Unidade VII – Análise de Correlação Canônica

7.1. Introdução

A análise de correlação canônica (ACC) é a técnica multivariada que estuda o relacionamento (linear) entre dois conjuntos de variáveis. Em muitos estudos são avaliados dois tipos de variáveis em cada unidade de pesquisa, por exemplo, um conjunto de variáveis de aptidão e um conjunto de variáveis de desempenho, um conjunto de variáveis de personalidade e um conjunto de medidas de habilidade, um conjunto de índices de preços e um conjunto de índices de produção, um conjunto de comportamentos dos alunos e um conjunto de comportamentos do professor, um conjunto de atributos psicológicos e um conjunto de atributos fisiológicos, um conjunto de variáveis ecológicas e um conjunto de variáveis ambientais, um conjunto de desempenho acadêmico e um conjunto de medidas de sucesso no trabalho, e um conjunto de variáveis de personalidade de alunos ao ingressarem na faculdade e as mesmas variáveis nos mesmos indivíduos quando idosos.
Na ACC procura-se identificar e quantificar a associação entre esses dois conjuntos de variáveis por meio do desenvolvimento de uma combinação linear das variáveis em cada um dos grupos, de modo que a correlação entre essas duas combinações seja maximizada. Essas combinações lineares são as variáveis canônicas e suas associações são denominadas de correlações canônicas.
A ACC objetiva encontrar combinações lineares que expressem bem as correlações entre os dois conjuntos de variáveis e obter a simplificação dos dados, ao descrever os dados em um número muito menor variáveis, apontando quais as variáveis originais são as mais importantes.

7.2. Modelo Teórico

Seja $\mathbf{X}$ um vetor de dimensão $(p+q)$x1que possui vetor de médias $\boldsymbol{\mu}$ e matriz de covariância $\boldsymbol{\Sigma}$, que é positiva definida. Consideram-se dois conjuntos de variáveis em $\mathbf{X}$. O 1º conjunto, $\mathbf{X}^{(1)}$, com p variáveis, e o 2º, $\mathbf{X}^{(2)}$, com q variáveis. Sem perda de generalidade assume-se que $p \le q$. A matriz $\mathbf{X}$, o vetor de médias $\boldsymbol{\mu}$ e a matriz de covariância $\boldsymbol{\Sigma}$ podem ser particionados da seguinte maneira:

$$\underset{1\times(p+q)}{\mathbf{X}'} = \left[\underset{(1\times p)}{\mathbf{X}^{(1)}} \vdots \underset{(1\times q)}{\mathbf{X}^{(2)}}\right], \qquad \underset{1\times(p+q)}{\boldsymbol{\mu}'} = \left[\underset{(1\times p)}{\boldsymbol{\mu}^{(1)}} \vdots \underset{(1\times q)}{\boldsymbol{\mu}^{(2)}}\right], \qquad \underset{(p+q)\times(p+q)}{\boldsymbol{\Sigma}} = E(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})' = \begin{bmatrix} \underset{(p\times p)}{\boldsymbol{\Sigma}_{11}} & \vdots & \underset{(p\times q)}{\boldsymbol{\Sigma}_{12}} \\ \cdots & \cdots & \cdots \\ \underset{(q\times p)}{\boldsymbol{\Sigma}_{21}} & \vdots & \underset{(q\times q)}{\boldsymbol{\Sigma}_{22}} \end{bmatrix}$$

onde
$\boldsymbol{\Sigma}_{11} = Cov(\mathbf{X}^{(1)})$ é a matriz de covariância para o 1º conjunto de dados.
$\boldsymbol{\Sigma}_{22} = Cov(\mathbf{X}^{(2)})$ é a matriz de covariância para o 2º conjunto de dados.
$\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}' = Cov(\mathbf{X}^{(1)}, \mathbf{X}^{(2)})$ é a matriz de covariância entre os dois conjuntos de dados, ela mede a associação entre os dois conjuntos.
Como o maior interesse está na matriz $\boldsymbol{\Sigma}_{12}$, mas ela pode ser de grande dimensão e, portanto, difícil de analisar, a ideia da técnica é estudar algumas poucas combinações lineares de variáveis pertencentes a $\mathbf{X}^{(1)}$ e $\mathbf{X}^{(2)}$, ao invés de usar a matriz de covariância $\boldsymbol{\Sigma}_{12}$.
São definidas como U e V as combinações lineares das variáveis de $\mathbf{X}^{(1)}$ e de $\mathbf{X}^{(2)}$:
$U = \mathbf{a}'\mathbf{X}^{(1)}$ e $V = \mathbf{b}'\mathbf{X}^{(2)}$
Sendo $\mathbf{a}$ e $\mathbf{b}$ vetores não nulos dos coeficientes dessas combinações lineares.
As variâncias e a covariância entre U e V são definidas como:

$$Var(U) = Cov(\mathbf{a}'\mathbf{X}^{(1)}) = \mathbf{a}'\boldsymbol{\Sigma}_{11}\mathbf{a}$$
$$Var(V) = Cov(\mathbf{b}'\mathbf{X}^{(2)}) = \mathbf{b}'\boldsymbol{\Sigma}_{22}\mathbf{b}$$
$$Cov(U,V) = Cov(\mathbf{a}'\mathbf{X}^{(1)}, \mathbf{b}'\mathbf{X}^{(2)}) = \mathbf{a}'Cov(\mathbf{X}^{(1)}, \mathbf{X}^{(2)})\mathbf{b} = \mathbf{a}'\boldsymbol{\Sigma}_{12}\mathbf{b}$$

E a correlação entre U e V é definida como:

$$Cor(U,V) = \rho_{U,V} = \frac{\mathbf{a}'\boldsymbol{\Sigma}_{12}\mathbf{b}}{\sqrt{\mathbf{a}'\boldsymbol{\Sigma}_{11}\mathbf{a}}\sqrt{\mathbf{b}'\boldsymbol{\Sigma}_{22}\mathbf{b}}} \tag{6.1}$$

Hotelling (1935) propôs estimar os vetores $\mathbf{a}$ e $\mathbf{b}$ maximizando $Cor(U,V)$ sujeita à restrição: $\mathrm{a}'\boldsymbol{\Sigma}_{11}\mathrm{a} = \mathrm{b}'\boldsymbol{\Sigma}_{22}\mathrm{b} = 1$. As variáveis U e V são chamadas variáveis canônicas e a correlação entre elas, correlação canônica.

Assumindo $p < q$, teremos p correlações canônicas que ao serem maximizadas irão gerar p variáveis canônicas em $\mathbf{X}^{(1)}$ chamadas de U_k, e p variáveis canônicas em $\mathbf{X}^{(2)}$ chamadas de V_k. Então, o primeiro par de variáveis canônicas (primeira função canônica), é o par de combinações lineares U_1 e V_1, com variâncias unitárias, que maximiza a correlação (6.1). A segunda função canônica, é o par de combinações lineares U_2 e V_2, com variâncias unitárias, que maximiza a correlação (6.1) dentre todas as possíveis funções canônicas não correlacionadas com a primeira E assim segue até o par p, pois são encontradas tantas funções canônicas quantos forem o menor número entre p e q. A p-ésima função canônica deve ser não correlacionada com as (p-1) funções canônicas anteriores.
Cada par de variáveis canônicas, U_k e V_k, que apresentam variância unitária, é definido pelos seus respectivos vetores, ($\mathbf{a}_k$ e $\mathbf{b}_k$), que maximizam $Cor(U_k, V_k) = \rho_k^*$.
O resultado da maximização são as combinações lineares do k-ésimo par de variáveis canônicas:

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

tal que $Max\ Cor(U_k, V_k) = \rho_k^* = \sqrt{\lambda_k} \Rightarrow \lambda_k = \left(\mathbf{a}_k'\mathbf{\Sigma}_{12}\mathbf{b}_k\right)^2$

Em que λ_k satisfaz: $\begin{cases} \left(\mathbf{\Sigma}_{11}^{-1/2}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1/2} - \lambda_k \boldsymbol{I}\right)\mathbf{e}_k = 0 \\ \left(\mathbf{\Sigma}_{22}^{-1/2}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1/2} - \lambda_k \boldsymbol{I}\right)\mathbf{f}_k = 0 \end{cases}$

Então $\rho_1^{*2} \geq \rho_2^{*2} \geq \cdots \geq \rho_p^{*2}$ são os p maiores autovalores da matriz $\mathbf{\Sigma}_{11}^{-1/2}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1/2}$ e $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_p$ são os autovetores associados. As quantidades $\rho_1^{*2}, \rho_2^{*2}, \cdots, \rho_p^{*2}$ são também os p maiores autovalores da matriz $\mathbf{\Sigma}_{22}^{-1/2}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1/2}$ com os correspondentes autovetores $\mathbf{f}_1, \mathbf{f}_2, \cdots, \mathbf{f}_p$. Cada $\mathbf{f}_i$ é proporcional a $\mathbf{\Sigma}_{22}^{-1/2}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1/2}\mathbf{e}_i$
As variáveis canônicas têm as seguintes propriedades:

$$Var(U_k) = Var(V_k) = 1$$
$$Cov(U_k, U_l) = Cov(V_k, V_l) = Cov(U_k, V_l) = Cov(U_l, V_k) = 0 \qquad (k \neq l)$$
$$Corr(U_k, U_l) = Corr(V_k, V_l) = Corr(U_k, V_l) = Corr(U_l, V_k) = 0 \qquad (k \neq l)$$

Para o caso amostral, a matriz de covariância particionada $\mathbf{\Sigma}$ deverá ser substituída pela sua estimativa amostral, $\mathbf{S}$.

Se as variáveis forem padronizadas não vai alterar a correlação canônica e o sistema para estimar os coeficientes $\mathbf{a}_k$ e $\mathbf{b}_k$ terá solução única se a matriz $\mathbf{\Sigma}$ tiver posto completo.

7.3. Interpretação das variáveis canônicas

Existem três formas de interpretação das variáveis canônicas. O nível de significância das raízes canônicas, a magnitude das correlações canônicas e as relações entre as variâncias das variáveis canônicas e variáveis originais. Essas relações podem ser de três tipos: os próprios pesos canônicos que estamos chamando de coeficientes canônicos, as cargas canônicas ou estrutura de correlação canônica e as cargas canônicas cruzadas.
As correlações entre as variáveis canônicas e seus respectivos conjuntos originais são chamadas de estrutura canônica e as correlações entre as variáveis canônicas e os conjuntos de variáveis opostas são chamadas de cargas cruzadas. Essas correlações devem ser interpretadas com cuidado porque elas trazem informações univariadas, ou seja, não indicam como a variável original contribui conjuntamente para a análise canônica. Por essa razão, muitos pesquisadores preferem avaliar a contribuição das variáveis originais diretamente pelos coeficientes canônicos calculados a partir das variáveis padronizadas (coeficientes canônicos padronizados).

Considerando as matrizes $\underset{(p\text{x}p)}{\mathbf{A}} = [\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_p]'$ e $\underset{(q\text{x}q)}{\mathbf{B}} = [\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_q]'$, cujas linhas são os vetores de coeficientes das variáveis canônicas, os vetores de variáveis canônicas são:

$$\underset{(p\text{x}1)}{\boldsymbol{U}} = \mathbf{AX}^{(1)} \text{ e } \underset{(q\text{x}1)}{\boldsymbol{V}} = \mathbf{BX}^{(2)} \qquad (6.2)$$

O interesse maior são as primeiras p variáveis canônicas em $\mathbf{V}$. Então:

$$Cov(\boldsymbol{U}, \mathbf{X}^{(1)}) = Cov(\mathbf{AX}^{(1)}, \mathbf{X}^{(1)}) = ACov(\mathbf{X}^{(1)}) = \mathbf{A\Sigma}_{11}$$

Como cada U_i tem variância unitária,

$$Corr\left(U_i, X_k^{(1)}\right) = \frac{Cov\left(U_i, X_k^{(1)}\right)}{\sqrt{Var(U_i)}\sqrt{Var\left(X_k^{(1)}\right)}} = \frac{Cov\left(U_i, X_k^{(1)}\right)}{\sqrt{Var\left(X_k^{(1)}\right)}} = \frac{Cov\left(U_i, X_k^{(1)}\right)}{\sigma_{kk}^{1/2}} = Cov\left(U_i, \sigma_{kk}^{-1/2}X_k^{(1)}\right)$$

Para colocar a correlação na forma matricial introduzimos a matriz diagonal p x p, $\mathbf{V}_{11}^{-1/2}$ que possui como k-ésimo elemento $\sigma_{kk}^{-1/2}$:

$$\underset{(p\mathrm{x}p)}{\boldsymbol{\rho}_{U,\mathbf{X}^{(1)}}} = Corr\left(\boldsymbol{U},\mathbf{X}^{(1)}\right) = Cov\left(\boldsymbol{U},\mathbf{V}_{11}^{-1/2}\mathbf{X}^{(1)}\right) = Cov\left(\mathbf{AX}^{(1)},\mathbf{V}_{11}^{-1/2}\mathbf{X}^{(1)}\right) = \mathbf{A}Cov\left(\mathbf{X}^{(1)},\mathbf{X}^{(1)}\right)\mathbf{V}_{11}^{-1/2} = \mathbf{A\Sigma}_{11}\mathbf{V}_{11}^{-1/2}$$

Cálculos similares podem ser realizados para determinar os outros pares de correlações:

$$\underset{(p\mathrm{x}p)}{\boldsymbol{\rho}_{U,\mathbf{X}^{(1)}}} = \mathbf{A\Sigma}_{11}\mathbf{V}_{11}^{-1/2} \qquad \underset{(q\mathrm{x}q)}{\boldsymbol{\rho}_{V,\mathbf{X}^{(2)}}} = \mathbf{B\Sigma}_{22}\mathbf{V}_{22}^{-1/2}$$

$$\underset{(p\mathrm{x}q)}{\boldsymbol{\rho}_{U,\mathbf{X}^{(2)}}} = \mathbf{A\Sigma}_{12}\mathbf{V}_{22}^{-1/2} \qquad \underset{(q\mathrm{x}p)}{\boldsymbol{\rho}_{V,\mathbf{X}^{(1)}}} = \mathbf{B\Sigma}_{21}\mathbf{V}_{11}^{-1/2}$$

onde $\mathbf{V}_{22}^{-1/2}$ é a matriz diagonal (q x q) que tem $Var\left(X_i^{(2)}\right)$ como o i-ésimo elemento da diagonal.

Se as variáveis forem padronizadas tem-se:

$$\underset{(p\mathrm{x}p)}{\boldsymbol{\rho}_{U,\mathbf{Z}^{(1)}}} = \mathbf{A}_z\boldsymbol{\rho}_{11} \qquad \underset{(q\mathrm{x}q)}{\boldsymbol{\rho}_{V,\mathbf{Z}^{(2)}}} = \mathbf{B}_z\boldsymbol{\rho}_{22}$$

$$\underset{(p\mathrm{x}q)}{\boldsymbol{\rho}_{U,\mathbf{Z}}} = \mathbf{A}_z\boldsymbol{\rho}_{12} \qquad \underset{(q\mathrm{x}p)}{\boldsymbol{\rho}_{V,\mathbf{Z}}} = \mathbf{B}_z\boldsymbol{\rho}_{21}$$

onde $\mathbf{A}_z$ e $\mathbf{B}_z$ são as matrizes cujas linhas contém os coeficientes canônicos para os conjuntos de variáveis padronizadas $\mathbf{Z}^{(1)}$ e $\mathbf{Z}^{(2)}$ respectivamente.

As correlações podem então ser classificadas como:

a) Estrutura canônica:

$$\underset{(p\mathrm{x}p)}{\boldsymbol{\rho}_{U,\mathbf{X}^{(1)}}} = \mathbf{A\Sigma}_{11}\mathbf{V}_{11}^{-1/2} \quad \text{e} \quad \underset{(q\mathrm{x}q)}{\boldsymbol{\rho}_{V,\mathbf{X}^{(2)}}} = \mathbf{B\Sigma}_{22}\mathbf{V}_{22}^{-1/2}$$

ou

$$\underset{(p\mathrm{x}p)}{\boldsymbol{\rho}_{U,\mathbf{Z}^{(1)}}} = \mathbf{A}_z\boldsymbol{\rho}_{11} \quad \text{e} \quad \underset{(q\mathrm{x}q)}{\boldsymbol{\rho}_{V,\mathbf{X}^{(2)}}} = \mathbf{B}_z\boldsymbol{\rho}_{22}$$

b) Cargas canônicas cruzadas:

$$\underset{(p\mathrm{x}q)}{\boldsymbol{\rho}_{U,\mathbf{X}^{(2)}}} = \mathbf{A\Sigma}_{12}\mathbf{V}_{22}^{-1/2} \quad \text{e} \quad \underset{(q\mathrm{x}p)}{\boldsymbol{\rho}_{V,\mathbf{X}^{(1)}}} = \mathbf{B\Sigma}_{21}\mathbf{V}_{11}^{-1/2}$$

ou

$$\underset{(p\mathrm{x}q)}{\boldsymbol{\rho}_{U,\mathbf{X}^{(2)}}} = \mathbf{A}_z\boldsymbol{\rho}_{12} \quad \text{e} \quad \underset{(q\mathrm{x}p)}{\boldsymbol{\rho}_{V,\mathbf{X}^{(1)}}} = \mathbf{B}_z\boldsymbol{\rho}_{21}$$

7.4. Proporção da Variância explicada e Índice de redundância

A correlação canônica elevada ao quadrado representa uma estimativa da variância conjunta (ou compartilhada) entre as variáveis canônicas. É uma medida que pode ser mal interpretada uma vez que mede a variância compartilhada entre as variáveis canônicas e não entre as variáveis originais.

A **proporção da variância explicada** é a proporção da variância de um conjunto de variáveis que é explicada pelas respectivas variáveis canônicas. Dessa forma, mesmo que as correlações canônicas sejam fortes podem não ter sido extraídas quantidades significativas da variância das variáveis originais.

O **índice de redundância** foi proposto para facilitar as interpretações. Ele é equivalente ao coeficiente de determinação da análise de regressão. É a média simples dos coeficientes de correlação múltiplo de um conjunto de variáveis com cada uma das variáveis do outro conjunto, que resulta num coeficiente de determinação médio. Essa medida mede a porcentagem da variância em um conjunto de variáveis que é explicada pelo outro conjunto. Deve-se notar que o valor máximo desse coeficiente não é 100% e sim a variância compartilhada entre os dois conjuntos.

O índice de redundância pode ser calculado para cada variável canônica e depois ser calculado o índice de redundância total, que nada mais é do que a quantidade de variância em um conjunto explicado pelas *r* primeiras variáveis canônicas do outro.
Quando as observações são padronizadas, as matrizes de covariâncias amostrais $\mathbf{S}_{kl}$ são as matrizes de correlação $\mathbf{R}_{kl}$. Os vetores de coeficientes canônicos são as linhas das matrizes $\widehat{\mathbf{A}}_z$ e $\widehat{\mathbf{B}}_z$ e as colunas de $\widehat{\mathbf{A}}_z^{-1}$ e $\widehat{\mathbf{B}}_z^{-1}$ são as correlações amostrais entre as variáveis canônicas e as variáveis originais padronizadas que as compõem.
Especificamente:

$$Cov\left(\mathbf{z}^{(1)}, \widehat{\boldsymbol{U}}\right) = Cov\left(\widehat{\mathbf{A}}_z^{-1}\widehat{\boldsymbol{U}}, \widehat{\boldsymbol{U}}\right) = \widehat{\mathbf{A}}_z^{-1}\mathbf{I} = \widehat{\mathbf{A}}_z^{-1}$$
$$Cov\left(\mathbf{z}^{(1)}, \widehat{\boldsymbol{V}}\right) = Cov\left(\widehat{\mathbf{B}}_z^{-1}\widehat{\boldsymbol{V}}, \widehat{\boldsymbol{V}}\right) = \widehat{\mathbf{B}}_z^{-1}\mathbf{I} = \widehat{\mathbf{B}}_z^{-1}$$

Então,

$$\widehat{\mathbf{A}}_z^{-1} = \left[\hat{\mathbf{a}}_z^{(1)}, \hat{\mathbf{a}}_z^{(2)}, \cdots, \hat{\mathbf{a}}_z^{(\boldsymbol{p})}\right] = \begin{bmatrix} r_{\widehat{U}_1,z_1^{(1)}} & r_{\widehat{U}_2,z_1^{(1)}} & \cdots & r_{\widehat{U}_p,z_1^{(1)}} \\ r_{\widehat{U}_1,z_2^{(1)}} & r_{\widehat{U}_2,z_2^{(1)}} & \cdots & r_{\widehat{U}_p,z_2^{(1)}} \\ \vdots & \vdots & \ddots & \vdots \\ r_{\widehat{U}_1,z_p^{(1)}} & r_{\widehat{U}_2,z_p^{(1)}} & \cdots & r_{\widehat{U}_p,z_p^{(1)}} \end{bmatrix}$$

e

$$\widehat{\mathbf{B}}_z^{-1} = \left[\hat{\mathbf{b}}_z^{(1)}, \hat{\mathbf{b}}_z^{(2)}, \cdots, \hat{\mathbf{b}}_z^{(\boldsymbol{p})}\right] = \begin{bmatrix} r_{\widehat{V}_1,z_1^{(2)}} & r_{\widehat{V}_2,z_1^{(2)}} & \cdots & r_{\widehat{V}_p,z_1^{(2)}} \\ r_{\widehat{V}_1,z_2^{(2)}} & r_{\widehat{V}_2,z_2^{(2)}} & \cdots & r_{\widehat{V}_p,z_2^{(2)}} \\ \vdots & \vdots & \ddots & \vdots \\ r_{\widehat{V}_1,z_p^{(2)}} & r_{\widehat{V}_2,z_p^{(2)}} & \cdots & r_{\widehat{V}_p,z_p^{(2)}} \end{bmatrix}$$

onde $r_{\widehat{U}_i,z_i^{(1)}}$ e $r_{\widehat{V}_i,z_i^{(2)}}$ são os coeficientes de correlação amostrais entre as variáveis originais e as variáveis canônicas, ou seja, os termos que compõem a matriz de estrutura canônica.
Para observações padronizadas, tem-se:
Variância total amostral (padronizada) no primeiro conjunto de variáveis =

$$= tr(\mathbf{R}_{11}) = tr\left[\hat{\mathbf{a}}_z^{(1)}\hat{\mathbf{a}}_z^{(1)\prime} + \hat{\mathbf{a}}_z^{(2)}\hat{\mathbf{a}}_z^{(2)\prime} + \cdots + \hat{\mathbf{a}}_z^{(p)}\hat{\mathbf{a}}_z^{(p)\prime}\right] = p$$

Variância total amostral (padronizada) no segundo conjunto de variáveis =

$$= tr(\mathbf{R}_{22}) = tr\left[\hat{\mathbf{b}}_z^{(1)}\hat{\mathbf{b}}_z^{(1)\prime} + \hat{\mathbf{b}}_z^{(2)}\hat{\mathbf{b}}_z^{(2)\prime} + \cdots + \hat{\mathbf{b}}_z^{(p)}\hat{\mathbf{b}}_z^{(p)\prime}\right] = q$$

Como as correlações nas primeiras $r < p$ colunas de $\widehat{\mathbf{A}}_z^{-1}$e $\widehat{\mathbf{B}}_z^{-1}$ envolvem só as variáveis canônicas amostrais, $\left[\widehat{U}_1, \widehat{U}_2, \cdots, \widehat{U}_r\right]$ e $\left[\widehat{V}_1, \widehat{V}_2, \cdots, \widehat{V}_r\right]$ respectivamente, definimos a contribuição das primeiras *r* variáveis canônicas à variância total amostral (padronizada) como:

$$tr\left[\hat{\mathbf{a}}_z^{(1)}\hat{\mathbf{a}}_z^{(1)\prime} + \hat{\mathbf{a}}_z^{(2)}\hat{\mathbf{a}}_z^{(2)\prime} + \cdots + \hat{\mathbf{a}}_z^{(r)}\hat{\mathbf{a}}_z^{(r)\prime}\right] = \sum_{i=1}^{r}\sum_{k=1}^{p} r^2_{\widehat{U}_i,z_k^{(1)}}$$
$$tr\left[\hat{\mathbf{b}}_z^{(1)}\hat{\mathbf{b}}_z^{(1)\prime} + \hat{\mathbf{b}}_z^{(2)}\hat{\mathbf{b}}_z^{(2)\prime} + \cdots + \hat{\mathbf{b}}_z^{(p)}\hat{\mathbf{b}}_z^{(p)\prime}\right] = \sum_{i=1}^{r}\sum_{k=1}^{p} r^2_{\widehat{V}_i,z_k^{(2)}}$$

Logo, a proporção da variância total amostral (padronizada) explicada pelas *r* primeiras variáveis canônicas $\left(\widehat{U}_1, \widehat{U}_2, \cdots, \widehat{U}_r\right)$ para o 1º conjunto é dada por:

$$\mathbf{R}^2_{\mathbf{z}^{(1)}|\widehat{U}_1,\widehat{U}_2,\cdots,\widehat{U}_r} = \frac{tr\left[\hat{\mathbf{a}}_z^{(1)}\hat{\mathbf{a}}_z^{(1)\prime} + \hat{\mathbf{a}}_z^{(2)}\hat{\mathbf{a}}_z^{(2)\prime} + \cdots + \hat{\mathbf{a}}_z^{(r)}\hat{\mathbf{a}}_z^{(r)\prime}\right]}{tr(\mathbf{R}_{11})} = \frac{\sum_{i=1}^{r}\sum_{k=1}^{p} r^2_{\widehat{U}_i,z_k^{(1)}}}{p}$$

que representa a soma de cada valor da matriz de estrutura canônica elevado ao quadrado dividida por *p* (que é o número de variáveis no primeiro conjunto).

De forma análoga, para o segundo conjunto de variáveis, a proporção da variância total amostral padronizada explicada por $\widehat{V}_1, \widehat{V}_2, \cdots, \widehat{V}_r$ é dada por:

$$\mathbf{R}^2_{\mathbf{z}^{(2)}|\widehat{V}_1,\widehat{V}_2,\cdots,\widehat{V}_r} = \frac{tr\left[\hat{\mathbf{b}}_z^{(1)}\hat{\mathbf{b}}_z^{(1)\prime} + \hat{\mathbf{b}}_z^{(2)}\hat{\mathbf{b}}_z^{(2)\prime} + \cdots + \hat{\mathbf{b}}_z^{(p)}\hat{\mathbf{b}}_z^{(p)\prime}\right]}{tr(\mathbf{R}_{22})} = \frac{\sum_{i=1}^{r}\sum_{k=1}^{p} r^2_{\widehat{V}_i,z_k^{(2)}}}{q}$$

que representa a soma de cada valor da matriz de estrutura canônica elevado ao quadrado dividida por q (que é o número de variáveis no segundo conjunto).

O índice de redundância das combinações lineares estimadas é determinado como:

$$IR_{U_i} = \hat{\rho}_i^{*2} \times \frac{\sum_{i=1}^{p} r^2_{\hat{U}_i, z_i^{(1)}}}{p} = \lambda_i \times \frac{\sum_{i=1}^{p} r^2_{\hat{U}_i, z_i^{(1)}}}{p}$$

ou seja, é o autovalor vezes a proporção da variância explicada pelo respetivo autovalor.

Da mesma forma para o segundo conjunto:

$$IR_{V_i} = \hat{\rho}_i^{*2} \times \frac{\sum_{i=1}^{p} r^2_{\hat{V}_i, z_i^{(2)}}}{q} = \lambda_i \times \frac{\sum_{i=1}^{p} r^2_{\hat{V}_i, z_i^{(2)}}}{q}$$

O índice de redundância total nada mais é do que a soma dos índices de redundância de cada combinação linear do mesmo conjunto e dados, ou seja:

$$IRT_U = \sum_{i=1}^{p} \hat{\rho}_i^{*2} \times \frac{\sum_{i=1}^{p} r^2_{\hat{U}_i, z_i^{(1)}}}{p} = \sum_{i=1}^{p} \lambda_i \times \frac{\sum_{i=1}^{p} r^2_{\hat{U}_i, z_i^{(1)}}}{p}$$

$$IRT_V = \sum_{i=1}^{p} \hat{\rho}_i^{*2} \times \frac{\sum_{i=1}^{p} r^2_{\hat{V}_i, z_i^{(2)}}}{q} = \sum_{i=1}^{p} \lambda_i \times \frac{\sum_{i=1}^{p} r^2_{\hat{V}_i, z_i^{(2)}}}{q}$$

7.5. Testes de significância

Quando se trabalha com grandes amostras há interesse em realizar inferências dos resultados da ACC. Como qualquer teste estatístico necessita-se saber qual a significância de cada correlação canônica. Existem testes de signjficância global e o mais utilizado é o teste de Rao. Para testar separadamente cada função canônica existem alguns testes. São eles: Lambda de Wilk (Wilk's Lambda), traço de Hotteling-Lawley (Hotteling's trace), Traço de Pillai (Pillai's trace) e maior raiz característica de Roy (Roy's gnc).

Quando $\mathbf{\Sigma}_{12} = 0$ ($\rho_1^* = \rho_2^* = \cdots = \rho_p^* = 0$), $\mathbf{a}'\mathbf{X}^{(1)}$ e $\mathbf{b}'\mathbf{X}^{(2)}$ têm covariância $\mathbf{a}'\mathbf{\Sigma}_{12}\mathbf{b} = 0$ para todos os vetores **a** e **b**. Consequentemente, todas as correlações canônicas serão zero, e não existe porque propor uma análise de correlação canônica. Pode-se então testar

H_0: $\mathbf{\Sigma}_{12} = 0$ *versus* H_1: $\mathbf{\Sigma}_{12} \neq 0$

utilizando o teste de razão de verossimilhança:

$$T = -2\ln(\Lambda) = n\ln\left(\frac{|\mathbf{S}_{11}||\mathbf{S}_{22}|}{|\mathbf{S}|}\right) = -n\ln\prod_{i=1}^{p}(1-\lambda_i) \sim \chi^2_{pq}$$

Rejeita-se a H_0 se o $T \geq T_c$ de uma distribuição qui-quadrado com *pq* graus de liberdade, dado nível de significância (α).

Se o teste global rejeitar H_0, pode ser de interesse avaliar se as *k* primeiras correlações são significativas e, então, se as variáveis canônicas seriam importantes para a caracterização dos dois conjuntos de dados. As hipóteses a serem testadas são:

$H_0^k = (\rho_1^*)^2 \neq 0, (\rho_2^*)^2 \neq 0, \cdots, (\rho_k^*)^2 \neq 0, (\rho_{k+1}^*)^2 \neq 0, \cdots, \left(\rho_p^*\right)^2 \neq 0$

$H_1^k = (\rho_i^*)^2 = 0$, para algum $i \geq k + 1$

Que podem ser testadas utilizando a estatística:

$$T = -\left(n - 1 - \frac{1}{2}(p + q + 1)\right)\ln\prod_{i=k+1}^{p}\left(1 - \rho_i^{*2}\right) \sim \chi^2_{(p-k+1)(q-k+1)}$$