10
Análise de correlação canônica

10.1. Introdução

A análise de correlação canônica é centrada na identificação e quantificação da associação entre dois grupos de variáveis. O foco da correlação canônica é direcionado para a correlação entre uma combinação linear das variáveis em um dos grupos com uma outra combinação linear das variáveis do outro grupo de variáveis. A idéia fundamental é, a princípio, determinar as combinações lineares dos dois grupos que possuem a maior correlação. No próximo estágio, é determinado o par de maior correlação que seja, ainda, não correlacionado com o par selecionado inicialmente. O processo continua até se esgotar as dimensões de ambos os grupos ou do menor grupo. Os pares de combinações lineares são denominados de variáveis canônicas e suas correlações são chamadas de correlações canônicas. A técnica de encontrar essas combinações lineares e suas respectivas correlações é devida a Hotelling (1935 e 1936).

A idéia fundamental é encontrar relações entre dois conjuntos de variáveis, em alta dimensão, em poucos pares de variáveis canônicas. Várias aplicações nas ciências humanas, na genética entre outras áreas são encontradas na literatura.

10.2. Variáveis canônicas e correlação canônica populacionais

Seja $\underset{\sim}{X}$ um vetor de dimensão (p+qx1), o qual possui matriz de covariância Σ e média $\underset{\sim}{\mu}$. Sejam os vetores $\underset{\sim}{X}^{(1)}$ (px1) e $\underset{\sim}{X}^{(2)}$ (qx1) definidos como sendo originados de uma partição do vetor original $\underset{\sim}{X}$, representando um grupo com p variáveis e outro com q, respectivamente. Sem perda de generalidade é assumido que p≤q. Pressupõe-se, também, que Σ possui elementos finitos e é positiva definida. Para o vetor aleatório $\underset{\sim}{X}$, os seguintes resultados são apresentados.

$$\underset{\sim}{X} = \begin{bmatrix} \underset{\sim}{X}^{(1)} \\ \underset{\sim}{X}^{(2)} \end{bmatrix} = \begin{bmatrix} X_1^{(1)} \\ X_2^{(1)} \\ \vdots \\ X_p^{(1)} \\ X_1^{(2)} \\ X_2^{(2)} \\ \vdots \\ X_q^{(2)} \end{bmatrix} \qquad (10.1)$$

Cuja média é:

$$\underset{\sim}{\mu} = E(\underset{\sim}{X}) = \begin{bmatrix} \underset{\sim}{\mu}^{(1)} \\ \underset{\sim}{\mu}^{(2)} \end{bmatrix} \quad (10.2)$$

E cuja matriz de covariâncias é:

$$\Sigma = E\left(\underset{\sim}{X} - \underset{\sim}{\mu}\right)\left(\underset{\sim}{X} - \underset{\sim}{\mu}\right)^t = \begin{array}{c c} & \begin{array}{cc} p & q \end{array} \\ \begin{array}{c} p \\ q \end{array} & \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \end{array} \quad (10.3)$$

Assim, para os vetores $\underset{\sim}{X}^{(1)}$ (px1) e $\underset{\sim}{X}^{(2)}$ (qx1) verifica-se que:

$$\begin{cases} E\left(\underset{\sim}{X}^{(1)}\right) = \underset{\sim}{\mu}^{(1)} \quad \mathrm{Cov}\left(\underset{\sim}{X}^{(1)}\right) = \Sigma_{11} \\ E\left(\underset{\sim}{X}^{(2)}\right) = \underset{\sim}{\mu}^{(2)} \quad \mathrm{Cov}\left(\underset{\sim}{X}^{(2)}\right) = \Sigma_{22} \\ \mathrm{Cov}\left(\underset{\sim}{X}^{(1)}, \underset{\sim}{X}^{(2)}\right) = \Sigma_{12} = \Sigma_{21}^t \end{cases} \quad (10.4)$$

As covariâncias entre pares de variáveis pertencentes aos dois grupos, uma de $\underset{\sim}{X}^{(1)}$ e outra de $\underset{\sim}{X}^{(2)}$, estão contidas em Σ_{12}. Dessa forma, os pq elementos de Σ_{12} medem a associação entre os dois grupos. Se ambos os valores de p e q são grandes, a interpretação simultânea desse conjunto de covariâncias é uma tarefa difícil e na maioria das vezes infrutífera. Como a finalidade, em geral, é de realizar predição ou realizar comparação, o interesse pode ser focado em combinações lineares das variáveis originais. A idéia é, portanto, concentrar a atenção em algumas poucas combinações lineares de variáveis pertencentes a $\underset{\sim}{X}^{(1)}$ e a $\underset{\sim}{X}^{(2)}$, ao invés de utilizar todas as pq covariâncias contidas em Σ_{12}.

Seguindo a notação normalmente utilizada na literatura especializada, sejam as variáveis U e V combinações lineares das variáveis de $\underset{\sim}{X}^{(1)}$ e de $\underset{\sim}{X}^{(2)}$, respectivamente, definidas por:

$$\begin{cases} U = \underset{\sim}{a}^t \underset{\sim}{X}^{(1)} \\ \\ V = \underset{\sim}{b}^t \underset{\sim}{X}^{(2)} \end{cases} \tag{10.5}$$

sendo $\underset{\sim}{a}$ e $\underset{\sim}{b}$ vetores não nulos dos coeficientes dessas combinações lineares.

Assim,

$$\begin{cases} \mathrm{Var}(U) = \mathrm{Cov}\left(\underset{\sim}{a}^t \underset{\sim}{X}^{(1)}\right) = \underset{\sim}{a}^t \Sigma_{11} \underset{\sim}{a} \\ \mathrm{Var}(V) = \mathrm{Cov}\left(\underset{\sim}{b}^t \underset{\sim}{X}^{(2)}\right) = \underset{\sim}{b}^t \Sigma_{22} \underset{\sim}{b} \\ \mathrm{Cov}(U, V) = \underset{\sim}{a}^t \mathrm{Cov}\left(\underset{\sim}{X}^{(1)}, \underset{\sim}{X}^{(2)}\right) \underset{\sim}{b} = \underset{\sim}{a}^t \Sigma_{12} \underset{\sim}{b} \end{cases} \tag{10.6}$$

A correlação entre U e V é definida por:

$$\mathrm{Corr}(U, V) = \rho_{U,V} = \frac{\underset{\sim}{a}^t \Sigma_{12} \underset{\sim}{b}}{\sqrt{\underset{\sim}{a}^t \Sigma_{11} \underset{\sim}{a}} \sqrt{\underset{\sim}{b}^t \Sigma_{22} \underset{\sim}{b}}} \tag{10.7}$$

Hotelling (1935 e 1936) propôs estabelecer os pares (U_i, V_i), i=1, 2, ..., p, determinando os vetores $\underset{\sim}{a}_i$ e $\underset{\sim}{b}_i$ que maximizam (10.7). As variáveis U_i e V_i são denominadas de variáveis canônicas e a correlação entre elas de correlação canônica. Na seqüência são apresentados os resultados necessários para a maximização de (10.7) e, portanto, para a obtenção das variáveis canônicas e de suas correlações.

Para determinar o máximo de $\rho_{U,V}$, inicialmente são impostas as restrições:

$$\underset{\sim}{a}^t \Sigma_{11} \underset{\sim}{a} = \underset{\sim}{b}^t \Sigma_{22} \underset{\sim}{b} = 1 \tag{10.8}$$

A mudança de escala imposta pelas restrições (10.8) não afeta a correlação (10.7). Para obter o máximo de $\rho_{U,V}$ é preciso deriva a equação (10.7) com relação aos vetores $\underset{\sim}{a}$ e $\underset{\sim}{b}$ e igualar as derivadas parciais a zero. As equações obtidas são:

$$\begin{cases} \dfrac{\partial \rho_{U,V}}{\partial \underset{\sim}{a}} = \left(\underset{\sim}{b}^t \Sigma_{22} \underset{\sim}{b}\right)^{-1/2} \left[\left(\underset{\sim}{a}^t \Sigma_{11} \underset{\sim}{a}\right)^{-1/2} \Sigma_{12} \underset{\sim}{b} - 2\left(\dfrac{-1}{2}\right) \left(\underset{\sim}{a}^t \Sigma_{12} \underset{\sim}{b}\right) \left(\underset{\sim}{a}^t \Sigma_{11} \underset{\sim}{a}\right)^{-3/2} \Sigma_{11} \underset{\sim}{a} \right] \\ \\ \dfrac{\partial \rho_{U,V}}{\partial \underset{\sim}{b}} = \left(\underset{\sim}{a}^t \Sigma_{11} \underset{\sim}{a}\right)^{-1/2} \left[\left(\underset{\sim}{b}^t \Sigma_{22} \underset{\sim}{b}\right)^{-1/2} \Sigma_{12}^t \underset{\sim}{a} - 2\left(\dfrac{-1}{2}\right) \left(\underset{\sim}{a}^t \Sigma_{12} \underset{\sim}{b}\right) \left(\underset{\sim}{b}^t \Sigma_{22} \underset{\sim}{b}\right)^{-3/2} \Sigma_{22} \underset{\sim}{b} \right] \end{cases} \tag{10.9}$$

Igualando as derivadas parciais de (10.9) a zero e impondo as restrições (10.8), rearranjando alguns termos, obtém-se:

$$\begin{cases} \left(\underset{\sim}{a}^t \Sigma_{12} \underset{\sim}{b}\right) \Sigma_{11} \underset{\sim}{a} + \Sigma_{12} \underset{\sim}{b} = \underset{\sim}{0} \\ \\ \Sigma_{12}^t \underset{\sim}{a} + \left(\underset{\sim}{a}^t \Sigma_{12} \underset{\sim}{b}\right) \Sigma_{22} \underset{\sim}{b} = \underset{\sim}{0} \end{cases} \tag{10.10}$$

É fácil observar que (10.7) sujeita as restrições (10.8) se torna igual a $\rho_{U,V} = \underset{\sim}{a}^t \Sigma_{12} \underset{\sim}{b}$, que é o valor máximo, então:

$$\begin{cases} \rho_{U,V} \Sigma_{11} \underset{\sim}{a} + \Sigma_{12} \underset{\sim}{b} = \underset{\sim}{0} \\ \\ \Sigma_{12}^t \underset{\sim}{a} + \rho_{U,V} \Sigma_{22} \underset{\sim}{b} = \underset{\sim}{0} \end{cases} \tag{10.11}$$

Assim, para solução de (10.11) é necessário que o determinante dos coeficientes do sistema de equações homogêneas seja nulo. Logo,

$$\begin{vmatrix} \rho_{U,V} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^t & \rho_{U,V} \Sigma_{22} \end{vmatrix} = 0 \tag{10.12}$$

Uma importante propriedade dos determinantes é reproduzida a seguir. Seja uma matriz A com as seguintes partições:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \tag{10.13}$$

O determinante de A, se A_{11} e A_{22} são não singulares, é dado por:

$$\begin{cases} |A| = |A_{11}||A_{22} - A_{21}A_{11}^{-1}A_{12}| \\ \text{ou} \\ |A| = |A_{22}||A_{11} - A_{12}A_{22}^{-1}A_{21}| \end{cases} \tag{10.14}$$

Utilizando o resultado (10.14) no determinante (10.12), obtém-se os seguintes resultados para a primeira equação:

$$|\rho_{U,V}\Sigma_{11}|\left|\rho_{U,V}\Sigma_{22} - \frac{1}{\rho_{U,V}}\Sigma_{12}^{t}\Sigma_{11}^{-1}\Sigma_{12}\right| = 0$$

Como $|\rho_{U,V}\Sigma_{11}|$ é diferente de zero, pois Σ_{11} é positiva definida, então, o determinante anterior só será zero se:

$$\left|\rho_{U,V}\Sigma_{22} - \frac{1}{\rho_{U,V}}\Sigma_{12}^{t}\Sigma_{11}^{-1}\Sigma_{12}\right| = 0$$

Como o resultado dessa equação é zero, não há alteração se ambos os termos da equação à esquerda da desigualdade for multiplicado por $(-\rho_{U,V})$. Da mesma forma pode-se proceder para a segunda equação do determinante de (10.14). O resultado final dessa derivação é:

$$\begin{cases} \left|\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^{t} - \rho_{U,V}^{2}\Sigma_{11}\right| = 0 \\ \\ \left|\Sigma_{12}^{t}\Sigma_{11}^{-1}\Sigma_{12} - \rho_{U,V}^{2}\Sigma_{22}\right| = 0 \end{cases} \tag{10.15}$$

Fazendo $\lambda = \rho_{U,V}^{2}$, verifica-se que as equações determinantais de (10.15) podem ser vistas como maximização de pares de formas quadráticas (capítulo 2) do tipo:

$$\lambda = \frac{\underset{\sim}{e}^{t} A \underset{\sim}{e}}{\underset{\sim}{e}^{t} B \underset{\sim}{e}}$$

restrito a $\underset{\sim}{e}^{t} B \underset{\sim}{e}$ =1.

Assim, os resultados de (10.15) podem ser reescritos (capítulo 2) da seguinte forma:

$$\begin{cases} \left(\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^{t} - \lambda\Sigma_{11}\right)\underset{\sim}{a} = \underset{\sim}{0} \ \ \text{(a)} \\ \\ \left(\Sigma_{12}^{t}\Sigma_{11}^{-1}\Sigma_{12} - \lambda\Sigma_{22}\right)\underset{\sim}{b} = \underset{\sim}{0} \ \ \text{(b)} \end{cases} \tag{10.16}$$

A resolução do sistema de equações pode ser feita aplicando uma transformação linear não singular. Isso é ilustrado doravante com a equação (a) de (10.16). Seja $\Sigma_{11}^{1/2}$ a matriz raiz quadrada de Σ_{11} e considere a transformação linear $\underset{\sim}{c} = \Sigma_{11}^{1/2}\underset{\sim}{a}$, então, $\underset{\sim}{a} = \Sigma_{11}^{-1/2}\underset{\sim}{c}$. Se a equação (a) for pré-multiplicada e $\underset{\sim}{a}$ for substituído por $\underset{\sim}{a} = \Sigma_{11}^{-1/2}\underset{\sim}{c}$, então:

$$\Sigma_{11}^{-1/2}\left(\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^{t}-\lambda\Sigma_{11}\right)\Sigma_{11}^{-1/2}\underset{\sim}{c}=\underset{\sim}{0}$$

$$\left(\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^{t}\Sigma_{11}^{-1/2}-\lambda\Sigma_{11}^{-1/2}\Sigma_{11}\Sigma_{11}^{-1/2}\right)\underset{\sim}{c}=\underset{\sim}{0}$$

Então a solução de (a) é dada pela solução do seguinte sistema de equações homogêneas:

$$\left(\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^{t}\Sigma_{11}^{-1/2}-\lambda_i I\right)\underset{\sim}{c}_i=\underset{\sim}{0} \qquad (10.17)$$

A solução de (10.17) é facilmente obtida pelo cálculo dos autovalores (λ_i) e autovetores $\left(\underset{\sim}{c}_i\right)$ de $\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^{t}\Sigma_{11}^{-1/2}$. Os autovalores ($\lambda_i$) dessa matriz são os mesmos do sistema não transformados por serem invariantes com relação a transformações não singulares, no entanto, os autovetores são afetados pela transformação. Dessa forma, os autovetores devem ser recuperados pela transformação linear inversa a efetuada. Assim,

$$\underset{\sim}{a}_i=\Sigma_{11}^{-1/2}\underset{\sim}{c}_i \qquad (10.18)$$

Tratamento igual é dado para a equação (b) de (10.16), agora efetuando a transformação linear $\underset{\sim}{d}=\Sigma_{22}^{1/2}\underset{\sim}{b}$. Então,

$$\left(\Sigma_{22}^{-1/2}\Sigma_{12}^{t}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1/2}-\lambda_i I\right)\underset{\sim}{d}_i=\underset{\sim}{0} \tag{10.19}$$

Os autovetores $\underset{\sim}{b}_i$, soluções almejadas, são recuperados por:

$$\underset{\sim}{b}_i=\Sigma_{22}^{-1/2}\underset{\sim}{d}_i \tag{10.20}$$

O máximo é obtido substituindo essas soluções em (10.7). Logo,

$$\underset{\underset{\sim}{a},\,\underset{\sim}{b}}{\mathrm{Max}}\left(\rho_{U,V}\right)=\frac{\underset{\sim}{a}^{t}\Sigma_{12}\underset{\sim}{b}}{\sqrt{\underset{\sim}{a}^{t}\Sigma_{11}\underset{\sim}{a}}\sqrt{\underset{\sim}{b}^{t}\Sigma_{22}\underset{\sim}{b}}}=\underset{\sim}{a}^{t}\Sigma_{12}\underset{\sim}{b}$$

Da equação (10.10), sabendo que $\rho_{U,V}=\underset{\sim}{a}^{t}\Sigma_{12}\underset{\sim}{b}=\sqrt{\lambda_i}$, verifica-se que $\lambda=\left(\underset{\sim}{a}^{t}\Sigma_{12}\underset{\sim}{b}\right)^2$, logo:

$$\underset{\underset{\sim}{a},\,\underset{\sim}{b}}{\mathrm{Max}}\left(\rho_{U,V}\right)=\sqrt{\lambda_i} \tag{10.21}$$

As variáveis canônicas têm as seguintes propriedades:

$$\mathrm{Var}(U_i)=\mathrm{Cov}\left(\underset{\sim}{a}_i^{t}\underset{\sim}{X}^{(1)}\right)=\underset{\sim}{a}_i^{t}\Sigma_{11}\underset{\sim}{a}_i=\underset{\sim}{c}_i^{t}\Sigma_{11}^{-1/2}\Sigma_{11}\Sigma_{11}^{-1/2}\underset{\sim}{c}_i=\underset{\sim}{c}_i^{t}\underset{\sim}{c}_i$$

Sabendo que $\underset{\sim}{c}_i$ é um autovetor de $\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^{t}\Sigma_{11}^{-1/2}$ com norma 1, e procedendo da mesma forma para Var(V_i) verifica-se que:

$$\operatorname{Var}\left(U_i\right) = \operatorname{Var}\left(V_i\right) = 1 \tag{10.22}$$

A $\operatorname{Cov}\left(U_k, U_\ell\right)$ com $(k \neq \ell)$ é dada por:

$$\begin{aligned}\operatorname{Cov}\left(U_k, U_\ell\right) &= \operatorname{Cov}\left(\underset{\sim}{a}_k^t \underset{\sim}{X}^{(1)}, \underset{\sim}{a}_\ell^t \underset{\sim}{X}^{(1)}\right) = \underset{\sim}{a}_k^t \Sigma_{11} \underset{\sim}{a}_\ell = \\ &= \underset{\sim}{c}_k^t \Sigma_{11}^{-1/2} \Sigma_{11} \Sigma_{11}^{-1/2} \underset{\sim}{c}_\ell = \underset{\sim}{c}_k^t I \underset{\sim}{c}_\ell = \underset{\sim}{c}_k^t \underset{\sim}{c}_\ell = 0 \ (k \neq \ell)\end{aligned}$$

Logo,

$$\begin{cases}\operatorname{Cov}\left(U_k, U_\ell\right) = \operatorname{Corr}\left(U_k, U_\ell\right) = 0 \quad \left(k \neq \ell\right) \\ \\ \operatorname{Cov}\left(V_k, V_\ell\right) = \operatorname{Corr}\left(V_k, V_\ell\right) = 0 \quad \left(k \neq \ell\right)\end{cases} \tag{10.23}$$

Finalmente, a covariância entre U_k e V_ℓ com $\left(k \neq \ell\right)$ é dada por:

$$\begin{aligned}\operatorname{Cov}\left(U_k, V_\ell\right) &= \operatorname{Cov}\left(\underset{\sim}{a}_k^t \underset{\sim}{X}^{(1)}, \underset{\sim}{b}_\ell^t \underset{\sim}{X}^{(2)}\right) = \underset{\sim}{a}_k^t \Sigma_{12} \underset{\sim}{b}_\ell = \\ &= \underset{\sim}{c}_k^t \Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1/2} \underset{\sim}{d}_\ell = 0 \qquad (k \neq \ell)\end{aligned}$$

Logo,

$$\mathrm{Cov}\left(U_k, V_\ell\right) = \mathrm{Corr}\left(U_k, V_\ell\right) = 0 \qquad \left(k \neq \ell\right) \tag{10.24}$$

Para variáveis padronizadas $\underset{\sim}{Z}^{(1)t} = \left[Z_1^{(1)}\ Z_2^{(1)} \cdots Z_p^{(1)}\right]$ e $\underset{\sim}{Z}^{(2)t} = \left[Z_1^{(1)}\ Z_2^{(2)} \cdots Z_q^{(2)}\right]$ as variáveis canônicas são dadas por:

$$\begin{cases} U_k = \underset{\sim}{a}_k^t \underset{\sim}{Z}^{(1)} = \underset{\sim}{c}_k^t \rho_{11}^{-1/2} \underset{\sim}{Z}^{(1)} \\ \\ V_k = \underset{\sim}{b}_k^t \underset{\sim}{Z}^{(2)} = \underset{\sim}{d}_k^t \rho_{22}^{-1/2} \underset{\sim}{Z}^{(2)} \end{cases} \tag{10.25}$$

em que $\underset{\sim}{c}_k$ e $\underset{\sim}{d}_k$ são os autovetores de norma 1 das matrizes $\rho_{11}^{-1/2}\rho_{12}\rho_{22}^{-1}\rho_{12}^t\rho_{11}^{-1/2}$ e $\rho_{22}^{-1/2}\rho_{12}^t\rho_{11}^{-1}\rho_{12}\rho_{22}^{-1/2}$, respectivamente. Os autovetores originais devem ser recuperados por:

$$\begin{cases} \underset{\sim}{a}_k = \rho_{11}^{-1/2} \underset{\sim}{c}_k \\ \\ \underset{\sim}{b}_k = \rho_{22}^{-1/2} \underset{\sim}{d}_k \end{cases} \tag{10.26}$$

em que: ρ_{11} (pxp), ρ_{12} (pxq) e ρ_{22} (qxq) são partições de ρ (p+q x p+q) dadas por:

$$\rho = E\left(\underset{\sim}{Z}\underset{\sim}{Z}^t\right) = \begin{array}{c c} & \begin{array}{cc} p & q \end{array} \\ \begin{array}{c} p \\ q \end{array} & \begin{bmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{bmatrix} \end{array} \tag{10.27}$$

de tal sorte que:

$$\begin{cases} E\left(\underset{\sim}{Z}^{(1)}\right) = \underset{\sim}{0} \quad Cov\left(\underset{\sim}{Z}^{(1)}\right) = \rho_{11} \\ E\left(\underset{\sim}{Z}^{(2)}\right) = \underset{\sim}{0} \quad Cov\left(\underset{\sim}{Z}^{(2)}\right) = \rho_{22} \\ Cov\left(\underset{\sim}{Z}^{(1)}, \underset{\sim}{Z}^{(2)}\right) = \rho_{12} = \rho_{21}^{t} \end{cases} \tag{10.28}$$

As correlações canônicas das combinações lineares padronizadas são dadas por:

$$Corr(U_k, V_k) = \frac{\underset{\sim}{a}_k^t \rho_{12} \underset{\sim}{b}_k}{\sqrt{\underset{\sim}{a}_k^t \rho_{11} \underset{\sim}{a}_k} \sqrt{\underset{\sim}{b}_k^t \rho_{22} \underset{\sim}{b}_k}} = \sqrt{\lambda_k} \tag{10.29}$$

em que λ_k é k-ésimo autovalor de $\rho_{11}^{-1/2}\rho_{12}\rho_{22}^{-1}\rho_{12}^{t}\rho_{11}^{-1/2}$, ou equivalentemente de $\rho_{22}^{-1/2}\rho_{12}^{t}\rho_{11}^{-1}\rho_{12}\rho_{22}^{-1/2}$.

Por se tratarem de variáveis artificiais, as variáveis canônicas não possuem significado físico. Se $\underset{\sim}{X}^{(1)}$ (px1) e $\underset{\sim}{X}^{(2)}$ (qx1) são utilizados, os coeficientes de $\underset{\sim}{a}$ e $\underset{\sim}{b}$ têm as unidades dos correspondentes coeficientes de $\underset{\sim}{X}^{(1)}$ e de $\underset{\sim}{X}^{(2)}$. Se as variáveis padronizadas forem utilizadas, então, os coeficientes canônicos não possuem unidades de mensuração e não dependem da escala das variáveis. Em geral, é dada uma interpretação subjetiva para as variáveis canônicas de acordo com a magnitude das correlações das variáveis originais com as variáveis canônicas em foco. Muitos pesquisadores preferem fazer tal relacionamento utilizando os coeficientes canônicos estandardizados.

Sejam A (pxp) e B (qxq) matrizes definidas pelos vetores canônicos:

$$A = \begin{bmatrix} \underset{\sim}{a}_1^t \\ \underset{\sim}{a}_2^t \\ \vdots \\ \underset{\sim}{a}_p^t \end{bmatrix} \text{ e } B = \begin{bmatrix} \underset{\sim}{b}_1^t \\ \underset{\sim}{b}_2^t \\ \vdots \\ \underset{\sim}{b}_q^t \end{bmatrix} \tag{10.30}$$

É possível definir os vetores de todas as p ou q variáveis canônicas simultaneamente por:

$$\underset{\sim}{U} = \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_p \end{bmatrix} = A\underset{\sim}{X}^{(1)} \quad \text{e} \quad \underset{\sim}{V} = \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_q \end{bmatrix} = B\underset{\sim}{X}^{(2)} \tag{10.31}$$

Logo,

$$\text{Cov}\left(\underset{\sim}{U}, \underset{\sim}{X}^{(1)}\right) = \text{Cov}\left(A\underset{\sim}{X}^{(1)}, \underset{\sim}{X}^{(1)}\right) = A\text{Cov}\left(\underset{\sim}{X}^{(1)}\right) = A\Sigma_{11} \tag{10.32}$$

A matriz de correlação entre as p variáveis originais de $\underset{\sim}{X}^{(1)}$ e as p variáveis canônicas de $\underset{\sim}{U}$ é dada pela "covariância" entre as p variáveis canônicas, as quais já são estandardizadas, e as p variáveis de $\underset{\sim}{X}^{(1)}$ padronizadas. A padronização de $\underset{\sim}{X}^{(1)}$ é dada por:

$$V_{11}^{-1/2}\underset{\sim}{X}^{(1)} = \begin{bmatrix} \dfrac{1}{\sqrt{\sigma_{11}^{(1)}}} & 0 & \cdots & 0 \\ 0 & \dfrac{1}{\sqrt{\sigma_{22}^{(1)}}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \dfrac{1}{\sqrt{\sigma_{pp}^{(1)}}} \end{bmatrix} \begin{bmatrix} X_1^{(1)} \\ X_2^{(1)} \\ \vdots \\ X_p^{(1)} \end{bmatrix} \tag{10.33}$$

Assim,

$$\rho_{\underset{\sim}{U},\underset{\sim}{X}^{(1)}} = \mathrm{Corr}\left(\underset{\sim}{U}, \underset{\sim}{X}^{(1)}\right) = \mathrm{Cov}\left(A\underset{\sim}{X}^{(1)}, V_{11}^{-1/2}\underset{\sim}{X}^{(1)}\right) = A\Sigma_{11}V_{11}^{-1/2} \tag{10.34}$$

Cálculo semelhante é realizado para os pares $\left(\underset{\sim}{U},\underset{\sim}{X}^{(2)}\right)$, $\left(\underset{\sim}{V},\underset{\sim}{X}^{(2)}\right)$ e $\left(\underset{\sim}{V},\underset{\sim}{X}^{(1)}\right)$ que resulta em:

$$\begin{cases} \rho_{\underset{\sim}{U},\underset{\sim}{X}^{(2)}} = A\Sigma_{12}V_{22}^{-1/2} \quad (p\times q) \\ \rho_{\underset{\sim}{V},\underset{\sim}{X}^{(2)}} = B\Sigma_{22}V_{22}^{-1/2} \quad (q\times q) \\ \rho_{\underset{\sim}{V},\underset{\sim}{X}^{(1)}} = B\Sigma_{12}^{t}V_{11}^{-1/2} \quad (q\times p) \end{cases} \tag{10.35}$$

em que $V_{22}^{-1/2}$ é uma matriz diagonal (qxq) com o i-ésimo elemento dado por $1/\sqrt{\sigma_{ii}^{(2)}}$.

Para as variáveis canônicas calculadas de matrizes de correlação ρ, a interpretação pode ser realizada alternativamente pelas correlações entre as variáveis canônicas e as variáveis padronizadas. Sejam A_Z (pxp) e B_Z (qxq) matrizes compostas dos coeficientes canônicos de $\underset{\sim}{Z}^{(1)}$ e $\underset{\sim}{Z}^{(2)}$, respectivamente. As correlações entre as variáveis canônicas e as variáveis padronizadas são dadas por:

$$\begin{cases} \rho_{\underset{\sim}{U},\underset{\sim}{Z}^{(1)}} = A_Z \rho_{11}; & \rho_{\underset{\sim}{V},\underset{\sim}{Z}^{(2)}} = B_Z \rho_{22} \\ \rho_{\underset{\sim}{U},\underset{\sim}{Z}^{(2)}} = A_Z \rho_{12}; & \rho_{\underset{\sim}{V},\underset{\sim}{Z}^{(1)}} = B_Z \rho_{12}^t \end{cases} \tag{10.36}$$

As matrizes de correlação (10.34), (10.35) com (10.36), apresentam, no entanto, os mesmos valores numéricos, como por exemplo $\rho_{\underset{\sim}{U},\underset{\sim}{Z}^{(1)}} = \rho_{\underset{\sim}{U},\underset{\sim}{X}^{(1)}}$, e assim por diante. Verifica-se facilmente isso por:

$$\rho_{\underset{\sim}{U},\underset{\sim}{X}^{(1)}} = A\Sigma_{11}V_{11}^{-1/2} = AV_{11}^{1/2}V_{11}^{-1/2}\Sigma_{11}V_{11}^{-1/2} = A_Z\rho_{11} = \rho_{\underset{\sim}{U},\underset{\sim}{Z}^{(1)}}$$

ou seja, a correlação não é afetada pela padronização (mudança de escala).

10.3. Variáveis e correlações canônicas amostrais

Uma amostra aleatória de tamanho n em cada conjunto de (p+q) variáveis aleatórias $\underset{\sim}{X}^{(1)}$ (px1) e $\underset{\sim}{X}^{(2)}$ (qx1), dada por $\underset{\sim}{X}_1^{(1)}, \underset{\sim}{X}_2^{(1)}, \cdots, \underset{\sim}{X}_n^{(1)}$ e $\underset{\sim}{X}_1^{(2)}, \underset{\sim}{X}_2^{(2)}, \cdots, \underset{\sim}{X}_n^{(2)}$ possui vetores de médias amostrais dados por:

$$\overline{\underset{\sim}{X}} = \begin{bmatrix} \overline{\underset{\sim}{X}}^{(1)} \\ \\ \overline{\underset{\sim}{X}}^{(2)} \end{bmatrix} = \begin{bmatrix} \overline{\underset{\sim}{X}}_1^{(1)} \\ \vdots \\ \overline{\underset{\sim}{X}}_p^{(1)} \\ \\ \overline{\underset{\sim}{X}}_1^{(2)} \\ \vdots \\ \overline{\underset{\sim}{X}}_q^{(2)} \end{bmatrix} \tag{10.37}$$

Em que:

$$\overline{\underset{\sim}{X}}^{(1)} = \frac{1}{n}\sum_{j=1}^{n} \underset{\sim}{X}_j^{(1)} \quad \text{e} \quad \overline{\underset{\sim}{X}}^{(2)} = \frac{1}{n}\sum_{j=1}^{n} \underset{\sim}{X}_j^{(2)} \tag{10.38}$$

A matriz de correlação amostral S (p+q x p+q) é dada por:

$$S = \begin{array}{c} \\ p \\ q \end{array}\!\!\begin{array}{c} \begin{array}{cc} p & \quad q \end{array} \\ \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \end{array} \tag{10.39}$$

em que $S_{k\ell} = \frac{1}{n-1}\sum\left(\underset{\sim}{X}_j^{(k)} - \overline{\underset{\sim}{X}}_j^{(k)}\right)\left(\underset{\sim}{X}_j^{(\ell)} - \overline{\underset{\sim}{X}}_j^{(\ell)}\right)^t$, $k, \ell = 1, 2$.

As k-ésimas variáveis canônicas amostrais são dadas pelas combinações lineares:

$$\begin{cases} \hat{U}_k = \underset{\sim}{\hat{a}}_k^t \underset{\sim}{X}^{(1)} \\ \\ \hat{V}_k = \underset{\sim}{\hat{b}}_k^t \underset{\sim}{X}^{(2)} \end{cases} \tag{10.40}$$

que maximizam a k-ésima correlação canônica amostral dada por:

$$r_{\hat{U}_k, \hat{V}_k} = \frac{\underset{\sim}{\hat{a}}_k^t S_{12} \underset{\sim}{\hat{b}}_k}{\sqrt{\underset{\sim}{\hat{a}}_k^t S_{11} \underset{\sim}{\hat{a}}_k}\sqrt{\underset{\sim}{\hat{b}}_k^t S_{22} \underset{\sim}{\hat{b}}_k}} \tag{10.41}$$

O processo de maximização de (10.41) segue estritamente os mesmos passos da maximização de (10.7), substituindo apenas Σ_{11}, Σ_{22} e Σ_{12} por S_{11}, S_{22} e S_{12}, respectivamente. As equações homogêneas correspondentes ao máximo são dadas por:

$$\begin{cases} \left(S_{12} S_{22}^{-1} S_{12}^t - \hat{\lambda}_k S_{11}\right) \underset{\sim}{\hat{a}}_k = \underset{\sim}{0} \quad \text{(a)} \\ \\ \left(S_{12}^t S_{11}^{-1} S_{12} - \hat{\lambda}_k S_{22}\right) \underset{\sim}{\hat{b}}_k = \underset{\sim}{0} \quad \text{(b)} \end{cases} \tag{10.42}$$

Em que o máximo de $r_{\hat{u}_k, \hat{v}_k}$ é dado por $\sqrt{\hat{\lambda}_k}$, para os autovetores $\underset{\sim}{\hat{a}}_k$ e $\underset{\sim}{\hat{b}}_k$ obtidos por:

$$\begin{cases} \hat{\underset{\sim}{a}}_k = S_{11}^{-1/2}\hat{\underset{\sim}{c}}_k & \text{(a)} \\ \\ \hat{\underset{\sim}{b}}_k = S_{22}^{-1/2}\hat{\underset{\sim}{d}}_k & \text{(b)} \end{cases} \tag{10.43}$$

sendo que $\hat{\underset{\sim}{c}}_k$ é k-ésimo autovetor de $S_{11}^{-1/2}S_{12}S_{22}^{-1}S_{12}^{t}S_{11}^{-1/2}$ e $\hat{\underset{\sim}{d}}_k$ o k-ésimo autovetor de $S_{22}^{-1/2}S_{12}^{t}S_{11}^{-1}S_{12}S_{22}^{-1/2}$; $\hat{\lambda}_k$ é o k-ésimo autovalor de ambas as matrizes, por serem idênticos; k=1, 2, ..., p≤q.

As variáveis canônicas amostrais têm as seguintes propriedades:

1. Variâncias amostrais unitárias

$$\text{V}\hat{\text{a}}\text{r}\left(\hat{U}_k\right) = \text{V}\hat{\text{a}}\text{r}\left(\hat{V}_k\right) = 1 \tag{10.44}$$

2. Correlações amostrais:

$$r_{\hat{U}_k;\hat{U}_\ell} = r_{\hat{V}_k;\hat{V}_\ell} = r_{\hat{U}_k;\hat{V}_\ell} = 0 \quad (k \neq \ell) \tag{10.45}$$

3. Correlação amostral máxima:

$$r_{\hat{U}_k;\hat{V}_k} = \sqrt{\hat{\lambda}_k} \tag{10.46}$$

Sejam as matrizes $\hat{A}\,(p \times p)$ e $\hat{B}\,(q \times q)$ definidas pelos vetores canônicos amostrais:

$$\hat{A} = \begin{bmatrix} \hat{\underset{\sim}{a}}_1^t \\ \hat{\underset{\sim}{a}}_2^t \\ \vdots \\ \hat{\underset{\sim}{a}}_p^t \end{bmatrix} \text{ e } \hat{B} = \begin{bmatrix} \hat{\underset{\sim}{b}}_1^t \\ \hat{\underset{\sim}{b}}_2^t \\ \vdots \\ \hat{\underset{\sim}{b}}_q^t \end{bmatrix} \tag{10.47}$$

Analogamente a (10.31) definem-se:

$$\hat{\underset{\sim}{U}} = \begin{bmatrix} \hat{U}_1 \\ \hat{U}_2 \\ \vdots \\ \hat{U}_p \end{bmatrix} = \hat{A}\underset{\sim}{X}^{(1)} \quad \text{e} \quad \hat{\underset{\sim}{V}} = \begin{bmatrix} \hat{V}_1 \\ \hat{V}_2 \\ \vdots \\ \hat{V}_q \end{bmatrix} = \hat{B}\underset{\sim}{X}^{(2)} \tag{10.48}$$

As correlações entre as variáveis canônicas amostrais e as variáveis originais de cada um dos grupos podem ser obtidas. Para isso definiu-se as matrizes diagonais $D_{11}^{-1/2} = \mathrm{Diag}\left(1/\sqrt{S_{ii}^{(1)}}\right)$, (pxp) e $D_{22}^{-1/2} = \mathrm{Diag}\left(1/\sqrt{S_{ii}^{(2)}}\right)$, (qxq).

1. Matriz de correlações entre $\hat{\underset{\sim}{U}}$ e $\underset{\sim}{X}^{(1)}$

$$R_{\hat{\underset{\sim}{U}},\underset{\sim}{X}^{(1)}} = \hat{A}S_{11}D_{11}^{-1/2} \tag{10.49}$$

2. Matriz de correlações entre $\hat{\underset{\sim}{U}}$ e $\underset{\sim}{X}^{(2)}$

$$R_{\hat{\underset{\sim}{U}},\underset{\sim}{X}^{(2)}} = \hat{A}S_{12}D_{22}^{-1/2} \tag{10.50}$$

3. Matriz de correlações entre $\hat{\underset{\sim}{V}}$ e $\underset{\sim}{X}^{(1)}$

$$R_{\hat{\underset{\sim}{V}},\underset{\sim}{X}^{(1)}} = \hat{B}S_{12}^{t}D_{11}^{-1/2} \tag{10.51}$$

4. Matriz de correlações entre $\hat{\underset{\sim}{V}}$ e $\underset{\sim}{X}^{(2)}$

$$R_{\hat{\underset{\sim}{V}},\underset{\sim}{X}^{(2)}} = \hat{B}S_{22}D_{22}^{-1/2} \tag{10.52}$$

Para variáveis padronizadas, as variáveis canônicas correspondentes são:

$$\hat{\underset{\sim}{U}} = \begin{bmatrix} \hat{U}_1 \\ \hat{U}_2 \\ \vdots \\ \hat{U}_p \end{bmatrix} = \hat{A}_Z \underset{\sim}{Z}^{(1)} \quad \text{e} \quad \hat{\underset{\sim}{V}} = \begin{bmatrix} \hat{V}_1 \\ \hat{V}_2 \\ \vdots \\ \hat{V}_q \end{bmatrix} = \hat{B}_Z \underset{\sim}{Z}^{(2)} \tag{10.53}$$

em que:

$$\hat{A}_Z = \hat{A}D^{1/2} \quad \text{e} \quad \hat{B}_Z = \hat{B}D^{1/2} \tag{10.54}$$

Sendo que $\hat{\underset{\sim}{a}}_z$ e $\hat{\underset{\sim}{b}}_z$ para as variáveis padronizadas são obtidos da mesma forma que os respectivos vetores para variáveis não padronizadas, substituindo-se nas expressões correspondentes S_{11}, S_{22} e S_{12} por R_{11}, R_{22} e R_{12}, respectivamente. A relação (10.54) se verifica para o caso de variáveis canônicas, mas não se pode estabelecer a mesma relação para os componentes principais de matriz de covariância e matriz de correlação, como apontado por Johnson e Wichern (1998). As matrizes de correlações entre as variáveis de cada grupo padronizadas e as respectivas variáveis canônicas são dadas por:

$$\begin{cases} R_{\hat{\underset{\sim}{U}},\underset{\sim}{Z}^{(1)}} = \hat{A}_Z R_{11} = \hat{A}_Z^{-1} & R_{\hat{\underset{\sim}{V}},\underset{\sim}{Z}^{(1)}} = \hat{B}_Z R_{12}^t \\ R_{\hat{\underset{\sim}{U}},\underset{\sim}{Z}^{(2)}} = \hat{A}_Z R_{12} & R_{\hat{\underset{\sim}{V}},\underset{\sim}{Z}^{(2)}} = \hat{B}_Z R_{22} = \hat{B}_Z^{-1} \end{cases} \tag{10.55}$$

Da mesma forma, é fácil verificar que as correlações não são afetadas pela padronização, ou seja, as correlações obtidas em (10.49) a (10.52) são as mesmas as correspondentes em (10.55).

Uma importante avaliação da qualidade do potencial das variáveis canônicas é medir o poder de resumo da variabilidade contida respectivo conjunto. Duas formas básicas são descritas: na primeira apresenta-se uma matriz de erro da aproximação e na segunda calcula-se a proporção da variância explicada pelas variáveis canônicas para cada grupo de variáveis.

As matrizes de erro são obtidas como se segue, admitindo as definições $\hat{\underset{\sim}{U}} = \hat{A}\underset{\sim}{X}^{(1)}$ e $\hat{\underset{\sim}{V}} = \hat{B}\underset{\sim}{X}^{(2)}$. Logo, é possível definir:

$$\underset{\sim}{X}^{(1)} = \hat{A}^{-1}\hat{\underset{\sim}{U}} \quad \text{e} \quad \underset{\sim}{X}^{(2)} = \hat{B}^{-1}\hat{\underset{\sim}{V}} \tag{10.56}$$

Como $\hat{A}$ e $\hat{B}$ são dadas por:

$$\hat{A} = \hat{P}^{(1)t} S_{11}^{-1/2} = \begin{bmatrix} \hat{\underset{\sim}{c}}_1^t \\ \hat{\underset{\sim}{c}}_2^t \\ \vdots \\ \hat{\underset{\sim}{c}}_p^t \end{bmatrix} S_{11}^{-1/2} \quad \text{e} \quad \hat{B} = \hat{P}^{(2)t} S_{22}^{-1/2} = \begin{bmatrix} \hat{\underset{\sim}{d}}_1^t \\ \hat{\underset{\sim}{d}}_2^t \\ \vdots \\ \hat{\underset{\sim}{d}}_p^t \end{bmatrix} S_{22}^{-1/2} \tag{10.57}$$

Então:

$$\hat{A}^{-1} = S_{11}^{1/2} \hat{P}^{(1)} \quad \text{e} \quad \hat{B}^{-1} = S_{22}^{1/2} \hat{P}^{(2)} \tag{10.58}$$

devido a $\hat{P}^{(1)}$ e $\hat{P}^{(2)}$ serem matrizes ortogonais de autovetores, é fácil perceber que $\left(\hat{P}^{(1)t}\right)^{-1} = \hat{P}^{(1)}$ e $\left(\hat{P}^{(2)t}\right)^{-1} = \hat{P}^{(2)}$.

Das definições de $\hat{\underset{\sim}{U}}$ e $\hat{\underset{\sim}{V}}$ sabe-se que a covariância entre eles é uma matriz diagonal $\hat{\Lambda}$ (pxq) com $\sqrt{\hat{\lambda}_k}$ na k-ésima diagonal para k=1, 2,...p, e cujas demais p-q colunas são formadas de zeros. Assim,

$$\begin{cases} \hat{Cov}\left(\hat{\underset{\sim}{U}}, \hat{\underset{\sim}{V}}\right) = \hat{A} S_{12} \hat{B}^t = \hat{P}^{(1)t} S_{11}^{-1/2} S_{12} S_{22}^{-1/2} \hat{P}^{(2)} = \hat{\Lambda} \\ \hat{Cov}\left(\hat{\underset{\sim}{U}}\right) = \hat{A} S_{11} \hat{A}^t = I \\ \hat{Cov}\left(\hat{\underset{\sim}{V}}\right) = \hat{B} S_{22} \hat{B}^t = I \end{cases} \tag{10.59}$$

Assim,

$$\hat{A}S_{12}\hat{B}^t = \hat{\Lambda}$$

$$S_{12}\hat{B}^t = \hat{A}^{-1}\hat{\Lambda}$$

$$S_{12} = \hat{A}^{-1}\hat{\Lambda}\left(\hat{B}^{-1}\right)^t$$

Da mesma forma:

$$S_{11} = \hat{A}^{-1}\left(\hat{A}^{-1}\right)^t \quad \text{e} \quad S_{22} = \hat{B}^{-1}\left(\hat{B}^{-1}\right)^t$$

A idéia é reter um número r menor ou igual a p de variáveis canônicas em cada grupo. O número r é escolhido de determinada forma que a covariância amostral dentro de grupo seja reproduzida de uma forma satisfatória. Da mesma forma é desejável uma boa aproximação das covariâncias entre grupos S_{12}. Sejam, então, as matrizes compostas das r (r≤p) primeiros autovalores e autovetores de $S_{11}^{-1/2}S_{12}S_{22}^{-1}S_{12}^tS_{11}^{-1/2}$ e de $S_{22}^{-1/2}S_{12}^tS_{11}^{-1}S_{12}S_{22}^{-1/2}$ definidas por:

$$\hat{A}_r = \hat{P}_r^{(1)t}S_{11}^{-1/2} = \begin{bmatrix} \hat{\underset{\sim}{c}}_1^t \\ \hat{\underset{\sim}{c}}_2^t \\ \vdots \\ \hat{\underset{\sim}{c}}_r^t \end{bmatrix} S_{11}^{-1/2} \qquad (10.60)$$

$$\hat{B}_r = \hat{P}_r^{(2)t} S_{22}^{-1/2} = \begin{bmatrix} \hat{\underset{\sim}{d}}_1^t \\ \hat{\underset{\sim}{d}}_2^t \\ \vdots \\ \hat{\underset{\sim}{d}}_r^t \end{bmatrix} S_{22}^{-1/2} \tag{10.61}$$

$$\hat{\Lambda}_r = \begin{bmatrix} \sqrt{\hat{\lambda}_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\hat{\lambda}_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\hat{\lambda}_r} \end{bmatrix} \tag{10.62}$$

Assim, definem-se as matrizes:

$$\hat{A}_r = S_{11}^{1/2} \hat{P}_r^{(1)} \quad \text{e} \quad \hat{B}_r = S_{22}^{1/2} \hat{P}_r^{(2)} \tag{10.63}$$

Considerando as matrizes de resíduos E_{11}, E_{22} e E_{12} das reproduções de S_{11}, S_{22} e S_{12}, respectivamente, têm-se:

$$\begin{cases} E_{11} = S_{11} - \left(\hat{A}_r^{-1}\right)\left(\hat{A}_r^{-1}\right)^t & \text{(a)} \\ E_{22} = S_{22} - \left(\hat{B}_r^{-1}\right)\left(\hat{B}_r^{-1}\right)^t & \text{(b)} \\ E_{12} = S_{12} - \left(\hat{A}_r^{-1}\right)\hat{\Lambda}_r \left(\hat{B}_r^{-1}\right)^t & \text{(c)} \end{cases} \tag{10.64}$$

A segunda alternativa relacionada a essa que apresenta em simples número a explicação do respectivo conjunto, em substituição aos p(p-1)/2, q(q-1)/2 ou pq valores de E_{11}, E_{22} e E_{12}. Como $\mathrm{tr}(S_{11}) = \mathrm{tr}\left[\left(\hat{A}_r^{-1}\right)\left(\hat{A}_r^{-1}\right)^t\right] + \mathrm{tr}(E_{11})$, e assim por diante para as demais matrizes, a explicação das r variáveis canônicas para o seu respectivo conjunto é dada por:

$$\begin{cases} \%\mathrm{Exp}\left(\hat{U}_1, \hat{U}_2, \cdots, \hat{U}_r\right) \text{ de } \underset{\sim}{X}^{(1)} = 100 \times \left(1 - \dfrac{\mathrm{tr}(E_{11})}{\mathrm{tr}(S_{11})}\right) \text{ (a)} \\ \\ \%\mathrm{Exp}\left(\hat{V}_1, \hat{V}_2, \cdots, \hat{V}_r\right) \text{ de } \underset{\sim}{X}^{(2)} = 100 \times \left(1 - \dfrac{\mathrm{tr}(E_{22})}{\mathrm{tr}(S_{22})}\right) \text{ (b)} \end{cases} \qquad (10.65)$$

10.4. Inferências para grandes amostras

Quando Σ_{12}=0 as variáveis canônicas $U = \underset{\sim}{a}^t \underset{\sim}{X}^{(1)}$ e $V = \underset{\sim}{b}^t \underset{\sim}{X}^{(2)}$ possuem covariância nula para todos os pares de vetores $\underset{\sim}{a}$ e $\underset{\sim}{b}$. Dessa forma, não existem vantagens em realizar uma análise de correlação canônica. Então, é evidente que um teste de hipótese de que (Σ_{12}) seja igual a uma matriz nula é primordial para a validação da análise de correlação canônica. A seguir é apresentado o teste para a hipótese:

$$H_0 : \Sigma_{12} = 0 \ (p \times q) \ \text{ vs } \ H_1 : \Sigma_{12} \neq 0 \qquad (10.66)$$

Seja o vetor aleatório normal de dimensão (p+qx1) com média μ e covariância Σ, dado por:

$$\underset{\sim}{X}_j = \begin{bmatrix} \underset{\sim}{X}_j^{(1)} \\ \underset{\sim}{X}_j^{(2)} \end{bmatrix}$$

cuja covariância pode ser particionada em:

$$\Sigma = \begin{array}{c} \\ p \\ q \end{array}\!\!\begin{array}{c} \begin{array}{cc} p & \;\; q \end{array} \\ \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \end{array}$$

O teste da razão de verossimilhança para a hipótese (10.66) dado por:

$$\chi_c^2 = -2\ln(\Lambda) = n\ln\left(\frac{|S_{11}||S_{22}|}{|S|}\right) = -n\ln\left[\prod_{i=1}^{p}\left(1-\hat{\lambda}_i\right)\right] \qquad (10.67)$$

tem distribuição qui-quadrado com ν=pq graus de liberdade.

O teste de razão de verossimilhança compara a variância amostral generalizada sob H_0:

$$\begin{vmatrix} S_{11} & 0 \\ 0 & S_{22} \end{vmatrix} = |S_{11}||S_{22}|$$

com a variância generalizada irrestrita, |S|. O primeiro caso com p(p-1)/2 + q(q-1)/2 parâmetros e o segundo com (p+q)(p+q-1)/2. A diferença é igual a ν=pq parâmetros, que é igual aos graus de liberdade do teste em questão. Bartlett (1939) sugere uma correção para uma melhor aproximação de qui-quadrado, substituindo n em (10.67) por n-1-(p+q+1)/2. O teste com a correção de Bartlett (1939) é dado por:

$$\chi_c^2 = \left[n - 1 - \frac{1}{2}(p+q+1)\right]\ln\left(\frac{|S_{11}||S_{22}|}{|S|}\right) = -\left[n - 1 - \frac{1}{2}(p+q+1)\right]\ln\left[\prod_{i=1}^{p}\left(1-\hat{\lambda}_i\right)\right] \quad (10.68)$$

Se a hipótese nula $H_0 : \Sigma_{12} = 0\left(\rho_1 = \rho_2 = \cdots = \rho_p = 0\right)$ for rejeitada, é natural buscar um número de correlações canônicas r que diferem significativamente de zero. Em que ρ_k é a notação abreviada de $\rho_{U_k;V_k}$. Bartlett (1938) sugere um teste seqüencial baseado na razão de verossimilhança. A princípio, testar a hipótese de que a primeira correlação canônica é não nula e as demais (p-1) são nulas; em seguida, testar que as duas primeiras são não nulas e as demais (p-2) são nulas; e assim por diante. Para o k-ésimo passo desse processo testar a hipótese $H_0^{(k)}$ dada por:

$$\begin{cases} H_0^{(k)} : \rho_1 \neq 0, \rho_2 \neq 0, \cdots, \rho_k \neq 0, \rho_{k+1} = \rho_{k+2} = \cdots = \rho_p = 0 \\ \\ H_1^{(k)} : \rho_i \neq 0 \ \text{ para algum } i \geq k+1 \end{cases} \quad (10.69)$$

O teste dessa hipótese incorporando a correção de Bartlett (1939) pode ser realizado por:

$$\chi_c^2 = -\left[n-1-\frac{1}{2}(p+q+1)\right]\ln\left[\prod_{i=k+1}^{p}\left(1-\hat{\lambda}_i\right)\right] \tag{10.70}$$

o qual possui distribuição de qui-quadrado com ν=(p-k)(q-k) graus de liberdade. O teste é realizado para k=1, 2, ..., (p-1).

Cada hipótese da seqüência H_0, $H_0^{(1)}$, $H_0^{(2)}$, etc. é testada uma de cada vez até que $H_0^{(k)}$ não seja rejeitada para algum k. O valor nominal da significância não é α, e possui difícil determinação. O teste é especialmente útil para os dados normais e deve ser interpretado com cautela, e possivelmente deva melhor ser usado como um guia não muito refinado de seleção do número r de variáveis canônicas a ser retido. As distribuições amostrais das variáveis canônicas possuem um estudo mais detalhado em Kshirsagar (1972).

Uma outra opção para esse teste é apresentada por Morrisson (1976) que afirma que a distribuição do maior autovalor segue a distribuição da maior raiz característica de Roy, com S=min(p,q), m=(|P-Q| -1)/2 e n=(n-p-q-2)/2.

O teste anterior foi generalizado por Wilks (1935) para avaliar a independência entre k grupos de variáveis. O teste de razão de verossimilhança para a hipótese de independência entre k-grupos da distribuição normal multivariada é apresentado doravante. Seja Σ, matriz de covariância para todas as variáveis, particionada em k grupos, cada um com p_i variáveis; a sub-matriz Σ_{ij} de dimensão $p_i x p_j$ ($i \neq j$=1, 2, ...,k) é uma partição de Σ que contem as correspondentes covariâncias entre as p_i variáveis do i-ésimo grupo com as p_j variáveis do j-ésimo grupo. A hipótese de interesse é:

$$\begin{cases} H_0 : \Sigma_{ij} = 0 \text{ para todo } i \neq j=1, 2, ..., k \\ \\ H_1 : \Sigma_{ij} \neq 0 \text{ para algum } i \neq j=1, 2, ..., k \end{cases} \tag{10.71}$$

Cujo teste apresentado por Wilks (1935) depende da quantidade:

$$\Lambda_c = \frac{|S|}{|S_{11}| \times |S_{22}| \times \cdots \times |S_{kk}|} \tag{10.72}$$

cuja distribuição é muito complicada. Mas Box (1949) obteve boa aproximação de qui-quadrado com ν graus de liberdade. O teste proposto é:

$$\chi_c^2 = -\frac{n-1}{C} \ln\left(\Lambda_c\right) \tag{10.73}$$

em que:

$$\begin{cases} C^{-1} = 1 - \dfrac{1}{12\nu(n-1)}\left(2\Gamma_3 + 3\Gamma_2\right) \\ \\ \nu = \dfrac{1}{2}\Gamma_2 \end{cases} \tag{10.74}$$

e

$$\Gamma_S = \left(\sum_{i=1}^{k} p_i\right)^S - \sum_{i=1}^{k} p_i^S; \qquad S = 2, 3 \tag{10.75}$$

Se k=2 com p_1=p e p_2=q, o teste (10.73) é exatamente o mesmo de (10.68). É conveniente que se saliente que se os testes forem aplicados sobre a matriz de correlação, os resultados são equivalentes aos obtidos para a matriz de covariâncias, substituindo-se S por R nas expressões anteriores.

10.5. Exercícios

10.5.1. Verifique que a derivação do máximo de (10.7) pode ser obtida a partir de (10.16) utilizando o fator de Cholesky F, na transformação linear de $\underset{\sim}{a} = \left(F_{11}^{-1}\right)^t \underset{\sim}{c}$ e de $\underset{\sim}{b} = \left(F_{22}^{-1}\right)^t \underset{\sim}{d}$ no lugar de $\underset{\sim}{a} = \Sigma_{11}^{-1/2}\underset{\sim}{c}$ e de $\underset{\sim}{b} = \Sigma_{22}^{-1/2}\underset{\sim}{d}$, respectivamente; em que, F_{11} e F_{22} são os fatores de Cholesky de Σ_{11} e de Σ_{22}, respectivamente.

10.5.2. Dois testes ($X_1^{(1)}$ e $X_2^{(1)}$) de leitura foram aplicados em n=140 crianças juntamente com dois testes de aritmética ($X_1^{(2)}$ e $X_2^{(2)}$). A matriz de correlação amostral obtida foi:

$$R_{11} = \begin{bmatrix} 1,0000 & 0,6328 \\ 0,6328 & 1,0000 \end{bmatrix};\ R_{22} = \begin{bmatrix} 1,0000 & 0,4248 \\ 0,4248 & 1,0000 \end{bmatrix};\ \text{e}\ R_{12} = \begin{bmatrix} 0,2412 & 0,0586 \\ -0,0553 & 0,0655 \end{bmatrix}$$

a) obtenha todas as variáveis canônicas amostrais e as respectivas correlações máximas.

b) realizar o teste da hipótese:

$$H_0 : \Sigma_{12} = \rho_{12} = 0\ \ (p \times q)\ \ \text{vs}\ \ H_1 : \Sigma_{12} = \rho_{12} \neq 0$$

Se H_0 for rejeitada realizar o teste da hipótese:

$$H_0 : \rho_1 \neq 0; \rho_2 = 0 \quad \text{Vs} \quad H_0 : \rho_2 \neq 0$$

discuta os resultados obtidos.

c) estime as matrizes E_{11}, E_{22} e E_{12} para o primeiro par de variáveis canônicas (r=1).

d) Determine a proporção da variação explicada pelo primeiro par de variáveis canônicas nos dois grupos.

e) calcule a correlação amostral entre $\underset{\sim}{Z}^{(1)}$ e $\underset{\sim}{Z}^{(2)}$ com $\underset{\sim}{U}$ e com $\underset{\sim}{V}$.

10.6. Referências

BARTLETT, M.S. Further aspects of the theory of multiple regression. **Proceedings of the Cambridge Philosophical Society,** v.34, p.33-40, 1938.

BARTLETT, M.S. a note on tests of significance in multivariate analysis. **Proceedings of the Cambridge Philosophical Society,** v.35, p.180-185, 1939.

BOX, G.E.P. A general distribution theory for a class of likelihood criteria, Biometrika. v.36, p.317346, 1949.

HOTELLING, H. The most predictable criterion. **Journal of Educational Psychology**. v.26, p.139-142, 1935.

HOTELLING, H. Relations between two sets of variables. **Biometrika**. v.28, p.321-377, 1936.

JOHNSON, R.A.; WICHERN, D.W. **Applied multivariate statistical analysis.** 4th edition. Prentice Hall, New Jersey, 1998. 816p.

KSHIRSAGAR, A.M. **Multivariate analysis**. New York: Marcel Dekker, 1972.

LAWLEY, D.N. Tests of significance in canonical analysis. **Biometrika.** v.46, p.59-66, 1959.

MORRISON, D.F. **Multivariate statistical methods.** New York: McGraw-Hill, 2d ed., 1976. 307p.

WIKS, S.S. On the independence of k sets of normally distributed statistical variables. **Econometrica.** v.3, p.309-326, 1935.