

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)

Planteo del problema



Definiciones



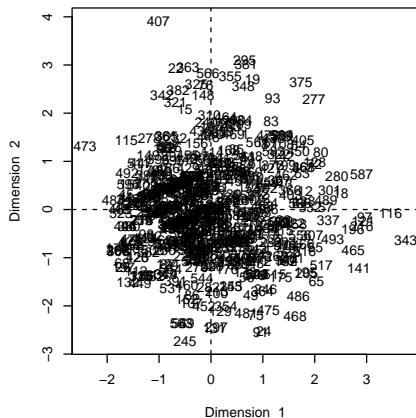
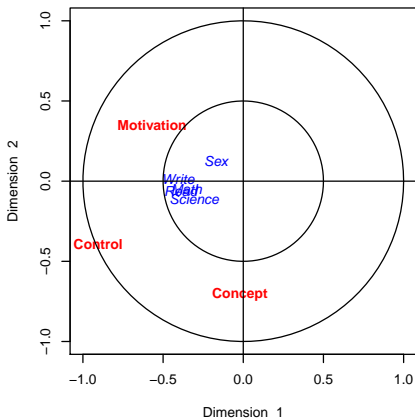
Estimación



Criterio de Wilks



Inferencia



Sea $\mathbf{z} = (z_1, \dots, z_{p+q})^T = \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \in \mathbb{R}^d$ 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

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

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

Supongamos $p = 1$

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

$$\rho_{1,(23\dots d)}^2 = \frac{\sigma_{21}^T \mathbf{\Sigma}_{22}^{-1} \sigma_{21}}{\sigma_{11}}$$

donde $\sigma_{21} = \text{Cov}(\mathbf{y}, z_1)$ y el máximo se alcanzaba en $\beta_0 = \mathbf{\Sigma}_{22}^{-1} \sigma_{21}$.

Caso General: $q < p$

Queremos medir la asociación entre \mathbf{x} y \mathbf{y} que estará dada por:

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

$$\max_{\alpha \neq \mathbf{0}_p, \beta \neq \mathbf{0}_q} \rho_{\alpha, \beta}^2$$

con

$$\rho_{\alpha, \beta}^2 = \frac{\text{COV}^2(\alpha^T \mathbf{x}, \beta^T \mathbf{y})}{\text{VAR}(\alpha^T \mathbf{x}) \text{VAR}(\beta^T \mathbf{y})} = \frac{(\alpha^T \Sigma_{12} \beta)^2}{\alpha^T \Sigma_{11} \alpha \beta^T \Sigma_{22} \beta}$$

Sea $\Sigma_{jj} = \mathbf{A}_j^T \mathbf{A}_j$,

$$\Psi_1 = (\mathbf{A}_1^T)^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \mathbf{A}_1^{-1} \quad \Psi_2 = (\mathbf{A}_2^T)^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \mathbf{A}_2^{-1}$$

entonces

$$\max_{\alpha \neq \mathbf{0}_p, \beta \neq \mathbf{0}_q} \rho_{\alpha, \beta}^2 = \rho_{\alpha_1, \beta_1}^2 = \rho_1^2$$

Sea $\Sigma_{jj} = \mathbf{A}_j^T \mathbf{A}_j$,

$$\Psi_1 = (\mathbf{A}_1^T)^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \mathbf{A}_1^{-1} \quad \Psi_2 = (\mathbf{A}_2^T)^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \mathbf{A}_2^{-1}$$

entonces

$$\max_{\alpha \neq \mathbf{0}_p, \beta \neq \mathbf{0}_q} \rho_{\alpha, \beta}^2 = \rho_{\alpha_1, \beta_1}^2 = \rho_1^2$$

- ρ_1^2 es el máximo autovalor de Ψ_1 , es decir, el máximo autovalor de $\Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$

Sea $\Sigma_{jj} = \mathbf{A}_j^T \mathbf{A}_j$,

$$\Psi_1 = (\mathbf{A}_1^T)^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \mathbf{A}_1^{-1} \quad \Psi_2 = (\mathbf{A}_2^T)^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \mathbf{A}_2^{-1}$$

entonces

$$\max_{\alpha \neq \mathbf{0}_p, \beta \neq \mathbf{0}_q} \rho_{\alpha, \beta}^2 = \rho_{\alpha_1, \beta_1}^2 = \rho_1^2$$

- ρ_1^2 es el máximo autovalor de Ψ_1 , es decir, el máximo autovalor de $\Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$
- α_1 es el autovector de $\Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$ asociado a ρ_1^2 tal que $\alpha_1^T \Sigma_{11} \alpha_1 = 1$.

Sea $\Sigma_{jj} = \mathbf{A}_j^T \mathbf{A}_j$,

$$\Psi_1 = (\mathbf{A}_1^T)^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \mathbf{A}_1^{-1} \quad \Psi_2 = (\mathbf{A}_2^T)^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \mathbf{A}_2^{-1}$$

entonces

$$\max_{\alpha \neq \mathbf{0}_p, \beta \neq \mathbf{0}_q} \rho_{\alpha, \beta}^2 = \rho_{\alpha_1, \beta_1}^2 = \rho_1^2$$

- ρ_1^2 es el máximo autovalor de Ψ_1 , es decir, el máximo autovalor de $\Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$
- α_1 es el autovector de $\Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$ asociado a ρ_1^2 tal que $\alpha_1^T \Sigma_{11} \alpha_1 = 1$.
- β_1 es el autovector de $\Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}$ asociado a ρ_1^2 tal que $\beta_1^T \Sigma_{22} \beta_1 = 1$.

$$\beta_1 = \frac{\Sigma_{22}^{-1} \Sigma_{21} \alpha_1}{\rho_1} \quad \alpha_1 = \frac{\Sigma_{11}^{-1} \Sigma_{12} \beta_1}{\rho_1}$$

Definición.

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

Definición.

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

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

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

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

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 $\rho_{\alpha,\beta}^2$ sea máxima. Pero, la reducción u_1 de \mathbf{x} puede no ser adecuada.

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

- $u_j = \alpha_j^T \mathbf{x}$, $v_j = \beta_j^T \mathbf{y}$
- u_1, \dots, u_m sean no correlacionados
- v_1, \dots, v_m sean no correlacionados
- $\text{CORR}(u_j, v_j)$ sea máxima en algún sentido

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

$$\alpha_j^T \Sigma_{11} \alpha_j = 1 \quad \beta_j^T \Sigma_{22} \beta_j = 1$$

Teorema

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

- $\mathbf{C} = (\mathbf{A}_1^T)^{-1} \Sigma_{12} \mathbf{A}_2^{-1}$
- $\Upsilon_1 = \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$
- $\Upsilon_2 = \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}$

Teorema

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

- $\mathbf{C} = (\mathbf{A}_1^T)^{-1} \Sigma_{12} \mathbf{A}_2^{-1}$
- $\Upsilon_1 = \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$
- $\Upsilon_2 = \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}$
- $1 > \rho_1^2 \geq \rho_2^2 \geq \dots \geq \rho_m^2 > 0$ con $m = \text{RANGO}(\Sigma_{12})$ los autovalores no nulos de Υ_1 (y de Υ_2)

Teorema

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

- $\mathbf{C} = (\mathbf{A}_1^T)^{-1} \Sigma_{12} \mathbf{A}_2^{-1}$
- $\Upsilon_1 = \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$
- $\Upsilon_2 = \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}$
- $1 > \rho_1^2 \geq \rho_2^2 \geq \dots \geq \rho_m^2 > 0$ con $m = \text{RANGO}(\Sigma_{12})$ los autovalores no nulos de Υ_1 (y de Υ_2)
- $\alpha_1, \dots, \alpha_m$ los autovectores de Υ_1 asociados a $\rho_1^2 \geq \rho_2^2 \geq \dots \geq \rho_m^2$ tales que $\alpha_j^T \Sigma_{11} \alpha_j = 1$
- β_1, \dots, β_m los autovectores de Υ_2 asociados a $\rho_1^2 \geq \rho_2^2 \geq \dots \geq \rho_m^2$ tales que $\beta_j^T \Sigma_{22} \beta_j = 1$

Teorema

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

- $\mathbf{C} = (\mathbf{A}_1^T)^{-1} \Sigma_{12} \mathbf{A}_2^{-1}$
- $\Upsilon_1 = \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$
- $\Upsilon_2 = \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}$
- $1 > \rho_1^2 \geq \rho_2^2 \geq \dots \geq \rho_m^2 > 0$ con $m = \text{RANGO}(\Sigma_{12})$ los autovalores no nulos de Υ_1 (y de Υ_2)
- $\alpha_1, \dots, \alpha_m$ los autovectores de Υ_1 asociados a $\rho_1^2 \geq \rho_2^2 \geq \dots \geq \rho_m^2$ tales que $\alpha_j^T \Sigma_{11} \alpha_j = 1$
- β_1, \dots, β_m los autovectores de Υ_2 asociados a $\rho_1^2 \geq \rho_2^2 \geq \dots \geq \rho_m^2$ tales que $\beta_j^T \Sigma_{22} \beta_j = 1$

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

$$\text{Cov}(\alpha^T \mathbf{x}, \alpha_j^T \mathbf{x}) = 0 \quad 1 \leq j \leq s$$

$$\text{Cov}(\beta^T \mathbf{y}, \beta_j^T \mathbf{y}) = 0 \quad 1 \leq j \leq s$$



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

$$\max_{\alpha \neq 0, \beta \neq 0} \rho_{\alpha, \beta}^2 = \rho_{s+1}^2 = \rho_{\alpha_{s+1}, \beta_{s+1}}^2$$

$$\text{Cov}(\alpha^T \mathbf{x}, \alpha_j^T \mathbf{x}) = 0, 1 \leq j \leq s$$

$$\text{Cov}(\beta^T \mathbf{y}, \beta_j^T \mathbf{y}) = 0, 1 \leq j \leq s$$

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

$$\max_{\alpha \neq 0, \beta \neq 0} \rho_{\alpha, \beta}^2 = \rho_{s+1}^2 = \rho_{\alpha_{s+1}, \beta_{s+1}}^2$$

$$\text{Cov}(\alpha_j^T \mathbf{x}, \alpha_j^T \mathbf{x}) = 0, 1 \leq j \leq s$$

$$\text{Cov}(\beta_j^T \mathbf{y}, \beta_j^T \mathbf{y}) = 0, 1 \leq j \leq s$$

- ii) $\text{Cov}(\alpha_j^T \mathbf{x}, \alpha_k^T \mathbf{x}) = 0$ si $j \neq k$ y $\text{Cov}(\beta_j^T \mathbf{y}, \beta_k^T \mathbf{y}) = 0$ si $j \neq k$

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

$$\max_{\alpha \neq 0, \beta \neq 0} \rho_{\alpha, \beta}^2 = \rho_{s+1}^2 = \rho_{\alpha_{s+1}, \beta_{s+1}}^2$$

$$\text{Cov}(\alpha_j^T \mathbf{x}, \alpha_j^T \mathbf{x}) = 0, 1 \leq j \leq s$$

$$\text{Cov}(\beta_j^T \mathbf{y}, \beta_j^T \mathbf{y}) = 0, 1 \leq j \leq s$$

- ii) $\text{COV}(\alpha_j^T \mathbf{x}, \alpha_k^T \mathbf{x}) = 0$ si $j \neq k$ y $\text{COV}(\beta_j^T \mathbf{y}, \beta_k^T \mathbf{y}) = 0$ si $j \neq k$
 iii) $\text{VAR}(\alpha_j^T \mathbf{x}) = \text{VAR}(\beta_j^T \mathbf{y}) = 1$

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

$$\max_{\alpha \neq 0, \beta \neq 0} \rho_{\alpha, \beta}^2 = \rho_{s+1}^2 = \rho_{\alpha_{s+1}, \beta_{s+1}}^2$$

$$\text{Cov}(\alpha^T \mathbf{x}, \alpha_j^T \mathbf{x}) = 0, 1 \leq j \leq s$$

$$\text{Cov}(\beta^T \mathbf{y}, \beta_j^T \mathbf{y}) = 0, 1 \leq j \leq s$$

ii) $\text{COV}(\alpha_j^T \mathbf{x}, \alpha_k^T \mathbf{x}) = 0$ si $j \neq k$ y $\text{COV}(\beta_j^T \mathbf{y}, \beta_k^T \mathbf{y}) = 0$ si $j \neq k$

iii) $\text{VAR}(\alpha_j^T \mathbf{x}) = \text{VAR}(\beta_j^T \mathbf{y}) = 1$

iv) Sea $m = \text{RANGO}(\Sigma_{12}) \leq \min(q, p) = q$ y tomemos

$$\mathbf{u} = \mathbf{A}^T \mathbf{x} \text{ con } \mathbf{A} = (\alpha_1, \dots, \alpha_m, \alpha_{m+1}, \dots, \alpha_p)$$

$$\mathbf{v} = \mathbf{B}^T \mathbf{y} \text{ con } \mathbf{B} = (\beta_1, \dots, \beta_q)$$

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

$$\text{VAR} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_q & \mathbf{0} & \mathbf{D}_\rho \\ \mathbf{0} & \mathbf{I}_{p-r} & \mathbf{0} \\ \mathbf{D}_\rho & \mathbf{0} & \mathbf{I}_q \end{pmatrix}$$

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

Definición.

- \mathbf{u} se llama las variables canónicas del espacio \mathbf{x} .
- \mathbf{v} se llama las variables canónicas del espacio \mathbf{y} .
- $\alpha_1, \dots, \alpha_m$ y β_1, \dots, β_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 $\rho_1^2, \dots, \rho_m^2$

Definición.

- \mathbf{u} se llama las variables canónicas del espacio \mathbf{x} .
- \mathbf{v} se llama las variables canónicas del espacio \mathbf{y} .
- $\alpha_1, \dots, \alpha_m$ y β_1, \dots, β_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 $\rho_1^2, \dots, \rho_m^2$

Como $\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\alpha_j = \rho_j^2\Sigma_{11}\alpha_j$ y $\alpha_j^T\Sigma_{11}\alpha_j = 1$ tenemos que

$$\rho_j^2 = \alpha_j^T \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \alpha_j$$

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

Otro enfoque

$$\Sigma_{jj} = \mathbf{A}_j^T \mathbf{A}_j \quad \mathbf{A}_j > 0 \quad \mathbf{C} = (\mathbf{A}_1^T)^{-1} \Sigma_{12} \mathbf{A}_2^{-1}$$

Por la descomposición de valores singulares

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

- $\mathbf{L} = (\ell_1, \dots, \ell_p) \in \mathbb{R}^{p \times p}$, $\mathbf{M} = (\mathbf{m}_1, \dots, \mathbf{m}_q) \in \mathbb{R}^{q \times q}$,
ortogonales, $\mathbf{m}_j \in \mathbb{R}^q$ y $\ell_j \in \mathbb{R}^p$.
- $\mathbf{D}_\rho = \text{DIAG}(\rho_1, \dots, \rho_q)$, $\rho_1 \geq \rho_2 \geq \dots \geq \rho_q \geq 0$,
 ρ_j es la raíz cuadrada de j -ésimo autovalor de $\mathbf{C}^T \mathbf{C} = \Psi_2$.

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

$$\ell_j = \frac{\mathbf{C} \mathbf{m}_j}{\rho_j} \quad 1 \leq j \leq q$$

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

Definamos

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

Entonces,

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

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

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

La cantidad de correlaciones no nulas es $m = \text{RANGO}(\mathbf{\Sigma}_{12})$

Las dos definiciones coinciden

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

$$\mathbf{u}_i = \mathbf{L}^T (\mathbf{A}_1^{-1})^T \mathbf{x}_i = (u_{1,i}, \dots, u_{p,i})^T \quad u_{j,i} = (\mathbf{A}_1^{-1} \ell_j)^T \mathbf{x}_i \quad i = 1, 2$$

Luego,

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

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

$$\mathbf{u}_i = \mathbf{L}^T (\mathbf{A}_1^{-1})^T \mathbf{x}_i = (u_{1,i}, \dots, u_{p,i})^T \quad u_{j,i} = (\mathbf{A}_1^{-1} \ell_j)^T \mathbf{x}_i \quad i = 1, 2$$

Luego,

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

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

$$\mathbf{v}_i = \mathbf{M}^T (\mathbf{A}_2^{-1})^T \mathbf{y}_i = (v_{1,i}, \dots, v_{q,i})^T \quad v_{j,i} = (\mathbf{A}_v^{-1} \mathbf{m}_j)^T \mathbf{y}_i \quad i = 1, 2$$

Luego,

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

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

$$\mathbf{u}_i = \mathbf{L}^T (\mathbf{A}_1^{-1})^T \mathbf{x}_i = (u_{1,i}, \dots, u_{p,i})^T \quad u_{j,i} = (\mathbf{A}_1^{-1} \ell_j)^T \mathbf{x}_i \quad i = 1, 2$$

Luego,

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

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

$$\mathbf{v}_i = \mathbf{M}^T (\mathbf{A}_2^{-1})^T \mathbf{y}_i = (v_{1,i}, \dots, v_{q,i})^T \quad v_{j,i} = (\mathbf{A}_v^{-1} \mathbf{m}_j)^T \mathbf{y}_i \quad i = 1, 2$$

Luego,

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

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

$$\Sigma_{ii}^{-1} \Sigma_{ij} \Sigma_{jj}^{-1} \Sigma_{ji} \quad i = 1, 2 \quad i \neq j$$

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

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

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 $u_j = \alpha_j^T \mathbf{x}$ y $v_j = \beta_j^T \mathbf{y}$ correspondientes.

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 $u_j = \alpha_j^T \mathbf{x}$ y $v_j = \beta_j^T \mathbf{y}$ correspondientes.
- Las correlaciones canónicas son invariantes ante transformaciones lineales no singulares de las variables.

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 $u_j = \alpha_j^T \mathbf{x}$ y $v_j = \beta_j^T \mathbf{y}$ 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.

$$\rho_1^2 \geq \text{CORR}^2(x_i, y_j) \quad \text{para todo} \quad 1 \leq i \leq p, 1 \leq j \leq q$$

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 $u_j = \alpha_j^T \mathbf{x}$ y $v_j = \beta_j^T \mathbf{y}$ 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.

$$\rho_1^2 \geq \text{CORR}^2(x_i, y_j) \quad \text{para todo} \quad 1 \leq i \leq p, 1 \leq j \leq q$$

- El coeficiente de correlación canónica ρ_j^2 es el cuadrado del coeficiente de correlación múltiple entre la variable $u_j = \alpha_j^T \mathbf{x}$ y el vector \mathbf{y} .

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 $u_j = \alpha_j^T \mathbf{x}$ y $v_j = \beta_j^T \mathbf{y}$ 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.

$$\rho_1^2 \geq \text{CORR}^2(x_i, y_j) \quad \text{para todo} \quad 1 \leq i \leq p, 1 \leq j \leq q$$

- El coeficiente de correlación canónica ρ_j^2 es el cuadrado del coeficiente de correlación múltiple entre la variable $u_j = \alpha_j^T \mathbf{x}$ y el vector \mathbf{y} .
- El coeficiente de correlación canónica ρ_j^2 es el cuadrado del coeficiente de correlación múltiple entre la variable $v_j = \beta_j^T \mathbf{y}$ 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 $\mathbf{u} = \mathbf{A}_s^T \mathbf{x}$ y $\mathbf{v} = \mathbf{B}_s^T \mathbf{y}$ con $s \leq m = \text{RANGO}(\boldsymbol{\Sigma}_{12})$

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

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

$$\mathbf{\Delta}_{\mathbf{x}} = \text{DIAG}(\text{VAR}(x_1), \dots, \text{VAR}(x_p))$$

$$\mathbf{\Delta}_{\mathbf{y}} = \text{DIAG}(\text{VAR}(y_1), \dots, \text{VAR}(y_q))$$

Se tiene que

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

es decir,

$$\text{CORR}(u_j, \mathbf{x}) = \mathbf{\Delta}_{\mathbf{x}}^{-\frac{1}{2}} \boldsymbol{\Sigma}_{11} \boldsymbol{\alpha}_j$$

$$\text{CORR}(v_j, \mathbf{y}) = \mathbf{\Delta}_{\mathbf{y}}^{-\frac{1}{2}} \boldsymbol{\Sigma}_{22} \boldsymbol{\beta}_j$$

Propiedades

Propiedad 1. Sean

$$\mathbf{\Delta}_{\mathbf{x}} = \text{DIAG}(\text{VAR}(x_1), \dots, \text{VAR}(x_p))$$

$$\mathbf{\Delta}_{\mathbf{y}} = \text{DIAG}(\text{VAR}(y_1), \dots, \text{VAR}(y_q))$$

Se tiene que

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

es decir,

$$\text{CORR}(u_j, \mathbf{x}) = \mathbf{\Delta}_{\mathbf{x}}^{-\frac{1}{2}} \boldsymbol{\Sigma}_{11} \boldsymbol{\alpha}_j$$

$$\text{CORR}(v_j, \mathbf{y}) = \mathbf{\Delta}_{\mathbf{y}}^{-\frac{1}{2}} \boldsymbol{\Sigma}_{22} \boldsymbol{\beta}_j$$

$$\text{CORR}(u_j, \mathbf{y}) = \mathbf{\Delta}_{\mathbf{y}}^{-\frac{1}{2}} \boldsymbol{\Sigma}_{21} \boldsymbol{\alpha}_j$$

$$\text{CORR}(v_j, \mathbf{x}) = \mathbf{\Delta}_{\mathbf{x}}^{-\frac{1}{2}} \boldsymbol{\Sigma}_{12} \boldsymbol{\beta}_j$$

Propiedades

Propiedad 1. Sean

$$\mathbf{\Delta}_{\mathbf{x}} = \text{DIAG}(\text{VAR}(x_1), \dots, \text{VAR}(x_p))$$

$$\mathbf{\Delta}_{\mathbf{y}} = \text{DIAG}(\text{VAR}(y_1), \dots, \text{VAR}(y_q))$$

Se tiene que

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

es decir,

$$\text{CORR}(u_j, \mathbf{x}) = \mathbf{\Delta}_{\mathbf{x}}^{-\frac{1}{2}} \boldsymbol{\Sigma}_{11} \boldsymbol{\alpha}_j$$

$$\text{CORR}(v_j, \mathbf{y}) = \mathbf{\Delta}_{\mathbf{y}}^{-\frac{1}{2}} \boldsymbol{\Sigma}_{22} \boldsymbol{\beta}_j$$

$$\text{CORR}(u_j, \mathbf{y}) = \mathbf{\Delta}_{\mathbf{y}}^{-\frac{1}{2}} \boldsymbol{\Sigma}_{21} \boldsymbol{\alpha}_j$$

$$\text{CORR}(v_j, \mathbf{x}) = \mathbf{\Delta}_{\mathbf{x}}^{-\frac{1}{2}} \boldsymbol{\Sigma}_{12} \boldsymbol{\beta}_j$$

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

Propiedades

Propiedad 3. Invarianza del análisis.

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

$$\mathbf{\Upsilon}_1 = \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \quad \mathbf{\Upsilon}_2 = \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}$$

tomamos

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

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

- Las correlaciones canónicas no cambian
- $\alpha_j^* = \Delta_{\mathbf{x}}^{1/2} \alpha_j$,
- $\beta_j^* = \Delta_{\mathbf{y}}^{1/2} \beta_j$

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 ($\Sigma > 0$), $\mathbf{x}_i \in \mathbb{R}^p$, $\mathbf{y}_i \in \mathbb{R}^q$. Sea

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

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

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

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

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

- $\mathbf{u}_i = (u_{1,i}, \dots, u_{q,i})^T$ con $u_{j,i} = \hat{\alpha}_j^T \tilde{\mathbf{x}}_i$ y
- $\mathbf{v}_i = (v_{1,i}, \dots, v_{q,i})^T$ $v_{j,i} = \hat{\beta}_j^T \tilde{\mathbf{y}}_i$

$$\begin{aligned} \mathbf{S}_{11}^{-1} \mathbf{S}_{12} \mathbf{S}_{22}^{-1} \mathbf{S}_{21} \hat{\alpha}_j &= r_j^2 \hat{\alpha}_j & \hat{\alpha}_j^T \mathbf{S}_{11} \hat{\alpha}_j &= 1 \\ \mathbf{S}_{22}^{-1} \mathbf{S}_{21} \mathbf{S}_{11}^{-1} \mathbf{S}_{12} \hat{\beta}_j &= r_j^2 \hat{\beta}_j & \hat{\beta}_j^T \mathbf{S}_{22} \hat{\beta}_j &= 1 \end{aligned}$$

Observemos que

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

La distribución exacta de r_j , aún en 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

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

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

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

- Además, $\mathbf{r}_k^{(2)}$ es independiente de r_j para $j > k$.
- Si $w_j = nr_j^2$ para $j > k$, entonces w_{k+1}, \dots, w_q son dependientes y no-normales. Su distribución asintótica conjunta es la distribución de los autovalores de una matriz $\mathbf{A} \in \mathbb{R}^{(q-k) \times (q-k)}$ tal que $\mathbf{AW}(\mathbf{I}_{q-k}, q - m, p - k)$. Su densidad asintótica es

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

Ejemplo

La matriz **S** es igual a

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

Ejemplo

La matriz **S** es igual a

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

$$\hat{\alpha}_j^T \mathbf{Q}_{11} \hat{\alpha}_j = 1 \quad \hat{\beta}_j^T \mathbf{Q}_{22} \hat{\beta}_j = 1$$

mientras que en CCA, están normalizados de modo que

$$\hat{\alpha}_j^T \mathbf{S}_{11} \hat{\alpha}_j = 1 \quad \hat{\beta}_j^T \mathbf{S}_{22} \hat{\beta}_j = 1$$

Estos últimos son los que presentamos

	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$
Control	-1.2538	-0.6215	-0.6617
Concept	0.3513	-1.1877	0.8267
Motivation	-1.2624	2.0273	2.0002
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
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

En CANCOR, los vectores canónicos están normalizados de modo que

$$\hat{\alpha}_j^T \mathbf{Q}_{11} \hat{\alpha}_j = 1 \quad \hat{\beta}_j^T \mathbf{Q}_{22} \hat{\beta}_j = 1$$

mientras que en CCA, están normalizados de modo que

$$\hat{\alpha}_j^T \mathbf{S}_{11} \hat{\alpha}_j = 1 \quad \hat{\beta}_j^T \mathbf{S}_{22} \hat{\beta}_j = 1$$

Estos últimos son los que presentamos

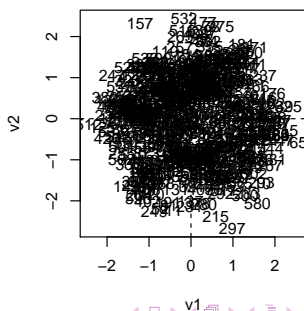
	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$
Control	-1.2538	-0.6215	-0.6617
Concept	0.3513	-1.1877	0.8267
Motivation	-1.2624	2.0273	2.0002
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
Read	-0.0446	-0.0049	0.0214
Write	-0.0359	0.0421	0.0913
Math	-0.0234	0.0042	0.0094
Science	-0.0050	-0.0852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

	$\text{CORR}(\mathbf{u}_1, \mathbf{x})$	$\text{CORR}(\mathbf{u}_2, \mathbf{x})$	$\text{CORR}(\mathbf{u}_3, \mathbf{x})$	$\text{CORR}(\mathbf{v}_1, \mathbf{x})$	$\text{CORR}(\mathbf{v}_2, \mathbf{x})$	$\text{CORR}(\mathbf{v}_3, \mathbf{x})$
<i>Control</i>	-0.9040	-0.3897	-0.1756	-0.4196	-0.0653	-0.0183
<i>Concept</i>	-0.0208	-0.7087	0.7052	-0.0097	-0.1187	0.0733
<i>Motivation</i>	-0.5672	0.3509	0.7451	-0.2632	0.0588	0.0775

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



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

- $\hat{\alpha}_j^* = \hat{\Delta}_x^{1/2} \hat{\alpha}_j$,
- $\hat{\beta}_j^* = \hat{\Delta}_y^{1/2} \hat{\beta}_j$

permiten comparaciones más simples entre variables. Estas direcciones cumplen

$$\hat{\alpha}_j^{*T} \hat{\mathbf{R}}_{11} \hat{\alpha}_j^* = 1 \quad \hat{\beta}_j^{*T} \hat{\mathbf{R}}_{22} \hat{\beta}_j^* = 1$$

donde $\hat{\mathbf{R}}_{11} = \hat{\Delta}_x^{-1/2} \mathbf{S}_{11} \hat{\Delta}_x^{-1/2}$ y $\hat{\mathbf{R}}_{22} = \hat{\Delta}_y^{-1/2} \mathbf{S}_{22} \hat{\Delta}_y^{-1/2}$

	$\hat{\alpha}_1^*$	$\hat{\alpha}_2^*$	$\hat{\alpha}_3^*$
Control	-0.8404	-0.4166	-0.4435
Concept	0.2479	-0.8379	0.5833
Motivation	-0.4327	0.6948	0.6855
	$\hat{\beta}_1^*$	$\hat{\beta}_2^*$	$\hat{\beta}_3^*$
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.

	$\hat{\alpha}_1^*$	$\hat{\alpha}_2^*$	$\hat{\alpha}_3^*$
Control	-0.8404	-0.4166	-0.4435
Concept	0.2479	-0.8379	0.5833
Motivation	-0.4327	0.6948	0.6855
	$\hat{\beta}_1^*$	$\hat{\beta}_2^*$	$\hat{\beta}_3^*$
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

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

independientes bajo la hipótesis nula de interés.

Más generalmente, tendremos

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

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 $\mathbf{H} = \sum_{j=1}^r \mathbf{z}_j \mathbf{z}_j^T \sim \mathcal{W}(\mathbf{\Sigma}, p, r)$

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 $\mathbf{H} = \sum_{j=1}^r \mathbf{z}_j \mathbf{z}_j^T \sim \mathcal{W}(\mathbf{\Sigma}, p, r)$

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} .

Criterio de Wilks

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

Definición Sean $\mathbf{U} \sim \mathcal{W}(\boldsymbol{\Sigma}, p, N - r)$ y $\mathbf{z}_j \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, $1 \leq j \leq r$ 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 $\mathbf{H} = \sum_{j=1}^r \mathbf{z}_j \mathbf{z}_j^T \sim \mathcal{W}(\boldsymbol{\Sigma}, p, r)$

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

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

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

b) Si $r < p$,

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

con $b_{11}^2, \dots, b_{rr}^2$ 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

$$f = N - \frac{p + r + 1}{2} \quad m = pr$$

$$\lambda = -\frac{pr}{2} + 1 \quad s = \frac{(p^2 r^2 - 4)^{\frac{1}{2}}}{(p^2 + r^2 - 5)^{\frac{1}{2}}}$$

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

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

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 $H_{01} : \Sigma = \text{diag}(\Sigma_{11}, \Sigma_{22})$

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

Queremos testear $\Sigma_{12} = 0$ o equivalentemente $H_{01} : \rho_1 = 0$

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

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

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

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

Queremos testear $\Sigma_{12} = 0$ 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

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

donde

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

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

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

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

Por otra parte,

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

con lo cual

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

donde r_j^2 son los autovalores de $\mathbf{Q}_{12} \mathbf{Q}_{22}^{-1} \mathbf{Q}_{21} \mathbf{Q}_{11}^{-1}$,

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

Por otra parte,

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

con lo cual

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

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

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

con $k = \exp(-\chi_{\nu, \alpha}^2 / m)$ con $m = n - (p + q + 3)/2$.

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

Por otra parte,

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

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

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

Luego,

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

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

Queremos aplicar el principio de unión intersección de Roy para testear $H_{01} : \Sigma_{12} = 0$.

Recordemos que

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

entonces $\text{COV}(\mathbf{a}^T \mathbf{x}, \mathbf{b}^T \mathbf{y}) = \mathbf{a}^T \Sigma_{12} \mathbf{b}$.

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

Queremos aplicar el principio de unión intersección de Roy para testear $H_{01} : \Sigma_{12} = 0$.

Recordemos que

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

entonces $\text{COV}(\mathbf{a}^T \mathbf{x}, \mathbf{b}^T \mathbf{y}) = \mathbf{a}^T \Sigma_{12} \mathbf{b}$. Definamos

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

Luego

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

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

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

Es decir, el test rechaza si

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

donde θ_{\max} es la máxima raíz de $|\mathbf{H} - \theta(\mathbf{U} + \mathbf{H})| = 0$, o sea, θ_{\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 rechazemos $H_{01} : \rho_1 = 0$ es posible que $\rho_2 = 0$, por lo tanto nos interesa testear

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

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 Σ_{12}

Si $\rho_1 \geq \dots \geq \rho_k > 0$ 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

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

se basa en

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

Determinando la cantidad de correlaciones canónicas

El test de cociente de verosimilitud para

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

se basa en

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

Además,

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

Determinando la cantidad de correlaciones canónicas

El test de cociente de verosimilitud para

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

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

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

$$f = n - 3/2 - (p + q)/2$$

$$m = (p - k)(q - k)$$

$$\lambda = -\frac{(p - k)(q - k)}{2} + 1$$

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

Determinando la cantidad de correlaciones canónicas

El test de cociente de verosimilitud para

$H_{0,k+1} : \rho_{k+1} = 0 \quad \rho_k > 0$ se basa en

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

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

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

Determinando la cantidad de correlaciones canónicas

El test de cociente de verosimilitud para

$H_{0,k+1} : \rho_{k+1} = 0 \quad \rho_k > 0$ se basa en

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

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

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

- Bartlett (1947) sugiere tomar $m = n - (p + q + 3)/2$, y usar la aproximación

$$-m \log(\gamma_{k+1}) \xrightarrow{D} \chi_{\nu_{k+1}}^2.$$

- Glynn y Muirhead (1978) sugieren la modificación

$$\ell_{k+1} = - \left[n - k - \frac{p + q + 3}{2} + \sum_{j=1}^k \frac{1}{r_j^2} \right] \log(\gamma_{k+1}) \xrightarrow{D} \chi_{\nu_{k+1}}^2.$$

Determinando la cantidad de correlaciones canónicas

Nos interesaba testear

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

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

Para determinar k testeamos la secuencia

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

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

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

Ejemplo

Deseamos saber cuantas correlaciones son significativas

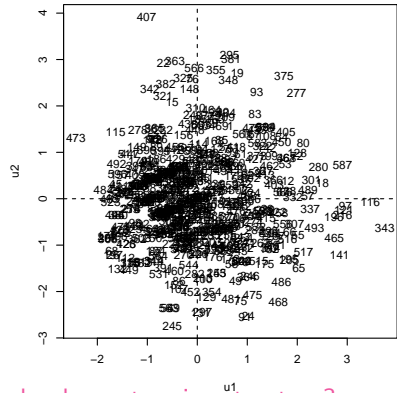
$H_{0,k+1}$	γ_{k+1}	\mathcal{F}	$\nu_{1,k+1}$	$\nu_{2,k+1}$	$p - valor$
$H_{01} : \rho_1 = 0$	0.7544	11.716	15	1635	$7.498e - 28$
$H_{02} : \rho_2 = 0(\rho_1 > 0)$	0.9614	2.944	8	1186	0.002905
$H_{03} : \rho_3 = 0(\rho_2 > 0)$	0.9892	2.165	3	594	0.09109

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

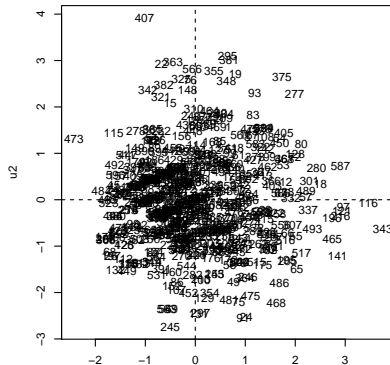
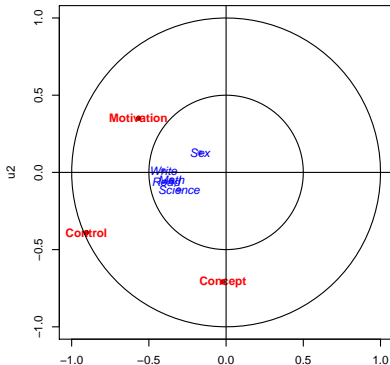
Alcanza con 2 variables canónicas.

b)



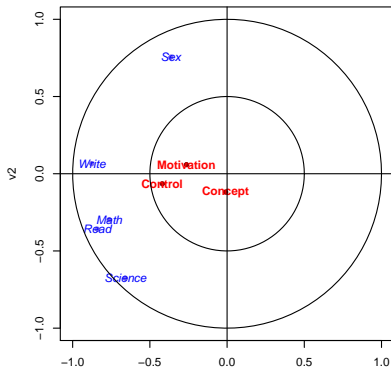
- Porqué el gráfico en b) es una nube de puntos sin estructura?

b)

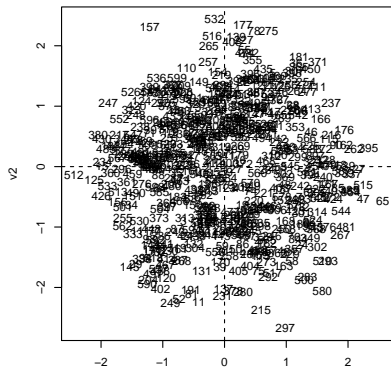


- Porqué el gráfico en b) es una nube de puntos sin estructura?
- En a) cada punto es la correlación de la variable indicada con los ejes u_1 , u_2 .
- En b), las distancias entre puntos del espacio de variables canónicas $(u_{1,i}, u_{2,i}) = (\hat{\alpha}_1^T \tilde{\mathbf{x}}_i, \hat{\alpha}_2^T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

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 $(v_{1,i}, v_{2,i}) = (\hat{\beta}_1^T \tilde{y}_i, \hat{\beta}_2^T \tilde{y}_i)$ dan aproximadamente la distancia de Mahalanobis entre observaciones \tilde{y}_i .

Planteo del problema

○○○
○

Definiciones

○○○○○○○
○○○○○○○○○

Estimación

○○○○
○○○○○○○

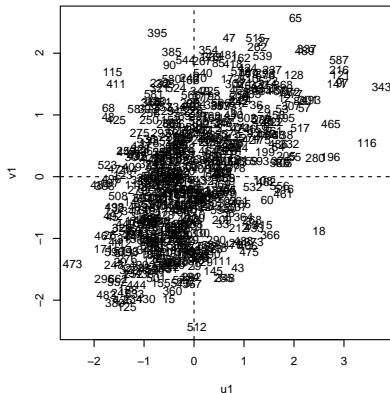
Criterio de Wilks

○○
○○○○

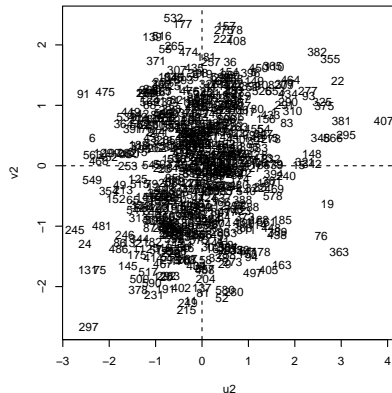
Inferencia

○○○○○
○○○
○○○●○○

a)



b)



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

Vectores canónicos estandarizados

Espacio x			Espacio y		
	$\hat{\alpha}_1^*$	$\hat{\alpha}_2^*$		$\hat{\beta}_1^*$	$\hat{\beta}_2^*$
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)
 - u_2 está influenciado por Concepto (-.84) and y Motivación (.69).

Vectores canónicos estandarizados

Espacio x			Espacio y		
	$\hat{\alpha}_1^*$	$\hat{\alpha}_2^*$		$\hat{\beta}_1^*$	$\hat{\beta}_2^*$
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)
 - u_2 está influenciado por Concepto (-.84) and y Motivación (.69).
- Para las variables académicas más sexo,
 - v_1 es un compromiso entre Lectura (.45), Escritura (.35) y Sexo (.32),
 - para 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,

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

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