Correlación Canónica

Graciela Boente

Un investigador recopiló datos sobre tres variables psicológicas, cuatro variables académicas (resultados de exámenes estandarizados) y de género para 600 estudiantes de primer año de universidad.

Está interesado en la forma en que el conjunto de variables psicológicas se relaciona con las variables académicas y el sexo.

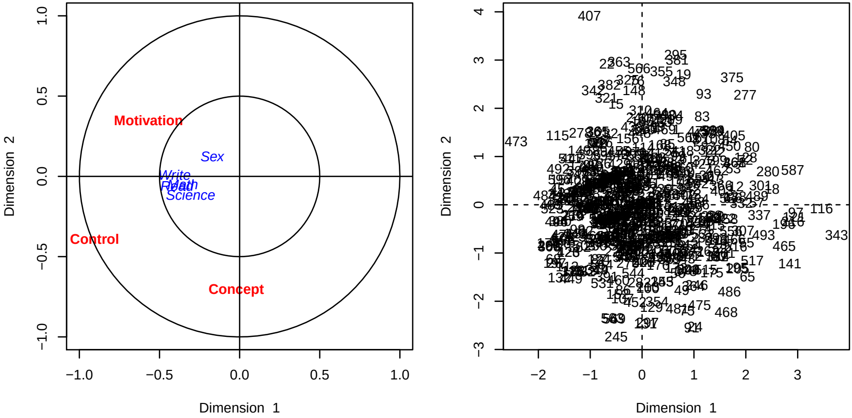
En particular, el investigador está interesado en saber cuántas dimensiones (variables canónicas) son necesarias para comprender la asociación entre los dos conjuntos de variables.

Las variables psicológicas son

- capacidad de control
- autoconcepto
- motivación

y las académicas son las pruebas estandarizadas de

- lectura (READ)
- escritura (WRITING)
- matemáticas (MATH)
- y ciencia (SCIENCE)


Motivation
Control
Concept
Sex
Write
Read
Math
Science
Dimension 1
Dimension 2
−1.0
−0.5
0.0
0.5
1.0
407
473
343
116
465
141
375
277
587
245
Dimension 1
Dimension 2
−2
−1
0
1
2
3
4
−3

Sea $\mathbf{z} = (z_1, \ldots, z_{p+q})^{\mathrm{T}} = \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \in \mathbb{R}^d$ donde $d = p + q$, $\mathbf{x} \in \mathbb{R}^p$, $\mathbf{y} \in \mathbb{R}^q$. Supongamos por simplicidad que $\mathbb{E}\mathbf{z} = \mathbf{0}_d$. Sea

$$\mathrm{Cov}(\mathbf{z}) = \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$$

donde $\boldsymbol{\Sigma}_{11} \in \mathbb{R}^{p\times p}$, $\boldsymbol{\Sigma}_{22} \in \mathbb{R}^{q\times q}$. Supongamos $\boldsymbol{\Sigma} > 0$.

Supongamos $p = 1$

Queremos medir la relación lineal entre z_1 y $\mathbf{y} = (z_2, \ldots, z_d)^{\mathrm{T}} \Rightarrow$ usamos el cuadrado del coeficiente de correlación múltiple, $\rho^2_{1,(23\ldots d)}$ que es *la máxima correlación al cuadrado entre z_1 y cualquier combinación lineal $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}$*

$$\rho^2_{1,(23\ldots d)} = \frac{\boldsymbol{\sigma}_{21}^{\mathrm{T}} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\sigma}_{21}}{\sigma_{11}}$$

donde $\boldsymbol{\sigma}_{21} = \mathrm{Cov}(\mathbf{y}, z_1)$ y el máximo se alcanzaba en $\boldsymbol{\beta}_0 = \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21}$.

Caso General: $q < p$

Queremos medir la asociación entre **x** y **y** que estará dada por:

la máxima correlación al cuadrado entre cualquier combinación lineal $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}$ *y cualquier combinación lineal* $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}$

$$\max_{\boldsymbol{\alpha}\neq\mathbf{0}_p,\boldsymbol{\beta}\neq\mathbf{0}_q} \rho^2_{\boldsymbol{\alpha},\boldsymbol{\beta}}$$

con

$$\rho^2_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \frac{\mathrm{Cov}^2\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}\right)}{\mathrm{Var}\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}\right)\mathrm{Var}\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}\right)} = \frac{\left(\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{12}\boldsymbol{\beta}\right)^2}{\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{11}\boldsymbol{\alpha}\ \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}}$$

Sea $\mathbf{\Sigma}_{jj} = \mathbf{A}_j^{\mathrm{T}} \mathbf{A}_j$,

$$\mathbf{\Psi}_1 = \left(\mathbf{A}_1^{\mathrm{T}}\right)^{-1} \mathbf{\Sigma}_{12} \mathbf{\Sigma}_{22}^{-1} \mathbf{\Sigma}_{21} \mathbf{A}_1^{-1} \qquad \mathbf{\Psi}_2 = \left(\mathbf{A}_2^{\mathrm{T}}\right)^{-1} \mathbf{\Sigma}_{21} \mathbf{\Sigma}_{11}^{-1} \mathbf{\Sigma}_{12} \mathbf{A}_2^{-1}$$

entonces

$$\max_{\boldsymbol{\alpha} \neq \mathbf{0}_p, \boldsymbol{\beta} \neq \mathbf{0}_q} \rho^2_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \rho^2_{\boldsymbol{\alpha}_1,\boldsymbol{\beta}_1} = \rho_1^2$$

Sea $\boldsymbol{\Sigma}_{jj} = \mathbf{A}_j^{\mathrm{T}} \mathbf{A}_j$,

$$\boldsymbol{\Psi}_1 = \left(\mathbf{A}_1^{\mathrm{T}}\right)^{-1} \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \mathbf{A}_1^{-1} \qquad \boldsymbol{\Psi}_2 = \left(\mathbf{A}_2^{\mathrm{T}}\right)^{-1} \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12} \mathbf{A}_2^{-1}$$

entonces

$$\max_{\boldsymbol{\alpha} \neq \mathbf{0}_p, \boldsymbol{\beta} \neq \mathbf{0}_q} \rho^2_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \rho^2_{\boldsymbol{\alpha}_1,\boldsymbol{\beta}_1} = \rho_1^2$$

- ρ_1^2 es el máximo autovalor de $\boldsymbol{\Psi}_1$, es decir, el máximo autovalor de $\boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21}$

Sea $\mathbf{\Sigma}_{jj} = \mathbf{A}_j^{\mathrm{T}} \mathbf{A}_j$,

$$\mathbf{\Psi}_1 = \left(\mathbf{A}_1^{\mathrm{T}}\right)^{-1} \mathbf{\Sigma}_{12} \mathbf{\Sigma}_{22}^{-1} \mathbf{\Sigma}_{21} \mathbf{A}_1^{-1} \qquad \mathbf{\Psi}_2 = \left(\mathbf{A}_2^{\mathrm{T}}\right)^{-1} \mathbf{\Sigma}_{21} \mathbf{\Sigma}_{11}^{-1} \mathbf{\Sigma}_{12} \mathbf{A}_2^{-1}$$

entonces

$$\max_{\boldsymbol{\alpha} \neq \mathbf{0}_p, \boldsymbol{\beta} \neq \mathbf{0}_q} \rho^2_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \rho^2_{\boldsymbol{\alpha}_1,\boldsymbol{\beta}_1} = \rho_1^2$$

- ρ_1^2 es el máximo autovalor de $\mathbf{\Psi}_1$, es decir, el máximo autovalor de $\mathbf{\Sigma}_{11}^{-1} \mathbf{\Sigma}_{12} \mathbf{\Sigma}_{22}^{-1} \mathbf{\Sigma}_{21}$
- $\boldsymbol{\alpha}_1$ es el autovector de $\mathbf{\Sigma}_{11}^{-1} \mathbf{\Sigma}_{12} \mathbf{\Sigma}_{22}^{-1} \mathbf{\Sigma}_{21}$ asociado a ρ_1^2 tal que $\boldsymbol{\alpha}_1^{\mathrm{T}} \mathbf{\Sigma}_{11} \boldsymbol{\alpha}_1 = 1$.

Sea $\mathbf{\Sigma}_{jj} = \mathbf{A}_j^{\mathrm{T}} \mathbf{A}_j$,

$$\mathbf{\Psi}_1 = \left(\mathbf{A}_1^{\mathrm{T}}\right)^{-1} \mathbf{\Sigma}_{12} \mathbf{\Sigma}_{22}^{-1} \mathbf{\Sigma}_{21} \mathbf{A}_1^{-1} \qquad \mathbf{\Psi}_2 = \left(\mathbf{A}_2^{\mathrm{T}}\right)^{-1} \mathbf{\Sigma}_{21} \mathbf{\Sigma}_{11}^{-1} \mathbf{\Sigma}_{12} \mathbf{A}_2^{-1}$$

entonces

$$\max_{\boldsymbol{\alpha} \neq \mathbf{0}_p, \boldsymbol{\beta} \neq \mathbf{0}_q} \rho^2_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \rho^2_{\boldsymbol{\alpha}_1,\boldsymbol{\beta}_1} = \rho_1^2$$

- ρ_1^2 es el máximo autovalor de $\mathbf{\Psi}_1$, es decir, el máximo autovalor de $\mathbf{\Sigma}_{11}^{-1} \mathbf{\Sigma}_{12} \mathbf{\Sigma}_{22}^{-1} \mathbf{\Sigma}_{21}$
- $\boldsymbol{\alpha}_1$ es el autovector de $\mathbf{\Sigma}_{11}^{-1} \mathbf{\Sigma}_{12} \mathbf{\Sigma}_{22}^{-1} \mathbf{\Sigma}_{21}$ asociado a ρ_1^2 tal que $\boldsymbol{\alpha}_1^{\mathrm{T}} \mathbf{\Sigma}_{11} \boldsymbol{\alpha}_1 = 1$.
- $\boldsymbol{\beta}_1$ es el autovector de $\mathbf{\Sigma}_{22}^{-1} \mathbf{\Sigma}_{21} \mathbf{\Sigma}_{11}^{-1} \mathbf{\Sigma}_{12}$ asociado a ρ_1^2 tal que $\boldsymbol{\beta}_1^{\mathrm{T}} \mathbf{\Sigma}_{22} \boldsymbol{\beta}_1 = 1$.

$$\boldsymbol{\beta}_1 = \frac{\mathbf{\Sigma}_{22}^{-1} \mathbf{\Sigma}_{21} \boldsymbol{\alpha}_1}{\rho_1} \qquad \boldsymbol{\alpha}_1 = \frac{\mathbf{\Sigma}_{11}^{-1} \mathbf{\Sigma}_{12} \boldsymbol{\beta}_1}{\rho_1}$$

Definición.

- $\rho_1 = \sqrt{\rho_1^2}$ se llama la primer correlación canónica entre $\mathbf{x}$ y $\mathbf{y}$.
- $u_1 = \boldsymbol{\alpha}_1^{\mathrm{T}}\mathbf{x}$, $v_1 = \boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{y}$ se llaman las primeras variables canónicas. Se cumple $\mathrm{VAR}(u_1) = \mathrm{VAR}(v_1) = 1$

Definición.

- $\rho_1 = \sqrt{\rho_1^2}$ se llama la primer correlación canónica entre **x** y **y**.
- $u_1 = \boldsymbol{\alpha}_1^{\mathrm{T}}\mathbf{x}$, $v_1 = \boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{y}$ se llaman las primeras variables canónicas. Se cumple $\mathrm{VAR}(u_1) = \mathrm{VAR}(v_1) = 1$

Como $\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\alpha}_1 = \rho_1^2\boldsymbol{\Sigma}_{11}\boldsymbol{\alpha}_1$ y $\boldsymbol{\alpha}_1^{\mathrm{T}}\boldsymbol{\Sigma}_{11}\boldsymbol{\alpha}_1 = 1$ tenemos que

$$\rho_1^2 = \boldsymbol{\alpha}_1^{\mathrm{T}}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\alpha}_1$$

O sea, ρ_1^2 es el cuadrado del coeficiente de correlación multiple entre la variable $u_1 = \boldsymbol{\alpha}_1^{\mathrm{T}}\mathbf{x}$ y el vector **y** ya que $\mathrm{VAR}(u_1) = 1$ y $\mathrm{COV}(u_1, \mathbf{y}) = \boldsymbol{\Sigma}_{21}\boldsymbol{\alpha}_1$.

Cómo seguimos?

Este procedimiento puede verse como una técnica de reducción de dimensión en la que $\mathbf{x}$ y $\mathbf{y}$ se reducen a u_1 y v_1 de modo que $\rho^2_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ sea máxima. Pero, la reducción u_1 de $\mathbf{x}$ puede no ser adecuada.

Buscamos $\mathbf{u} = (u_1, \ldots, u_m)^{\mathrm{T}}$ y $\mathbf{v} = (v_1, \ldots, v_m)$ tales que

- $u_j = \boldsymbol{\alpha}_j^{\mathrm{T}}\mathbf{x}$, $v_j = \boldsymbol{\beta}_j^{\mathrm{T}}\mathbf{y}$
- $u_1, \ldots, u_m$ sean no correlacionados
- $v_1, \ldots, v_m$ sean no correlacionados
- $\mathrm{Corr}(u_j, v_j)$ sea máxima en algún sentido

Observemos primero que como hablamos de correlaciones podemos suponer que $\mathrm{Var}(u_j) = \mathrm{Var}(v_j) = 1$, es decir,

$$\boldsymbol{\alpha}_j^{\mathrm{T}}\boldsymbol{\Sigma}_{11}\boldsymbol{\alpha}_j = 1 \qquad \boldsymbol{\beta}_j^{\mathrm{T}}\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}_j = 1$$

Teorema

Supongamos $\boldsymbol{\Sigma}_{11} > 0$, $\boldsymbol{\Sigma}_{22} > 0$, $q < p$, $\boldsymbol{\Sigma}_{jj} = \mathbf{A}_j^{\mathrm{T}} \mathbf{A}_j$. Sean

- $\mathbf{C} = \left(\mathbf{A}_1^{\mathrm{T}}\right)^{-1} \boldsymbol{\Sigma}_{12} \mathbf{A}_2^{-1}$
- $\boldsymbol{\Upsilon}_1 = \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21}$
- $\boldsymbol{\Upsilon}_2 = \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12}$

Teorema

Supongamos $\mathbf{\Sigma}_{11} > 0$, $\mathbf{\Sigma}_{22} > 0$, $q < p$, $\mathbf{\Sigma}_{jj} = \mathbf{A}_j^{\mathrm{T}} \mathbf{A}_j$. Sean

- $\mathbf{C} = (\mathbf{A}_1^{\mathrm{T}})^{-1} \mathbf{\Sigma}_{12} \mathbf{A}_2^{-1}$
- $\mathbf{\Upsilon}_1 = \mathbf{\Sigma}_{11}^{-1} \mathbf{\Sigma}_{12} \mathbf{\Sigma}_{22}^{-1} \mathbf{\Sigma}_{21}$
- $\mathbf{\Upsilon}_2 = \mathbf{\Sigma}_{22}^{-1} \mathbf{\Sigma}_{21} \mathbf{\Sigma}_{11}^{-1} \mathbf{\Sigma}_{12}$
- $1 > \rho_1^2 \geq \rho_2^2 \geq \cdots \geq \rho_m^2 > 0$ con $m = \mathrm{rango}(\mathbf{\Sigma}_{12})$ los autovalores no nulos de $\mathbf{\Upsilon}_1$ (y de $\mathbf{\Upsilon}_2$)

Teorema

Supongamos $\mathbf{\Sigma}_{11} > 0$, $\mathbf{\Sigma}_{22} > 0$, $q < p$, $\mathbf{\Sigma}_{jj} = \mathbf{A}_j^{\mathrm{T}}\mathbf{A}_j$. Sean

- $\mathbf{C} = (\mathbf{A}_1^{\mathrm{T}})^{-1}\mathbf{\Sigma}_{12}\mathbf{A}_2^{-1}$
- $\mathbf{\Upsilon}_1 = \mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}$
- $\mathbf{\Upsilon}_2 = \mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}$
- $1 > \rho_1^2 \geq \rho_2^2 \geq \cdots \geq \rho_m^2 > 0$ con $m = \mathrm{rango}(\mathbf{\Sigma}_{12})$ los autovalores no nulos de $\mathbf{\Upsilon}_1$ (y de $\mathbf{\Upsilon}_2$)
- $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_m$ los autovectores de $\mathbf{\Upsilon}_1$ asociados a $\rho_1^2 \geq \rho_2^2 \geq \cdots \geq \rho_m^2$ tales que $\boldsymbol{\alpha}_j^{\mathrm{T}}\mathbf{\Sigma}_{11}\boldsymbol{\alpha}_j = 1$
- $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m$ los autovectores de $\mathbf{\Upsilon}_2$ asociados a $\rho_1^2 \geq \rho_2^2 \geq \cdots \geq \rho_m^2$ tales que $\boldsymbol{\beta}_j^{\mathrm{T}}\mathbf{\Sigma}_{22}\boldsymbol{\beta}_j = 1$

Teorema

Supongamos $\boldsymbol{\Sigma}_{11} > 0$, $\boldsymbol{\Sigma}_{22} > 0$, $q < p$, $\boldsymbol{\Sigma}_{jj} = \mathbf{A}_j^{\mathrm{T}}\mathbf{A}_j$. Sean

- $\mathbf{C} = (\mathbf{A}_1^{\mathrm{T}})^{-1}\boldsymbol{\Sigma}_{12}\mathbf{A}_2^{-1}$
- $\boldsymbol{\Upsilon}_1 = \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$
- $\boldsymbol{\Upsilon}_2 = \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$
- $1 > \rho_1^2 \geq \rho_2^2 \geq \cdots \geq \rho_m^2 > 0$ con $m = \mathrm{rango}(\boldsymbol{\Sigma}_{12})$ los autovalores no nulos de $\boldsymbol{\Upsilon}_1$ (y de $\boldsymbol{\Upsilon}_2$)
- $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_m$ los autovectores de $\boldsymbol{\Upsilon}_1$ asociados a $\rho_1^2 \geq \rho_2^2 \geq \cdots \geq \rho_m^2$ tales que $\boldsymbol{\alpha}_j^{\mathrm{T}}\boldsymbol{\Sigma}_{11}\boldsymbol{\alpha}_j = 1$
- $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m$ los autovectores de $\boldsymbol{\Upsilon}_2$ asociados a $\rho_1^2 \geq \rho_2^2 \geq \cdots \geq \rho_m^2$ tales que $\boldsymbol{\beta}_j^{\mathrm{T}}\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}_j = 1$

Sea $s \leq m - 1$ y sean $\boldsymbol{\alpha} \in \mathbb{R}^p$ y $\boldsymbol{\beta} \in \mathbb{R}^q$ tales que

$$\begin{aligned} \mathrm{Cov}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}, \boldsymbol{\alpha}_j^{\mathrm{T}}\mathbf{x}) &= 0 \qquad 1 \leq j \leq s \\ \mathrm{Cov}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}, \boldsymbol{\beta}_j^{\mathrm{T}}\mathbf{y}) &= 0 \qquad 1 \leq j \leq s \end{aligned}$$

$$\Longrightarrow$$

i) La máxima correlación al cuadrado entre $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}$ y $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}$ está dada por ρ^2_{s+1} y ocurre cuando $\boldsymbol{\alpha} = \boldsymbol{\alpha}_{s+1}$ y $\boldsymbol{\beta} = \boldsymbol{\beta}_{s+1}$, o sea,

$$\max_{\substack{\boldsymbol{\alpha}\neq \mathbf{0},\boldsymbol{\beta}\neq\mathbf{0} \\ \mathrm{Cov}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x},\boldsymbol{\alpha}_j^{\mathrm{T}}\mathbf{x})=0,\ 1\leq j\leq s \\ \mathrm{Cov}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y},\boldsymbol{\beta}_j^{\mathrm{T}}\mathbf{y})=0,\ 1\leq j\leq s}} \rho^2_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \rho^2_{s+1} = \rho^2_{\boldsymbol{\alpha}_{s+1},\boldsymbol{\beta}_{s+1}}$$

i) La máxima correlación al cuadrado entre $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}$ y $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}$ está dada por ρ^2_{s+1} y ocurre cuando $\boldsymbol{\alpha} = \boldsymbol{\alpha}_{s+1}$ y $\boldsymbol{\beta} = \boldsymbol{\beta}_{s+1}$, o sea,

$$\max_{\substack{\boldsymbol{\alpha} \neq \mathbf{0}, \boldsymbol{\beta} \neq \mathbf{0} \\ \mathrm{Cov}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}, \boldsymbol{\alpha}_j^{\mathrm{T}}\mathbf{x}) = 0,\ 1 \leq j \leq s \\ \mathrm{Cov}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}, \boldsymbol{\beta}_j^{\mathrm{T}}\mathbf{y}) = 0,\ 1 \leq j \leq s}} \rho^2_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \rho^2_{s+1} = \rho^2_{\boldsymbol{\alpha}_{s+1},\boldsymbol{\beta}_{s+1}}$$

ii) $\mathrm{Cov}(\boldsymbol{\alpha}_j^{\mathrm{T}}\mathbf{x}, \boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{x}) = 0$ si $j \neq k$ y $\mathrm{Cov}(\boldsymbol{\beta}_j^{\mathrm{T}}\mathbf{y}, \boldsymbol{\beta}_k^{\mathrm{T}}\mathbf{y}) = 0$ si $j \neq k$

i) La máxima correlación al cuadrado entre $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}$ y $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}$ está dada por ρ^2_{s+1} y ocurre cuando $\boldsymbol{\alpha} = \boldsymbol{\alpha}_{s+1}$ y $\boldsymbol{\beta} = \boldsymbol{\beta}_{s+1}$, o sea,

$$\max_{\substack{\boldsymbol{\alpha} \neq \mathbf{0}, \boldsymbol{\beta} \neq \mathbf{0} \\ \mathrm{Cov}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}, \boldsymbol{\alpha}_j^{\mathrm{T}}\mathbf{x}) = 0,\ 1 \leq j \leq s \\ \mathrm{Cov}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}, \boldsymbol{\beta}_j^{\mathrm{T}}\mathbf{y}) = 0,\ 1 \leq j \leq s}} \rho^2_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \rho^2_{s+1} = \rho^2_{\boldsymbol{\alpha}_{s+1},\boldsymbol{\beta}_{s+1}}$$

ii) $\mathrm{Cov}(\boldsymbol{\alpha}_j^{\mathrm{T}}\mathbf{x}, \boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{x}) = 0$ si $j \neq k$ y $\mathrm{Cov}(\boldsymbol{\beta}_j^{\mathrm{T}}\mathbf{y}, \boldsymbol{\beta}_k^{\mathrm{T}}\mathbf{y}) = 0$ si $j \neq k$

iii) $\mathrm{Var}(\boldsymbol{\alpha}_j^{\mathrm{T}}\mathbf{x}) = \mathrm{Var}(\boldsymbol{\beta}_j^{\mathrm{T}}\mathbf{y}) = 1$

i) La máxima correlación al cuadrado entre $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}$ y $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}$ está dada por ρ_{s+1}^2 y ocurre cuando $\boldsymbol{\alpha} = \boldsymbol{\alpha}_{s+1}$ y $\boldsymbol{\beta} = \boldsymbol{\beta}_{s+1}$, o sea,

$$\max_{\substack{\boldsymbol{\alpha} \neq \mathbf{0}, \boldsymbol{\beta} \neq \mathbf{0} \\ \mathrm{Cov}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}, \boldsymbol{\alpha}_j^{\mathrm{T}}\mathbf{x}) = 0,\ 1 \leq j \leq s \\ \mathrm{Cov}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}, \boldsymbol{\beta}_j^{\mathrm{T}}\mathbf{y}) = 0,\ 1 \leq j \leq s}} \rho_{\boldsymbol{\alpha},\boldsymbol{\beta}}^2 = \rho_{s+1}^2 = \rho_{\boldsymbol{\alpha}_{s+1},\boldsymbol{\beta}_{s+1}}^2$$

ii) $\mathrm{Cov}(\boldsymbol{\alpha}_j^{\mathrm{T}}\mathbf{x}, \boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{x}) = 0$ si $j \neq k$ y $\mathrm{Cov}(\boldsymbol{\beta}_j^{\mathrm{T}}\mathbf{y}, \boldsymbol{\beta}_k^{\mathrm{T}}\mathbf{y}) = 0$ si $j \neq k$

iii) $\mathrm{Var}(\boldsymbol{\alpha}_j^{\mathrm{T}}\mathbf{x}) = \mathrm{Var}(\boldsymbol{\beta}_j^{\mathrm{T}}\mathbf{y}) = 1$

iv) Sea $m = \mathrm{rango}(\boldsymbol{\Sigma}_{12}) \leq \min(q, p) = q$ y tomemos

$$\mathbf{u} = \mathbf{A}^{\mathrm{T}}\mathbf{x} \text{ con } \mathbf{A} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_m, \boldsymbol{\alpha}_{m+1}, \ldots, \boldsymbol{\alpha}_p)$$
$$\mathbf{v} = \mathbf{B}^{\mathrm{T}}\mathbf{y} \text{ con } \mathbf{B} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_q)$$

Luego, si $\mathbf{D}_{\boldsymbol{\rho}} = \mathrm{DIAG}(\rho_1, \ldots, \rho_q)$ con $\rho_i = 0$ si $i > m$

$$\mathrm{Var}\begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_q & \mathbf{0} & \mathbf{D}_{\boldsymbol{\rho}} \\ \mathbf{0} & \mathbf{I}_{p-q} & \mathbf{0} \\ \mathbf{D}_{\boldsymbol{\rho}} & \mathbf{0} & \mathbf{I}_q \end{pmatrix}$$

o sea, $\mathrm{Cov}(u_i, v_j) = 0$ si $i \neq j$ y $\mathrm{Cov}(u_i, v_i) = \rho_i$

Definición.

- $\mathbf{u}$ se llama las variables canónicas del espacio $\mathbf{x}$.
- $\mathbf{v}$ se llama las variables canónicas del espacio $\mathbf{y}$.
- $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_m$ y $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m$ son los vectores canónicos
- u_j es la $j-$ésima variable canónica en el espacio $\mathbf{x}$
- v_j es la $j-$ésima variable canónica en el espacio $\mathbf{y}$

La relación entre $\mathbf{x}$ y $\mathbf{y}$ queda expresada por las correlaciones canónicas $\rho_1^2, \ldots, \rho_m^2$

Como $\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\alpha}_j = \rho_j^2\boldsymbol{\Sigma}_{11}\boldsymbol{\alpha}_j$ y $\boldsymbol{\alpha}_j^{\mathrm{T}}\boldsymbol{\Sigma}_{11}\boldsymbol{\alpha}_j = 1$ tenemos que

$$\rho_j^2 = \boldsymbol{\alpha}_j^{\mathrm{T}}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\alpha}_j$$

O sea, ρ_j^2 es el cuadrado del coeficiente de correlación multiple entre la variable $u_j = \boldsymbol{\alpha}_j^{\mathrm{T}}\mathbf{x}$ y el vector $\mathbf{y}$, pues $\mathrm{VAR}(u_j) = 1$ y $\mathrm{COV}(u_j, \mathbf{y}) = \boldsymbol{\Sigma}_{21}\boldsymbol{\alpha}_j$.

Otro enfoque

$$\mathbf{\Sigma}_{jj} = \mathbf{A}_j^{\mathrm{T}}\mathbf{A}_j \qquad \mathbf{A}_j > 0 \qquad\qquad \mathbf{C} = \left(\mathbf{A}_1^{\mathrm{T}}\right)^{-1}\mathbf{\Sigma}_{12}\mathbf{A}_2^{-1}$$

Por la descomposición de valores singulares

$$\mathbf{C} = \mathbf{L}\begin{pmatrix} \mathbf{D}_\rho \\ \mathbf{0}_{(p-q)\times q} \end{pmatrix}\mathbf{M}^{\mathrm{T}}$$

- $\mathbf{L} = (\mathbf{l}_1, \cdots, \mathbf{l}_p) \in \mathbb{R}^{p\times p}$, $\mathbf{M} = (\mathbf{m}_1, \cdots, \mathbf{m}_q) \in \mathbb{R}^{q\times q}$, ortogonales, $\mathbf{m}_j \in \mathbb{R}^q$ y $\mathbf{l}_j \in \mathbb{R}^p$.
- $\mathbf{D}_{\boldsymbol{\rho}} = \mathrm{DIAG}(\rho_1, \ldots, \rho_q)$, $\rho_1 \geq \rho_2 \geq \cdots \geq \rho_q \geq 0$, ρ_j es la raíz cuadrada de $j-$ésimo autovalor de $\mathbf{C}^{\mathrm{T}}\mathbf{C} = \Psi_2$.

Los vectores $\{\mathbf{m}_1, \cdots, \mathbf{m}_q\}$ son los autovectores de $\mathbf{C}^{\mathrm{T}}\mathbf{C} = \Psi_2$ y

$$\mathbf{l}_j = \frac{\mathbf{C}\mathbf{m}_j}{\rho_j} \qquad 1 \leq j \leq q$$

son los autovectores ortonormales de $\mathbf{C}\mathbf{C}^{\mathrm{T}} = \Psi_1$

Definamos

$$\begin{aligned} \mathbf{u} &= \mathbf{L}^{\mathrm{T}} \left(\mathbf{A}_1^{-1}\right)^{\mathrm{T}} \mathbf{x} = (u_1, \ldots, u_p)^{\mathrm{T}} \qquad u_j = \left(\mathbf{A}_1^{-1}\mathbf{l}_j\right)^{\mathrm{T}} \mathbf{x} \\ \mathbf{v} &= \mathbf{M}^{\mathrm{T}} \left(\mathbf{A}_2^{-1}\right)^{\mathrm{T}} \mathbf{x} = (v_1, \ldots, v_q)^{\mathrm{T}} \qquad v_j = \left(\mathbf{A}_2^{-1}\mathbf{m}_j\right)^{\mathrm{T}} \mathbf{x} \end{aligned}$$

Entonces,

$$\mathrm{VAR}\begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_p & \begin{pmatrix} \mathbf{D}_\rho \\ \mathbf{0} \end{pmatrix} \\ (\mathbf{D}_\rho \quad \mathbf{0}) & \mathbf{I}_q \end{pmatrix}$$

- $\mathrm{VAR}(u_i) = \mathrm{VAR}(u_j) = 1$
- $\mathrm{CORR}(u_i, v_j) = \rho_i\, \delta_{ij}$, $1 \leq i \leq p$, $1 \leq j \leq q$.

Definición. Las variables $u_1, \ldots, u_p$ y $v_1, \ldots, v_q$ se llaman las variables canónicas y los números ρ_j, $\rho_1 \geq \rho_2 \geq \cdots \geq \rho_q \geq 0$ son las correlaciones canónicas.

La cantidad de correlaciones no nulas es $m = \mathrm{rango}(\boldsymbol{\Sigma}_{12})$

Las dos definiciones coinciden

Sean dos puntos $\mathbf{x}_1 \in \mathbb{R}^p$, $\mathbf{x}_2 \in \mathbb{R}^p$ y definamos

$$\mathbf{u}_i = \mathbf{L}^{\mathrm{T}} \left(\mathbf{A}_1^{-1}\right)^{\mathrm{T}} \mathbf{x}_i = (u_{1,i}, \ldots, u_{p,i})^{\mathrm{T}} \qquad u_{j,i} = \left(\mathbf{A}_1^{-1}\mathbf{l}_j\right)^{\mathrm{T}} \mathbf{x}_i \quad i = 1, 2$$

Luego,

$$\|\mathbf{u}_1 - \mathbf{u}_2\|^2 = \|\mathbf{L}^{\mathrm{T}} \left(\mathbf{A}_1^{-1}\right)^{\mathrm{T}} (\mathbf{x}_1 - \mathbf{x}_2)\|^2 = (\mathbf{x}_1 - \mathbf{x}_2)^{\mathrm{T}} \boldsymbol{\Sigma}_{11}^{-1} (\mathbf{x}_1 - \mathbf{x}_2)$$

Sean dos puntos $\mathbf{y}_1 \in \mathbb{R}^p$, $\mathbf{y}_2 \in \mathbb{R}^p$ y definamos

$$\mathbf{v}_i = \mathbf{M}^{\mathrm{T}} \left(\mathbf{A}_2^{-1}\right)^{\mathrm{T}} \mathbf{y}_i = (v_{1,i}, \ldots, v_{q,i})^{\mathrm{T}} \qquad v_{j,i} = \left(\mathbf{A}_2^{-1}\mathbf{m}_j\right)^{\mathrm{T}} \mathbf{y}_i \quad i = 1, 2$$

Luego,

$$\|\mathbf{v}_1 - \mathbf{v}_2\|^2 = \|\mathbf{M}^{\mathrm{T}} \left(\mathbf{A}_2^{-1}\right)^{\mathrm{T}} (\mathbf{y}_1 - \mathbf{y}_2)\|^2 = (\mathbf{y}_1 - \mathbf{y}_2)^{\mathrm{T}} \boldsymbol{\Sigma}_{22}^{-1} (\mathbf{y}_1 - \mathbf{y}_2)$$

O sea, distancias entre puntos del espacio de las variables canónicas representan distancias de Mahalanobis en el espacio original.

Al usar, la reducción de s variables canónicas, las distancias entre puntos del espacio de variables canónicas dan aproximadamente la distancia de Mahalanobis en el espacio original.

Resumen de propiedades de las variables y correlaciones canónicas

- Las variables canónicas son indicadores de los dos conjuntos de variables que se definen por pares, con la condición de máxima correlación
- Los coeficientes de las variables canónicas son los autovectores asociados al mismo autovalor de las matrices

$$\boldsymbol{\Sigma}_{ii}^{-1}\boldsymbol{\Sigma}_{ij}\boldsymbol{\Sigma}_{jj}^{-1}\boldsymbol{\Sigma}_{ji} \qquad i = 1,2 \qquad i \neq j$$

- Si $\boldsymbol{\alpha}_j^{\mathrm{T}}\mathbf{x}$ es una variable canónica también lo es $-\boldsymbol{\alpha}_j^{\mathrm{T}}\mathbf{x}$. Los signos de las variables canónicas suelen tomarse para que la correlación entre las variables canónicas $u_j = \boldsymbol{\alpha}_j^{\mathrm{T}}\mathbf{x}$ y $v_j = \boldsymbol{\beta}_j^{\mathrm{T}}\mathbf{y}$ sean positiva.
- Los cuadrados de las correlaciones canónicas ρ_j^2 son los autovalores de $\boldsymbol{\Upsilon}_1 = \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$ y $\boldsymbol{\Upsilon}_2 = \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$, o sea las raíces de

$$|\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} - \lambda\mathbf{I}_p| = 0$$

Resumen de propiedades de las variables y correlaciones canónicas

- Los cuadrados de las correlaciones canónicas ρ_j^2 son el cuadrado del coeficiente de correlación entre las dos variables canónicas $u_j = \boldsymbol{\alpha}_j^{\mathrm{T}}\mathbf{x}$ y $v_j = \boldsymbol{\beta}_j^{\mathrm{T}}\mathbf{y}$ correspondientes.
- Las correlaciones canónicas son invariantes ante transformaciones lineales no singulares de las variables.
- La primer correlación canónica ρ_1^2 es mayor o igual que el mayor coeficiente de correlación al cuadrado entre una variable de cada conjunto.

  $$\rho_1^2 \geq \mathrm{Corr}^2(x_i, y_j) \qquad \text{para todo} \qquad 1 \leq i \leq p, 1 \leq j \leq q$$
- El coeficiente de correlación canónica ρ_j^2 es el cuadrado del coeficiente de correlación multiple entre la variable $u_j = \boldsymbol{\alpha}_j^{\mathrm{T}}\mathbf{x}$ y el vector $\mathbf{y}$.
- El coeficiente de correlación canónica ρ_j^2 es el cuadrado del coeficiente de correlación multiple entre la variable $v_j = \boldsymbol{\beta}_j^{\mathrm{T}}\mathbf{y}$ y el vector $\mathbf{x}$.

Resumen de propiedades de las variables y correlaciones canónicas

- Las variables canónicas son predictores óptimos en el siguiente sentido:

Queremos hallar $2s$ combinaciones lineales $\mathbf{u} = \mathbf{A}_s^{\mathrm{T}}\mathbf{x}$ y $\mathbf{v} = \mathbf{B}_s^{\mathrm{T}}\mathbf{y}$ con $s \leq m = \mathrm{rango}(\boldsymbol{\Sigma}_{12})$

- $\mathbf{A}_s = (\mathbf{a}_1, \ldots, \mathbf{a}_m)$, $\mathbf{a}_1, \ldots, \mathbf{a}_s$ linealmente independientes (o sea, $\mathrm{rango}(\mathbf{A}_s) = s$)
- $\mathbf{B}_s = (\mathbf{b}_1, \ldots, \mathbf{b}_s)$, $\mathbf{b}_1, \ldots, \mathbf{b}_s$ linealmente independientes (o sea, $\mathrm{rango}(\mathbf{B}_s) = s$)
- $\mathbf{A}_s^{\mathrm{T}}\boldsymbol{\Sigma}_{11}\mathbf{A}_s = \mathbf{I}_s$, $\mathbf{B}_s^{\mathrm{T}}\boldsymbol{\Sigma}_{22}\mathbf{B}_s = \mathbf{I}_s$

tales que $\mathbb{E}\|\mathbf{u} - \mathbf{v}\|^2$ sea mínima.

El mínimo se alcanza si $\mathbf{a}_j = \boldsymbol{\alpha}_j$ y $\mathbf{b}_j = \boldsymbol{\beta}_j$.

Propiedades

Propiedad 1. Sean

$$\mathbf{\Delta_x} = \text{DIAG}\left(\text{VAR}(x_1), \ldots, \text{VAR}(x_p)\right)$$
$$\mathbf{\Delta_x} = \text{DIAG}\left(\text{VAR}(y_1), \ldots, \text{VAR}(y_q)\right)$$

Se tiene que

- $\text{CORR}(u_j, x_\ell) = \mathbb{E}\left(x_\ell \mathbf{x}^{\mathrm{T}}\right) \boldsymbol{\alpha}_j / \sqrt{\text{VAR}(x_\ell)}$
- $\text{CORR}(v_j, y_\ell) = \mathbb{E}\left(y_\ell \mathbf{y}^{\mathrm{T}}\right) \boldsymbol{\beta}_j / \sqrt{\text{VAR}(y_\ell)}$

es decir,

$$\text{CORR}(u_j, \mathbf{x}) = \mathbf{\Delta_x}^{-\frac{1}{2}} \mathbf{\Sigma}_{11} \boldsymbol{\alpha}_j \qquad \text{CORR}(v_j, \mathbf{y}) = \mathbf{\Delta_y}^{-\frac{1}{2}} \mathbf{\Sigma}_{22} \boldsymbol{\beta}_j$$
$$\text{CORR}(u_j, \mathbf{y}) = \mathbf{\Delta_y}^{-\frac{1}{2}} \mathbf{\Sigma}_{21} \boldsymbol{\alpha}_j \qquad \text{CORR}(v_j, \mathbf{x}) = \mathbf{\Delta_x}^{-\frac{1}{2}} \mathbf{\Sigma}_{12} \boldsymbol{\beta}_j$$

Propiedad 2. Si $\mathbf{z} \sim N(\mathbf{0}, \mathbf{\Sigma})$, $\mathbf{x}$ es independiente de $\mathbf{y}$ si y sólo si $\rho_1 = 0$.

Propiedades

Propiedad 3. Invarianza del análisis.

Si llamamos $\mathbf{z}^\star = \begin{pmatrix} \boldsymbol{\Delta}_{\mathbf{x}}^{-\frac{1}{2}}\mathbf{x} \\ \boldsymbol{\Delta}_{\mathbf{y}}^{-\frac{1}{2}}\mathbf{y} \end{pmatrix}$ y efectuamos el análisis de correlación canónica de $\mathbf{z}^\star$, o sea, en lugar de

$$\boldsymbol{\Upsilon}_1 = \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \quad \boldsymbol{\Upsilon}_2 = \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$$

tomamos

$$\boldsymbol{\Upsilon}_1^\star = \mathbf{R}_{11}^{-1}\mathbf{R}_{12}\mathbf{R}_{22}^{-1}\mathbf{R}_{21} \quad \boldsymbol{\Upsilon}_2^\star = \mathbf{R}_{22}^{-1}\mathbf{R}_{21}\mathbf{R}_{11}^{-1}\mathbf{R}_{12}$$

con $\mathbf{R}_{ij}$ las matrices de correlación entonces

- Las correlaciones canónicas no cambian
- $\boldsymbol{\alpha}_j^\star = \boldsymbol{\Delta}_{\mathbf{x}}^{1/2}\boldsymbol{\alpha}_j$,
- $\boldsymbol{\beta}_j^\star = \boldsymbol{\Delta}_{\mathbf{y}}^{1/2}\boldsymbol{\beta}_j$

Por lo tanto, las variables canónicas u_j y v_j son las mismas.

Sean $\mathbf{z}_i = \begin{pmatrix} \mathbf{x}_i \\ \mathbf{y}_i \end{pmatrix} \in \mathbb{R}^d$ i.i.d. con densidad ($\boldsymbol{\Sigma} > 0$), $\mathbf{x}_i \in \mathbb{R}^p$, $\mathbf{y}_i \in \mathbb{R}^q$. Sea

- $\widetilde{\mathbf{x}}_i = \mathbf{x}_i - \overline{\mathbf{x}}$
- $\widetilde{\mathbf{y}}_i = \mathbf{y}_i - \overline{\mathbf{y}}$
- $\widetilde{\mathbf{X}} = \begin{pmatrix} \widetilde{\mathbf{x}}_1^{\mathrm{T}} \\ \vdots \\ \widetilde{\mathbf{x}}_n^{\mathrm{T}} \end{pmatrix} \quad \widetilde{\mathbf{Y}} = \begin{pmatrix} \widetilde{\mathbf{y}}_1^{\mathrm{T}} \\ \vdots \\ \widetilde{\mathbf{y}}_n^{\mathrm{T}} \end{pmatrix}$
- $\mathbf{Q}_{11} = \widetilde{\mathbf{X}}^{\mathrm{T}}\widetilde{\mathbf{X}}$, $\mathbf{Q}_{22} = \widetilde{\mathbf{Y}}^{\mathrm{T}}\widetilde{\mathbf{Y}}$, $\mathbf{Q}_{12} = \widetilde{\mathbf{X}}^{\mathrm{T}}\widetilde{\mathbf{Y}}$.
- $\mathbf{S}_{ij} = \mathbf{Q}_{ij}/(n-1)$
- $\mathbf{S} = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{pmatrix}$

Supongamos $n > d = p + q$ y $q < p$ luego $\mathbf{S}$ es definida positiva con probabilidad 1 y $\mathrm{rango}(\mathbf{Q}_{12}) = q$.

Definición. Se definen las correlaciones canónicas muestrales r_j, $r_1 > r_2 > \cdots > r_q > 0$ (con prob. 1) como $r_j = \sqrt{r_j^2}$ donde $r_1^2 > r_2^2 > \cdots > r_q^2 > 0$ (con prob. 1) son los autovalores de

$$\widehat{\boldsymbol{\Upsilon}}_1 = \mathbf{S}_{11}^{-1}\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21} \quad \text{o de} \quad \widehat{\boldsymbol{\Upsilon}}_2 = \mathbf{S}_{22}^{-1}\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}$$

o sea, la solución de

$$|\mathbf{S}_{11}^{-1}\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21} - \lambda\mathbf{I}_p| = 0 \quad \text{o} \quad |\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21} - \lambda\mathbf{S}_{11}| = 0$$

Definición. Se definen las variables canónicas muestrales como

- $\mathbf{u}_i = (u_{1,i}, \ldots, u_{q,i})^{\mathrm{T}}$ con $u_{j,i} = \widehat{\boldsymbol{\alpha}}_j^{\mathrm{T}} \widetilde{\mathbf{x}}_i$ y
- $\mathbf{v}_i = (v_{1,i}, \ldots, v_{q,i})^{\mathrm{T}}$ $v_{j,i} = \widehat{\boldsymbol{\beta}}_j^{\mathrm{T}} \widetilde{\mathbf{y}}_i$

$$\begin{array}{ll} \mathbf{S}_{11}^{-1}\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21}\widehat{\boldsymbol{\alpha}}_j = r_j^2 \widehat{\boldsymbol{\alpha}}_j & \qquad \widehat{\boldsymbol{\alpha}}_j^{\mathrm{T}}\mathbf{S}_{11}\widehat{\boldsymbol{\alpha}}_j = 1 \\ \mathbf{S}_{22}^{-1}\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}\widehat{\boldsymbol{\beta}}_j = r_j^2 \widehat{\boldsymbol{\beta}}_j & \qquad \widehat{\boldsymbol{\beta}}_j^{\mathrm{T}}\mathbf{S}_{22}\widehat{\boldsymbol{\beta}}_j = 1 \end{array}$$

Observemos que

$$\begin{aligned} 1 &= \widehat{\boldsymbol{\alpha}}_j^{\mathrm{T}}\mathbf{S}_{11}\widehat{\boldsymbol{\alpha}}_j = \frac{1}{n-1}\sum_{i=1}^{n} u_{j,i}^2 \\ 1 &= \widehat{\boldsymbol{\beta}}_j^{\mathrm{T}}\mathbf{S}_{22}\widehat{\boldsymbol{\beta}}_j = \frac{1}{n-1}\sum_{i=1}^{n} v_{j,i}^2 \end{aligned}$$

La distribución exacta de r_j, aún el el caso $\mathbf{z}_i \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ es complicada (ver Muirhead, 1982, sección 11.3.4). La distribución asintótica puede verse en Bilodeau & Brenner (1999, sección 11.5).

Distribución asintótica de r_j^2

Proposición 1. Sean $\mathbf{z}_i \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ i.i.d., tales que $\rho_1 > \rho_2 > \cdots > \rho_k > \rho_{k+1} = \cdots = \rho_p = 0$, luego

$$\sqrt{n}\,(r_j^2 - \rho_j^2) \xrightarrow{D} N\left(0, 4\,\rho_j^2\,(1-\rho_j^2)^2\right), \quad j = 1, \ldots, k$$

- Más aún, si $\mathbf{r}_k^{(2)} = (r_1^2, \ldots, r_k^2)^{\mathrm{T}}$ y $\boldsymbol{\rho}_k^{(2)} = (\rho_1^2, \ldots, \rho_k^2)$ entonces

$$\sqrt{n}\,(\mathbf{r}_k^{(2)} - \boldsymbol{\rho}_k^{(2)}) \xrightarrow{D} N\left(\mathbf{0}_q, 4\,\mathrm{DIAG}\left(\rho_1^2\,(1-\rho_1^2)^2, \ldots, \rho_k^2\,(1-\rho_k^2)^2\right)\right)$$

- Además, $\mathbf{r}_k^{(2)}$ es independiente de r_j para $j > k$.

Distribución asintótica de r_j^2

- Si $w_j = nr_j^2$ para $j > k$, entonces $w_{k+1}, \ldots, w_q$ son dependientes y no–normales.

- La distribución asintótica conjunta de $\mathbf{w} = (w_{k+1}, \ldots, w_q)^{\mathrm{T}}$ es la distribución de los autovalores de una matriz $\mathbf{A} \in \mathbb{R}^{(q-k)\times(q-k)}$ tal que $\mathbf{A} \sim \mathcal{W}(\mathbf{I}_{q-k}, q-m, p-k)$.

- La densidad asintótica de $\mathbf{w}$ es

$$C \exp\{-\frac{1}{2} \sum_{j=k+1}^{q} w_j\} \prod_{j=k+1}^{q} w_j^{\frac{p-q-1}{2}} \prod_{k+1 \leq i < j \leq q} (w_i - w_j)$$

Ejemplo

La matriz **S** es igual a

	Control	*Concept*	*Motiv.*	*Read*	*Write*	*Math*	*Science*	*Sex*
Control	0.449	0.081	0.056	2.530	2.340	2.128	2.112	0.038
Concept	0.081	0.498	0.070	0.432	0.133	0.356	0.478	−0.044
Motiv.	0.056	0.070	0.117	0.729	0.848	0.629	0.385	0.017
Read	2.530	0.432	0.729	102.070	61.769	64.611	67.730	−0.210
Write	2.340	0.133	0.848	61.769	94.604	57.935	53.732	1.184
Math	2.128	0.356	0.629	64.611	57.935	88.637	59.354	−0.226
Science	2.112	0.478	0.385	67.730	53.732	59.354	94.210	−0.668
Sex	0.038	−0.044	0.017	−0.210	1.184	−0.226	−0.668	0.248

Las correlaciones canónicas son

j	1	2	3
r_j	0.464	0.167	0.104

En CANCOR, los vectores canónicos están normalizados de modo que

$$\widehat{\boldsymbol{\alpha}}_j^{\mathrm{T}} \mathbf{Q}_{11} \widehat{\boldsymbol{\alpha}}_j = 1 \qquad \widehat{\boldsymbol{\beta}}_j^{\mathrm{T}} \mathbf{Q}_{22} \widehat{\boldsymbol{\beta}}_j = 1$$

mientras que en CCA, están normalizados de modo que

$$\widehat{\boldsymbol{\alpha}}_j^{\mathrm{T}} \mathbf{S}_{11} \widehat{\boldsymbol{\alpha}}_j = 1 \qquad \widehat{\boldsymbol{\beta}}_j^{\mathrm{T}} \mathbf{S}_{22} \widehat{\boldsymbol{\beta}}_j = 1$$

Estos últimos son los que presentamos

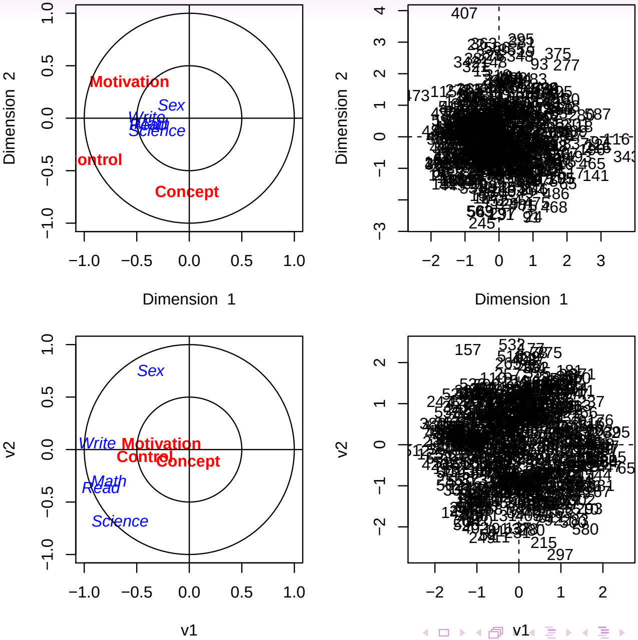
	$\widehat{\boldsymbol{\alpha}}_1$	$\widehat{\boldsymbol{\alpha}}_2$	$\widehat{\boldsymbol{\alpha}}_3$
Control	-1.2538	-0.6215	-0.6617
Concept	0.3513	-1.1877	0.8267
Motivation	-1.2624	2.0273	2.0002
	$\widehat{\boldsymbol{\beta}}_1$	$\widehat{\boldsymbol{\beta}}_2$	$\widehat{\boldsymbol{\beta}}_3$
Read	-0.0446	-0.0049	0.0214
Write	-0.0359	0.0421	0.0913
Math	-0.0234	0.0042	0.0094
Science	-0.0050	-0.0852	-0.1098
Sex	-0.6321	1.0846	-1.7946

Los coeficientes canónicos se interpretan en forma análoga a los coeficientes de regresión, o sea, por ejemplo para la variable *Read* una unidad de incremento en Lectura lleva a un 0.0446 de decrecimiento en la primer variable canónica $\mathbf{v}_1$ cuando todas las demás variables permanecen constantes.

Por otra parte

	$\textsc{Corr}(\mathbf{u}_1, \mathbf{x})$	$\textsc{Corr}(\mathbf{u}_2, \mathbf{x})$	$\textsc{Corr}(\mathbf{u}_3, \mathbf{x})$	$\textsc{Corr}(\mathbf{v}_1, \mathbf{x})$	$\textsc{Corr}(\mathbf{v}_2, \mathbf{x})$	$\textsc{Corr}(\mathbf{v}_3, \mathbf{x})$
Control	−0.9040	−0.3897	−0.1756	−0.4196	−0.0653	−0.0183
Concept	−0.0208	−0.7087	0.7052	−0.0097	−0.1187	0.0733
Motivation	−0.5672	0.3509	0.7451	−0.2632	0.0588	0.0775

	$\textsc{Corr}(\mathbf{u}_1, \mathbf{y})$	$\textsc{Corr}(\mathbf{u}_2, \mathbf{y})$	$\textsc{Corr}(\mathbf{u}_3, \mathbf{y})$	$\textsc{Corr}(\mathbf{v}_1, \mathbf{y})$	$\textsc{Corr}(\mathbf{v}_2, \mathbf{y})$	$\textsc{Corr}(\mathbf{v}_3, \mathbf{y})$
Read	−0.3900	−0.0601	0.0141	−0.8404	−0.3588	0.1354
Write	−0.4068	0.0109	0.0265	−0.8765	0.0648	0.2546
Math	−0.3545	−0.0499	0.0154	−0.7639	−0.2979	0.1478
Science	−0.3056	−0.1134	−0.0240	−0.6584	−0.6768	−0.2304
Sex	−0.1690	0.1265	−0.0565	−0.3641	0.7549	−0.5434


Dimension 2
Dimension 1
Motivation
Sex
Write
Read
Science
ontrol
Concept
407
375
277
587
116
465
141
486
468
245
v2
v1
Sex
Write
Motivation
Control
Concept
Math
Read
Science
157
580
215
297

Cuando las variables en el modelo tienen desvíos estándar muy distintos, se usan los coeficientes estandarizados que se definen como

- $\widehat{\boldsymbol{\alpha}}_j^\star = \widehat{\boldsymbol{\Delta}}_{\mathbf{x}}^{1/2} \widehat{\boldsymbol{\alpha}}_j$,
- $\widehat{\boldsymbol{\beta}}_j^\star = \widehat{\boldsymbol{\Delta}}_{\mathbf{y}}^{1/2} \widehat{\boldsymbol{\beta}}_j$

permiten comparaciones más simples entre variables. Estas direcciones cumplen

$$\widehat{\boldsymbol{\alpha}}_j^{\star\,\mathrm{T}} \widehat{\mathbf{R}}_{11} \widehat{\boldsymbol{\alpha}}_j^\star = 1 \qquad \widehat{\boldsymbol{\beta}}_j^{\star\,\mathrm{T}} \widehat{\mathbf{R}}_{22} \widehat{\boldsymbol{\beta}}_j^\star = 1$$

donde $\widehat{\mathbf{R}}_{11} = \widehat{\boldsymbol{\Delta}}_{\mathbf{x}}^{-1/2} \mathbf{S}_{11} \widehat{\boldsymbol{\Delta}}_{\mathbf{x}}^{-1/2}$ y $\widehat{\mathbf{R}}_{22} = \widehat{\boldsymbol{\Delta}}_{\mathbf{y}}^{-1/2} \mathbf{S}_{22} \widehat{\boldsymbol{\Delta}}_{\mathbf{y}}^{-1/2}$

	$\widehat{\boldsymbol{\alpha}}_1^\star$	$\widehat{\boldsymbol{\alpha}}_2^\star$	$\widehat{\boldsymbol{\alpha}}_3^\star$
Control	-0.8404	-0.4166	-0.4435
Concept	0.2479	-0.8379	0.5833
Motivation	-0.4327	0.6948	0.6855
	$\widehat{\boldsymbol{\beta}}_1^\star$	$\widehat{\boldsymbol{\beta}}_2^\star$	$\widehat{\boldsymbol{\beta}}_3^\star$
Read	-0.4508	-0.0496	0.2160
Write	-0.3490	0.4092	0.8881
Math	-0.2205	0.0398	0.0885
Science	-0.0488	-0.8266	-1.0661
Sex	-0.3150	0.5406	-0.8944

- En este caso, la interpretación es como sigue: para la variable *Read* un aumento de 1 desvío estándar en Lectura lleva a un decrecimiento de 0.45 desvíos estándar en la primer variable canónica $\mathbf{v}_1$ cuando todas las demás variables permanecen constantes.
- Para las variables psicológicas, el primer vector canónico está fuertemente influenciado por Control (.84) y para $\mathbf{u}_2$ por Concepto (-.84) and y Motivación (.69).
- Para las variables académicas más sexo, $\mathbf{v}_1$ es un compromiso entre Lectura (.45), Escritura (.35) y Sexo (.32), mientras que para $\mathbf{v}_2$ Escritura (.41), Ciencia (-.83) y Sexo (.54) son las variables dominantes.

Criterio de Wilks

Tanto en el caso del test de independencia como en el modelo lineal multivariado es posible hallar dos matrices

- $\mathbf{U} \sim \mathcal{W}(\mathbf{\Sigma}, p, N-r)$ y
- $\mathbf{H} \sim \mathcal{W}(\mathbf{\Sigma}, p, r)$

independientes bajo la hipótesis nula de interés.

Más generalmente, tendremos

- $\mathbf{U} \sim \mathcal{W}(\mathbf{\Sigma}, p, N-r)$ y
- $\mathbf{z}_j \sim N(\boldsymbol{\mu}, \mathbf{\Sigma})$, $1 \leq j \leq r$, $\mathbf{H} = \sum_{j=1}^{r} \mathbf{z}_j \mathbf{z}_j^{\mathrm{T}}$

donde muchas veces $r < p$ pero $N - r > p$.

El estadístico de Wilks se utiliza para testear cualquier hipótesis equivalente a $\boldsymbol{\mu} = \mathbf{0}$. Rechazaremos si el Wilks es pequeño.

Criterio de Wilks

En analogía con el caso univariado Wilks (1932) definió el estadístico de Wilks.

Definición Sean $\mathbf{U} \sim \mathcal{W}(\mathbf{\Sigma}, p, N-r)$ y $\mathbf{z}_j \sim N(\mathbf{0}, \mathbf{\Sigma})$, $1 \leq j \leq r$ independientes entre sí, el criterio de Wilks se define como

$$\Lambda(N, p, r) = \frac{|\mathbf{U}|}{|\mathbf{U} + \mathbf{H}|}$$

donde $|\mathbf{A}|$ indica el determinante de la matriz $\mathbf{A} \in \mathbb{R}^{p \times p}$ y $\mathbf{H} = \sum_{j=1}^{r} \mathbf{z}_j \mathbf{z}_j^{\mathrm{T}} \sim \mathcal{W}(\mathbf{\Sigma}, p, r)$

N, p y r son los parámetros del Wilks Λ y corresponden respectivamente a los grados de libertad de $\mathbf{U} + \mathbf{H}$, la dimensión de las matrices y los grados de libertad de $\mathbf{H}$.

Por otra parte, si $\mathbf{U}$ es inversible $\Lambda(N, p, r)$ depende sólo de los autovalores de $\mathbf{H}\mathbf{U}^{-1}$.

Distribución del criterio de Wilks

a) Si $r \geq p$, $\mathbf{H} \sim \mathcal{W}(\mathbf{\Sigma}, p, r)$ tiene densidad y

$$\Lambda(N, p, r) \sim \prod_{j=1}^{p} b_{ii}^2$$

con $b_{11}^2, \ldots, b_{pp}^2$ son independientes $b_{ii}^2 \sim \mathcal{B}\left(\frac{N-r+1-i}{2}, \frac{r}{2}\right)$

b) Si $r < p$,

$$\Lambda(N, p, r) \sim \prod_{j=1}^{r} b_{ii}^2$$

con $b_{11}^2, \ldots, b_{rr}^2$ son independientes $b_{ii}^2 \sim \mathcal{B}\left(\frac{N-p+1-i}{2}, \frac{p}{2}\right)$

Es decir, $\Lambda(N, p, r) \sim \Lambda(N, r, p)$

Corolario

a) Si $p = 1$,

$$\frac{1 - \Lambda(N, 1, r)}{\Lambda(N, 1, r)} \, \frac{N - r}{r} \sim \mathcal{F}_{r, N-r}$$

b) Si $r = 1$,

$$\frac{1 - \Lambda(N, p, 1)}{\Lambda(N, p, 1)} \, \frac{N - p}{p} \sim \mathcal{F}_{p, N-p}$$

c) Si $p = 2$,

$$\frac{1 - \Lambda(N, 2, r)^{\frac{1}{2}}}{\Lambda(N, 2, r)^{\frac{1}{2}}} \, \frac{N - r - 1}{r} \sim \mathcal{F}_{2r, 2(N-r-1)}$$

d) Si $r = 2$,

$$\frac{1 - \Lambda(N, p, 2)^{\frac{1}{2}}}{\Lambda(N, p, 2)^{\frac{1}{2}}} \, \frac{N - p - 1}{p} \sim \mathcal{F}_{2p, 2(N-p-1)}$$

Aproximaciones a la Distribución del criterio de Wilks

- Rao (1951) mostró que

$$\frac{(fs+\lambda)}{m}\,\frac{(1-\Lambda(N,p,r)^{\frac{1}{s}})}{\Lambda(N,p,r)^{\frac{1}{s}}} \approx \mathcal{F}_{m,fs+\lambda}$$

donde

$$\begin{aligned} f &= N - \frac{p+r+1}{2} \qquad & m = pr \\ \lambda &= -\frac{pr}{2}+1 \qquad & s = \frac{(p^2r^2-4)^{\frac{1}{2}}}{(p^2+r^2-5)^{\frac{1}{2}}} \end{aligned}$$

Otros criterios

Hay otros criterios que se utilizan

- **Criterio de Roy** o de la máxima raíz. Considera la máxima raíz θ_{max} de $|\mathbf{H} - \theta(\mathbf{U} + \mathbf{H})| = 0$. Luego

$$\theta_{\text{max}} = \frac{\lambda_1}{1 + \lambda_1}$$

y rechazo si θ_{max} es grande. Los percentiles de la distribución de θ_{max} están dados en el Apéndice D14 de Seber (1984).

Test de Independencia $H_{01} : \boldsymbol{\Sigma} = \mathrm{diag}(\boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{22})$

Supongamos $\mathbf{z} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Queremos testear $\boldsymbol{\Sigma}_{12} = 0$ o equivalentemente $H_{01} : \rho_1 = 0$

El test de cociente de máxima verosimilitud para H_0 está basado en

$$\gamma_1 = \frac{|\mathbf{Q}|}{|\mathbf{Q}_{11}| \, |\mathbf{Q}_{22}|} \, .$$

Vimos que

$$-n \log(\gamma_1) \xrightarrow{D} \chi^2_{\nu_1}$$

donde

$$\nu_1 = \frac{1}{2} \left(d^2 - (p^2 + q^2) \right) = pq \, .$$

La aproximación mejora si tomamos en lugar de n la corrección de Bartlett, $m = n - (p + q + 3)/2$, es decir,

$$-m \log(\gamma_1) \xrightarrow{D} \chi^2_{\nu_1} \, .$$

Test de Independencia $H_{01} : \mathbf{\Sigma} = \text{diag}(\mathbf{\Sigma}_{11}, \mathbf{\Sigma}_{22})$

Por otra parte,

$$|\mathbf{Q}| \quad = \quad = |\mathbf{Q}_{22}|\,|\mathbf{Q}_{11.2}| = |\mathbf{Q}_{11}|\,|\mathbf{Q}_{22}|\,|\mathbf{I}_p - \mathbf{Q}_{11}^{-1}\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\mathbf{Q}_{21}|$$

con lo cual

$$\gamma_1 = |\mathbf{I}_p - \mathbf{Q}_{11}^{-1}\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\mathbf{Q}_{21}| = \prod_{j=1}^{q}(1 - r_i^2)$$

donde r_j^2 son los autovalores de $\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\mathbf{Q}_{21}\mathbf{Q}_{11}^{-1}$, o sea, el cuadrado de las correlaciones canónicas muestrales.

Es decir, rechazo con nivel α si

$$\prod_{j=1}^{q}(1 - r_i^2) < k << 1$$

con $k = \exp\left(-\chi^2_{\nu_1,\alpha}/m\right)$ con $m = n - (p + q + 3)/2$.

Test de Independencia $H_{01} : \boldsymbol{\Sigma} = \mathrm{diag}(\boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{22})$

Por otra parte,

$$\gamma_1 = \frac{|\mathbf{Q}_{11.2}|}{|\mathbf{Q}_{11}|} = \frac{|\mathbf{U}|}{|\mathbf{U} + \mathbf{H}|}$$

donde $\boldsymbol{\Sigma}_{11.2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$. Ahora bien,

- $\mathbf{U} = \mathbf{Q}_{11.2} \sim \mathcal{W}(\boldsymbol{\Sigma}_{11.2}, q, n-1-p) = \mathcal{W}(\boldsymbol{\Sigma}_{11}, p, n-1-q)$ bajo H_{01}
- $\mathbf{H} = \mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\mathbf{Q}_{21} \sim \mathcal{W}(\boldsymbol{\Sigma}_{11.2}, p, q) = \mathcal{W}(\boldsymbol{\Sigma}_{11}, p, q)$ bajo H_{01}
- $\mathbf{H}$ y $\mathbf{U}$ son independientes bajo H_{01}.

Luego,

$$\gamma_1 = \Lambda(n-1, p, q)$$

Test de Independencia $H_{01} : \mathbf{\Sigma} = \mathrm{diag}(\mathbf{\Sigma}_{11}, \mathbf{\Sigma}_{22})$

Queremos aplicar el principio de unión interseción de Roy para testear $H_{01} : \mathbf{\Sigma}_{12} = 0$.

Recordemos que

$$\mathbf{z}_i = \begin{pmatrix} \mathbf{x}_i \\ \mathbf{y}_i \end{pmatrix}$$

entonces $\mathrm{Cov}(\mathbf{a}^{\mathrm{T}}\mathbf{x}, \mathbf{b}^{\mathrm{T}}\mathbf{y}) = \mathbf{a}^{\mathrm{T}}\mathbf{\Sigma}_{12}\mathbf{b}$. Definamos

$$H_{0,\mathbf{ab}} : \mathbf{a}^{\mathrm{T}}\mathbf{\Sigma}_{12}\mathbf{b} = 0$$

Luego

$$H_{01} = \cap_{\mathbf{a}\neq\mathbf{0}} \cap_{\mathbf{b}\neq\mathbf{0}} H_{0,\mathbf{ab}}$$

Test de Independencia $H_{01} : \mathbf{\Sigma} = \text{diag}(\mathbf{\Sigma}_{11}, \mathbf{\Sigma}_{22})$

Aplicando el principio de unión intersección, se obtiene el criterio de Roy.

Es decir, el test rechaza si

$$\theta_{\max} > k_\alpha.$$

donde $\theta_{\max}$ es la máxima raíz de $|\mathbf{H} - \theta(\mathbf{U} + \mathbf{H})| = 0$, o sea, $\theta_{\max}$ es el máximo autovalor de $\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\mathbf{Q}_{21}\mathbf{Q}_{11}^{-1}$, es decir, rechaza si

$$r_1 > k_\alpha$$

Determinando la cantidad de correlaciones canónicas

Aunque rechacemos $H_{01} : \rho_1 = 0$ es posible que $\rho_2 = 0$, por lo tanto nos interesa testear

$$H_{0,k+1} : \rho_{k+1} = 0 \qquad \rho_k > 0$$

que da la dimensión de la relación entre $\mathbf{x}$ y $\mathbf{y}$.

- Si $\rho_1 > 0$ y $\rho_2 = 0$ la relación es lineal.
- Si $\rho_1 > \rho_2 > 0$ y $\rho_3 = 0$ la relación es planar.

El número de correlaciones no nulas da el rango de $\mathbf{\Sigma}_{12}$

Si $\rho_1 \geq \cdots \geq \rho_k > 0$ y $\rho_{k+1} = 0$, las variables canónicas asociadas a ρ_{k+1} se llaman funciones nulas y se usan en economía.

Determinando la cantidad de correlaciones canónicas

El test de cociente de verosimilitud para

$$H_{0,k+1} : \rho_{k+1} = 0 \qquad \rho_k > 0$$

se basa en

$$\gamma_{k+1} = \prod_{j=k+1}^{q} (1 - r_j^2)$$

Además,

$$\gamma_{k+1} = \Lambda(n-1-k, p-k, q-k)$$

Por la aproximación de Rao

$$\frac{(fs+\lambda)}{m} \frac{(1-\gamma_{k+1}^{\frac{1}{s}})}{\gamma_{k+1}^{\frac{1}{s}}} \approx \mathcal{F}_{m,fs+\lambda} = \mathcal{F}_{\nu_{1,k+1},\nu_{2,k+1}}$$

$$f = n - 3/2 - (p+q)/2 \qquad m = (p-k)(q-k)$$

$$\lambda = -\frac{(p-k)(q-k)}{2} + 1 \qquad s = \frac{((p-k)^2(q-k)^2-4)^{\frac{1}{2}}}{((p-k)^2+(q-k)^2-5)^{\frac{1}{2}}}$$

Determinando la cantidad de correlaciones canónicas

El test de cociente de verosimilitud para
$H_{0,k+1} : \rho_{k+1} = 0 \qquad \rho_k > 0$ se basa en

$$\gamma_{k+1} = \prod_{j=k+1}^{q} (1 - r_j^2)$$

Si $H_{0,k+1}$ es cierta,

$$-n \log(\gamma_{k+1}) \xrightarrow{D} \chi^2_{\nu_{k+1}} \qquad \nu_{k+1} = (p-k)(q-k)$$

- Bartlett (1947) sugiere tomar $m = n - (p+q+3)/2$, y usar la aproximación
$$-m \log(\gamma_{k+1}) \xrightarrow{D} \chi^2_{\nu_{k+1}} .$$
- Glynn y Muirhead (1978) sugieren la modificación
$$\ell_{k+1} = -\left[n - k - \frac{p+q+3}{2} + \sum_{j=1}^{k} \frac{1}{r_j^2} \right] \log(\gamma_{k+1}) \xrightarrow{D} \chi^2_{\nu_{k+1}} .$$

Determinando la cantidad de correlaciones canónicas

Nos interesaba testear

$$H_{0,k+1} : \rho_{k+1} = 0 \qquad \rho_k > 0$$

que da la dimensión de la relación entre **x** y **y**.

Para determinar k testeamos la secuencia

$$H_{01}, H_{02}, \ldots$$

hasta encontrar un test no significativo para digamos H_{0r} entonces elegimos $k = r - 1$.

Ejemplo

Deseamos saber cuantas correlaciones son significativas

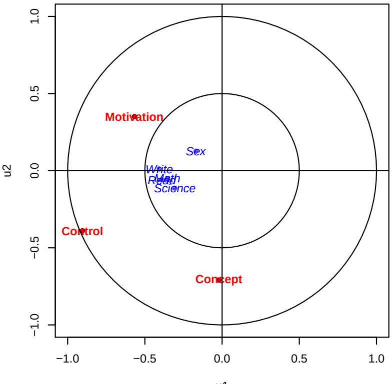
$H_{0,k+1}$	γ_{k+1}	$\mathcal{F}$	$\nu_{1,k+1}$	$\nu_{2,k+1}$	$p-valor$
$H_{01} : \rho_1 = 0$	0.7544	11.716	15	1635	$7.498e-28$
$H_{02} : \rho_2 = 0(\rho_1 > 0)$	0.9614	2.944	8	1186	0.002905
$H_{03} : \rho_3 = 0(\rho_2 > 0)$	0.9892	2.165	3	594	0.09109

Aproximación Asintótica

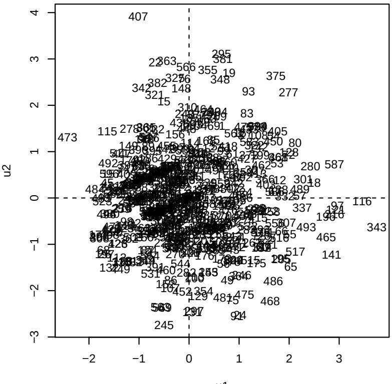
$H_{0,k+1}$	ℓ_{k+1}	ν_{k+1}	$p-valor$
$H_{01} : \rho_1 = 0$	167.580	15	0.00000000
$H_{02} : \rho_2 = 0(\rho_1 > 0)$	23.162	8	0.003163
$H_{03} : \rho_3 = 0(\rho_2 > 0)$	6.004	3	0.111401

Alcanza con 2 variables canónicas.

a)

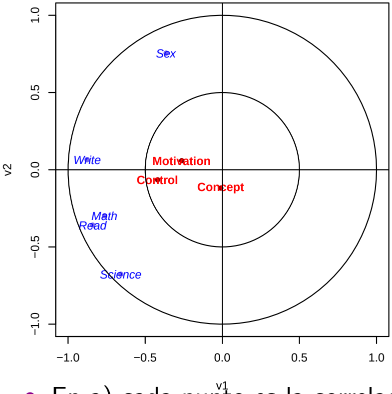


b)

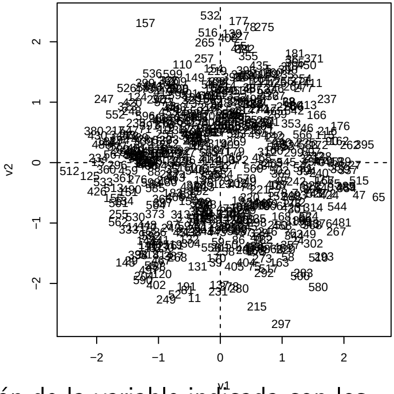


- Porqué el gráfico en b) es una nube de puntos sin estructura?
- En a) cada punto es la correlación de la variable indicada con los ejes u_1, u_2.
- En b), las distancias entre puntos del espacio de variables canónicas $(u_{1,i}, u_{2,i}) = (\widehat{\alpha}_1^{\mathrm{T}}\widetilde{\mathbf{x}}_i, \widehat{\alpha}_2^{\mathrm{T}}\widetilde{\mathbf{x}}_i)$ dan aproximadamente la distancia de Mahalanobis entre observaciones $\widetilde{\mathbf{x}}_i$.

a)

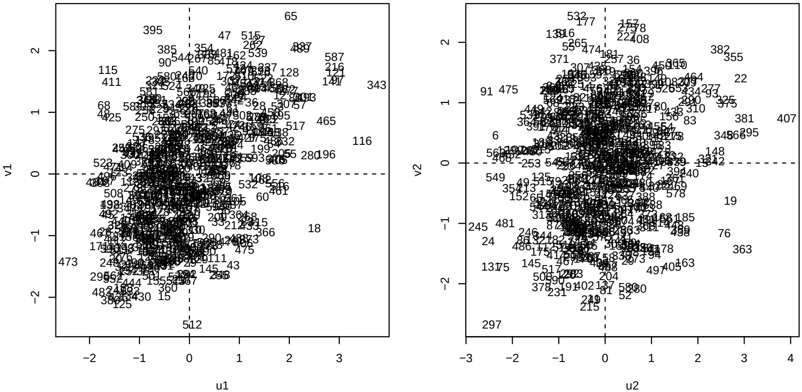


b)



- En a) cada punto es la correlación de la variable indicada con los ejes v_1, v_2.
- En b), las distancias entre puntos del espacio de variables canónicas $(v_{1,i}, v_{2,i}) = (\widehat{\boldsymbol{\beta}}_1^{\mathrm{T}} \widetilde{\mathbf{y}}_i, \widehat{\boldsymbol{\beta}}_2^{\mathrm{T}} \widetilde{\mathbf{y}}_i)$ dan aproximadamente la distancia de Mahalanobis entre observaciones $\widetilde{\mathbf{y}}_i$.

a) b)



La relación entre u_j y v_j decrece con j, como esperábamos.

Vectores canónicos estandarizados

Espacio $\mathbf{x}$			Espacio $\mathbf{y}$		
	$\widehat{\boldsymbol{\alpha}}_1^\star$	$\widehat{\boldsymbol{\alpha}}_2^\star$		$\widehat{\boldsymbol{\beta}}_1^\star$	$\widehat{\boldsymbol{\beta}}_2^\star$
Control	-0.8404	-0.4166	Read	-0.4508	-0.0496
Concept	0.2479	-0.8379	Write	-0.3490	0.4092
Motivation	-0.4327	0.6948	Math	-0.2205	0.0398
			Science	-0.0488	-0.8266
			Sex	-0.3150	0.5406

- Para las variables psicológicas,
  - el primer vector canónico está fuertemente influenciado por Control (.84)
  - $\mathbf{u}_2$ está influenciado por Concepto (-.84) and y Motivación (.69).
- Para las variables académicas más sexo,
  - $\mathbf{v}_1$ es un compromiso entre Lectura (.45), Escritura (.35) y Sexo (.32),
  - para $\mathbf{v}_2$ Escritura (.41), Ciencia (-.83) y Sexo (.54) son las variables dominantes.

Relación con coordenadas discriminantes

Si tenemos k grupos y el vector $\mathbf{y} \in \mathbb{R}^{k-1}$ es el indicador de grupo o sea,

$$\begin{aligned} \mathbf{y}_i &= (y_{i,1}, \ldots, y_{i,k-1})^{\mathrm{T}} \\ y_{i,j} &= \begin{cases} 1 & \text{si } \mathbf{x}_i \text{ pertenece al grupo } j \\ 0 & \text{otro caso} \end{cases} \end{aligned}$$

entonces el análisis de correlación canónica entre $(\mathbf{x}_i, \mathbf{y}_i)$ resulta en las coordenadas discriminantes.