Análise de Correlação Canônica: parte 1

Prof. Caio Azevedo

Motivação

- Como mencionado diversas vezes (aqui, por exemplo), modelar/compreender/estudar a estrutura de dependência é um dos objetivos da análise multivariada.
- A redução de demensionalidade, muitas vezes, nos leva à uma melhor compreensão do problema (não somente em relação à estrutura de dependência, mas também a respeito de outros elementos, como: caracterização de grupos e de unidades experimentais/amostrais).

Motivação

- Vimos também que utilizar modelos de regressão multivariados pode nos levar à inferências mais precisas mas, ao mesmo tempo, também à necessidade de utilização de metodologias mais complexas, bem como a resultados mais difícieis de serem interpretados/obtidos.
- Em particular, com o (res)surgimento dos conceitos de "Ciências de Dados" e "Big Data", a simplificação da estrutura dos dados, bem como uma melhor compreensão dela, são elementos fundamentais. Ademais, alguma vezes, temos o interesse em estudar a relação entre grupos de variáveis, definidas de acordo com as características do problema e/ou do interesse de pesquisadores.

(voltando ao) 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.
- Temos dois "grupos naturais" de variáveis, sendo que um deles está ligado às pétalas (estrutura de reprodução) e outro as sépalas (estrutura de proteção).

Exemplo 1: (iris de Fisher)

- Seria possível (interessante) representar cada grupo (de interesse) 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, identificar estruturas de interesse etc.

Idéia

- 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 (entre grupos) de combinações lineares (intra grupos) que tenha a maior correlação.

Idéia

- 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 o comportamento/relacionameento entre as variáveis intra e entre grupos.
- Gerar subsídios (informações) sobre aspectos de interesse.
- Se, por exemplo, um desses grupo for de variáveis resposta e, o outro, de variáveis explicativas, pode-se usar (p) modelos de regressão univariados em que p é o número de pares de variáveis canônicas, ao invés de se usar um modelo de regressão multivariado.
- De uma forma mais geral, possibilitar a utilização de métodos univariados, ao invés de multivariados, com perda "mínima" de informação.

ACC populacional (vetores aleatórios)

- Definamos os conjuntos de variáveis de interesse por $\boldsymbol{X}^{(1)}_{(p\times 1)}$ e $\boldsymbol{X}^{(2)}_{(q\times 1)}$ (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 da existência dos dois primeiros momentos de cada um.
- Defina $\mathcal{E}(\boldsymbol{X}^{(1)}) = \boldsymbol{\mu}^{(1)}$, $\mathcal{E}(\boldsymbol{X}^{(2)}) = \boldsymbol{\mu}^{(2)}$, $Cov(\boldsymbol{X}^{(1)}) = \Sigma_{11}$, $Cov(\boldsymbol{X}^{(2)}) = \Sigma_{22}$ e $Cov(\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}) = \Sigma_{12} = \Sigma_{21}'$.
- Para facilitar os desenvolvimentos, seja: $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}^{(1)} \\ --- \\ \boldsymbol{X}^{(2)} \end{bmatrix}$

Cont.

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

Cont.

- Também,

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

em que $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}'_{21}$.

Cont.

- Podemos notar que a matriz $\mathbf{\Sigma}_{12}$ comporta a estrutura de dependência entre as variáveis dos dois conjuntos (entre grupos).
- Pode ser complicado 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, comparação, explicação de variabilidade, mensuração de influência).

Cont.

- Combinações lineares de interesse:

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

- Podemos provar que (exercício)

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

- Nosso objetivo é encontrar coeficientes $(\boldsymbol{a}', \boldsymbol{b}')'$ tais que:

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

seja máxima.

Cont.

- Para viabilizar o processo de obtenção das variáveis canônicas (veja também o processo de obtenção das componentes principais), impõe-se que $\boldsymbol{a}'\boldsymbol{\Sigma}_{11}\boldsymbol{a} = \boldsymbol{b}'\boldsymbol{\Sigma}_{22}\boldsymbol{b} = 1$ para todas as variáveis canônicas.
- Tal restrição não afeta a correlação entre as variáveis canônicas (transformações lineares não afetam a correlação de Pearson, exercício).
- Sem a aplicação de tal restrição poderíamos obter um número infinito de soluções (como na análise de componentes principais).
- 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.

Procedimento

- O primeiro par de variáveis canônicas é aquele correspondente a combinações lineares (U_1, V_1), tendo variâncias unitárias, e que maximiza (1).
- O segundo par de variáveis canônicas é aquele correspondente a 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.

Procedimento (cont.)

- (No passo k): O $k-$ésimo par de variáveis canônicas é aquele correspondente a 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 $\boldsymbol{A}$ é uma matriz, digamos $\boldsymbol{A}^{1/2}$, tal que $\boldsymbol{A} = \boldsymbol{A}^{1/2}\boldsymbol{A}^{1/2}$. Uma forma de obtê-la é $\boldsymbol{A}^{1/2} = \boldsymbol{E}\Lambda^{1/2}\boldsymbol{E}'$ (ou seja, usando uma função da respectiva decomposição espectral).

Resultado

- Descrição:
  - Suponha que $p \leq q$ e sejam $\boldsymbol{X}^{(1)}$ e $\boldsymbol{X}^{(2)}$ os vetores aleatórios definidos anteriormente (em que $\boldsymbol{\Sigma}$ tem posto completo). Defina as combinações lineares $U = \boldsymbol{a}'\boldsymbol{X}^{(1)}$ e $V = \boldsymbol{b}'\boldsymbol{X}^{(2)}$, com $\boldsymbol{a}_{(p\times 1)}$ e $\boldsymbol{b}_{(q\times 1)}$.
  - Então: $\max_{\boldsymbol{a},\boldsymbol{b}} Corre(U, V) = \rho_1^*$ é obtido pelas combinações lineares (primeiro par de variáveis canônicas):
$$U_1 = \boldsymbol{e}_1'\boldsymbol{\Sigma}_{11}^{-1/2}\boldsymbol{X}^{(1)}; \; V_1 = \boldsymbol{f}_1'\boldsymbol{\Sigma}_{22}^{-1/2}\boldsymbol{X}^{(2)}.$$
  - O $k-$ésimo par de variáveis canônicas, k=2,3,...,p,
$$U_k = \boldsymbol{e}_k'\boldsymbol{\Sigma}_{11}^{-1/2}\boldsymbol{X}^{(1)}; \; V_k = \boldsymbol{f}_k'\boldsymbol{\Sigma}_{22}^{-1/2}\boldsymbol{X}^{(2)}.$$

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

Cont.

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

Propriedades das variáveis canônicas

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

$$\begin{cases} \mathcal{V}(U_k) = \mathcal{V}(V_k) = 1 \\ Cov(U_k, U_l) = Corre(U_k, U_l) = 0, k \neq l \\ Cov(V_k, V_l) = Corre(V_k, V_l) = 0, k \neq l \\ Cov(U_k, V_l) = Corre(U_k, V_l) = 0, k \neq l \end{cases}$$

Cont.

- Note que

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

- Por outro lado

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

Cont.

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

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

Cont.

- Note que

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

Cont.

- Portanto

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

Cont.

- Além disso

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

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

Cont.

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

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

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

Propriedades das variáveis canônicas

- Defina

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}_1' \\ \boldsymbol{a}_2' \\ \vdots \\ \boldsymbol{a}_p' \end{bmatrix} = \begin{bmatrix} \boldsymbol{e}_1'\boldsymbol{\Sigma}_{11}^{-1/2} \\ \boldsymbol{e}_2'\boldsymbol{\Sigma}_{11}^{-1/2} \\ \vdots \\ \boldsymbol{e}_p'\boldsymbol{\Sigma}_{11}^{-1/2} \end{bmatrix} \text{ e } \boldsymbol{B} = \begin{bmatrix} \boldsymbol{b}_1' \\ \boldsymbol{b}_2' \\ \vdots \\ \boldsymbol{b}_q' \end{bmatrix} = \begin{bmatrix} \boldsymbol{f}_1'\boldsymbol{\Sigma}_{22}^{-1/2} \\ \boldsymbol{f}_2'\boldsymbol{\Sigma}_{22}^{-1/2} \\ \vdots \\ \boldsymbol{f}_q'\boldsymbol{\Sigma}_{22}^{-1/2} \end{bmatrix}.$$

Propriedades das variáveis canônicas

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

Resultado

- Se usarmos as variáveis padronizadas, ou seja, $Z_i^{(k)} = \frac{X_i^{(k)} - \mu_i^{(k)}}{\sqrt{\sigma_{ii}^{(k)}}}, k = 1, 2; i = 1, 2, ..., p\,(q)$ (em que $\mathcal{E}(X_i^{(k)}) = \mu_i^{(k)}$ e $\mathcal{V}(X_i^{(k)}) = \sigma_{ii}^{(k)}$), ou seja, se tivermos $\boldsymbol{Z}^{(1)} = (Z_1^{(1)}, Z_2^{(1)}, ..., Z_p^{(1)})'$ e $\boldsymbol{Z}^{(2)} = (Z_1^{(2)}, Z_2^{(2)}, ..., Z_q^{(2)})'$, os resultados anteriores continuam válidos.
- Contudo, neste caso, temos que $Cov(\boldsymbol{Z}^{(1)}) = Corre(\boldsymbol{X}^{(1)}) = \boldsymbol{\rho}_{11}$, $Cov(\boldsymbol{Z}^{(2)}) = Corre(\boldsymbol{X}^{(2)}) = \boldsymbol{\rho}_{22}$ e $Cov(\boldsymbol{Z}^{(1)}, \boldsymbol{Z}^{(2)}) = Corre(\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}) = \boldsymbol{\rho}_{12} = \boldsymbol{\rho}_{21}'$.

Cont.

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

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

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

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

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

$$\boldsymbol{\rho} = \left[\begin{array}{c|c} \boldsymbol{\rho}_{11(p\times p)} & \boldsymbol{\rho}_{12(p\times q)} \\ --- & --- \\ \boldsymbol{\rho}_{21(q\times p)} & \boldsymbol{\rho}_{22(q\times q)} \end{array}\right]$$

Cont.

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

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

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

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

Cont.

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

Cont.

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

Resumo das Características (ACC)

- Lembremos que: $\boldsymbol{U}_{(p\times 1)} = \boldsymbol{A}\boldsymbol{X}^{(1)}$ e $\boldsymbol{V}_{(q\times 1)} = \boldsymbol{B}\boldsymbol{X}^{(2)}$.
- Temos que $Cov(\boldsymbol{U}, \boldsymbol{X}^{(1)}) = Cov(\boldsymbol{A}\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(1)}) = \boldsymbol{A}\Sigma_{11}$.
- Analogamente, temos que

  $Cov(\boldsymbol{V}, \boldsymbol{X}^{(2)}) = Cov(\boldsymbol{B}\boldsymbol{X}^{(2)}, \boldsymbol{X}^{(2)}) = \boldsymbol{B}\Sigma_{22}$.
- Além disso,

$$\begin{aligned} Corre(\boldsymbol{U}, \boldsymbol{X}^{(1)}) &= Cov(\boldsymbol{U}, \boldsymbol{V}_{11}^{-1/2}\boldsymbol{X}^{(1)}) = Cov(\boldsymbol{A}\boldsymbol{X}^{(1)}, \boldsymbol{V}_{11}^{-1/2}\boldsymbol{X}^{(1)}) \\ &= \boldsymbol{A}Cov(\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(1)})\boldsymbol{V}_{11}^{-1/2} = \boldsymbol{A}\Sigma_{11}\boldsymbol{V}_{11}^{-1/2} \end{aligned}$$

- Analogamente,

$$\begin{aligned} Corre(\boldsymbol{V}, \boldsymbol{X}^{(2)}) &= Cov(\boldsymbol{V}, \boldsymbol{V}_{22}^{-1/2}\boldsymbol{X}^{(2)}) = Cov(\boldsymbol{B}\boldsymbol{X}^{(2)}, \boldsymbol{V}_{22}^{-1/2}\boldsymbol{X}^{(2)}) \\ &= \boldsymbol{B}Cov(\boldsymbol{X}^{(2)}, \boldsymbol{X}^{(2)})\boldsymbol{V}_{22}^{-1/2} = \boldsymbol{B}\Sigma_{22}\boldsymbol{V}_{22}^{-1/2} \end{aligned}$$

Resumo das Características (ACC)

- Pode-se provar ainda que (exercício):

$$
\begin{aligned}
Cov(\boldsymbol{V}, \boldsymbol{X}^{(1)}) &= \boldsymbol{B}\Sigma_{21}, Cov(\boldsymbol{U}, \boldsymbol{X}^{(2)}) = \boldsymbol{A}\Sigma_{12}, \\
Corre(\boldsymbol{V}, \boldsymbol{X}^{(1)}) &= \boldsymbol{B}\Sigma_{21}\boldsymbol{V}_{11}^{-1/2}, Corre(\boldsymbol{U}, \boldsymbol{X}^{(2)}) = \boldsymbol{A}\Sigma_{12}\boldsymbol{V}_{22}^{-1/2}
\end{aligned}
$$

- Defina $\boldsymbol{U}^{(Z)} = \boldsymbol{A}_Z\boldsymbol{Z}^{(1)}$ e $\boldsymbol{V}^{(Z)} = \boldsymbol{B}_Z\boldsymbol{Z}^{(2)}$, em que $\boldsymbol{A}_Z$ e $\boldsymbol{B}_Z$ são matrizes cujas linhas são os coeficientes que determinam as variáveis canônicas para $\boldsymbol{Z}^{(1)}$ e $\boldsymbol{Z}^{(2)}$, respectivamente.

Resumo das Características (ACC)

- Como

$$
\begin{aligned}
Cov(\mathbf{Z}^{(1)}) &= Corre(\mathbf{Z}^{(1)}) = \boldsymbol{\rho}_{11} \\
Cov(\mathbf{Z}^{(2)}) &= Corre(\mathbf{Z}^{(2)}) = \boldsymbol{\rho}_{22},
\end{aligned}
$$

então

$$
\begin{aligned}
Cov(\mathbf{U}^{(Z)}, \mathbf{Z}^{(1)}) &= Corre(\mathbf{U}^{(Z)}, \mathbf{Z}^{(1)}), \\
Cov(\mathbf{U}^{(Z)}, \mathbf{Z}^{(2)}) &= Corre(\mathbf{U}^{(Z)}, \mathbf{Z}^{(2)}), \\
Cov(\mathbf{V}^{(Z)}, \mathbf{Z}^{(1)}) &= Corre(\mathbf{V}^{(Z)}, \mathbf{Z}^{(1)}), \\
Cov(\mathbf{V}^{(Z)}, \mathbf{Z}^{(2)}) &= Corre(\mathbf{V}^{(Z)}, \mathbf{Z}^{(2)})
\end{aligned}
$$

Resumo das Características (ACC)

- Além disso (exercício):

$$
\begin{aligned}
Corre(\boldsymbol{U}^{(Z)}, \boldsymbol{Z}^{(1)}) &= \boldsymbol{A}_Z \boldsymbol{\rho}_{11}, \\
Corre(\boldsymbol{U}^{(Z)}, \boldsymbol{Z}^{(2)}) &= \boldsymbol{A}_Z \boldsymbol{\rho}_{12} \\
Corre(\boldsymbol{V}^{(Z)}, \boldsymbol{Z}^{(2)}) &= \boldsymbol{B}_Z \boldsymbol{\rho}_{22} \\
Corre(\boldsymbol{V}^{(Z)}, \boldsymbol{Z}^{(1)}) &= \boldsymbol{B}_Z \boldsymbol{\rho}_{21}
\end{aligned}
$$

- Finalmente, podemos provar que (exercício):

$$
\begin{aligned}
Corre(\boldsymbol{U}^{(Z)}, \boldsymbol{Z}^{(1)}) &= Corre(\boldsymbol{U}, \boldsymbol{X}^{(1)}), \\
Corre(\boldsymbol{U}^{(Z)}, \boldsymbol{Z}^{(2)}) &= Corre(\boldsymbol{U}, \boldsymbol{X}^{(2)}) \\
Corre(\boldsymbol{V}^{(Z)}, \boldsymbol{Z}^{(2)}) &= Corre(\boldsymbol{V}, \boldsymbol{X}^{(2)}) \\
Corre(\boldsymbol{V}^{(Z)}, \boldsymbol{Z}^{(1)}) &= Corre(\boldsymbol{V}, \boldsymbol{X}^{(1)})
\end{aligned}
$$

Resumo das Características (ACC)

- Resumindo:

$$\begin{cases} Corre(\boldsymbol{U}^{(Z)}, \boldsymbol{Z}^{(1)}) = Corre(\boldsymbol{U}, \boldsymbol{X}^{(1)}) = \boldsymbol{A}\Sigma_{11}\boldsymbol{V}_{11}^{-1/2} = \boldsymbol{A}_Z\boldsymbol{\rho}_{11}, \\ Corre(\boldsymbol{U}^{(Z)}, \boldsymbol{Z}^{(2)}) = Corre(\boldsymbol{U}, \boldsymbol{X}^{(2)}) = \boldsymbol{A}\Sigma_{12}\boldsymbol{V}_{12}^{-1/2} = \boldsymbol{A}_Z\boldsymbol{\rho}_{12}, \\ Corre(\boldsymbol{V}^{(Z)}, \boldsymbol{Z}^{(2)}) = Corre(\boldsymbol{V}, \boldsymbol{X}^{(2)}) = \boldsymbol{B}\Sigma_{22}\boldsymbol{V}_{22}^{-1/2} = \boldsymbol{B}_Z\boldsymbol{\rho}_{22}, \\ Corre(\boldsymbol{V}^{(Z)}, \boldsymbol{Z}^{(1)}) = Corre(\boldsymbol{V}, \boldsymbol{X}^{(1)}) = \boldsymbol{B}\Sigma_{21}\boldsymbol{V}_{21}^{-1/2} = \boldsymbol{B}_Z\boldsymbol{\rho}_{21}. \end{cases}$$

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

Correlações canônicas como generalização de outros tipos de correlações

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

$$|Corre(X_1^{(1)}, X_1^{(2)})| = |Corre(aX^{(1)}, bX^{(2)})| = \rho_1^*$$

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 $\boldsymbol{X}^{(1)}$ e $\boldsymbol{X}^{(2)}$, tiver mais de uma componente e fizermos $\boldsymbol{a}' = [0, ..., 0, 1, 0, ..., 0]$ (valor 1 na i-ésima posição) e $\boldsymbol{b}' = [0, ..., 0, 1, 0, ..., 0]$ (valor 1 na k-ésima posição), vem que:

$$|Corre(X_i^{(1)}, X_k^{(2)})| = |Corre(\boldsymbol{a}'\boldsymbol{X}^{(1)}, \boldsymbol{b}'\boldsymbol{X}^{(2)})|$$

$$\leq max_{\boldsymbol{a},\boldsymbol{b}} Corre(\boldsymbol{a}'\boldsymbol{X}^{(1)}, \boldsymbol{b}'\boldsymbol{X}^{(2)}),$$

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

Cont.

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

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

- 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 $\boldsymbol{X}^{(2)}$ ou a correlação múltipla de $X_i^{(2)}$ com $\boldsymbol{X}^{(1)}$.

Cont.

- Finalmente, note que

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

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

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

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 interpretações (apropriadas) 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, a maior parte possível da (estrutura de) variabilidade dos dados originais.

Cont.

- Em relação à este último item, note que $\boldsymbol{U} = \boldsymbol{A}\boldsymbol{X}^{(1)} \rightarrow \boldsymbol{X}^{(1)} = \boldsymbol{A}^{-1}\boldsymbol{U}$. Logo (multiplicação de matrizes blocos),

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

- Analogamente,

$$Cov(\boldsymbol{X}^{(2)}) = \boldsymbol{B}^{-1}(\boldsymbol{B}^{-1})' = \boldsymbol{b}^{(1)}(\boldsymbol{b}^{(1)})' + \boldsymbol{b}^{(2)}(\boldsymbol{b}^{(2)})' + ... + \boldsymbol{b}^{(q)}(\boldsymbol{b}^{(q)})'$$

- Neste caso estamos denotanto as colunas das matrizes $\boldsymbol{A}^{-1}$ e $\boldsymbol{B}^{-1}$ por $\boldsymbol{a}^{(\cdot)}$ e $\boldsymbol{b}^{(\cdot)}$, respectivamente.

Cont.

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

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

- Portanto, vem que:

$$\begin{aligned} Cov(\boldsymbol{X}^{(1)}) &\approx \boldsymbol{a}^{(1)}(\boldsymbol{a}^{(1)})' + \boldsymbol{a}^{(2)}(\boldsymbol{a}^{(2)})' + \ldots + \boldsymbol{a}^{(r)}(\boldsymbol{a}^{(r)})' = \boldsymbol{A}_z^{-1}(\boldsymbol{A}_z^{-1})' \\ Cov(\boldsymbol{X}^{(2)}) &\approx \boldsymbol{b}^{(1)}(\boldsymbol{b}^{(1)})' + \boldsymbol{b}^{(2)}(\boldsymbol{b}^{(2)})' + \ldots + \boldsymbol{b}^{(r)}(\boldsymbol{b}^{(r)})' = \boldsymbol{B}_z^{-1}(\boldsymbol{B}_z^{-1})' \end{aligned} \tag{2}$$

Cont.

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

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

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

$$\boldsymbol{\rho}_{11} - (\boldsymbol{A}_r^{(Z)})^{-1}\left(\boldsymbol{A}_r^{(Z)'}\right)^{-1} \approx \boldsymbol{0}_{(p\times p)}; \boldsymbol{\rho}_{22} - (\boldsymbol{B}_r^{(Z)})^{-1}\left(\boldsymbol{B}_r^{(Z)'}\right)^{-1} \approx \boldsymbol{0}_{(q\times q)}$$

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

Cont.

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

$$
\begin{aligned}
tr(\boldsymbol{\Sigma}_{11}) &\approx tr(\boldsymbol{A}_r^{-1}(\boldsymbol{A}_r')^{-1}) \\
tr(\boldsymbol{\Sigma}_{22}) &\approx tr(\boldsymbol{B}_r^{-1}(\boldsymbol{B}_r')^{-1}) \\
det(\boldsymbol{\Sigma}_{11}) &\approx det(\boldsymbol{A}_r^{-1}(\boldsymbol{A}_r')^{-1}) \\
det(\boldsymbol{\Sigma}_{22}) &\approx det(\boldsymbol{B}_r^{-1}(\boldsymbol{B}_r')^{-1}) \\
tr(\boldsymbol{\rho}_{11}) &\approx tr((\boldsymbol{A}_r^{(Z)})^{-1}(\boldsymbol{A}_r^{(Z)'})^{-1}) \\
tr(\boldsymbol{\rho}_{22}) &\approx tr((\boldsymbol{B}_r^{(Z)})^{-1}(\boldsymbol{B}_r^{(Z)'})^{-1}) \\
det(\boldsymbol{\rho}_{11}) &\approx det((\boldsymbol{A}_r^{(Z)})^{-1}(\boldsymbol{A}_r^{(Z)'})^{-1}) \\
det(\boldsymbol{\rho}_{22}) &\approx det((\boldsymbol{B}_r^{(Z)})^{-1}(\boldsymbol{B}_r^{(Z)'})^{-1})
\end{aligned}
$$

Cont.

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

$$
\begin{aligned}
Cov(\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}) &= Cov(\boldsymbol{A}^{-1}\boldsymbol{U}, \boldsymbol{B}^{-1}\boldsymbol{V}) = \boldsymbol{A}^{-1}Cov(\boldsymbol{U}, \boldsymbol{V})(\boldsymbol{B}')^{-1} \\
&= \boldsymbol{A}^{-1}\left[\begin{array}{cccc|c}
\rho_1^* & 0 & \dots & 0 & \\
0 & \rho_2^* & \dots & 0 & \\
\vdots & \vdots & \ddots & \vdots & \boldsymbol{0}_{(p\times(q-p))} \\
0 & 0 & \dots & \rho_p^* &
\end{array}\right](\boldsymbol{B}^{-1})' \\
&= \rho_1^*\boldsymbol{a}^{(1)}(\boldsymbol{b}^{(1)})' + \rho_2^*\boldsymbol{a}^{(2)}(\boldsymbol{b}^{(2)})' + \dots + \rho_p^*\boldsymbol{a}^{(p)}(\boldsymbol{b}^{(p)})'
\end{aligned}
$$

Cont.

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

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

- Em se trabalhando com as variáveis padronizadas, vem que

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

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

Cont.

- Equações:

$$
\begin{aligned}
U_1^{(Z)} &= a_{11}^{(Z)} Z_1^{(1)} + a_{12}^{(Z)} Z_2^{(1)} + \ldots + a_{1p}^{(Z)} Z_p^{(1)} \\
U_2^{(Z)} &= a_{21}^{(Z)} Z_1^{(1)} + a_{22}^{(Z)} Z_2^{(1)} + \ldots + a_{2p}^{(Z)} Z_p^{(1)} \\
&\vdots \\
U_p^{(Z)} &= a_{p1}^{(Z)} Z_1^{(1)} + a_{p2}^{(Z)} Z_2^{(1)} + \ldots + a_{pp}^{(Z)} Z_p^{(1)} \\
\\
V_1^{(Z)} &= b_{11}^{(Z)} Z_1^{(2)} + b_{12}^{(Z)} Z_2^{(2)} + \ldots + b_{1q}^{(Z)} Z_q^{(2)} \\
V_2^{(Z)} &= b_{21}^{(Z)} Z_1^{(2)} + b_{22}^{(Z)} Z_2^{(2)} + \ldots + b_{2q}^{(Z)} Z_q^{(2)} \\
&\vdots \\
V_q^{(Z)} &= b_{q1}^{(Z)} Z_1^{(2)} + b_{q2}^{(Z)} Z_2^{(2)} + \ldots + b_{qq}^{(Z)} Z_q^{(2)}
\end{aligned}
$$

Estimação das variáveis canônicas

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

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

Cont.

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

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

Cont.

- Equações ($i = 1, 2, ..., n$):

$$
\begin{aligned}
\widehat{U}_{i1}^{(Z)} &= \widehat{a}_{11}^{(Z)} Z_{i1}^{(1)} + \widehat{a}_{12}^{(Z)} Z_{i2}^{(1)} + ... + \widehat{a}_{1p}^{(Z)} Z_{ip}^{(1)} \\
\widehat{U}_{i2}^{(Z)} &= \widehat{a}_{21}^{(Z)} Z_{i1}^{(1)} + \widehat{a}_{22}^{(Z)} Z_{i2}^{(1)} + ... + \widehat{a}_{2p}^{(Z)} Z_{ip}^{(1)} \\
&\vdots \\
\widehat{U}_{ip}^{(Z)} &= \widehat{a}_{p1}^{(Z)} Z_{i1}^{(1)} + \widehat{a}_{p2}^{(Z)} Z_{i2}^{(1)} + ... + \widehat{a}_{pp}^{(Z)} Z_{ip}^{(1)} \\
\widehat{V}_{i1}^{(Z)} &= \widehat{b}_{11}^{(Z)} Z_{i1}^{(2)} + \widehat{b}_{12}^{(Z)} Z_{2}^{(2)} + ... + \widehat{b}_{1q}^{(Z)} Z_{iq}^{(2)} \\
\widehat{V}_{i2}^{(Z)} &= \widehat{b}_{21}^{(Z)} Z_{i1}^{(2)} + \widehat{b}_{22}^{(Z)} Z_{i2}^{(2)} + ... + \widehat{b}_{2q}^{(Z)} Z_{iq}^{(2)} \\
&\vdots \\
\widehat{V}_{iq}^{(Z)} &= \widehat{b}_{q1}^{(Z)} Z_{i1}^{(2)} + \widehat{b}_{q2}^{(Z)} Z_{i2}^{(2)} + ... + \widehat{b}_{qq}^{(Z)} Z_{iq}^{(2)}
\end{aligned}
$$