)**

Análise de correlações canônicas

Objetivos:

- **Compreender e quantificar a associação linear** entre dois conjuntos de variáveis.
- **Resumir** a informação de cada conjunto de variáveis nas chamadas variáveis canônicas, para **avaliar a correlação entre elas** (**correlação canônica**)

Análise de correlações canônicas

Origem: Hotelling, 1935.

Exemplo: avaliar a associação linear entre notas no vestibular e no primeiro semestre da faculdade.

Na prática, temos que resolver um problema de otimização restrita, como veremos a seguir.

Análise de correlações canônicas

Origem: Hotelling, 1935.

Exemplo: avaliar a associação linear entre notas no vestibular e no primeiro semestre da faculdade.

Na prática, temos que resolver um problema de otimização restrita, como veremos a seguir.

Modelo de Análise de correlações canônicas via matriz de correlação

Sejam $\underset{\sim}{X}_{p\times 1}$ e $\underset{\sim}{Y}_{q\times 1}$ vetores aleatórios com

$$\begin{aligned}
&\mathsf{E}(\underset{\sim}{X}) = \underset{\sim}{\mu}_X, \ \mathsf{Var}(\underset{\sim}{X}) = \Sigma_X, \\
&\mathsf{E}(\underset{\sim}{Y}) = \underset{\sim}{\mu}_Y, \ \mathsf{Var}(\underset{\sim}{Y}) = \Sigma_Y \text{ e} \\
&\mathsf{Cov}(\underset{\sim}{X}, \underset{\sim}{Y}) = \Sigma_{XY} \\
&\mathsf{Cov}(\underset{\sim}{Y}, \underset{\sim}{X}) = \Sigma_{YX} = \Sigma_{XY}^\top
\end{aligned}$$

Variáveis canônicas

Definimos

$$U_1 = \underset{\sim}{a}_1^\top \underset{\sim}{X} \text{ e}$$
$$V_1 = \underset{\sim}{b}_1^\top \underset{\sim}{Y}$$

em que $dim(\underset{\sim}{a}_1) = p \times 1$ e $dim(\underset{\sim}{b}_1) = q \times 1$. Queremos encontrar vetores de constantes $\underset{\sim}{a}_1$ e $\underset{\sim}{b}_1$ tais que a correlação entre as variáveis U_1 e V_1 seja máxima, com a restrição de que

$$\text{Var}(U_1) = 1 = \text{Var}(V_1)$$

(U_1, V_1) é denominado o **primeiro par de variáveis canônicas**.

Variáveis canônicas

Um **segundo par de variáveis canônicas** (U_2, V_2) poderia ser definido como

$$U_2 = \underset{\sim}{a}_2^\top \underset{\sim}{X} \text{ e}$$
$$V_2 = \underset{\sim}{b}_2^\top \underset{\sim}{Y}$$

em que $dim(\underset{\sim}{a}_2) = p \times 1$ e $dim(\underset{\sim}{b}_2) = q \times 1$. Queremos encontrar vetores de constantes $\underset{\sim}{a}_2$ e $\underset{\sim}{b}_2$ tais que a correlação entre as variáveis U_2 e V_2 a máxima possível, com as restrições de que

$$\text{Var}(U_2) = 1 = \text{Var}(V_2) \text{ e}$$

$$\text{Cov}(U_1, U_2) = \text{Cov}(U_1, V_2) = \text{Cov}(U_2, V_1) = \text{Cov}(U_2, V_2) = 0$$

e assim por diante, até o par de variáveis canônicas (U_k, V_k), com $k = min(p, q)$.

Variáveis canônicas

Ou seja, para obter o primeiro par de variáveis canônicas (U_1, V_1) queremos obter o

$$\begin{aligned}&\max \text{ Cor}(U_1, V_1)\\ &\text{sujeito a Var}(U_1) = \text{Var}(V_1) = 1 \quad (\star)\end{aligned}$$

Temos que

$$\begin{aligned}\text{Cor}(U_1, V_1) &= \frac{\text{Cov}(U_1, V_1)}{\sqrt{\text{Var}(U_1)}\sqrt{\text{Var}(V_1)}} \overset{(\star)}{=} \text{Cov}(U_1, V_1)\\ &= \text{Cov}(\underset{\sim}{a}_1^\top \underset{\sim}{X}, \underset{\sim}{b}_1^\top \underset{\sim}{Y}) = \underset{\sim}{a}_1^\top \Sigma_{XY} \underset{\sim}{b}_1.\end{aligned}$$

Análise de correlações canônicas

Isto é, temos que obter

$$\text{max } \underset{\sim}{a}_1^\top \Sigma_{XY} \underset{\sim}{b}_1$$
$$\text{sujeito a } \underset{\sim}{a}_1^\top \Sigma_X \underset{\sim}{a}_1 = \underset{\sim}{b}_1^\top \Sigma_Y \underset{\sim}{b}_1 = 1$$

Podemos considerar, por exemplo, o **Método dos multiplicadores de Lagrange**, com

$$\mathscr{L} = \underset{\sim}{a}_1^\top \Sigma_{XY} \underset{\sim}{b}_1 - \frac{\alpha}{2}\left(\underset{\sim}{a}_1^\top \Sigma_X \underset{\sim}{a}_1 - 1\right) - \frac{\beta}{2}\left(\underset{\sim}{b}_1^\top \Sigma_Y \underset{\sim}{b}_1 - 1\right)$$

em que α e β são multiplicadores de Lagrange.

Análise de correlações canônicas

Isto é, temos que obter

$$\begin{aligned}&\text{max } \underset{\sim}{a}_1^\top \Sigma_{XY} \underset{\sim}{b}_1\\ &\text{sujeito a } \underset{\sim}{a}_1^\top \Sigma_X \underset{\sim}{a}_1 = \underset{\sim}{b}_1^\top \Sigma_Y \underset{\sim}{b}_1 = 1\end{aligned}$$

Podemos considerar, por exemplo, o **Método dos multiplicadores de Lagrange**, com

$$\mathscr{L} = \underset{\sim}{a}_1^\top \Sigma_{XY} \underset{\sim}{b}_1 - \frac{\alpha}{2}\left(\underset{\sim}{a}_1^\top \Sigma_X \underset{\sim}{a}_1 - 1\right) - \frac{\beta}{2}\left(\underset{\sim}{b}_1^\top \Sigma_Y \underset{\sim}{b}_1 - 1\right)$$

em que α e β são multiplicadores de Lagrange.

Análise de correlações canônicas

Para encontrar o máximo, fazemos

$$\frac{\partial \mathscr{L}}{\partial \underset{\sim}{a}_1} = \Sigma_{XY}\underset{\sim}{b}_1 - \alpha\Sigma_X \underset{\sim}{a}_1 = \underset{\sim}{0} \quad (i)$$

$$\frac{\partial \mathscr{L}}{\partial \underset{\sim}{b}_1} = \Sigma_{XY}^\top \underset{\sim}{a}_1 - \beta\Sigma_Y \underset{\sim}{b}_1 = \underset{\sim}{0} \quad (ii)$$

Multiplicando (i) à esquerda por $\underset{\sim}{a}_1^\top$ e (ii) à esquerda por $\underset{\sim}{b}_1^\top$ temos

Análise de correlações canônicas

Para encontrar o máximo, fazemos

$$\frac{\partial \mathscr{L}}{\partial \underset{\sim}{a}_1} = \Sigma_{XY}\underset{\sim}{b}_1 - \alpha \Sigma_X \underset{\sim}{a}_1 = \underset{\sim}{0} \quad (i)$$

$$\frac{\partial \mathscr{L}}{\partial \underset{\sim}{b}_1} = \Sigma_{XY}^\top \underset{\sim}{a}_1 - \beta \Sigma_Y \underset{\sim}{b}_1 = \underset{\sim}{0} \quad (ii)$$

Multiplicando (i) à esquerda por $\underset{\sim}{a}_1^\top$ e (ii) à esquerda por $\underset{\sim}{b}_1^\top$ temos

Análise de correlações canônicas

$$\underset{\sim}{a}_1^\top \Sigma_{XY} \underset{\sim}{b}_1 - \alpha \underset{\sim}{a}_1^\top \Sigma_X \underset{\sim}{a}_1 = \underset{\sim}{0}$$

$$\underset{\sim}{b}_1^\top \Sigma_{XY}^\top \underset{\sim}{a}_1 - \beta \underset{\sim}{b}_1^\top \Sigma_Y \underset{\sim}{b}_1 = \underset{\sim}{0}$$

$\Longrightarrow \alpha = \underset{\sim}{a}_1^\top \Sigma_{XY} \underset{\sim}{b}_1$ e $\beta = \underset{\sim}{b}_1^\top \Sigma_{XY}^\top \underset{\sim}{a}_1$,

já que $\underset{\sim}{a}_1^\top \Sigma_X \underset{\sim}{a}_1 = \underset{\sim}{b}_1^\top \Sigma_Y \underset{\sim}{b}_1 = 1$.

Mas

$$\alpha = \underset{\sim}{a}_1^\top \Sigma_{XY} \underset{\sim}{b}_1 = \underset{\sim}{b}_1^\top \Sigma_{XY}^\top \underset{\sim}{a}_1 = \beta \doteq \eta$$

Análise de correlações canônicas

$$\underset{\sim}{a}_1^\top \Sigma_{XY} \underset{\sim}{b}_1 - \alpha \underset{\sim}{a}_1^\top \Sigma_X \underset{\sim}{a}_1 = \underset{\sim}{0}$$

$$\underset{\sim}{b}_1^\top \Sigma_{XY}^\top \underset{\sim}{a}_1 - \beta \underset{\sim}{b}_1^\top \Sigma_Y \underset{\sim}{b}_1 = \underset{\sim}{0}$$

$$\Longrightarrow \alpha = \underset{\sim}{a}_1^\top \Sigma_{XY} \underset{\sim}{b}_1 \text{ e } \beta = \underset{\sim}{b}_1^\top \Sigma_{XY}^\top \underset{\sim}{a}_1,$$

já que $\underset{\sim}{a}_1^\top \Sigma_X \underset{\sim}{a}_1 = \underset{\sim}{b}_1^\top \Sigma_Y \underset{\sim}{b}_1 = 1$.

Mas

$$\alpha = \underset{\sim}{a}_1^\top \Sigma_{XY} \underset{\sim}{b}_1 = \underset{\sim}{b}_1^\top \Sigma_{XY}^\top \underset{\sim}{a}_1 = \beta \doteq \eta$$

Análise de correlações canônicas

$$\underset{\sim}{a}_1^\top \Sigma_{XY} \underset{\sim}{b}_1 - \alpha \underset{\sim}{a}_1^\top \Sigma_X \underset{\sim}{a}_1 = \underset{\sim}{0}$$

$$\underset{\sim}{b}_1^\top \Sigma_{XY}^\top \underset{\sim}{a}_1 - \beta \underset{\sim}{b}_1^\top \Sigma_Y \underset{\sim}{b}_1 = \underset{\sim}{0}$$

$\implies \alpha = \underset{\sim}{a}_1^\top \Sigma_{XY} \underset{\sim}{b}_1$ e $\beta = \underset{\sim}{b}_1^\top \Sigma_{XY}^\top \underset{\sim}{a}_1$,

já que $\underset{\sim}{a}_1^\top \Sigma_X \underset{\sim}{a}_1 = \underset{\sim}{b}_1^\top \Sigma_Y \underset{\sim}{b}_1 = 1.$

Mas

$$\alpha = \underset{\sim}{a}_1^\top \Sigma_{XY} \underset{\sim}{b}_1 = \underset{\sim}{b}_1^\top \Sigma_{XY}^\top \underset{\sim}{a}_1 = \beta \doteq \eta$$

Análise de correlações canônicas

Obs:

$(i) - (ii)$ podem ser reescritas como

$$\begin{cases} \Sigma_{XY} \underset{\sim}{b}_1 - \eta \Sigma_X \underset{\sim}{a}_1 = \underset{\sim}{0} \\ \\ \Sigma_{XY}^\top \underset{\sim}{a}_1 - \eta \Sigma_Y \underset{\sim}{b}_1 = \underset{\sim}{0} \end{cases}$$

ou, matricialmente,

$$\begin{pmatrix} -\eta \Sigma_X & \Sigma_{XY} \\ \Sigma_{XY}^\top & -\eta \Sigma_Y \end{pmatrix} \begin{pmatrix} \underset{\sim}{a}_1 \\ \underset{\sim}{b}_1 \end{pmatrix} = \begin{pmatrix} \underset{\sim}{0} \\ \underset{\sim}{0} \end{pmatrix}$$

Análise de correlações canônicas

Daí, multiplicando a primeira equação por η e a segunda por Σ_Y^{-1}, temos

$$\begin{cases} \eta\Sigma_{XY}\underset{\sim}{b}_1 - \eta^2\Sigma_X\underset{\sim}{a}_1 = \underset{\sim}{0} \quad (\star) \\ \\ \Sigma_Y^{-1}\Sigma_{XY}^\top\underset{\sim}{a}_1 = \eta\underset{\sim}{b}_1 \quad (\star\star) \end{cases}$$

Em seguida, substituímos $(\star\star)$ em $(\star)$ e obtemos

Análise de correlações canônicas

$\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{XY}^\top \underset{\sim}{a}_1 - \eta^2\Sigma_X \underset{\sim}{a}_1 = \underset{\sim}{0}$, ou

$(\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{XY}^\top - \eta^2\Sigma_X)\underset{\sim}{a}_1 = \underset{\sim}{0} \quad (iii)$

e, analogamente, teríamos também

$(\Sigma_{XY}^\top\Sigma_X^{-1}\Sigma_{XY} - \eta^2\Sigma_Y)\underset{\sim}{b}_1 = \underset{\sim}{0} \quad (iv),$

Análise de correlações canônicas

$\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{XY}^\top \underset{\sim}{a}_1 - \eta^2\Sigma_X \underset{\sim}{a}_1 = \underset{\sim}{0}$, ou

$(\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{XY}^\top - \eta^2\Sigma_X)\underset{\sim}{a}_1 = \underset{\sim}{0} \quad (iii)$

e, analogamente, teríamos também

$(\Sigma_{XY}^\top\Sigma_X^{-1}\Sigma_{XY} - \eta^2\Sigma_Y)\underset{\sim}{b}_1 = \underset{\sim}{0} \quad (iv),$

Análise de correlações canônicas

Vamos denotar por $\lambda = \eta^2$ as soluções de (iii) e (iv).

Em particular, tomando λ o maior autovalor de

$$\Sigma_X^{-1}\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{XY}^\top,$$

ou, equivalentemente, de

$$\Sigma_Y^{-1}\Sigma_{XY}^\top\Sigma_X^{-1}\Sigma_{XY},$$

teremos a maior correlação entre U_1 e V_1, que é $\underset{\sim}{a}_1^\top\Sigma_{XY}\underset{\sim}{b}_1$.

Análise de correlações canônicas

De forma mais geral, para encontrar $\underset{\sim}{a}_k$ e $\underset{\sim}{b}_k$, para $k = 1, \ldots, min(p, q)$, podemos resolver o sistema de equações

$$\begin{cases} (\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{XY}^\top - \lambda_k\Sigma_X)\underset{\sim}{a}_k = \underset{\sim}{0} \\ \\ (\Sigma_{XY}^\top\Sigma_X^{-1}\Sigma_{XY} - \lambda_k\Sigma_Y)\underset{\sim}{b}_k = \underset{\sim}{0} \end{cases}$$

com λ_k satisfazendo

$$\begin{cases} |\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{XY}^\top - \lambda_k\Sigma_X| = \underset{\sim}{0} \\ \\ |\Sigma_{XY}^\top\Sigma_X^{-1}\Sigma_{XY} - \lambda_k\Sigma_Y| = \underset{\sim}{0} \end{cases}$$

isto é, λ_k é o k-ésimo maior autovalor da matriz $\Sigma_X^{-1}\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{XY}^\top$ ou equivalentemente da matriz $\Sigma_Y^{-1}\Sigma_{XY}^\top\Sigma_X^{-1}\Sigma_{XY}$.

Análise de correlações canônicas

A **correlação canônica** é a correlação em valor absoluto entre U_k e V_k e é igual a $\sqrt{\lambda_k}$, isto é,

$$\rho_k^{\star 2} = \lambda_k = (Cor(U_k, V_k))^2 = \frac{(\underset{\sim}{a}_k^\top \Sigma_{XY} \underset{\sim}{b}_k)^2}{(\underset{\sim}{a}_k \Sigma_X \underset{\sim}{a}_k)(\underset{\sim}{b}_k \Sigma_Y \underset{\sim}{b}_k)}.$$

As **cargas canônicas** são as correlações entre as variáveis originais e as variáveis canônicas.

Correlações canônicas a partir da matriz de correlações

Da mesma forma, as variáveis canônicas também podem ser obtidas a partir de variáveis padronizadas. Nesse caso, devemos resolver o sistema de equações

$$\begin{cases} (P_{XY} P_Y^{-1} P_{XY}^\top - \lambda_k P_X) \underset{\sim}{a}_k = \underset{\sim}{0} \\ \\ (P_{XY}^\top P_X^{-1} P_{XY} - \lambda_k P_Y) \underset{\sim}{b}_k = \underset{\sim}{0} \end{cases}$$

com λ_k satisfazendo

$$\begin{cases} |P_{XY} P_Y^{-1} P_{XY}^\top - \lambda_k P_X| = \underset{\sim}{0} \\ \\ |P_{XY}^\top P_X^{-1} P_{XY} - \lambda_k P_Y| = \underset{\sim}{0}. \end{cases}$$

Correlações canônicas a partir da matriz de correlações

Da mesma forma, as variáveis canônicas também podem ser obtidas a partir de variáveis padronizadas. Nesse caso, devemos resolver o sistema de equações

$$\begin{cases} (P_{XY}P_Y^{-1}P_{XY}^\top - \lambda_k P_X)\underset{\sim}{a}_k = \underset{\sim}{0} \\ \\ (P_{XY}^\top P_X^{-1}P_{XY} - \lambda_k P_Y)\underset{\sim}{b}_k = \underset{\sim}{0} \end{cases}$$

com λ_k satisfazendo

$$\begin{cases} |P_{XY}P_Y^{-1}P_{XY}^\top - \lambda_k P_X| = \underset{\sim}{0} \\ \\ |P_{XY}^\top P_X^{-1}P_{XY} - \lambda_k P_Y| = \underset{\sim}{0}. \end{cases}$$

Estimação das variáveis canônicas

Dadas as amostras aleatórias de $\underset{\sim}{X}$ e $\underset{\sim}{Y}$, as **matrizes de variâncias e covariâncias populacionais** Σ_X, Σ_Y, Σ_{XY}, $\Sigma_{YX} = \Sigma_{XY}^\top$ são estimadas pelas correspondentes **matrizes de variâncias e covariâncias amostrais** S_X, S_Y, S_{XY} e $S_{YX} = S_{XY}^\top$.

Caso as matrizes de variâncias e covariâncias populacionais sejam desconhecidas, a análise de correlações canônicas pode então ser desenvolvida substituindo as matrizes Σ por S.

Da mesma forma, a análise de correlações canônicas pode ser desenvolvida substituindo as matrizes P pelas correspondentes **matrizes de correlações amostrais** R no slide anterior.

Proporção da variância total explicada

A proporção da variância total explicada pelas variáveis canônicas é dada por

$$\text{PVTE}_{U_k} = \frac{\sum_{i=1}^{p}(Cor(U_k, X_i))^2}{p} \times 100\%$$

$$\text{PVTE}_{V_k} = \frac{\sum_{i=1}^{q}(Cor(V_k, Y_i))^2}{q} \times 100\%.$$

Inferência sobre as correlações canônicas

Queremos avaliar se $H_0 : \Sigma_{XY} = 0$ contra $H_1 : \Sigma_{XY} \neq 0$.

Em casos de amostras relativamente grandes, podemos considerar a estatística

$$Q_1 = -n \log \prod_{i=1}^{p} (1 - \lambda_i)$$

e rejeitamos H_0 se $Q_1 > c$, com c o valor crítico tal que $P(\chi^2_{pq} > c) = \alpha$.

Se as amostras forem pequenas, Bartlett sugere que rejeitemos H_0 se

$$Q_2 = -(n - 1 - \frac{1}{2}(p + q + 1) \log \prod_{i=1}^{p} (1 - \lambda_i))$$

for maior que o valor crítico c tal que $P(\chi^2_{pq} > c) = \alpha$.

Inferência sobre as correlações canônicas

Queremos avaliar se $H_0 : \Sigma_{XY} = 0$ contra $H_1 : \Sigma_{XY} \neq 0$.

Em casos de amostras relativamente grandes, podemos considerar a estatística

$$Q_1 = -n \log \prod_{i=1}^{p} (1 - \lambda_i)$$

e rejeitamos H_0 se $Q_1 > c$, com c o valor crítico tal que $P(\chi^2_{pq} > c) = \alpha$.

Se as amostras forem pequenas, Bartlett sugere que rejeitemos H_0 se

$$Q_2 = -(n - 1 - \frac{1}{2}(p + q + 1) \log \prod_{i=1}^{p} (1 - \lambda_i))$$

for maior que o valor crítico c tal que $P(\chi^2_{pq} > c) = \alpha$.

Inferência sobre as correlações canônicas

Queremos avaliar se $H_0 : \Sigma_{XY} = 0$ contra $H_1 : \Sigma_{XY} \neq 0$.

Em casos de amostras relativamente grandes, podemos considerar a estatística

$$Q_1 = -n \log \prod_{i=1}^{p} (1 - \lambda_i)$$

e rejeitamos H_0 se $Q_1 > c$, com c o valor crítico tal que $P(\chi^2_{pq} > c) = \alpha$.

Se as amostras forem pequenas, Bartlett sugere que rejeitemos H_0 se

$$Q_2 = -(n - 1 - \frac{1}{2}(p + q + 1) \log \prod_{i=1}^{p} (1 - \lambda_i))$$

for maior que o valor crítico c tal que $P(\chi^2_{pq} > c) = \alpha$.