Análise de Correlação Canônica

Prof. George von Borries

Departamento de Estatística
Universidade de Brasília

2023

Introdução

Observação

O material seguinte é baseado em

- *Slides de aula de Júlia M. Pavan Soler, Instituto de Matemática e Estatística, USP. (Autorizado pela autora)*
- *Johnson, R.A. e Wichern, D.W. (2007) Applied Multivariate Analysis, 6th Edition. Capítulo 10.*

Introdução

- Correlação Canônica tem o objetivo de verifica a relação linear **entre** dois **conjuntos** de variáveis.
- Na ACP e AF consideramos a variação **dentro** de um conjunto de variáveis.
- Assumimos que os dois conjuntos de variáveis são medidos na mesma unidade amostral.
- A ideia é determinar o pares de combinações lineares com maior correlação e não correlacionados com pares anteriormente selecionados. Os pares de combinações lineares são chamados de variáveis canônicas e suas correlações de correlação canônica.
- Exemplos:
  1. Relacionar variáveis de política governamental com variáveis econômicas ao longo do tempo.
  2. Relacionar variáveis de desempenho escolar no ensino médio e superior.
  3. Relacionar variáveis de satisfação numa profissão com variáveis de características da profissão.

- Seja $\Re^{p+q} = \Re^p \times \Re^q$ tal que $\mathbf{Y}_{p\times 1} = \{\mathrm{y}_1, \ldots, \mathrm{y}_p\}$ e $\mathbf{X}_{q\times 1} = \{\mathrm{x}_1, \ldots, \mathrm{y}_q\}$ são dois grupos de variáveis aleatórias, i.e.,

$$\mathbf{M} = \left(\frac{\mathbf{Y}}{\mathbf{X}}\right) = \begin{pmatrix} \mathrm{y}_1 \\ \mathrm{y}_2 \\ \vdots \\ \mathrm{y}_p \\ \hline \mathrm{x}_1 \\ \mathrm{x}_2 \\ \vdots \\ \mathrm{x}_q \end{pmatrix}$$

- Seguindo a notação de Johnson e Wichern (2007), assumimos $p \leq q$.

- Sejam ainda,

$$\boldsymbol{\mu} = \mathrm{E}(\mathbf{M}) = \left(\frac{\mathrm{E}(\mathbf{Y})}{\mathrm{E}(\mathbf{X})}\right) = \left(\frac{\boldsymbol{\mu}_{\mathrm{y}}}{\boldsymbol{\mu}_{\mathrm{x}}}\right)$$

e

$$\begin{aligned}
\boldsymbol{\Sigma} &= \mathrm{E}[(\mathbf{M}-\boldsymbol{\mu})(\mathbf{M}-\boldsymbol{\mu})^{\mathsf{T}}] \\
&= \left(\begin{array}{c|c}
\mathrm{E}[(\mathbf{Y}-\boldsymbol{\mu}_{\mathrm{y}})(\mathbf{Y}-\boldsymbol{\mu}_{\mathrm{y}})^{\mathsf{T}}] & \mathrm{E}[(\mathbf{Y}-\boldsymbol{\mu}_{\mathrm{y}})(\mathbf{X}-\boldsymbol{\mu}_{\mathrm{x}})^{\mathsf{T}}] \\ \hline
\mathrm{E}[(\mathbf{X}-\boldsymbol{\mu}_{\mathrm{x}})(\mathbf{Y}-\boldsymbol{\mu}_{\mathrm{y}})^{\mathsf{T}}] & \mathrm{E}[(\mathbf{X}-\boldsymbol{\mu}_{\mathrm{x}})(\mathbf{X}-\boldsymbol{\mu}_{\mathrm{x}})^{\mathsf{T}}]
\end{array}\right) \\
&= \left(\begin{array}{c|c}
\boldsymbol{\Sigma}_{\mathrm{yy}} & \boldsymbol{\Sigma}_{\mathrm{yx}} \\ \hline
\boldsymbol{\Sigma}_{\mathrm{xy}} & \boldsymbol{\Sigma}_{\mathrm{xx}}
\end{array}\right)
\end{aligned}$$

- $\boldsymbol{\Sigma}_{\mathrm{yx}} = \boldsymbol{\Sigma}_{\mathrm{xy}}^{\mathsf{T}}$ e mede a associação entre os dois conjuntos de variáveis.
- As covariâncias em $\boldsymbol{\Sigma}_{\mathrm{yx}}$ são medidas em pares, i.e., temos $p \times q$ covariâncias.

Combinações Lineares

- A correlação canônica busca obter combinações lineares das variáveis que resumem "o melhor possível" a associação entre as variáveis.
- Sejam então, $\mathrm{U} = \mathbf{a}^{\top}\mathbf{Y}$ e $\mathrm{V} = \mathbf{b}^{\top}\mathbf{X}$. Então,

$$\begin{aligned}
\mathrm{Var}(\mathrm{U}) &= \mathbf{a}^{\top}\mathrm{Cov}(\mathbf{Y})\mathbf{a} = \mathbf{a}^{\top}\boldsymbol{\Sigma}_{\mathrm{yy}}\mathbf{a} \\
\mathrm{Var}(\mathrm{V}) &= \mathbf{b}^{\top}\mathrm{Cov}(\mathbf{X})\mathbf{b} = \mathbf{b}^{\top}\boldsymbol{\Sigma}_{\mathrm{xx}}\mathbf{b} \\
\mathrm{Cov}(\mathrm{U},\mathrm{V}) &= \mathbf{a}^{\top}\mathrm{Cov}(\mathbf{Y},\mathbf{X})\mathbf{b} = \mathbf{a}^{\top}\boldsymbol{\Sigma}_{\mathrm{yx}}\mathbf{b}
\end{aligned}$$

- Procuramos coeficientes **a** e **b** tais que

$$\mathrm{Cor}(\mathrm{U},\mathrm{V}) = \frac{\mathbf{a}^{\top}\boldsymbol{\Sigma}_{\mathrm{yx}}\mathbf{b}}{\sqrt{\mathbf{a}^{\top}\boldsymbol{\Sigma}_{\mathrm{yy}}\mathbf{a}}\sqrt{\mathbf{b}^{\top}\boldsymbol{\Sigma}_{\mathrm{xx}}\mathbf{b}}}$$

  seja o maior possível.
- O primeiro par de variáveis canonicas U_1, V_1, tem variânica unitária e maximiza a correlação.

- O segundo par de variáveis canonicas U_2, V_2, tem variânica unitária e maximiza a correlação entre todas as possíveis escolhas que são não correlacionadas com o primeiro par de variáveis canônicas.
- O processo continua até obtermos p correlações de variáveis canônicas.

Resultado 1: Suponha $\Sigma_{(p+q)\times(p+q)}$ como descrito anteriormente e de rank completo. Seja ainda $p \leq q$. Então

$$\max_{\mathbf{a},\mathbf{b}} \ \mathrm{Cor}(\mathrm{U},\mathrm{V}) = \rho_1$$

obtido pelo primeiro par de variáveis canônicas

$$\mathrm{U}_1 = \mathbf{a}_1^\top \mathbf{Y} = \mathbf{e}_1^\top \mathbf{\Sigma}_{\mathrm{yy}}^{-1/2}\mathbf{Y} \quad \text{e} \quad \mathrm{V}_1 = \mathbf{b}_1^\top \mathbf{X} = \mathbf{f}_1^\top \mathbf{\Sigma}_{\mathrm{xx}}^{-1/2}\mathbf{X}$$

e ρ_k obtido pelo primeiro par de variáveis canônicas

$$\mathrm{U}_k = \mathbf{a}_k^\top \mathbf{Y} = \mathbf{e}_k^\top \mathbf{\Sigma}_{\mathrm{yy}}^{-1/2}\mathbf{Y} \quad \text{e} \quad \mathrm{V}_k = \mathbf{b}_k^\top \mathbf{X} = \mathbf{f}_k^\top \mathbf{\Sigma}_{\mathrm{xx}}^{-1/2}\mathbf{X}$$

entre as combinações lineares não correlacionadas com os pares $1, 2, \ldots, k-1$ de variáveis canônicas.

Resultado 1: (continuação)

As correlações ρ_1^2 e $\mathbf{e}_1$ são o maior autovalor e o respectivo autovetor de

$$\boldsymbol{\Sigma}_{\mathrm{yy}}^{-1/2}\boldsymbol{\Sigma}_{\mathrm{yx}}\boldsymbol{\Sigma}_{\mathrm{xx}}^{-1}\boldsymbol{\Sigma}_{\mathrm{xy}}\boldsymbol{\Sigma}_{\mathrm{yy}}^{-1/2}$$

e as correlações ρ_1^2 e $\mathbf{f}_1$ são o maior autovalor e o respectivo autovetor de

$$\boldsymbol{\Sigma}_{\mathrm{xx}}^{-1/2}\boldsymbol{\Sigma}_{\mathrm{xy}}\boldsymbol{\Sigma}_{\mathrm{yy}}^{-1}\boldsymbol{\Sigma}_{\mathrm{yx}}\boldsymbol{\Sigma}_{\mathrm{xx}}^{-1/2}.$$

Generalizando,

- $\rho_1^2 \geq \rho_2^2 \geq \ldots \geq \rho_p^2$ são autovalores de

$$\boldsymbol{\Sigma}_{\mathrm{yy}}^{-1/2}\boldsymbol{\Sigma}_{\mathrm{yx}}\boldsymbol{\Sigma}_{\mathrm{xx}}^{-1}\boldsymbol{\Sigma}_{\mathrm{xy}}\boldsymbol{\Sigma}_{\mathrm{yy}}^{-1/2}$$

- $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_p$ são os autovetores correspondentes $(p \times 1)$.
- $\rho_1^2 \geq \rho_2^2 \geq \ldots \geq \rho_p^2$ são também autovalores de

$$\boldsymbol{\Sigma}_{\mathrm{xx}}^{-1/2}\boldsymbol{\Sigma}_{\mathrm{xy}}\boldsymbol{\Sigma}_{\mathrm{yy}}^{-1}\boldsymbol{\Sigma}_{\mathrm{yx}}\boldsymbol{\Sigma}_{\mathrm{xx}}^{-1/2}.$$

- $\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_p$ são os autovetores correspondentes $(q \times 1)$.
- $\mathbf{f}_i$ é proporcional a

$$\boldsymbol{\Sigma}_{\mathrm{xx}}^{-1/2}\boldsymbol{\Sigma}_{\mathrm{xy}}\boldsymbol{\Sigma}_{\mathrm{yy}}^{-1}\mathbf{e}_i$$

Se as variáveis são padronizadas,

$$\mathrm{Cor}(\mathrm{U}_k^*, \mathrm{V}_k^*) = \frac{\mathbf{a}_k^{*\mathsf{T}} \boldsymbol{\rho}_{\mathrm{yx}} \mathbf{b}_k^*}{\sqrt{\mathbf{a}_k^{*\mathsf{T}} \boldsymbol{\rho}_{\mathrm{yy}} \mathbf{a}_k^*} \sqrt{\mathbf{b}_k^{*\mathsf{T}} \boldsymbol{\rho}_{\mathrm{xx}} \mathbf{b}_k^*}} = \rho_k$$

- **Resultado 2:** As correlações canônicas são invariantes a padronização dos dados.
- $\rho_1^2 \geq \rho_2^2 \geq \ldots \geq \rho_p^2$ são autovalores de

$$\boldsymbol{\rho}_{\mathrm{yy}}^{-1/2} \boldsymbol{\rho}_{\mathrm{yx}} \boldsymbol{\rho}_{\mathrm{xx}}^{-1} \boldsymbol{\rho}_{\mathrm{xy}} \boldsymbol{\rho}_{\mathrm{yy}}^{-1/2}$$

- $\mathbf{e}_1^*, \mathbf{e}_2^*, \ldots, \mathbf{e}_p^*$ são os autovetores correspondentes $(p \times 1)$.
- $\rho_1^2 \geq \rho_2^2 \geq \ldots \geq \rho_p^2$ são também autovalores de

$$\boldsymbol{\rho}_{\mathrm{xx}}^{-1/2} \boldsymbol{\rho}_{\mathrm{xy}} \boldsymbol{\rho}_{\mathrm{yy}}^{-1} \boldsymbol{\rho}_{\mathrm{yx}} \boldsymbol{\rho}_{\mathrm{xx}}^{-1/2}.$$

- $\mathbf{f}_1^*, \mathbf{f}_2^*, \ldots, \mathbf{f}_p^*$ são os autovetores correspondentes $(q \times 1)$.

- Os coeficientes canônicos, em geral, não tem interpretação física.
- Os coeficientes canônicos das variáveis originais (U, V) tem unidades de medida proporcionais àquelas das variàveis originais nos dois grupos.
- Os coeficientes canônicos das variáveis padronizadas $(\mathrm{U}^*, \mathrm{V}^*)$ não tem unidades de medida e devem ser interpretados em função das variáveis padronizadas (média zero e variância unitária).
- Os coeficientes canônicos das variáveis padronizadas podem ser obtidos diretamente dos coeficientes das variáveis originais.
- O coeficiente canônico das variáveis originais e das variáveis padronizadas é o mesmo, i.e., o coeficiente de correlação canônico é invariante sob padronização das variáveis.

Interpretação Geométrica

$$\mathrm{U}_1 = \mathbf{a}_1^\mathsf{T}\mathbf{Y} = \mathbf{e}_1^\mathsf{T}\boldsymbol{\Sigma}_{\mathrm{yy}}^{-1/2}\mathbf{Y} = \mathbf{e}_1^\mathsf{T}\mathbf{P}_1\underbrace{\mathbf{D}^{-1/2}\overbrace{\mathbf{P}_1\mathbf{Y}}^{\text{CP de } \mathbf{Y}}}_{\text{Fator Comum de } \mathbf{Y}}$$

- A variável canônica U_1 resulta de uma rotação ortogonal (via $\mathbf{P}_1$ e determinada por $\boldsymbol{\Sigma}_{\mathrm{yy}}$) do CP padronizado (Fator Comum de $\mathbf{Y}$) seguida por outra rotação ortogonal (via $\mathbf{e}_1$ e determinada por $\boldsymbol{\Sigma}_{\mathrm{yy}}^{-1/2}\boldsymbol{\Sigma}_{\mathrm{yx}}\boldsymbol{\Sigma}_{\mathrm{xx}}^{-1}\boldsymbol{\Sigma}_{\mathrm{xy}}\boldsymbol{\Sigma}_{\mathrm{yy}}^{-1/2}$).

- Se λ é autovalor de $\boldsymbol{\Sigma}_{\mathrm{yy}}^{-1/2}\boldsymbol{\Sigma}_{\mathrm{yx}}\boldsymbol{\Sigma}_{\mathrm{xx}}^{-1}\boldsymbol{\Sigma}_{\mathrm{xy}}\boldsymbol{\Sigma}_{\mathrm{yy}}^{-1/2}$ com $\mathbf{e}$ o correpondente autovetor, então λ é também autovalor de $\boldsymbol{\Sigma}_{\mathrm{yy}}^{-1/2}\boldsymbol{\Sigma}_{\mathrm{yx}}\boldsymbol{\Sigma}_{\mathrm{xx}}^{-1}\boldsymbol{\Sigma}_{\mathrm{xy}}$ com $\boldsymbol{\Sigma}_{\mathrm{yy}}^{-1/2}\mathbf{e}$ o correpondente autovetor.
- Assim, as variáveis e os coeficientes canônicos podem ser obtidos diretamente da decomposição em valores singulares de $\boldsymbol{\Sigma}_{\mathrm{yy}}^{-1/2}\boldsymbol{\Sigma}_{\mathrm{yx}}\boldsymbol{\Sigma}_{\mathrm{xx}}^{-1}\boldsymbol{\Sigma}_{\mathrm{xy}}$.

Exemplos de Johnson e Wichern - Capítulo 10:
`J&W-Exemplos-Cap10.R`

- Exemplo 10.1 - pág. 543 a 545.
- Exemplo 10.2 - pág. 546 a 547.
- Exemplo 10.3 - pág. 549.
- Exemplo 10.4 - pág. 552 e 553.
- Exemplo 10.5 - pág. 553 a 555.