

Análise de Correlação Canônica

Prof. Caio Azevedo

Exemplo 1: (iris de Fisher)

- Temos quatro variáveis: duas relativas à características morfológicas da sépala e duas relativas à essas características da pétala. Portanto, temos dois grupos de variáveis.
- As variáveis apresentam correlações (aparentemente) significativas.

Exemplo 1: (iris de Fisher)

- Seria possível (interessante) representar cada grupo de variáveis por uma única (p.e., através de uma combinação linear), sendo que a correlação entre elas fosse máxima e tivessem interpretação em termos das variáveis originais?
- Caracterizar melhor (ou de forma mais simples) o conjunto de dados, possibilidade de análises mais simples, preservação da estrutura de correlação (ainda que parcialmente), reduzir dimensionalidade dos dados.

Ideia

- A análise de correlação canônica (ACC) busca quantificar a associação entre dois conjuntos de variáveis.
- ACC foca na correlação entre combinações lineares de cada um desses conjuntos.
- A ideia é primeiro determinar o par de combinações lineares que tenha a maior correlação.
- Depois, procura-se obter um segundo par que tenha a maior correlação entre todos os pares de combinações lineares que sejam não correlacionadas com o primeiro par, e assim por diante.
- Tais pares de combinações lineares são chamadas de variáveis canônicas e as respectivas correlações, de correlações canônicas.

Motivação

- Caracterizar melhor o conjunto de dados.
- Entender melhor comportamento entre as variáveis intra e entre grupos.
- Ao invés de usar um modelo de regressão linear multivariado, pode-se usar (p) modelos de regressão linear univariados em que p é o número de pares de variáveis canônicas.

ACC populacional

- Definemos os conjuntos de variáveis de interesse por $\mathbf{X}_{(p \times 1)}^{(1)}$ e $\mathbf{X}_{(q \times 1)}^{(2)}$ (por conveniência, vamos assumir que $p \leq q$, devido ao número máximo de variáveis canônicas que podem ser obtidas).
- Não faremos suposições acerca das distribuições desses vetores aleatórios, com exceção dos dois primeiros momentos de cada um.
- Defina $\mathcal{E}(\mathbf{X}^{(1)}) = \boldsymbol{\mu}^{(1)}$, $\mathcal{E}(\mathbf{X}^{(2)}) = \boldsymbol{\mu}^{(2)}$, $\text{Cov}(\mathbf{X}^{(1)}) = \boldsymbol{\Sigma}_{11}$, $\text{Cov}(\mathbf{X}^{(2)}) = \boldsymbol{\Sigma}_{22}$ e $\text{Cov}(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) = \boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}'_{21}$.
- Para facilitar os desenvolvimentos, vamos definir $\mathbf{X} = \begin{bmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{bmatrix}$

Cont.

■ Assim, $\mathbf{X} = \begin{bmatrix} X_1^{(1)} \\ X_2^{(1)} \\ \vdots \\ X_p^{(1)} \\ \dots \\ X_1^{(2)} \\ X_2^{(2)} \\ \vdots \\ X_q^{(2)} \end{bmatrix}$; $\mathcal{E}(\mathbf{X}) = \boldsymbol{\mu} = \begin{bmatrix} \mu_1^{(1)} \\ \mu_2^{(1)} \\ \vdots \\ \mu_p^{(1)} \\ \dots \\ \mu_1^{(2)} \\ \mu_2^{(2)} \\ \vdots \\ \mu_q^{(2)} \end{bmatrix}$

Cont.

- Também,

$$\begin{aligned}\Sigma &= \mathcal{E}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'] \\ &= \begin{bmatrix} \mathcal{E}[(\mathbf{X}^{(1)} - \boldsymbol{\mu}^{(1)})(\mathbf{X}^{(1)} - \boldsymbol{\mu}^{(1)})'] & \mathcal{E}[(\mathbf{X}^{(1)} - \boldsymbol{\mu}^{(1)})(\mathbf{X}^{(2)} - \boldsymbol{\mu}^{(2)})'] \\ \mathcal{E}[(\mathbf{X}^{(2)} - \boldsymbol{\mu}^{(2)})(\mathbf{X}^{(1)} - \boldsymbol{\mu}^{(1)})'] & \mathcal{E}[(\mathbf{X}^{(2)} - \boldsymbol{\mu}^{(2)})(\mathbf{X}^{(2)} - \boldsymbol{\mu}^{(2)})'] \end{bmatrix} \\ &= \begin{bmatrix} \Sigma_{11(p \times p)} & \Sigma_{12(p \times q)} \\ \Sigma_{21(q \times p)} & \Sigma_{22(q \times q)} \end{bmatrix}\end{aligned}$$

em que $\Sigma_{12} = \Sigma_{21}'$.

Cont.

- Podemos notar que a matriz Σ_{12} comporta a estrutura de dependência entre as variáveis dos dois conjuntos (entre grupos).
- Pode ser complicado tentar interpretar tal estrutura, principalmente se a quantidade de variáveis em cada grupo for grande.
- Além disso, muitas vezes, o interesse maior (em termos do problema a ser analisado) consiste em estudar combinações lineares das variáveis originais (por exemplo, para predição ou comparação).

Cont.

- Combinações lineares de interesse:

$$U = \mathbf{a}'\mathbf{X}^{(1)}; V = \mathbf{b}'\mathbf{X}^{(2)}$$

- Podemos provar que (exercício)

$$\mathcal{V}(U) = \mathbf{a}'\boldsymbol{\Sigma}_{11}\mathbf{a}; \mathcal{V}(V) = \mathbf{b}'\boldsymbol{\Sigma}_{22}\mathbf{b}; \text{Cov}(U, V) = \mathbf{a}'\boldsymbol{\Sigma}_{12}\mathbf{b}$$

- Nosso objetivo é encontrar coeficientes tais que:

$$\text{Corre}(U, V) = \frac{\mathbf{a}'\boldsymbol{\Sigma}_{12}\mathbf{b}}{\sqrt{\mathbf{a}'\boldsymbol{\Sigma}_{11}\mathbf{a}}\sqrt{\mathbf{b}'\boldsymbol{\Sigma}_{22}\mathbf{b}}} \quad (1)$$

seja máxima.

Cont.

- Para viabilizar o processo de obtenção das variáveis canônicas (veja também o processo de obtenção das componentes principais), impõe-se que $\mathbf{a}'\boldsymbol{\Sigma}_{11}\mathbf{a} = \mathbf{b}'\boldsymbol{\Sigma}_{22}\mathbf{b} = 1$ para todas as variáveis canônicas.
- Tal restrição não afeta a correlação entre as variáveis canônicas (transformações lineares não afetam a correlação de Pearson).
- Sem a aplicação de tal restrição poderíamos obter um número infinito de soluções.
- Além disso, a utilização da mesma escala para todas as variáveis canônicas permite interpretar os resultados e fazer comparações de interesse, de modo mais fácil e apropriado.

Define-se o seguinte

- O primeiro par de variáveis canônicas é o par de combinações lineares (U_1, V_1) , tendo variâncias unitárias, e que maximiza (1).
- O segundo par de variáveis canônicas é o par de combinações lineares (U_2, V_2) , tendo variâncias unitárias, e que maximiza (1), entre todas as escolhas que sejam não correlacionadas com o primeiro par de variáveis canônicas.

Define-se o seguinte

- (No passo k): O k -ésimo par de variáveis canônicas é o par de combinações lineares, tendo variâncias unitárias, e que maximiza (1), entre todas as escolhas que sejam não correlacionadas com as $k - 1$ variáveis canônicas anteriores.
- A correlação entre os elementos do k -ésimo par de variáveis canônicas é chamada de k -ésima correlação canônica.
- Definição: a raiz quadrada de uma matriz \mathbf{A} é uma matriz, digamos $\mathbf{A}^{1/2}$, tal que $\mathbf{A} = \mathbf{A}^{1/2} \mathbf{A}^{1/2}$. Uma forma de obtê-la é $\mathbf{A}^{1/2} = \mathbf{E} \mathbf{\Lambda}^{1/2} \mathbf{E}'$ (ou seja, usando uma função da respectiva decomposição espectral).

Resultado

- Define-se o seguinte

- Suponha que $p \leq q$ e sejam $\mathbf{X}^{(1)}$ e $\mathbf{X}^{(2)}$ os vetores aleatórios definidos anteriormente (em que $\mathbf{\Sigma}$ tem posto completo). Defina as combinações lineares $U = \mathbf{a}\mathbf{X}^{(1)}$ e $V = \mathbf{b}\mathbf{X}^{(2)}$, com $\mathbf{a}_{(p \times 1)}$ e $\mathbf{b}_{(q \times 1)}$.

- Então: $\max_{\mathbf{a}, \mathbf{b}} \text{Corre}(U, V) = \rho_1^*$ é obtido pelas combinações lineares (primeiro par de variáveis canônicas):

$$U_1 = \mathbf{e}'_1 \mathbf{\Sigma}_{11}^{-1/2} \mathbf{X}^{(1)}; V_1 = \mathbf{f}'_1 \mathbf{\Sigma}_{22}^{-1/2} \mathbf{X}^{(2)}.$$

- O k -ésimo par de variáveis canônicas, $k=2,3,\dots,p$,

$$U_k = \mathbf{e}'_k \mathbf{\Sigma}_{11}^{-1/2} \mathbf{X}^{(1)}; V_k = \mathbf{f}'_k \mathbf{\Sigma}_{22}^{-1/2} \mathbf{X}^{(2)}.$$

maximizam $\text{Corre}(U_k, V_k) = \rho_k^*$, entre todas as combinações lineares não correlacionadas com as variáveis canônicas anteriores: $1, 2, \dots, k-1$.

Cont.

- Temos que $\rho_1^{*2} \geq \rho_2^{*2} \geq \dots \geq \rho_p^{*2}$ são os autovalores de $\mathbf{E}_1 = \boldsymbol{\Sigma}_{11}^{-1/2} \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1/2}$, enquanto que $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_p$ são os respectivos autovetores.
- As quantidades $(\rho_1^{*2}, \rho_2^{*2}, \dots, \rho_p^{*2})$ também são os p maiores autovalores da matriz $\mathbf{E}_2 = \boldsymbol{\Sigma}_{22}^{-1/2} \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1/2}$ com $\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_p$ sendo os respectivos autovetores.
- Cada \mathbf{f}_i é proporcional à $\boldsymbol{\Sigma}_{22}^{-1/2} \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1/2} \mathbf{e}_i, i = 1, 2, \dots, p$.
- Seja $\mathbf{C} = \boldsymbol{\Sigma}_{11}^{-1/2} \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1/2}$. Note, então, que $\mathbf{C}\mathbf{C}' = \mathbf{E}_1$ e $\mathbf{C}'\mathbf{C} = \mathbf{E}_2$. Assim, os auto-valores não nulos de \mathbf{E}_1 e \mathbf{E}_2 , são iguais.

Propriedades das variáveis canônicas

- Para $k, l = 1, 2, \dots, p$, temos que

$$\left\{ \begin{array}{l} \mathcal{V}(U_k) = \mathcal{V}(V_k) = 1 \\ \text{Cov}(U_k, U_l) = \text{Corre}(U_k, U_l) = 0, k \neq l \\ \text{Cov}(V_k, V_l) = \text{Corre}(V_k, V_l) = 0, k \neq l \\ \text{Cov}(U_k, V_l) = \text{Corre}(U_k, V_l) = 0, k \neq l \end{array} \right.$$

Cont.

- Note que

$$\begin{aligned} \mathbf{A}^{-1/2} \mathbf{A} \mathbf{A}^{-1/2} &= (\mathbf{E} \mathbf{\Delta}^{1/2} \mathbf{E}')^{-1} \mathbf{E} \mathbf{\Delta} \mathbf{E}' (\mathbf{E} \mathbf{\Delta}^{1/2} \mathbf{E}')^{-1} \\ &= (\mathbf{E}')^{-1} \mathbf{\Delta}^{-1/2} \mathbf{E}^{-1} \mathbf{E} \mathbf{\Delta} \mathbf{E}' (\mathbf{E}')^{-1} \mathbf{\Delta}^{-1/2} \mathbf{E}^{-1} \\ &= (\mathbf{E}')^{-1} \mathbf{\Delta}^{-1/2} \mathbf{\Delta} \mathbf{\Delta}^{-1/2} \mathbf{E}^{-1} = (\mathbf{E}')^{-1} \mathbf{E}^{-1} \\ &= (\mathbf{E} \mathbf{E}')^{-1} = \mathbf{I} \end{aligned}$$

Cont.

- Por outro lado

$$\begin{aligned}\mathcal{V}(U_k) &= \mathcal{V}(\mathbf{e}'_k \boldsymbol{\Sigma}_{11}^{-1/2} \mathbf{X}^{(1)}) = \mathbf{e}'_k \boldsymbol{\Sigma}_{11}^{-1/2} \text{Cov}(\mathbf{X}^{(1)}) \boldsymbol{\Sigma}_{11}^{-1/2} \mathbf{e}_k \\ &= \mathbf{e}'_k \boldsymbol{\Sigma}_{11}^{-1/2} \boldsymbol{\Sigma}_{11} \boldsymbol{\Sigma}_{11}^{-1/2} \mathbf{e}_k = \mathbf{e}'_k \mathbf{e}_k = 1\end{aligned}$$

Cont.

- Além disso, temos que (lembrando que $\mathbf{f}_k \propto \boldsymbol{\Sigma}_{22}^{-1/2} \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1/2} \mathbf{e}_k$)

$$\begin{aligned} \text{Cov}(U_k, V_k) &= \text{Cov}(\mathbf{e}'_k \boldsymbol{\Sigma}_{11}^{-1/2} \mathbf{X}^{(1)}, \mathbf{f}'_k \boldsymbol{\Sigma}_{22}^{-1/2} \mathbf{X}^{(2)}) \\ &= \mathbf{e}'_k \boldsymbol{\Sigma}_{11}^{-1/2} \text{Cov}(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) \boldsymbol{\Sigma}_{22}^{-1/2} \mathbf{f}_k \\ &= \mathbf{e}'_k \boldsymbol{\Sigma}_{11}^{-1/2} \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1/2} \mathbf{f}_k \\ &\propto \mathbf{e}'_k \boldsymbol{\Sigma}_{11}^{-1/2} \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1/2} \boldsymbol{\Sigma}_{22}^{-1/2} \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1/2} \mathbf{e}_k \\ &= \mathbf{e}'_k \boldsymbol{\Sigma}_{11}^{-1/2} \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1/2} \mathbf{e}_k = \mathbf{e}'_k \mathbf{E} \boldsymbol{\Lambda} \mathbf{E}' \mathbf{e}_k \end{aligned}$$

Cont.

- Note que

$$\begin{aligned} \mathbf{e}'_k \mathbf{E} &= \mathbf{e}'_k [\mathbf{e}_1 \ \mathbf{e}_2 \ \dots \ \mathbf{e}_k \ \dots \ \mathbf{e}_p] \\ &= [\mathbf{e}'_k \mathbf{e}_1 \ \mathbf{e}'_k \mathbf{e}_2 \ \dots \ \mathbf{e}'_k \mathbf{e}_k \ \dots \ \mathbf{e}'_k \mathbf{e}_p] \\ &= [0 \ 0 \ \dots \ \underbrace{1}_{\text{k-ésima posição}} \ \dots \ 0] \end{aligned}$$

Cont.

- Portanto

$$\text{Cov}(U_k, V_k) \propto [0 \ 0 \ \dots \ 1 \ \dots \ 0] \begin{bmatrix} \rho_1^{*2} & 0 & \dots & 0 \\ 0 & \rho_2^{*2} & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & \rho_p^{*2} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \rho_k^{*2}$$

Cont.

- Além disso

$$\begin{aligned}\mathcal{V}(V_k) &= \mathcal{V}(\mathbf{f}'_k \boldsymbol{\Sigma}_{22}^{-1/2} \mathbf{X}^{(2)}) \propto \mathcal{V}(\mathbf{e}'_k \boldsymbol{\Sigma}_{11}^{-1/2} \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1/2} \boldsymbol{\Sigma}_{22}^{-1/2} \mathbf{X}^{(2)}) \\ &= \mathbf{e}'_k \boldsymbol{\Sigma}_{11}^{-1/2} \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \text{Cov}(\mathbf{X}^{(2)}) \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1/2} \mathbf{e}_k \\ &= \mathbf{e}'_k \boldsymbol{\Sigma}_{11}^{-1/2} \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{22} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1/2} \mathbf{e}_k \\ &= \mathbf{e}'_k \boldsymbol{\Sigma}_{11}^{-1/2} \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1/2} \mathbf{e}_k \\ &= \mathbf{e}'_k \mathbf{E} \boldsymbol{\Lambda} \mathbf{E}' \mathbf{e}_k = \rho_k^{*2} = \text{Cov}(U_k, V_k)\end{aligned}$$

- Assim, $\text{Corre}(U_k, V_k) = \frac{\rho_k^{*2}}{\rho_k^*} = \rho_k^*$.

Cont.

- Por outro lado (se $k \neq l$) (exercício),

$$\text{Cov}(U_k, V_l) = [0 \ 0 \ \dots \ \underbrace{1}_{k\text{-ésima posição}} \ \dots \ 0] \begin{bmatrix} \rho_1^{*2} & 0 & \dots & 0 \\ 0 & \rho_2^{*2} & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & \rho_p^{*2} \end{bmatrix}$$

$$\times \begin{bmatrix} 0 \\ \vdots \\ \underbrace{1}_{l\text{-ésima posição}} \\ \vdots \\ 0 \end{bmatrix} = 0$$

Propriedades das variáveis canônicas

■ Defina $\mathbf{A} = \begin{bmatrix} \mathbf{a}'_1 \\ \mathbf{a}'_2 \\ \vdots \\ \mathbf{a}'_p \end{bmatrix} = \begin{bmatrix} \mathbf{e}'_1 \boldsymbol{\Sigma}_{11}^{-1/2} \\ \mathbf{e}'_2 \boldsymbol{\Sigma}_{11}^{-1/2} \\ \vdots \\ \mathbf{e}'_p \boldsymbol{\Sigma}_{11}^{-1/2} \end{bmatrix}$ e

$$\mathbf{B} = \begin{bmatrix} \mathbf{b}'_1 \\ \mathbf{b}'_2 \\ \vdots \\ \mathbf{b}'_q \end{bmatrix} = \begin{bmatrix} \mathbf{f}'_1 \boldsymbol{\Sigma}_{22}^{-1/2} \\ \mathbf{f}'_2 \boldsymbol{\Sigma}_{22}^{-1/2} \\ \vdots \\ \mathbf{f}'_q \boldsymbol{\Sigma}_{22}^{-1/2} \end{bmatrix}.$$

Propriedades das variáveis canônicas

- Temos, assim, que $\mathbf{U} = \mathbf{A}\mathbf{X}^{(1)}$ e $\mathbf{V} = \mathbf{B}\mathbf{X}^{(2)}$.
- Pelos resultados anteriores, temos que (exercício): $\text{Cov}(\mathbf{U}) = \mathbf{I}_p$ e $\text{Cov}(\mathbf{V}) = \mathbf{I}_q$.
- Exercício: calcular de forma matricial $\text{Cov}(\mathbf{U}, \mathbf{V})$ e $\text{Corre}(\mathbf{U}, \mathbf{V})$.

Resultado

- Se usarmos as variáveis padronizadas

$$Z_i^{(k)} = \frac{X_i^{(k)} - \mu_i^{(k)}}{\sqrt{\sigma_{ii}^{(k)}}}, k = 1, 2; i = 1, 2, \dots, p \text{ (q)} \text{ (em que } \mathcal{E}(X_i^{(k)}) = \mu_i^{(k)}$$

e $\mathcal{V}(X_i^{(k)}) = \sigma_{ii}^{(k)}$), ou seja, se tivermos $\mathbf{Z}^{(1)} = (Z_1^{(1)}, Z_2^{(1)}, \dots, Z_p^{(1)})'$

e $\mathbf{Z}^{(2)} = (Z_1^{(2)}, Z_2^{(2)}, \dots, Z_q^{(2)})'$, os resultados anteriores continuam válidos.

- Contudo, teremos que $\text{Cov}(\mathbf{Z}^{(1)}) = \text{Corre}(\mathbf{X}^{(1)}) = \boldsymbol{\rho}_{11}$,

$$\text{Cov}(\mathbf{Z}^{(2)}) = \text{Corre}(\mathbf{X}^{(2)}) = \boldsymbol{\rho}_{22} \text{ e}$$

$$\text{Cov}(\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)}) = \text{Corre}(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) = \boldsymbol{\rho}_{12} = \boldsymbol{\rho}'_{21}.$$

- Então: $\max_{ab} \text{Corre}(U^{(Z)}, V^{(Z)}) = \rho_1^*$ é obtido pelas combinações lineares (primeiro par de variáveis canônicas):

$$U_1^{(Z)} = \mathbf{e}'_1 \boldsymbol{\rho}_{11}^{-1/2} \mathbf{Z}^{(1)}; V_1^{(Z)} = \mathbf{f}'_1 \boldsymbol{\rho}_{22}^{-1/2} \mathbf{Z}^{(2)}.$$

- O k -ésimo par de variáveis canônicas, $k=2,3,\dots,p$,

$$U_k^{(Z)} = \mathbf{e}'_k \boldsymbol{\rho}_{11}^{-1/2} \mathbf{Z}^{(1)}; V_k^{(Z)} = \mathbf{f}'_k \boldsymbol{\rho}_{22}^{-1/2} \mathbf{Z}^{(2)}.$$

maximizam $\text{Corre}(U_k^{(Z)}, V_k^{(Z)}) = \rho_k^*$, entre todas as combinações lineares não correlacionadas com as variáveis canônicas anteriores $1,2,\dots,k-1$, em que (ρ é matriz de correlações)

$$\boldsymbol{\rho} = \begin{bmatrix} \boldsymbol{\rho}_{11(p \times p)} & \boldsymbol{\rho}_{12(p \times q)} \\ \boldsymbol{\rho}_{21(q \times p)} & \boldsymbol{\rho}_{22(q \times q)} \end{bmatrix}$$

Cont.

- Temos que $\rho_1^{*2} \geq \rho_2^{*2} \geq \dots \geq \rho_p^{*2}$ são os autovalores de $\mathbf{E}_1^* = \rho_{11}^{-1/2} \rho_{12} \rho_{22}^{-1} \rho_{21} \rho_{11}^{-1/2}$, enquanto que $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_p$ são os respectivos autovetores.
- As quantidades $(\rho_1^{*2}, \rho_2^{*2}, \dots, \rho_p^{*2})$ também são os p maiores autovalores da matriz $\mathbf{E}_2^* = \rho_{22}^{-1/2} \rho_{21} \rho_{11}^{-1} \rho_{12} \rho_{22}^{-1/2}$ com $\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_p$ sendo os respectivos autovetores.
- Cada \mathbf{f}_i é proporcional à $\rho_{22}^{-1/2} \rho_{21} \rho_{11}^{-1/2} \mathbf{e}_i, i = 1, 2, \dots, p.$

Relação entre as variáveis canônicas obtidas via \mathbf{X} e \mathbf{Z}

- Note que $\text{Cov}(\mathbf{X}^{(1)} - \boldsymbol{\mu}^{(1)}) = \text{Cov}(\mathbf{X}^{(1)})$.
- Além disso, temos que

$$\begin{aligned} \mathbf{a}'_k (\mathbf{X}^{(1)} - \boldsymbol{\mu}^{(1)}) &= a_{k1}(X_1^{(1)} - \mu_1^{(1)}) + a_{k2}(X_2^{(1)} - \mu_2^{(1)}) + \\ &\dots a_{kp}(X_p^{(1)} - \mu_p^{(1)}) \\ &= a_{k1} \sqrt{\sigma_{11}^{(1)}} \frac{(X_1^{(1)} - \mu_1^{(1)})}{\sqrt{\sigma_{11}^{(1)}}} + a_{k2} \sqrt{\sigma_{22}^{(1)}} \frac{(X_2^{(1)} - \mu_2^{(1)})}{\sqrt{\sigma_{22}^{(1)}}} + \\ &\dots + a_{kp} \sqrt{\sigma_{pp}^{(1)}} \frac{(X_p^{(1)} - \mu_p^{(1)})}{\sqrt{\sigma_{pp}^{(1)}}} \\ &= a_{k1} \sqrt{\sigma_{11}^{(1)}} Z_1^{(1)} + a_{k2} \sqrt{\sigma_{22}^{(1)}} Z_2^{(1)} + \\ &\dots + a_{kp} \sqrt{\sigma_{pp}^{(1)}} Z_p^{(1)} \end{aligned}$$

Cont.

- Assim, os coeficientes que determinam as variáveis canônicas a partir de \mathbf{X} e de \mathbf{Z} estão relacionados.
- Seja \mathbf{a}'_k é o vetor de coeficientes associadas às variáveis canônicas obtidas para \mathbf{X} , relativas à $\mathbf{X}^{(1)}$.
- Portanto, os coeficientes análogos para \mathbf{Z} , relativos à $\mathbf{Z}^{(1)}$, são dados por $\mathbf{a}^{*'}_k = \mathbf{a}'_k \mathbf{V}_{11}^{1/2}$, em que $\mathbf{V}_{11}^{1/2}$ é uma matriz diagonal com o i -ésimo elemento da diagonal igual à $\sigma_{ii}^{(1)}$.

Cont.

- Seja \mathbf{b}'_k é o vetor de coeficientes associadas às variáveis canônicas obtidas para \mathbf{X} , relativas à $\mathbf{X}^{(2)}$.
- Analogamente para $\mathbf{X}^{(2)}$ e $\mathbf{Z}^{(2)}$, temos que: $\mathbf{b}^{*'}_k = \mathbf{b}_k \mathbf{V}_{22}^{1/2}$, em que $\mathbf{V}_{22}^{1/2}$ é uma matriz diagonal com o i -ésimo elemento da diagonal igual à $\sigma_{ii}^{(2)}$.
- As correlações canônicas $(\rho_1^*, \rho_2^*, \dots, \rho_p^*)$ são as mesmas (utilizando-se \mathbf{X} e \mathbf{Z}).

Características

- Lembremos que: $\mathbf{U}_{(p \times 1)} = \mathbf{A}\mathbf{X}^{(1)}$ e $\mathbf{V}_{(q \times 1)} = \mathbf{B}\mathbf{X}^{(2)}$.
- Temos que $\text{Cov}(\mathbf{U}, \mathbf{X}^{(1)}) = \text{Cov}(\mathbf{A}\mathbf{X}^{(1)}, \mathbf{X}^{(1)}) = \mathbf{A}\boldsymbol{\Sigma}_{11}$.
- Analogamente, temos que
$$\text{Cov}(\mathbf{V}, \mathbf{X}^{(2)}) = \text{Cov}(\mathbf{B}\mathbf{X}^{(2)}, \mathbf{X}^{(2)}) = \mathbf{B}\boldsymbol{\Sigma}_{22}.$$
- Além disso, $\text{Corre}(\mathbf{U}, \mathbf{X}^{(1)}) = \text{Cov}(\mathbf{U}, \mathbf{V}_{11}^{-1/2} \mathbf{X}^{(1)}) =$
$$\text{Cov}(\mathbf{A}\mathbf{X}^{(1)}, \mathbf{V}_{11}^{-1/2} \mathbf{X}^{(1)}) = \mathbf{A}\text{Cov}(\mathbf{X}^{(1)}, \mathbf{X}^{(1)})\mathbf{V}_{11}^{-1/2} = \mathbf{A}\boldsymbol{\Sigma}_{11}\mathbf{V}_{11}^{-1/2}.$$
- Analogamente, $\text{Corre}(\mathbf{V}, \mathbf{X}^{(2)}) = \text{Cov}(\mathbf{V}, \mathbf{V}_{22}^{-1/2} \mathbf{X}^{(2)}) =$
$$\text{Cov}(\mathbf{B}\mathbf{X}^{(2)}, \mathbf{V}_{22}^{-1/2} \mathbf{X}^{(2)}) = \mathbf{B}\text{Cov}(\mathbf{X}^{(2)}, \mathbf{X}^{(2)})\mathbf{V}_{22}^{-1/2} = \mathbf{B}\boldsymbol{\Sigma}_{22}\mathbf{V}_{22}^{-1/2}.$$

Características

- Pode-se provar ainda que (exercício): $Cov(\mathbf{V}, \mathbf{X}^{(1)}) = \mathbf{B}\Sigma_{21}$,
 $Cov(\mathbf{U}, \mathbf{X}^{(2)}) = \mathbf{A}\Sigma_{12}$, $Corre(\mathbf{V}, \mathbf{X}^{(1)}) = \mathbf{B}\Sigma_{21}\mathbf{V}_{11}^{-1/2}$,
 $Corre(\mathbf{U}, \mathbf{X}^{(2)}) = \mathbf{A}\Sigma_{12}\mathbf{V}_{22}^{-1/2}$.
- Defina $\mathbf{U}^{(Z)} = \mathbf{A}_Z\mathbf{Z}^{(1)}$ e $\mathbf{V}^{(Z)} = \mathbf{B}_Z\mathbf{Z}^{(2)}$, em que \mathbf{A}_Z e \mathbf{B}_Z são matrizes cujas linhas são os coeficientes que determinam as variáveis canônicas para $\mathbf{Z}^{(1)}$ e $\mathbf{Z}^{(2)}$, respectivamente.
- Como $Cov(\mathbf{Z}^{(1)}) = Corre(\mathbf{Z}^{(1)}) = \rho_{11}$ e $Cov(\mathbf{Z}^{(2)}) = Corre(\mathbf{Z}^{(2)}) = \rho_{22}$, então
 $Cov(\mathbf{U}^{(Z)}, \mathbf{Z}^{(1)}) = Corre(\mathbf{U}^{(Z)}, \mathbf{Z}^{(1)})$, $Cov(\mathbf{U}^{(Z)}, \mathbf{Z}^{(2)}) =$
 $Corre(\mathbf{U}^{(Z)}, \mathbf{Z}^{(2)})$, $Cov(\mathbf{V}^{(Z)}, \mathbf{Z}^{(1)}) = Corre(\mathbf{V}^{(Z)}, \mathbf{Z}^{(1)})$ e
 $Cov(\mathbf{V}^{(Z)}, \mathbf{Z}^{(2)}) = Corre(\mathbf{V}^{(Z)}, \mathbf{Z}^{(2)})$.

Características

- Além disso (exercício): $Corre(\mathbf{U}^{(Z)}, \mathbf{Z}^{(1)}) = \mathbf{A}_Z \boldsymbol{\rho}_{11}$,
 $Corre(\mathbf{U}^{(Z)}, \mathbf{Z}^{(2)}) = \mathbf{A}_Z \boldsymbol{\rho}_{12}$, $Corre(\mathbf{V}^{(Z)}, \mathbf{Z}^{(2)}) = \mathbf{B}_Z \boldsymbol{\rho}_{22}$ e
 $Corre(\mathbf{V}^{(Z)}, \mathbf{Z}^{(1)}) = \mathbf{B}_Z \boldsymbol{\rho}_{21}$.
- Finalmente, podemos provar que (exercício):
 $Corre(\mathbf{U}^{(Z)}, \mathbf{Z}^{(1)}) = Corre(\mathbf{U}, \mathbf{X}^{(1)})$,
 $Corre(\mathbf{U}^{(Z)}, \mathbf{Z}^{(2)}) = Corre(\mathbf{U}, \mathbf{X}^{(2)})$,
 $Corre(\mathbf{V}^{(Z)}, \mathbf{Z}^{(2)}) = Corre(\mathbf{V}, \mathbf{X}^{(2)})$ e
 $Corre(\mathbf{U}^{(Z)}, \mathbf{Z}^{(1)}) = Corre(\mathbf{V}, \mathbf{X}^{(1)})$.

Características

- Resumindo:

$$\left\{ \begin{array}{l} \text{Corre}(\mathbf{U}^{(Z)}, \mathbf{Z}^{(1)}) = \text{Corre}(\mathbf{U}, \mathbf{X}^{(1)}) = \mathbf{A}\boldsymbol{\Sigma}_{11} \mathbf{V}_{11}^{-1/2} = \mathbf{A}_Z \boldsymbol{\rho}_{11}, \\ \text{Corre}(\mathbf{U}^{(Z)}, \mathbf{Z}^{(2)}) = \text{Corre}(\mathbf{U}, \mathbf{X}^{(2)}) = \mathbf{A}\boldsymbol{\Sigma}_{12} \mathbf{V}_{12}^{-1/2} = \mathbf{A}_Z \boldsymbol{\rho}_{12}, \\ \text{Corre}(\mathbf{V}^{(Z)}, \mathbf{Z}^{(2)}) = \text{Corre}(\mathbf{V}, \mathbf{X}^{(2)}) = \mathbf{B}\boldsymbol{\Sigma}_{22} \mathbf{V}_{22}^{-1/2} = \mathbf{B}_Z \boldsymbol{\rho}_{22}, \\ \text{Corre}(\mathbf{V}^{(Z)}, \mathbf{Z}^{(1)}) = \text{Corre}(\mathbf{V}, \mathbf{X}^{(1)}) = \mathbf{B}\boldsymbol{\Sigma}_{21} \mathbf{V}_{21}^{-1/2} = \mathbf{B}_Z \boldsymbol{\rho}_{21}. \end{array} \right.$$

- Note que $\mathbf{A}_Z = \mathbf{V}_{11}^{1/2} \mathbf{A}$ e $\mathbf{B}_Z = \mathbf{V}_{22}^{1/2} \mathbf{B}$.

Correlações canônicas como generalização de outros tipos de correlações

- Se cada um dos vetores, $\mathbf{X}^{(1)}$ e $\mathbf{X}^{(2)}$, tiver apenas uma única componente, temos que

$$|\text{Corre}(X_1^{(1)}, X_1^{(2)})| = |\text{Corre}(aX_1^{(1)}, bX_1^{(2)})| = \rho_1^*$$

ou seja, as “variáveis canônicas” $U = X_1^{(1)}$, $V = X_1^{(2)}$ possuem correlação ρ_1^* , em que a e b são os coeficientes definidores das variáveis canônicas.

Cont.

- Se cada um dos vetores $\mathbf{X}^{(1)}$ e $\mathbf{X}^{(2)}$, tiver mais de uma componente e fizermos $\mathbf{a}' = [0, \dots, 0, 1, 0, \dots, 0]$ (valor 1 na i -ésima posição) e $\mathbf{b}' = [0, \dots, 0, 1, 0, \dots, 0]$ (valor 1 na k -ésima posição), vem que:

$$\begin{aligned} |\text{Corre}(X_i^{(1)}, X_k^{(2)})| &= |\text{Corre}(\mathbf{a}'\mathbf{X}^{(1)}, \mathbf{b}'\mathbf{X}^{(2)})| \\ &\leq \max_{\mathbf{a}, \mathbf{b}} \text{Corre}(\mathbf{a}'\mathbf{X}^{(1)}, \mathbf{b}'\mathbf{X}^{(2)}), \end{aligned}$$

ou seja, a primeira correlação canônica é maior do que o valor absoluto do que qualquer entrada da matriz $\rho_{12} = \text{Corre}(\mathbf{X}^{(1)}, \mathbf{X}^{(2)})$

Cont.

- O coeficiente de correlação múltipla (CCM) ($\rho_{1(\mathbf{X}^{(2)})}^* = 1 - \frac{SQR}{SQT}$) (obtidos através do ajuste de um modelo de regressão linear de $X_1^{(1)}$ em função de $\mathbf{X}^{(2)}$) é um caso especial da correlação canônica, quando $\mathbf{X}^{(1)}$ tem um único elemento, digamos $X_1^{(1)}$ ($p = 1$).

Lembremos que

$$\rho_{1(\mathbf{X}^{(2)})} = \max_{\mathbf{b}} \text{Corre}(X_1^{(1)}, \mathbf{b}' \mathbf{X}^{(2)}) = \rho_1^*.$$

Note que o CCM não, necessariamente, será igual à ρ_1^* , em geral $CCM < \rho_1^*$. Quando $p > 1$, ρ_1^* é maior do que cada correlação múltipla de $X_i^{(1)}$ com $\mathbf{X}^{(2)}$ ou a correlação múltipla de $X_i^{(2)}$ com $\mathbf{X}^{(1)}$.

Cont.

- Finalmente, note que

$$\rho_{U_k(\mathbf{X}^{(2)})} = \max_{\mathbf{b}} \text{Corre}(U_k, \mathbf{b}' \mathbf{X}^{(2)}) = \text{Corre}(U_k, V_k) = \rho_k^*$$

$$\rho_{V_k(\mathbf{X}^{(1)})} = \max_{\mathbf{a}} \text{Corre}(\mathbf{a}' \mathbf{X}^{(1)}, V_k) = \text{Corre}(U_k, V_k) = \rho_k^*$$

- Ou seja, as correlações canônicas também são os coeficientes de correlação múltipla de U_k com $\mathbf{X}^{(2)}$ bem como os coeficientes de correlação múltipla de V_k com $\mathbf{X}^{(1)}$.
- Devido à esse resultados, temos que ρ_k é a proporção da variância da variável canônica $U_k(V_k)$ “explicada” pelo vetor $\mathbf{X}^{(2)}(\mathbf{X}^{(1)})$.

Determinação do número de pares de variáveis canônicas

- Em geral, trabalha-se com as variáveis canônicas obtidas a partir das variáveis padronizadas: interpretação, menor influência das médias, menor influência das variâncias.
- Interpretação: dependendo do objetivo, as variáveis canônicas retidas têm de ter interpretação em termos do problema.
- Correlações canônicas: o interesse reside em pares de variáveis canônicas que tenham correlação “significativa” (ρ_k^*).
- Estrutura de covariância: as variáveis canônicas tem de preservar, de modo satisfatório, uma parte da variabilidade dos dados originais.

Cont.

- Em relação à este último item, note que

$\mathbf{U} = \mathbf{A}\mathbf{X}^{(1)} \rightarrow \mathbf{X}^{(1)} = \mathbf{A}^{-1}\mathbf{U}$. Logo (multiplicação de matrizes blocos),

$$\begin{aligned}\text{Cov}(\mathbf{X}^{(1)}) &= \mathbf{A}^{-1}\text{Cov}(\mathbf{U})(\mathbf{A}^{-1})' = \mathbf{A}^{-1}\mathbf{I}_p(\mathbf{A}^{-1})' = \mathbf{A}^{-1}(\mathbf{A}^{-1})' \\ &= \mathbf{a}^{(1)}(\mathbf{a}^{(1)})' + \mathbf{a}^{(2)}(\mathbf{a}^{(2)})' + \dots + \mathbf{a}^{(p)}(\mathbf{a}^{(p)})'\end{aligned}$$

- Analogamente,

$$\text{Cov}(\mathbf{X}^{(2)}) = \mathbf{B}^{-1}(\mathbf{B}^{-1})' = \mathbf{b}^{(1)}(\mathbf{b}^{(1)})' + \mathbf{b}^{(2)}(\mathbf{b}^{(2)})' + \dots + \mathbf{b}^{(q)}(\mathbf{b}^{(q)})'$$

- Neste caso estamos denotando as colunas das matrizes \mathbf{A}^{-1} e \mathbf{B}^{-1}



Cont.

- Logo, se considerarmos $r < p$ pares de variáveis canônicas, teremos:

$$\mathbf{X}^{(1)} \approx [\mathbf{a}^{(1)} | \mathbf{a}^{(2)} | \dots | \mathbf{a}^{(r)}] \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_r \end{bmatrix}; \mathbf{X}^{(2)} \approx [\mathbf{b}^{(1)} | \mathbf{b}^{(2)} | \dots | \mathbf{b}^{(r)}] \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_r \end{bmatrix}$$

- Portanto, vem que:

$$\begin{aligned} \text{Cov}(\mathbf{X}^{(1)}) &\approx \mathbf{a}^{(1)}(\mathbf{a}^{(1)})' + \mathbf{a}^{(2)}(\mathbf{a}^{(2)})' + \dots + \mathbf{a}^{(r)}(\mathbf{a}^{(r)})' = \mathbf{A}_z^{-1}(\mathbf{A}_z^{-1})' \\ \text{Cov}(\mathbf{X}^{(2)}) &\approx \mathbf{b}^{(1)}(\mathbf{b}^{(1)})' + \mathbf{b}^{(2)}(\mathbf{b}^{(2)})' + \dots + \mathbf{b}^{(r)}(\mathbf{b}^{(r)})' = \mathbf{B}_z^{-1}(\mathbf{B}_z^{-1})' \end{aligned} \quad (2)$$

Cont.

- Logo, um outro critério para auxiliar na escolha do número de variáveis canônicas é se

$$\boldsymbol{\Sigma}_{11} - \mathbf{A}_r^{-1} (\mathbf{A}'_r)^{-1} \approx \mathbf{0}_{(p \times p)}; \boldsymbol{\Sigma}_{22} - \mathbf{B}_r^{-1} (\mathbf{B}'_r)^{-1} \approx \mathbf{0}_{(q \times q)}$$

- Se trabalharmos com as variáveis padronizadas o procedimento é análogo mas agora, objetivamos ter:

$$\rho_{11} - (\mathbf{A}_r^{(Z)})^{-1} (\mathbf{A}_r^{(Z)'})^{-1} \approx \mathbf{0}_{(p \times p)}; \rho_{22} - (\mathbf{B}_r^{(Z)})^{-1} (\mathbf{B}_r^{(Z)'})^{-1} \approx \mathbf{0}_{(q \times q)}$$

- Em que $(\mathbf{A}_r^{(Z)})^{-1} (\mathbf{A}_r^{(Z)'})^{-1}$ e $(\mathbf{B}_r^{(Z)})^{-1} (\mathbf{B}_r^{(Z)'})^{-1}$ são análogas à (2) com \mathbf{A} e \mathbf{B} substituídas por $\mathbf{A}^{(Z)}$ e $\mathbf{B}^{(Z)}$, respectivamente.

Cont.

- Por simplicidade, também podemos considerar a soma das variâncias e as variâncias generalizadas, assim desejamos que:

$$tr(\boldsymbol{\Sigma}_{11}) \approx tr(\mathbf{A}_r^{-1} (\mathbf{A}'_r)^{-1}); tr(\boldsymbol{\Sigma}_{22}) \approx tr(\mathbf{B}_r^{-1} (\mathbf{B}'_r)^{-1})$$

$$det(\boldsymbol{\Sigma}_{11}) \approx det(\mathbf{A}_r^{-1} (\mathbf{A}'_r)^{-1}); det(\boldsymbol{\Sigma}_{22}) \approx det(\mathbf{B}_r^{-1} (\mathbf{B}'_r)^{-1})$$

$$tr(\boldsymbol{\rho}_{11}) \approx tr((\mathbf{A}_r^{(Z)})^{-1} (\mathbf{A}_r^{(Z)'})^{-1}); tr(\boldsymbol{\rho}_{22}) \approx tr((\mathbf{B}_r^{(Z)})^{-1} (\mathbf{B}_r^{(Z)'})^{-1})$$

$$det(\boldsymbol{\rho}_{11}) \approx det((\mathbf{A}_r^{(Z)})^{-1} (\mathbf{A}_r^{(Z)'})^{-1}); det(\boldsymbol{\rho}_{22}) \approx det((\mathbf{B}_r^{(Z)})^{-1} (\mathbf{B}_r^{(Z)'})^{-1})$$

Cont.

- Pode-se também quantificar a perda na representação de $\text{Cov}(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) = \boldsymbol{\Sigma}_{12}$, notando que:

$$\begin{aligned}\text{Cov}(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) &= \text{Cov}(\mathbf{A}^{-1}\mathbf{U}, \mathbf{B}^{-1}\mathbf{V}) = \mathbf{A}^{-1}\text{Cov}(\mathbf{U}, \mathbf{V})(\mathbf{B}')^{-1} \\ &= \mathbf{A}^{-1} \begin{bmatrix} \rho_1^* & 0 & \dots & 0 \\ 0 & \rho_2^* & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \rho_p^* \end{bmatrix} \mathbf{0}_{(p \times (q-p))} (\mathbf{B}^{-1})' \\ &= \rho_1^* \mathbf{a}^{(1)}(\mathbf{b}^{(1)})' + \rho_2^* \mathbf{a}^{(2)}(\mathbf{b}^{(2)})' + \dots + \rho_p^* \mathbf{a}^{(p)}(\mathbf{b}^{(p)})'\end{aligned}$$

Cont.

- Assim, analogamente aos desenvolvimentos anteriores, em se considerando somente r pares de variáveis canônicas, teríamos que

$$\text{Cov}(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) \approx \rho_1^* \mathbf{a}^{(1)}(\mathbf{b}^{(1)})' + \rho_2^* \mathbf{a}^{(2)}(\mathbf{b}^{(2)})' + \dots + \rho_r^* \mathbf{a}^{(r)}(\mathbf{b}^{(r)})'$$

- Em se trabalhando com as variáveis padronizadas, vem que

$$\text{Cov}(\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)}) \approx \rho_1^* \mathbf{a}_Z^{(1)}(\mathbf{b}_Z^{(1)})' + \rho_2^* \mathbf{a}_Z^{(2)}(\mathbf{b}_Z^{(2)})' + \dots + \rho_r^* \mathbf{a}_Z^{(r)}(\mathbf{b}_Z^{(r)})'$$

em que $\mathbf{a}_Z^{(\cdot)}$ e $\mathbf{b}_Z^{(\cdot)}$ são análogas à $\mathbf{a}^{(\cdot)}$ e $\mathbf{b}^{(\cdot)}$ e obtidas através de $\mathbf{A}^{(Z)}$ e $\mathbf{B}^{(Z)}$.

■ Equações:

$$U_1^{(Z)} = a_{11}^{(Z)} Z_1^{(1)} + a_{12}^{(Z)} Z_2^{(1)} + \dots + a_{1p}^{(Z)} Z_p^{(1)}$$

$$U_2^{(Z)} = a_{21}^{(Z)} Z_1^{(1)} + a_{22}^{(Z)} Z_2^{(1)} + \dots + a_{2p}^{(Z)} Z_p^{(1)}$$

⋮

$$U_p^{(Z)} = a_{p1}^{(Z)} Z_1^{(1)} + a_{p2}^{(Z)} Z_2^{(1)} + \dots + a_{pp}^{(Z)} Z_p^{(1)}$$

$$V_1^{(Z)} = b_{11}^{(Z)} Z_1^{(2)} + b_{12}^{(Z)} Z_2^{(2)} + \dots + b_{1q}^{(Z)} Z_q^{(2)}$$

$$V_2^{(Z)} = b_{21}^{(Z)} Z_1^{(2)} + b_{22}^{(Z)} Z_2^{(2)} + \dots + b_{2q}^{(Z)} Z_q^{(2)}$$

⋮

$$V_q^{(Z)} = b_{q1}^{(Z)} Z_1^{(2)} + b_{q2}^{(Z)} Z_2^{(2)} + \dots + b_{qq}^{(Z)} Z_q^{(2)}$$

Estimação das variáveis canônicas

- Dada uma matriz de dados (relativas à variáveis $\mathbf{X}^{(1)}$)

Indivíduo	Variável 1	Variável 2	...	Variável p
1	$X_{11}^{(1)}$	$X_{12}^{(1)}$...	$X_{1p}^{(1)}$
2	$X_{21}^{(1)}$	$X_{22}^{(1)}$...	$X_{2p}^{(1)}$
⋮	⋮	⋮	⋮	⋮
n	$X_{n1}^{(1)}$	$X_{n2}^{(1)}$...	$X_{np}^{(1)}$

Cont.

- Dada uma matriz de dados (relativas à variáveis $\mathbf{X}^{(2)}$)

Indivíduo	Variável 1	Variável 2	...	Variável q
1	$X_{11}^{(2)}$	$X_{12}^{(2)}$...	$X_{1q}^{(2)}$
2	$X_{21}^{(2)}$	$X_{22}^{(2)}$...	$X_{2q}^{(2)}$
\vdots	\vdots	\vdots	\ddots	\vdots
n	$X_{n1}^{(2)}$	$X_{n2}^{(2)}$...	$X_{nq}^{(2)}$

estima-se a matriz de covariâncias ($\tilde{\Sigma}$) ou a matriz de correlações amostrais ($\tilde{\rho}$), conforme visto anteriormente e considera-se as submatrizes de interesse.

■ Equações ($i = 1, 2, \dots, n$):

$$\widehat{U}_{i1}^{(Z)} = \widehat{a}_{11}^{(Z)} Z_{i1}^{(1)} + \widehat{a}_{12}^{(Z)} Z_{i2}^{(1)} + \dots + \widehat{a}_{1p}^{(Z)} Z_{ip}^{(1)}$$

$$\widehat{U}_{i2}^{(Z)} = \widehat{a}_{21}^{(Z)} Z_{i1}^{(1)} + \widehat{a}_{22}^{(Z)} Z_{i2}^{(1)} + \dots + \widehat{a}_{2p}^{(Z)} Z_{ip}^{(1)}$$

⋮

$$\widehat{U}_{ip}^{(Z)} = \widehat{a}_{p1}^{(Z)} Z_{i1}^{(1)} + \widehat{a}_{p2}^{(Z)} Z_{i2}^{(1)} + \dots + \widehat{a}_{pp}^{(Z)} Z_{ip}^{(1)}$$

$$\widehat{V}_{i1}^{(Z)} = \widehat{b}_{11}^{(Z)} Z_{i1}^{(2)} + \widehat{b}_{12}^{(Z)} Z_{i2}^{(2)} + \dots + \widehat{b}_{1q}^{(Z)} Z_{iq}^{(2)}$$

$$\widehat{V}_{i2}^{(Z)} = \widehat{b}_{21}^{(Z)} Z_{i1}^{(2)} + \widehat{b}_{22}^{(Z)} Z_{i2}^{(2)} + \dots + \widehat{b}_{2q}^{(Z)} Z_{iq}^{(2)}$$

⋮

$$\widehat{V}_{iq}^{(Z)} = \widehat{b}_{q1}^{(Z)} Z_{i1}^{(2)} + \widehat{b}_{q2}^{(Z)} Z_{i2}^{(2)} + \dots + \widehat{b}_{qq}^{(Z)} Z_{iq}^{(2)}$$