Treball final de grau

GRAU DE MATEMÀTIQUES

Facultat de Matemàtiques
Universitat de Barcelona

ANÀLISI DE CORRELACIONS CANÒNIQUES

Autor: Lorena Muelas Gorriz

Director:	**Dr. Josep Fortiana**
Co-Director:	**Dra. Carme Florit**
Realitzat a:	**Departament de Probabilitat, Lògica i Estadística**
Barcelona,	**27 de juny de 2016**

Abstract

Canonical Correlations Analysis (CCA) is the most general of multivariate statistical techniques. It can be considered an extension of the concepts of analysis simple and multiple correlation.

In this work we study CCA according to the classical definition of Hotelling (1936), adding a geometric perspective. Furthermore we explore relationship between CCA and other multivariate tecniques.

Resum

L'anàlisi de correlacions canòniques és la més general de les tècniques estadístiques multivariants. Es pot considerar una extensió dels conceptes de correlació simple i múltiple.

En aquest treball estudiem la correlació canònica segons la definició clàssica de Hotelling (1936) i hi afegim una visió geomètrica. A més, veurem les relacions amb altres tècniques multivariants.

Continguts

1 Introducció **4**

2 Dades Multivariants **5**

2.1 Matriu de dades 5

2.2 Diagonalització de matrius 6

2.3 Descomposició singular 8

2.4 Inversa generalitzada 9

2.5 Projecció Ortogonal 10

3 Anàlisi de correlacions canòniques **12**

3.1 Correlació simple 12

3.2 Correlació multiple 12

3.3 Correlacions canòniques 15

3.4 Correlacions canòniques i descomposició singular 20

3.5 Significació de les correlacions canòniques 24

4 Visió geomètrica de les correlacions canòniques **26**

5 Relació amb altres tècniques multivariants **31**

5.1 Anàlisi de components principals 31

5.2 Anàlisi canònica de poblacions 35

5.3 Anàlisi de correspondències 36

6 Exemples **41**

7 Conclusions **44**

1 Introducció

Què és l'anàlisi de correlacions canòniques?

L'anàlisi de correlacions canòniques és un mètode d'anàlisi multivariant desenvolupat per Harold Hotelling. El seu objectiu és buscar les relacions lineals que hi pugui haver entre dos grups de variables.
Es diferència de l'anàlisi de correlació múltiple en el fet que aquesta només prediu una variable (resposta) a partir d'un grup de variables (predictores), mentre que l'anàlisi de correlacions canòniques busca relacions entre dos conjunts de variables.

Motivació

La motivació per aquest Treball Final de Grau va sorgir pel meu interés per l'Estadística i, en concret, per l'anàlisi de dades i, en particular, l'anàlisi de correlacions canòniques.

L'Estadística sempre ha sigut un tema pel qual he mostrat molta atenció i em semblava interessant aprofundir més en aquest àmbit.

Objectius

L'objectiu d'aquest treball és fer un estudi clàssic de l'anàlisi de correlacions canòniques i tenir una visió geomètrica de la primera correlació canònica.

Una vegada estudiades les correlacions canòniques veurem les relacions que hi ha entre l'anàlisi de correlacions canòniques i altres tècniques multivariants, en concret, l'anàlisi de components principals, l'ànalisi canònica de poblacions i per últim l'anàlisi de correspondències.

2 Dades Multivariants

Abans de començar amb la teoria del tema, es defineixen alguns conceptes previs.

2.1 Matriu de dades

Suposem que sobre els individus $\omega_1, \ldots, \omega_n$ s'han observat les variables $X_1, \ldots, X_p$. Sigui $x_{ij} = X_j(\omega_i)$ l' observació de la variable X_j sobre l'individu ω_i. La matriu de dades multivariants és:

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1j} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i1} & \cdots & x_{ij} & \cdots & x_{ip} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nj} & \cdots & x_{np} \end{pmatrix}.$$

Les files de X s'identifiquen amb els individus i les columnes de X amb les variables.

Siguin,

1. x_i la fila i-èsima de X.

2. X_j la columna j-èsima de X.

3. $\overline{x} = (\overline{x}_1, \ldots, \overline{x}_j, \ldots, \overline{x}_p)$ el vector fila de les mitjanes de les variables tal que

$$\overline{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}.$$

4. La matriu simètrica $p \times p$ de covariàncies és:

$$S = \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{pmatrix},$$

on

$$s_{jj'} = \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \overline{x}_j)(x_{ij'} - \overline{x}_{j'}),$$

és la covariància entre les variables j, j' i $\overline{x}$ i S són mesures multivariants de tendència central i dispersió, respectivament.

5. La matriu simètrica $p \times p$ de correlacions és,

$$R = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{pmatrix},$$

on $r_{jj'} = cor(X_j, X_{j'})$ és el coeficient de correlació entre les variables X_j, $X_{j'}$. Aquest coeficient ve donat per

$$r_{jj'} = \frac{s_{jj'}}{s_j s_{j'}},$$

on s_j, $s_{j'}$ són les desviacions típiques.

6. La relació entre la matriu de covariàncies i la matriu de correlacions ve donada per

$$R = D_s^{-1/2} \cdot S \cdot D_s^{1/2},$$

on D_s és la diagonal de S.

2.2 Diagonalització de matrius

Sigui A una matriu quadrada d'ordre n

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}.$$

El vector columna $\mathbf{v} = (v_1, \ldots, v_n)'$ és un vector propi d'A de valor propi λ si

$$A\mathbf{v} = \lambda\mathbf{v}.$$

Els valors propis s'obtenen resolvent *l'equació característica*:

$$det(A - \lambda I) = 0.$$

Algunes propietats bàsiques dels valors propis són,

1. La suma dels valors propis és igual a la traça d'A.

2. El producte dels valors propis és igual al determinant d'A.

3. Si T és la matriu-columna amb els vectors propis i D_λ és la matriu diagonal amb els valors propis, aleshores

$$AT = TD_\lambda.$$

A l'anàlisi multivariant són especialment importants les matrius simètriques donat que les matrius de covariàncies i correlacions ho són. Les propietats més importants en aquest cas són,

1. Els valors propis són sempre reals.

2. Dos vectors propis $\mathbf{u}$, $\mathbf{v}$, associats a valors propis diferents són sempre ortogonals, és a dir,

$$A\mathbf{u} = \lambda_i\mathbf{u}, \qquad A\mathbf{v} = \lambda_j\mathbf{v}, \qquad \lambda_i \neq \lambda_j \Rightarrow \mathbf{u'} \cdot \mathbf{v} = 0.$$

3. Tota matriu simètrica pot expressar-se com el producte:

$$A = TD_\lambda T',$$

on T és una matriu ortogonal, és a dir, $T \cdot T' = T' \cdot T = I$ amb els vectors propis normalitzats i D_λ és la matriu diagonal amb els valors propis.

Observem que si A és una matriu definida positiva, és a dir, amb tots els seus valors propis positius, aleshores existeix

$$D_\lambda^{1/2} = \mathrm{diag}(\sqrt{\lambda_1}, \cdots, \sqrt{\lambda_n}),$$

i agafant $B = TD_\lambda^{1/2}$ es verifica $A = B \cdot B'$.

A més, si A és definida positiva podem escriure,

$$A^{-1} = TD_\lambda^{-1}T,$$

$$A^{-1/2} = TD_\lambda^{-1/2}T.$$

2.3 Descomposició singular

Per a tota matriu A, $n \times p$. Suposarem $n \geq p$ sense pèrdua de generalitat i $rang(A) = r \leq p$.

Existeix:

1. U una matriu $n \times r$ tal que $U' \cdot U = I_r$. Les seves r columnes són vectors ortonormals i vectors propis d' $A \cdot A'$.
2. V una matriu ortogonal d'ordre $p \times r$. Les seves columnes són vectors propis d' $A' \cdot A$ tal que $V \cdot V' = I_r$.
3. $D = diag(s_1, \ldots, s_r)$ és una matriu diagonal tal que

$$s_1 > s_2 > \ldots > s_r > 0,$$

amb la propietat que:

$$A = U \cdot D \cdot V'.$$

Aquesta descomposició s'anomena *descomposició en valors singulars* d'A.

Els nombres $s_1, \ldots, s_r$ s'anomenen valors singulars d'A que verifiquen

$$s_i = \sqrt{\lambda_i}, \quad i = 1, \ldots, p,$$

on $\lambda_1, \ldots, \lambda_p$ són els valors propis d'$A' \cdot A$.

Algunes propietats són,

1. El nombre de valors singulars d'A és igual al rang d'A.

2. Si A és simètrica, $n = p$, els valors singulars coincideixen amb els valors propis d'A, aleshores, $U = V$.

2.4 Inversa generalitzada

Si A és una matriu quadrada d'ordre $n \times n$ no singular, és a dir, $\text{rang}(A) = n$, existeix la matriu inversa A^{-1} tal que

$$AA^{-1} = A^{-1}A = I_n.$$

Si el $\text{rang}(A) = r < n$, o A no és una matriu quadrada, la inversa no existeix però existeixen inverses generalitzades o g-inverses A^-.

Sigui A una matriu d'ordre $n \times r$, amb $n \geq r$. S'anomena *inversa generalitzada* o *g-inversa* d'A a una matriu A^- que verifica,

$$AA^-A = A \quad \text{i} \quad A^-AA^- = A^-.$$

La g-inversa no és unica, però si A^- verifica addicionalment

$$(AA^-)' = AA^-, \qquad (A^-A)' = A^-A,$$

aleshores és única i s'anomena *pseudo-inversa* o *g-inversa de Moore-Penrose*.

Sigui el $\text{rang}(A) = r$ i $A = U \cdot D \cdot V'$ la descomposició singular d'A, amb

$$D = \text{diag}(s_1, \ldots, s_r),$$

aleshores,

$$D^- = \text{diag}(s_1^{-1}, \ldots, s_r^{-1}),$$

i la matriu $p \times n$

$$A^- = V \cdot D^- \cdot U'$$

és una g-inversa de Moore Penrose d'A.

En efecte

$$\begin{aligned} AA^-A &= UDV'VD^-U'UDV' = A, \\ A^-AA^- &= VD^-U'UDV'VD^-U' = A^-. \end{aligned}$$

2.5 Projecció Ortogonal

Siguin V i W siguin dos subespais linelament disjunts de l'espai ε_n de dimensió n tal que $V \oplus W$. Llavors qualsevol vector $\mathbf{x} \in \varepsilon_n$ es pot descompondre de manera única com $\mathbf{x} = \mathbf{y} + \mathbf{z}$ on $\mathbf{y} \in V$ and $\mathbf{z} \in W$.
La correspondència de $\mathbf{x}$ a $\mathbf{y}$ determina una transformació lineal $\mathbf{P}$ d'ε_n a V que s'expressa per,

$$\mathbf{y} = \mathbf{P}\mathbf{x}.$$

Anàlogament, la correspondència de $\mathbf{x}$ a $\mathbf{z}$ determina una transformació lineal $\mathbf{Q}$ d'ε_n a W que s'expressa per,

$$\mathbf{z} = \mathbf{Q}\mathbf{x}.$$

Les matrius $\mathbf{P}$ i $\mathbf{Q}$ són matrius quadrardes d'ordre n i s'anomenen *matrius de projecció* i algunes de les propietats que compleixen són:

- $\mathbf{P}$ i $\mathbf{Q}$ són lineals i úniques en ε_n.
- $\mathbf{P}$ i $\mathbf{Q}$ són matrius idempotents, és a dir, $\mathbf{P}^2 = \mathbf{P}$ i $\mathbf{Q}^2 = \mathbf{Q}$.

Les matrius $\mathbf{P}$ i $\mathbf{Q}$ són matrius de *projecció ortogonal* si a més compleixen,

$$\mathbf{P}' = \mathbf{P}, \qquad \mathbf{Q}' = \mathbf{Q}.$$

Amb les propietats anteriors sabem que una matriu de projecció ortogonal és idempotent i simètrica, però reciprocament, tota matriu real, idempotent i simètrica és una matriu de projecció ortogonal.

L'expressió per una matriu de projecció ortogonal sobre un subespai V és,

$$\mathbf{P} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}',$$

amb $\mathbf{A}$ una matriu on les seves columnes són una base d'V.

3 Anàlisi de correlacions canòniques

En aquest capítol estudiarem la relació multivariant entre vectors aleatoris. Introduirem i estudiarem les correlacions canòniques, que són generalitzacions de les correlacions simple i múltiple.

Tenim tres possibilitats per relacionar dos variables,

- La correlació simple si X, Y són dos variables aleatòries.
- La correlació múltiple si Y és una variable aleatòria i $\mathbf{X} = (X_1, \ldots, X_p)$ és un vector aleatòri.
- Les correlacions canoniques si $\mathbf{X} = (X_1, \ldots, X_p)$ i $\mathbf{Y} = (Y_1, \ldots, Y_q)$ són dos vectors aleatoris.

3.1 Correlació simple

Siguin X i Y dos variables aleatòries. Siguin σ_X^2 i σ_Y^2 les seves variàncies respectives i σ_{XY} la seva covariància. S'anomena el *coeficient de correlació simple* entre ambdues variables a,

$$\rho_{XY} = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2 \sigma_Y^2}}.$$

El coeficient de correlació és un nombre que indica el grau de relació lineal entre les variables X i Y. El seu valor varia entre -1 i $+1$.

- Si $r = -1$ la relació és inversament proporcional.
- Si $r = +1$ la relació és directament proporcional.
- Si $r = 0$ no existeix cap relació lineal entre les variables.

3.2 Correlació multiple

Volem relacionar una variable Y amb p variables explicatives $X_1, \ldots, X_p$, que suposem centrades. El model de regressió múltiple consisteix a trobar la combinació lineal

$$\hat{Y} = \beta_1 X_1 + \ldots + \beta_p X_p,$$

que millor s'ajusti a la variable Y.

Sigui Σ la matriu de covariancies de $\mathbf{X}$ i sigui $\delta = (\delta_1, \ldots, \delta_p)'$ el vector columna amb les covariancies $\delta_i = \mathrm{cov}(Y, X_j)$, $j = 1, \ldots, p$. El criteri d'ajust és el de minims quadrats.

Teorema 3.2.1. *Els coeficients $\hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_p)$ que minimitzen $E(Y - \hat{Y})^2$ verifiquen l'equació*

$$\hat{\beta} = \Sigma^{-1}\delta,$$

o equivalentment

$$\Sigma\hat{\beta} = \delta.$$

Demostració. Sigui,

$$E(Y - \hat{Y})^2 = E(Y)^2 + E(\hat{Y})^2 - 2E(Y\hat{Y}) = \mathrm{var}(Y) + \beta'\Sigma\beta - 2\beta'\delta.$$

Derivant vectorialment respecte β i igualant a 0 tenim,

$$2\Sigma\beta - 2\delta = 0.$$

□

El valor ajustat de Y és:

$$\hat{Y} = \mathbf{X}\hat{\beta} = \hat{\beta}_1 X_1 + \ldots + \hat{\beta}_p X_p.$$

Si posem

$$Y = \hat{Y} + \tilde{Y},$$

aleshores $\tilde{Y}$ és la *variable residual.*

La correlació múltiple entre Y i $X_1, \ldots, X_p$ és la correlació simple entre Y i la millor predicció $\hat{Y} = \mathbf{X}\hat{\beta}$.

S'indica per $R = cor(Y, \hat{Y})$ que verifica

1. $0 \leq R \leq 1$.

2. $R = 1$ si Y és combinació lineal de $X_1, \ldots, X_p$.

3. $R = 0$ si Y està incorrelacionada amb cadascuna de les variables X_i.

Teorema 3.2.2. *La variables $\hat{Y}$, $\tilde{Y}$ i la correlació múltiple R compleixen:*

1. *$\hat{Y}$ i $\tilde{Y}$ són variables incorrelacionades.*

2. *$var(Y) = var(\hat{Y}) + var(\tilde{Y})$.*

3. *$R^2 = var(\hat{Y})/var(Y)$.*

Demostració.

1. És consecuencia de $\Sigma\beta = \delta$. En efecte,

$$cov(\hat{Y}, \tilde{Y}) = E(\hat{Y}\tilde{Y}) = E(\hat{\beta}'\mathbf{X}(Y - \hat{\beta}'\mathbf{X})) = \hat{\beta}'\delta - \hat{\beta}'\Sigma\hat{\beta} = 0.$$

2. Aplicant la fórmula de la variància d'una suma

$$var(X + Y) = var(X) + var(Y) + 2cov(X, Y),$$

tenim,

$$var(Y) = var(\hat{Y} + \tilde{Y}) = var(\hat{Y}) + var(\tilde{Y}) + 2cov(\hat{Y}, \tilde{Y}) = var(\hat{Y}) + var(\tilde{Y}).$$

3. De

$$cov(Y, \hat{Y}) = cov(Y, \sum_{i=1}^{p} \hat{\beta}_i X_i = \sum_{i=1}^{p} \hat{\beta}_i \delta_i = \hat{\beta}'\delta = \hat{\beta}'\Sigma\hat{\beta} = var(\hat{Y}),$$

obtenim

$$R^2 = \frac{cov^2(Y, \hat{Y})}{var(Y)var(\hat{Y})} = \frac{var(\hat{Y})}{var(Y)}.$$

□

3.3 Correlacions canòniques

Siguin $\mathbf{X} = (X_1, \ldots, X_p)$, $\mathbf{Y} = (Y_1, \ldots, Y_q)$ dos vectors aleatoris de dimensions p i q amb vectors d'esperances $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)$ i $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_q)$ respectivament. Es planteja el problema de trobar dues variables compostes,

$$U = \mathbf{X}\mathbf{a} = a_1X_1 + \cdots + a_pX_p, \qquad V = \mathbf{Y}\mathbf{b} = b_1Y_1 + \cdots + b_qY_q,$$

on $\mathbf{a} = (a_1, \ldots a_p)'$, $\mathbf{b} = (b_1, \ldots b_q)'$ són dos vectors, tals que la correlació:

$$cor(U, V) = \frac{\mathbf{a}'\Sigma_{XY}\mathbf{b}}{\sqrt{\mathbf{a}'\Sigma_{XX}\mathbf{a}}\sqrt{\mathbf{b}'\Sigma_{YY}\mathbf{b}}},$$

sigui màxima.

Siguin Σ_{XX}, Σ_{YY} les matrius de covariàncies del primer i segon conjunt, és a dir, dels vectors $\mathbf{X}$, $\mathbf{Y}$, respectivament,

$$E(\mathbf{X} - \boldsymbol{\mu})'(\mathbf{X} - \boldsymbol{\mu}) = \Sigma_{XX}, \qquad E(\mathbf{Y} - \boldsymbol{\nu})'(\mathbf{Y} - \boldsymbol{\nu}) = \Sigma_{YY},$$

i sigui Σ_{XY} la matriu $p \times q$ amb les covariàncies de les variables $\mathbf{X}$ amb les variables $\mathbf{Y}$, és a dir,

$$E(\mathbf{X} - \boldsymbol{\mu})'(\mathbf{Y} - \boldsymbol{\nu}) = \Sigma_{XY} = \Sigma_{YX},$$

és a dir, la matriu de covariàncies conjunta de (X, Y) té la següent forma en caixes:

	$\mathbf{X}$	$\mathbf{Y}$
$\mathbf{X}$	Σ_{XX}	Σ_{XY}
$\mathbf{Y}$	Σ_{YX}	Σ_{YY}

.

Suposarem que Σ_{XX}, Σ_{YY} són no singulars.

$$var(U) = \mathbf{a}'\Sigma_{XX}\mathbf{a} = 1, \qquad var(V) = \mathbf{b}'\Sigma_{YY}\mathbf{b} = 1.$$

Així el problema es redueix a

$$\text{maximitzar } \mathbf{a}'\Sigma_{XY}\mathbf{b} \text{ restringit a } \mathbf{a}'\Sigma_{XX}\mathbf{a} = 1,\ \mathbf{b}'\Sigma_{YY}\mathbf{b} = 1.$$

Definició. Utilitzant les notacions anteriors

1. Els vectors de coeficients $\mathbf{a}$, $\mathbf{b}$ són els **vectors canònics**.

2. Les variables U,V són les **variables canòniques**.

3. La màxima correlació entre U i V és la **primera correlació canònica** r_1.

Teorema 3.3.1. *Els primers vectors canònics satisfan les equacions*

$$\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\mathbf{a} = \lambda\Sigma_{XX}\mathbf{a}, \tag{3.1}$$

$$\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\mathbf{b} = \lambda\Sigma_{YY}\mathbf{b}. \tag{3.2}$$

Demostració. Considerem la funció lagrangiana:

$$\phi(\mathbf{a},\mathbf{b}) = \mathbf{a}'\Sigma_{XY}\mathbf{b} - \frac{\delta}{2}(\mathbf{a}'\Sigma_{XX}\mathbf{a} - 1) - \frac{\gamma}{2}(\mathbf{b}'\Sigma_{YY}\mathbf{b} - 1),$$

on δ i γ són multiplicadors de Lagrange.

Aleshores de:

$$\frac{\partial\phi}{\partial\mathbf{a}} = \frac{\partial\phi}{\partial\mathbf{b}} = 0,$$

obtenim dues equacions,

$$\Sigma_{XY}\mathbf{b} - \delta\Sigma_{XX}\mathbf{a} = 0, \qquad \Sigma_{YX}\mathbf{a} - \gamma\Sigma_{YY}\mathbf{b} = 0. \tag{3.3}$$

Multiplicant a l'esquerra la primera per $\mathbf{a}'$ i la segona per $\mathbf{b}'$, tenim

$$\mathbf{a}'\Sigma_{XY}\mathbf{b} = \delta\mathbf{a}'\Sigma_{XX}\mathbf{a}, \qquad \mathbf{b}'\Sigma_{YX}\mathbf{a} = \gamma\mathbf{b}'\Sigma_{YY}\mathbf{b},$$

donant lloc a que $\delta = \gamma$. Així doncs de (3.3) obtenim,

$$\mathbf{a} = \delta^{-1}\Sigma_{XX}^{-1}\Sigma_{XY}\mathbf{b}, \qquad \mathbf{b} = \delta^{-1}\Sigma_{YY}^{-1}\Sigma_{YX}\mathbf{a},$$

que, substituint a l'equació (3.3), obtenim:

$$\delta^{-1}\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\mathbf{a} - \delta\Sigma_{XX}\mathbf{a} = 0,$$
$$\delta^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\mathbf{b} - \delta\Sigma_{YY}\mathbf{b} = 0.$$

Multiplicant per δ donat que és un factor multiplicatiu obtenim,

$$\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\mathbf{a} = \delta^2\Sigma_{XX}\mathbf{a},$$
$$\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\mathbf{b} = \delta^2\Sigma_{YY}\mathbf{b}.$$

Amb $\delta^2 = \lambda$ obtenim (3.1) i (3.2) respectivament. $\square$

Amb relació amb els valors i vectors propis relatius, podem tenir una diagonalització simètrica o asimètrica.

En el cas d'una diagonalització asimètrica, multipliquem a l'esquerra per $\Sigma_{XX}^{-1/2}$ a l'equació (3.1) i per $\Sigma_{YY}^{-1/2}$ a l'equació (3.2):

$$\Sigma_{XX}^{-1/2}\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\mathbf{a} = \lambda\Sigma_{XX}^{-1/2}\Sigma_{XX}\mathbf{a},$$
$$\Sigma_{YY}^{-1/2}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\mathbf{b} = \lambda\Sigma_{YY}^{-1/2}\Sigma_{YY}\mathbf{b},$$

equivalentment,

$$\Sigma_{XX}^{-1/2}\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\mathbf{a} = \lambda\Sigma_{XX}^{1/2}\mathbf{a} \quad \text{tal que,} \quad \mathbf{a}'\Sigma_{XX}\mathbf{a} = 1$$
$$\Sigma_{YY}^{-1/2}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\mathbf{b} = \lambda\Sigma_{YY}^{1/2}\mathbf{b} \quad \text{tal que,} \quad \mathbf{b}'\Sigma_{YY}\mathbf{b} = 1$$

fent el canvi de variables,

$$\boldsymbol{\alpha} = \Sigma_{XX}^{1/2}\mathbf{a} \quad \text{i per tant,} \quad \mathbf{a} = \Sigma_{XX}^{-1/2}\boldsymbol{\alpha},$$
$$\boldsymbol{\beta} = \Sigma_{YY}^{1/2}\mathbf{b} \quad \text{i per tant,} \quad \mathbf{b} = \Sigma_{YY}^{-1/2}\boldsymbol{\beta},$$

tenim,

$$\Sigma_{XX}^{-1/2}\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1/2}\boldsymbol{\alpha} = \lambda\boldsymbol{\alpha} \quad \text{tal que,} \quad \boldsymbol{\alpha}'\boldsymbol{\alpha} = 1$$
$$\Sigma_{YY}^{-1/2}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1/2}\boldsymbol{\beta} = \lambda\boldsymbol{\beta} \quad \text{tal que,} \quad \boldsymbol{\beta}'\boldsymbol{\beta} = 1$$

Per tant , ens quedem amb un problema simètric de la forma $Ax = \lambda x$.
D'altra banda, en el cas d'una diagonalització simètrica, multipliquem a l'esquerra per Σ_{XX}^{-1} a l'equació (3.1) i per Σ_{YY}^{-1} a l'equació (3.2):

$$\Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\mathbf{a} = \lambda\Sigma_{XX}^{1}\Sigma_{XX}\mathbf{a},$$
$$\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\mathbf{b} = \lambda\Sigma_{YY}^{-1}\Sigma_{YY}\mathbf{b},$$

equivalentment,

$$\Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\mathbf{a} = \lambda\mathbf{a},$$
$$\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\mathbf{b} = \lambda\mathbf{b}.$$

Per tant , també ens quedem amb un problema simètric de la forma $Ax = \lambda x$.

Teorema 3.3.2. *Els vectors canònics,* $\mathbf{a}$ *i* $\mathbf{b}$*, normalitzats per,* $\mathbf{a}'\Sigma_{XX}\mathbf{a} = 1$ *i* $\mathbf{b}'\Sigma_{YY}\mathbf{b} = 1$ *estàn relacionats per,*

$$\mathbf{a} = \lambda^{-1/2}\Sigma_{XX}^{-1}\Sigma_{XY}\mathbf{b},$$
$$\mathbf{b} = \lambda^{-1/2}\Sigma_{YY}^{-1}\Sigma_{YX}\mathbf{a}.$$

A més la primera correlació canònica és $r_1 = \sqrt{\lambda_1}$*, on* λ_1 *és el valor propi més gran de* $\Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}$*, és a dir:*

$$\left|\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX} - \lambda_1\Sigma_{XX}\right| = 0. \tag{3.4}$$

Demostració. De (3.3) tenim que $\mathbf{a} = \eta\Sigma_{XX}^{-1}\Sigma_{XY}\mathbf{b}$ on η és una constant que haurem de determinar.

A partir de $\mathbf{a}'\Sigma_{XX}\mathbf{a} = 1$ tenim,

$$\begin{aligned}\mathbf{a}'\Sigma_{XX}\mathbf{a} &= \eta\mathbf{a}'\Sigma_{XX}\Sigma_{XX}^{-1}\Sigma_{XY}\mathbf{b}\\ &= \eta\mathbf{a}'\Sigma_{XY}\mathbf{b}\\ &= \eta\lambda^{-1/2}\mathbf{a}'\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\mathbf{a}\\ &= \eta\lambda^{-1/2}\lambda\mathbf{a}'\Sigma_{XX}\mathbf{a}\\ &= \eta\lambda^{-1/2}\lambda.\end{aligned}$$

Com $\mathbf{a}'\Sigma_{XX}\mathbf{a} = 1 \Rightarrow \eta\lambda^{-1/2}\lambda = 1 \Leftrightarrow \eta = \lambda^{-1/2}$.

Anàlogament, de $\mathbf{b}'\Sigma_{YY}\mathbf{b} = 1$,

$$\begin{aligned}
\mathbf{b}'\Sigma_{YY}\mathbf{b} &= \tau\mathbf{b}'\Sigma_{YY}\Sigma_{YY}^{-1}\Sigma_{YX}\mathbf{a} \\
&= \tau\mathbf{b}'\Sigma_{YX}\mathbf{a} \\
&= \tau\lambda^{-1/2}\mathbf{b}'\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\mathbf{b} \\
&= \tau\lambda^{-1/2}\lambda\mathbf{b}'\Sigma_{YY}\mathbf{b} \\
&= \tau\lambda^{-1/2}\lambda
\end{aligned}$$

Com $\mathbf{b}'\Sigma_{YY}\mathbf{b} = 1 \Rightarrow \tau\lambda^{-1/2}\lambda = 1 \Leftrightarrow \tau = \lambda^{-1/2}$.

La correlació és $r_1 = \mathbf{a}'\Sigma_{XY}\mathbf{b}$ i com $1 = \lambda^{-1/2}\mathbf{a}'\Sigma_{XY}\mathbf{b}$ deduïm que $r_1^2 = \lambda_1$, és a dir, $r_1 = \sqrt{\lambda_1}$. □

El procès descrit pot continuar-se, buscant (U_2, V_2) variables canòniques de manera que U_2 i V_2 tinguin la màxima correlació possible entre elles, siguin combinacions lineals de les variables originals $\mathbf{X}$ i $\mathbf{Y}$ i a més siguin ortogonals a U_1 i V_1 respectivament.

Així podem obtenir $2m$ variables canòniques $(U_1, U_2, \cdots, U_m)$, $(V_1, V_2, \cdots, V_m)$ on $m = min\{p, q\}$ amb els vectors canònics $\mathbf{a_1}, \mathbf{b_1}, \ldots, \mathbf{a_m}, \mathbf{b_m}$ que proporcionen les variables i correlacions canòniques següents,

$$\begin{array}{lll}
U_1 = \mathbf{Xa_1}, & V_1 = \mathbf{Yb_1}, & r_1 = \text{cor}(U_1, V_1), \\
U_2 = \mathbf{Xa_2}, & V_2 = \mathbf{Yb_2}, & r_2 = \text{cor}(U_2, V_2), \\
\quad\vdots & \quad\vdots & \quad\vdots \\
U_m = \mathbf{Xa_m}, & V_m = \mathbf{Yb_m}, & r_m = \text{cor}(U_m, V_m).
\end{array}$$

Teorema 3.3.3. *Suposem $r_1 > r_2 > \ldots > r_m$. Aleshores,*

1. *Les variables canòniques $U_1, \ldots, U_m$ són incorrelacionades i les variables canòniques $V_1, \ldots, V_m$ també són incorrelacionades.*

2. *$cor(U_i, V_j) = 0$ si $i \neq j$.*

Demostració.

Sigui $i \neq j$. Expresem (3.1) per a $\mathbf{a_k}, \lambda_k, k = i, j$

$$\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\mathbf{a_i} = \lambda_i\Sigma_{XX}\mathbf{a_i},$$
$$\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\mathbf{a_j} = \lambda_j\Sigma_{XX}\mathbf{a_j}.$$

Multipliquem per $\mathbf{a_j}'$ i per $\mathbf{a_i}'$:

$$\mathbf{a_j}'\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\mathbf{a_i} = \lambda_i\mathbf{a_j}'\Sigma_{XX}\mathbf{a_i},$$
$$\mathbf{a_i}'\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\mathbf{a_j} = \lambda_j\mathbf{a_i}'\Sigma_{XX}\mathbf{a_j}.$$

Restant les dues equacions,

$$(\lambda_i - \lambda j)\mathbf{a_i}'\Sigma_{XX}\mathbf{a_j} = 0 \;\Rightarrow\; \mathbf{a_i}'\Sigma_{XX}\mathbf{a_j} = 0 \;\Rightarrow\; cor(U_i, U_j) = 0.$$

D'altra banda, expressant (3.1) com,

$$\Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\mathbf{a_i} = \lambda_i\mathbf{a_i},$$
$$\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\mathbf{b_j} = \lambda_j\mathbf{b_j},$$

i multiplicant per $\mathbf{b_j}'\Sigma_{YX}$ i per $\mathbf{a_i}'\Sigma_{XY}$ arribem a,

$$b_j'\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\mathbf{a_i} = \lambda_i\mathbf{b_j}'\Sigma_{YX}\mathbf{a_i},$$
$$a_i'\Sigma_{YX}\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\mathbf{b_j} = \lambda_j\mathbf{a_i}'\Sigma_{XY}\mathbf{b_j}.$$

Restant les dues equacions,

$$(\lambda_i - \lambda j)\mathbf{a_i}'\Sigma_{XY}\mathbf{b_j} = 0 \;\Rightarrow\; \mathbf{a_i}'\Sigma_{XY}\mathbf{b_j} = 0 \;\Rightarrow\; cor(U_i, V_j) = 0.$$

$\square$

3.4 Correlacions canòniques i descomposició singular

Podem formular una expressió conjunta per als vectors canònics utilitzant la descomposició singular d'una matriu. Suposem $p \geq q$, considerem la matriu $p \times q$,

$$\mathbf{K} = \Sigma_{XX}^{-1/2}\Sigma_{XY}\Sigma_{YY}^{-1/2},$$

alesores, el nombre de correlacions canòniques és $rang(\mathbf{K}) = rang(\Sigma_{XY}) = q$. Busquem la descomposició singular de $\mathbf{K}$,

$$\mathbf{K} = UD_\lambda V',$$

tal que

- U és una matriu $p \times q$ amb columnes ortonormals, $\mathbf{u} = (u_1, \ldots, u_q)$.
- V és una matriu $q \times q$ ortogonal, $\mathbf{v} = (v_1, \ldots, v_q)$.
- D_λ és una matriu diagonal amb els valors singulars de $\mathbf{K}$.

Teorema 3.4.1. *Els vectors canònics i correlacions canòniques són,*

$$\mathbf{a_i} = \Sigma_{XX}^{-1/2}\mathbf{u_i}, \quad \mathbf{b_i} = \Sigma_{YY}^{-1/2}\mathbf{v_i}, \quad r_i = \lambda_i, \quad 1 \leq i \leq q.$$

Demostració. Siguin,

$$\begin{aligned}\mathbf{N_1} = \mathbf{K}\mathbf{K}' =&\Sigma_{XX}^{-1/2}\Sigma_{XY}\Sigma_{YY}^{-1/2}\Sigma_{YY}^{-1/2}\Sigma_{YX}\Sigma_{XX}^{-1/2} = \\ &\Sigma_{XX}^{-1/2}\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1/2} = \\ &UD_\lambda^2U',\end{aligned}$$

$$\begin{aligned}\mathbf{N_2} = \mathbf{K}'\mathbf{K} =&\Sigma_{YY}^{-1/2}\Sigma_{YX}\Sigma_{XX}^{-1/2}\Sigma_{XX}^{-1/2}\Sigma_{XY}\Sigma_{YY}^{-1/2} = \\ &\Sigma_{YY}^{-1/2}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1/2} = \\ &VD_\lambda^2V',\end{aligned}$$

que compleixen,

$$\begin{aligned}\mathbf{N_1}\mathbf{u_i} &= \lambda_i^2\mathbf{u_i}, \\ \mathbf{N_2}\mathbf{v_i} &= \lambda_i^2\mathbf{v_i}.\end{aligned}$$

Multiplicant per $\Sigma_{XX}^{-1/2}$ tenim,

$$\begin{aligned}\Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}(\Sigma_{XX}^{-1/2}\mathbf{u_i}) &= \lambda_i^2(\Sigma_{XX}^{-1/2}\mathbf{u_i}), \\ \Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}(\Sigma_{YY}^{-1/2}\mathbf{v_i}) &= \lambda_i^2(\Sigma_{YY}^{-1/2}\mathbf{v_i}).\end{aligned}$$

Definint, $\mathbf{a_i} = \Sigma_{XX}^{-1/2}\mathbf{u_i}$, i $\mathbf{b_i} = \Sigma_{YY}^{-1/2}\mathbf{v_i}$, i observant que

$$\mathbf{M_1} = \Sigma_{XX}^{-1/2}\mathbf{N_1}\Sigma_{XX}^{1/2} = \Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX},$$
$$\mathbf{M_2} = \Sigma_{YY}^{-1/2}\mathbf{N_2}\Sigma_{YY}^{1/2} = \Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY},$$

obtenim finalment,

$$\mathbf{M_1}\mathbf{a_i} = \lambda_i^2\mathbf{a_i},$$
$$\mathbf{M_2}\mathbf{b_i} = \lambda_i^2\mathbf{b_i}.$$

□

Les correlacions canòniques tenen la propietat de ser invariant per transformacions linials de les variables.

Siguin $\mathbf{X} = (X_1, \ldots, X_p)'$, $\mathbf{Y} = (Y_1, \ldots, Y_q)'$ dos vectors aleatoris de dimensions p i q respectivament, i siguin $\mathbf{A} = (a_{ij})$, $\mathbf{B} = (b_{ij})$ dues matrius $p \times p$ i $q \times q$, respectivament, no singulars, i $\mathbf{c} = (c_1, \ldots, c_p)'$, $\mathbf{d} = (d_1, \ldots, d_q)'$ dos vectors.

Teorema 3.4.2. *Siguin* $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_p)'$, $\tilde{\boldsymbol{Y}} = (\tilde{Y}_1, \ldots, \tilde{Y}_q)'$ *dues variables tal que,*

$$\tilde{\boldsymbol{X}} = \boldsymbol{A}\boldsymbol{X} + \mathbf{c}, \qquad \tilde{\boldsymbol{Y}} = \boldsymbol{B}\boldsymbol{Y} + \mathbf{d},$$

es verifica,

1. *Les correlacions canòniques entre* $\tilde{\boldsymbol{X}}$, $\tilde{\boldsymbol{Y}}$ *són les mateixes que entre* $\boldsymbol{X}$, $\boldsymbol{Y}$.

2. *Els vectors canònics que defineixen les variables canòniques es transformen en,*

$$\tilde{\mathbf{a}}_i = \boldsymbol{A}^{-1}\mathbf{a}_i, \qquad \tilde{\mathbf{b}}_i = \boldsymbol{B}^{-1}\mathbf{b}_i.$$

Demostració. La matriu de covariàncies de les variables $\tilde{\mathbf{X}}$, $\tilde{\mathbf{Y}}$ seran de la forma,

$$\begin{pmatrix} \mathbf{A}'\Sigma_{XX}\ \mathbf{A} & \mathbf{A}'\Sigma_{XY}\mathbf{B} \\ \mathbf{B}'\Sigma_{YX}\ \mathbf{A} & \mathbf{B}'\Sigma_{YY}\mathbf{B} \end{pmatrix}.$$

Si $\tilde{r}_i$, $0 \leq i \leq q$ és la i-èsima correlació canònica, aleshores,

$$\begin{aligned}\left|\mathbf{A}'\Sigma_{XY}\mathbf{B}(\mathbf{B}'\Sigma_{YY}\mathbf{B})^{-1}\mathbf{B}'\Sigma_{YX}\mathbf{A} - \tilde{r}_i^2\mathbf{A}'\Sigma_{XX}\mathbf{A}\right| &= \left|\mathbf{A}'\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\mathbf{A} - \tilde{r}_i^2\mathbf{A}'\Sigma_{XX}\mathbf{A}\right| \\ &= \left|\mathbf{A}'(\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX} - \tilde{r}_i^2\Sigma_{XX})\mathbf{A}\right| \\ &= |\mathbf{A}'| \cdot \left|\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX} - \tilde{r}_i^2\Sigma_{XX}\right| \cdot |\mathbf{A}| = 0.\end{aligned}$$

Com $|\mathbf{A}| \neq 0$ aquesta equació implica (3.4).

Per tant,

$$\tilde{r}_i = r_i.$$

D'altra banda, el corresponent vector propi serà,

$$\mathbf{A}'\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\mathbf{A}\tilde{\mathbf{a}}_\mathbf{i} = r_i^2\mathbf{A}'\Sigma_{XX}\mathbf{A}\tilde{\mathbf{a}}_\mathbf{i}.$$

Multiplicant per $(\mathbf{A}')^{-1}$,

$$\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}(\mathbf{A}\tilde{\mathbf{a}}_\mathbf{i}) = r_i^2\Sigma_{XX}(\mathbf{A}\tilde{\mathbf{a}}_\mathbf{i}),$$

d'on $\mathbf{A}\tilde{\mathbf{a}}_\mathbf{i} = \mathbf{a}_\mathbf{i}$. □

Una conseqüència d'aquest teorema és que les correlacions canòniques es poden calcular a partir de les matrius de correlacions.

Siguin R_{XX} R_{YY} les matrius de correlacions del primer i segon conjunt, és a dir, de **X** i **Y** respectivament i sigui R_{XY} la matriu p x q amb els coeficients de correlacions simples de les variables **X** amb **Y**, és a dir,

	X	**Y**
X	R_{XX}	R_{XY}
Y	R_{YX}	R_{YY}

que per el teorema d'invariància es compleix,

$$\left|R_{XY}R_{YY}^{-1}R_{YX} - r_i^2R_{XX}\right| = 0,$$

i la relació entre els vectors $\mathbf{a}_\mathbf{i}$ amb $\tilde{\mathbf{a}}_\mathbf{i}$ i $\mathbf{b}_\mathbf{i}$ amb $\tilde{\mathbf{b}}_\mathbf{i}$ és,

$$\tilde{\mathbf{a}}_{\mathbf{i}} = diag(\Sigma_{XX})^{1/2}\mathbf{a}_{\mathbf{i}},$$
$$\tilde{\mathbf{b}}_{\mathbf{i}} = diag(\Sigma_{YY})^{1/2}\mathbf{b}_{\mathbf{i}}.$$

Una segona conseqüència és que el càlcul de les correlacions canòniques es pot reduir a un càlcul amb una matriu de covariàncies de la forma,

	X	**Y**
X	Id_{XX}	Σ_{XY}
Y	Σ_{YX}	Id_{YY}

3.5 Significació de les correlacions canòniques

Hem construït les correlacions canòniques i les variables canòniques amb matrius de covariancies poblacionals. En les aplicacions pràctiques es parteix de mostres de mida n i s'utilitzen les matrius de covariàncies mostrals.

Siguin, $\mathbf{X} = (X_1, \ldots, X_p)$, $\mathbf{Y} = (Y_1, \ldots, Y_q)$ dos vectors aleatoris de dimensió p i q, respectivament.
Siguin S_{XX}, S_{YY} les matrius de covariàncies mostrals del primer i segon conjunt, és a dir, de $\mathbf{X}$, $\mathbf{Y}$, respectivament, i sigui S_{XY} la matriu $p \times q$ amb les covariancies de les variables $\mathbf{X}$ amb les variables $\mathbf{Y}$. És a dir,

	X	**Y**
X	S_{XX}	S_{XY}
Y	S_{YX}	S_{YY}

Siguin,

$$r_1 \geq r_2 \geq \cdots \geq r_m,$$

les $m = min\{p, q\}$ correlacions canòniques obtingudes a partir de S_{XX}, S_{XY}, S_{YY} solucions de,

$$|S_{XY}S_{YY}^{-1}S_{YX} - r^2 S_{XX}| = 0.$$

Si volem decidir quines són significatives, suposem normalitat multivariant, és a dir, $(\mathbf{X}, \mathbf{Y}) \sim N_{p+q}((\boldsymbol{\mu}, \boldsymbol{\nu}), S)$ on $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)$ i $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_q)$ són els vectors d'esperances de $\mathbf{X}$, $\mathbf{Y}$, respectivament, i plantejem el següent test:

$$H_0^k : r_k > r_{k+1} = \cdots = r_m = 0, \quad (r_0 = 1, \ k = 0, 1, \ldots, m).$$

El test de Bartlett-Lawley (Mardia 1979, pp. 288-289) demostra que si H_0^k és certa, aleshores,

$$L_k = -\left[n - 1 - k - \frac{1}{2}(p + q + 1) + \sum_{i=1}^{k} r_i^{-2}\right] \cdot \log\left[\prod_{i=k+1}^{m} (1 - r_i^2)\right],$$

es distribueix asimptòticament com una khi-quadrat amb $(m - k)(p - k)$ graus de llibertat.

Aquest test s'aplica seqüencialment de la forma següent, si L_i és significatiu per a $i = 0, 1, \ldots, k - 1$, però L_k no és significatiu, aleshores s'accepta H_0^k, és a dir, les primeres k correlacions canòniques són positives, les restants són nul·les.

4 Visió geomètrica de les correlacions canòniques

En aquest capítol il·lustrarem el concepte de correlació canònica geomètricament, per això farem correspondre una variable aleatòria amb un vector de n components. És a dir, donat dos conjunts de n observacions $x_1, x_2, \ldots, x_n$, i $y_1, y_2, \ldots, y_n$ definim dos vectors x_r i y_r d'ordre n,

$$x_r = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad y_r = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

Geomètricament, x_r i y_r estan continguts a l'espai ε_n de dimensió n. La mitjana dels vectors x i y ve representada per,

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i,$$

i definim els vectors x i y com

$$x = \begin{bmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{bmatrix}, \quad y = \begin{bmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{bmatrix}.$$

Per els vectors x, y de dimensió $n \times 1$, definim la norma al quadrat com,

$$\|x\|^2 = (x, x) = \sum_{i=1}^{n} (x_i - \bar{x})^2,$$
$$\|y\|^2 = (y, y) = \sum_{i=1}^{n} (y_i - \bar{y})^2.$$

Dividint per n, s'obtenen les variancies dels vectors x i y,

$$s_x^2 = \frac{1}{n}\|x\|^2 = \frac{1}{n}(x, x) = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2,$$
$$s_y^2 = \frac{1}{n}\|y\|^2 = \frac{1}{n}(y, y) = \frac{1}{n}\sum_{i=1}^{n} (y_i - \bar{y})^2.$$

La covariancia dels vectors x i y es denota per,

$$s_{xy} = \frac{1}{n}(x, y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})^2,$$

i el coeficient de correlació ve donat per,

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{\frac{1}{n}(x, y)}{\frac{1}{\sqrt{n}}\|x\|\frac{1}{\sqrt{n}}\|y\|} = \frac{(x, y)}{\|x\|\,\|y\|}.$$

Definim el cosinus de l'angle com

$$cos\theta_{xy} = \frac{(x, y)}{\|x\|\,\|y\|} = r_{xy},$$

el que implica que el coeficient de correlació ve donat pel cosinus de l'angle que formen els vectors x i y.

Siguin $\mathbf{X} = (x_1, \ldots, x_p)$, $\mathbf{Y} = (y_1, \ldots, y_q)$ dos conjunts de vectors a l'espai ε_n de dimensions p i q respectivament. Aquests vectors formen els subespais W_X i W_Y. Definim la matriu de covariàncies de p i q variables com

$$\Sigma_{XX} = \frac{1}{n}X'X = \frac{1}{n}\begin{bmatrix} (x_1, x_1) & (x_1, x_2) & \cdots & (x_1, x_p) \\ (x_2, x_1) & (x_2, x_2) & \cdots & (x_2, x_p) \\ \vdots & \vdots & \ddots & \vdots \\ (x_p, x_1) & (x_p, x_2) & \cdots & (x_p, x_p) \end{bmatrix} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix},$$

$$\Sigma_{YY} = \frac{1}{n}Y'Y = \frac{1}{n}\begin{bmatrix} (y_1, y_1) & (y_1, y_2) & \cdots & (y_1, y_q) \\ (y_2, y_1) & (y_2, y_2) & \cdots & (y_2, y_q) \\ \vdots & \vdots & \ddots & \vdots \\ (y_q, y_1) & (y_q, y_2) & \cdots & (y_q, y_q) \end{bmatrix} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1q} \\ s_{21} & s_{22} & \cdots & s_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ s_{q1} & s_{q2} & \cdots & s_{qq} \end{bmatrix}.$$

Considerem ara les dues combinacions lineals,

$$\mathbf{f} = \mathbf{Xa} = a_1x_1 + a_2x_2 + \ldots + a_px_p \in W_Y,$$
$$\mathbf{g} = \mathbf{Yb} = b_1y_1 + b_2y_2 + \ldots + b_qx_q \in W_Y,$$

i la recerca de dos vectors $\mathbf{a} = (a_1, \ldots, a_p)'$ i $\mathbf{b} = (b_1, \ldots, b_q)'$ de tal manera que les dues variables $\mathbf{f}$ i $\mathbf{g}$ tinguin la màxima correlació.

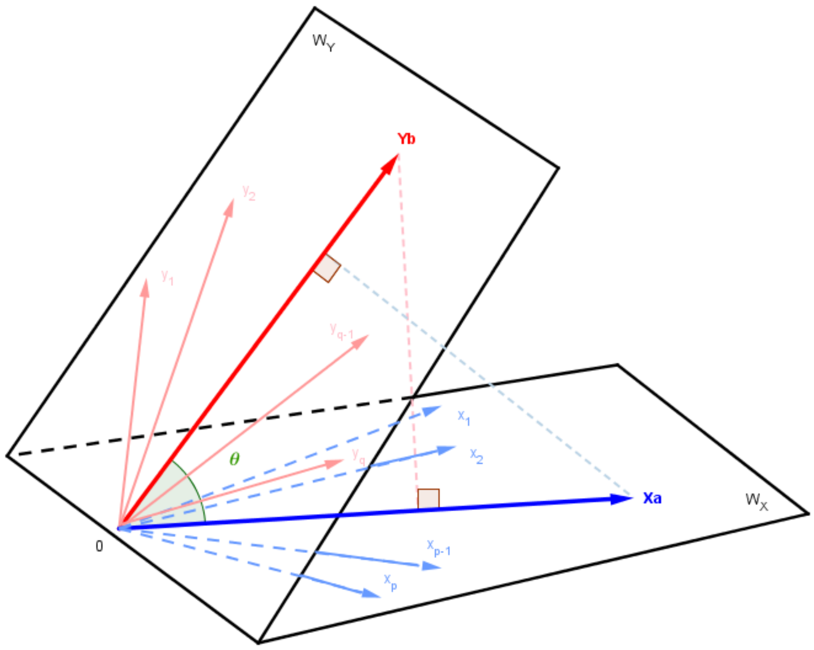

Figura 1: Representació de la correlació canònica en termes de vectors.

El coeficient de màxima correlació es defineix com el cosinus de l'angle entre els vectors $\mathbf{f}$ i $\mathbf{g}$ (Veure Figura 1) i s'anomena el coeficient de correlació canònica entre $\mathbf{X}$ i $\mathbf{Y}$.

Per obtenir aquests resultats necessitem el concepte de projector ortogonal definit al capítol 2.

Suposem doncs, que es dona $\mathbf{g} = \mathbf{Yb}$. La projecció ortogonal de $\mathbf{g}$ en $\mathbf{Y}$ ve donada per,

$$P_X \mathbf{g} = P_X \mathbf{Yb},$$

és a dir, la projecció de $\mathbf{g}$ sobre W_X.

D'altra banda donat $\mathbf{f} = \mathbf{Xa}$, la projecció ortogonal de $\mathbf{f}$ en $\mathbf{Y}$ ve donada per,

$$P_Y\mathbf{f} = P_Y\mathbf{Xa},$$

és a dir, la projecció de $\mathbf{f}$ sobre W_Y.

Sense pèrdua de generalitat,

$$s_f^2 = \frac{1}{N}\left\|\mathbf{Xa}\right\|^2 = \frac{1}{N}(\mathbf{Xa}, \mathbf{Xa}) = \mathbf{a}'\Sigma_{XX}\mathbf{a} = 1, \tag{4.1}$$

$$s_g^2 = \frac{1}{N}\left\|\mathbf{Yb}\right\|^2 = \frac{1}{N}(\mathbf{Yb}, \mathbf{Yb}) = \mathbf{b}'\Sigma_{YY}\mathbf{b} = 1. \tag{4.2}$$

Introduim l'escalar λ_1 on $\overrightarrow{OH} = \lambda_1\mathbf{f} = \lambda_1\mathbf{Xa}$ d'on sobté,

$$P_X\mathbf{Yb} = \lambda_1\mathbf{Xa}. \tag{4.3}$$

Multiplicant per $\mathbf{X}'$ a l'esquerra i dividint per N obtenim,

$$\Sigma_{XY}\mathbf{b} = \lambda_1\Sigma_{XX}\mathbf{a}. \tag{4.4}$$

D'altra banda si λ_2 és l'escalar tal que $\overrightarrow{OH'} = \lambda_2\mathbf{g} = \lambda_2\mathbf{Yb}$ d'on sobté,

$$P_Y\mathbf{Xa} = \lambda_2\mathbf{Yb}. \tag{4.5}$$

Multiplicant per $\mathbf{Y}'$ a l'esquerra i dividint per N obtenim,

$$\Sigma_{YX}\mathbf{a} = \lambda_2\Sigma_{YY}\mathbf{b}. \tag{4.6}$$

Amb aquestes condicions de $\mathbf{Xa}$ i $\mathbf{Yb}$ aplicades simultàniament obtenim dues equacions per $\mathbf{a}$, $\mathbf{b}$, λ_1, λ_2.

Multiplicant (4.3) per $\mathbf{a}'$ i amb l'observació de (4.1) obtenim,

$$\mathbf{a}'\Sigma_{XY}\mathbf{b} = \lambda_1\mathbf{a}'\Sigma_{XX}\mathbf{a} = \lambda_1.$$

De manera similar multipliquem (4.6) per $\mathbf{b}'$,

$$\mathbf{b}'\Sigma_{YX}\mathbf{a} = \lambda_2\mathbf{b}'\Sigma_{YY}\mathbf{b} = \lambda_2.$$

Com $(\mathbf{a}'\Sigma_{XY}\mathbf{b})' = \mathbf{b}'\Sigma_{YX}\mathbf{a} = \mathbf{a}'\Sigma_{XY}\mathbf{b}$ obtenim $\lambda_1 = \lambda_2$. Podem suposar $\lambda_1 = \lambda_2 \geq 0$ i posar $\sqrt{\lambda} = \lambda_1 = \lambda_2$.

Multiplicant (4.6) per Σ_{YY}^{-1} tenim,

$$\mathbf{b} = \frac{\Sigma_{YY}^{-1}\Sigma_{YX}}{\sqrt{\lambda}}.$$

Substiuint a l'equació (4.4) obtenim,

$$\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\mathbf{a} = \lambda\Sigma_{XX}\mathbf{a}, \tag{4.7}$$

que dona,

$$\Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\mathbf{a} = \lambda\mathbf{a}.$$

S'obté el vector corresponent al màxim valor propi que s'atisfà,

$$\begin{gathered}\left|\Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX} - \lambda I_p\right| = 0,\\ \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\mathbf{a} = \lambda\Sigma_{XX}\mathbf{a}.\end{gathered}$$

Una representació alternativa de (4.7) és,

$$(P_X P_Y)\mathbf{Xa} = \lambda\mathbf{Xa},$$

que segueix mitjançant la substitució (4.5) en (4.3).

D'altra banda la substitució de (4.3) en (4.5) queda determinada per,

$$(P_Y P_X)\mathbf{Yb} = \lambda\mathbf{Yb}.$$

Es veu que el coeficient de correlació entre $\mathbf{f} = \mathbf{Xa}$ i $\mathbf{g} = \mathbf{Yb}$ s'obté de la relació,

$$R_{Xa,Yb} = cos\theta_{Xa,Yb} = \frac{(\mathbf{Xa},\mathbf{Yb})}{\|\mathbf{Xa}\|\cdot\|\mathbf{Yb}\|} = (\mathbf{a},\Sigma_{XY}\mathbf{b}) = \mathbf{a}'\Sigma_{XY}\mathbf{b} = \sqrt{\lambda},$$

que es diu el *coeficient de la correlació canònica*.

D'aqui es dedueix que,

$$\begin{gathered}\overrightarrow{OH} = \|\lambda_1\mathbf{Xa}\| = \lambda_1\|\mathbf{Xa}\| = \sqrt{N}\lambda_1 = \sqrt{N}\sqrt{\lambda},\\ \overrightarrow{OH'} = \|\lambda_2\mathbf{Yb}\| = \lambda_2\|\mathbf{Yb}\| = \sqrt{N}\lambda_2 = \sqrt{N}\sqrt{\lambda},\end{gathered}$$

el que demostra que el coeficient de correlació canònica es igual a $\frac{1}{\sqrt{N}}$ vegades la longitud de $\overrightarrow{OH}$ o $\overrightarrow{OH'}$.

5 Relació amb altres tècniques multivariants

5.1 Anàlisi de components principals

Quan es recull informació d'una mostra de dades, el més freqüent és agafar el nombre màxim de variables que es disposa. Però si agafem massa variables per exemple 20 variables, haurem de considerar $\binom{20}{2} = 180$ possibles coeficients de correlació; si són 40 variables aquest nombre augmenta fins a 780.

En aquests casos és difícil visualitzar relacions entre les variables. Un altre problema que es presenta és la forta correlació que moltes vegades es presenta entre les variables, és a dir, si agafem moltes variables la situació més comú es que les variables estiguin relacionades o que mesurin el mateix però en diferents punts de vista.

És necessari doncs, reduir el nombre de variables.

Per estudiar les relacions que es presenten entre p variables correlacionades es pot transformar el conjunt original de variables en un altre conjunt de noves variables incorrelacionades entre sí, anomenat conjunt de components principals.

Càlcul de les components principals

Es consideren variables $(x_1, \ldots, x_p)$ sobre un grup d'objectes o individus i es tracta de calcular, a partir d'elles, un nou conjunt de variables $z_1, \ldots, z_p$ incorrelacionades entre sí.

Cada z_j, $j = 1, \ldots, p$ és una combinació lineal de les $x_1, \ldots, x_p$ originals, és a dir,

$$z_j = a_{j1}x_1 + a_{j2}x_2 + \cdots + a_{jp}x_p = \mathbf{a_j'x},$$

on $\mathbf{a_j'} = (a_{1j}, a_{2j}, \cdots, a_{pj})$ un vector de constants.

Per mantenir la ortogonalitat de la transformació s'imposa que el mòdul del vector $\mathbf{a_j'}$ sigui 1, és a dir,

$$\mathbf{a_j'a_j} = \sum_{k=1}^{p} a_{kj}^2 = 1.$$

La primera component es calcula escollint a_1 de manera que maximitzi la variància de z_1 amb la restricció $\mathbf{a_j'a_j} = 1$,

$$Var(z_1) = Var(\mathbf{a_1'}x) = \mathbf{a_1'}\Sigma\mathbf{a_1}.$$

on Σ és la matriu de covariàncies de les observacions.
El mètode habitual per maximitzar una funció de variables restringida és el mètode de multiplicadors de Lagrange. Construim la funció L,

$$L(\mathbf{a_1}) = \mathbf{a_1'}\Sigma\mathbf{a_1} - \lambda(\mathbf{a_1'}\mathbf{a_1} - 1),$$

i busquem el màxim de la funció, derivant i igualant a 0,

$$\frac{\partial L}{\partial \mathbf{a}_1} = 2\Sigma\mathbf{a_1} - 2\lambda I\mathbf{a_1} = 0,$$
$$(\Sigma - \lambda I)\mathbf{a_1} = 0.$$

Perque el sistema tingui solució la matriu $(\Sigma - \lambda I)$ ha de ser singular, i això implica que,

$$|\Sigma - \lambda I| = 0.$$

D'aquesta forma λ és un valor propi de Σ. La matriu de covariàncies Σ és d'ordre p i tindrà p valors propis positius, $\lambda_1, \ldots, \lambda_p$.

Desenvolupant l'expressió anterior

$$(\Sigma - \lambda I)\mathbf{a_1} = 0,$$
$$\Sigma\mathbf{a_1} - \lambda I\mathbf{a_1} = 0,$$
$$\Sigma\mathbf{a_1} = \lambda I\mathbf{a_1},$$

aleshores,

$$Var(z_1) = Var(\mathbf{a_1'}x) = \mathbf{a_1'}\Sigma\mathbf{a_1} = \mathbf{a_1'}\lambda I\mathbf{a_1} = \lambda\mathbf{a_1'}\mathbf{a_1} = \lambda \cdot 1 = \lambda.$$

Per maximitzar la variància de z_1 s'ha d'agafar el valor propi més gran, per exemple λ_1 i el corresponent vector propi a_1.

Aquest vector propi ens dóna la combinació de les variables originals que te major variància, és a dir, si $a_1 = (a_{11}, \ldots, a_{1p})$,

$$z_1 = \mathbf{a_1'x} = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1p}x_p.$$

La segona component principal, $z_2 = \mathbf{a_2'x}$ s'obté mitjançant un argument similar. A més es vol que y_2 estigui incorrelacionat amb l'anterior component z_1, és a dir $Cor(z_2, z_1) = 0$. Per tant,

$$Cov(z_2, z_1) = Cov(\mathbf{a_2'x}, \mathbf{a_1'x}) = \mathbf{a_2'}\Sigma\mathbf{a_1},$$

és a dir, es requereix que $\mathbf{a_2'}\Sigma\mathbf{a_1} = 0$ i això equival a que $\mathbf{a_2'a_1} = 0$, és a dir que els vectors siguin ortogonals.

Utilitzant els mateixos raonaments que amb la primera component, escollim λ com el segon valor propi més gran de la matriu Σ amb el seu vector propi $\mathbf{a_2}$.

Aquests raonaments es poden estendre de forma que l'i-èsima component li correspondria l'i-èsim vector propi.

Finalment, totes les components $\mathbf{z}$ es poden expressar de la següent manera,

$$\mathbf{z} = \mathbf{xA},$$

on,

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{pmatrix}.$$

Com

$$
\begin{aligned}
Var(z_1) &= \lambda_1, \\
Var(z_2) &= \lambda_2, \\
&\vdots \\
Var(z_p) &= \lambda_p,
\end{aligned}
$$

la matriu de covariàncies de $\mathbf{z}$ serà,

$$
D_\lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{pmatrix},
$$

perque $z_1, \ldots, z_p$ s'han construït com variables incorrelacionades.
Es té,

$$
D_\lambda = Var(\mathbf{z}) = \mathbf{A}' Var(\mathbf{x}) \mathbf{A} = \mathbf{A}' \Sigma \mathbf{A},
$$

o bé,

$$
\Sigma = A D_\lambda A',
$$

ja que $\mathbf{A}$ és una matriu ortogonal, $\mathbf{A}\mathbf{A}' = I$.

Siguin,

$$
\mathbf{X} = U_X D_{\lambda_X} V_X', \qquad \mathbf{Y} = U_Y D_{\lambda_Y} V_Y',
$$

les descomposicions singulars de les matrius $\mathbf{X}$, $\mathbf{Y}$, respectivament, i

$$
\Sigma_{XX} = \mathbf{X}'\mathbf{X} = V_X D^2_{\lambda_X} V_X', \qquad \Sigma_{YY} = \mathbf{Y}'\mathbf{Y} = V_Y D^2_{\lambda_Y} V_Y'.
$$

Calculem,

$$
Z_X = \mathbf{X}\mathbf{A} = \mathbf{X} V_X = U_X D_{\lambda_X}, \qquad Z_Y = \mathbf{Y}\mathbf{A} = \mathbf{Y} V_Y = U_Y D_{\lambda_Y}.
$$

Les matrius U_X i U_Y són matrius amb els vectors columnes ortonormals que es corresponen amb les corresponents components principals del conjunt $\mathbf{X}$ i $\mathbf{Y}$.

Gràcies al teorema d'invariancia podem calcular les correlacions canòniques a partir de la matriu de covariàncies o correlacions que seran de la forma,

	X	**Y**
X	Id_{XX}	Σ_{XY}
Y	Σ_{YX}	Id_{YY}

	X	**Y**
X	Id_{XX}	R_{XY}
Y	R_{YX}	Id_{YY}

5.2 Anàlisi canònica de poblacions

L'anàlisi canònica de poblacions és una tècnica que s'aplica si es tenen diverses matrius de dades, com a resultat d'observar les variables en diverses poblacions, i el que es vol és representar les poblacions.

Es suposa que de l'observació de p variables quantitatives $X_1, ..., X_p$ en g poblacions s'obtenen g matrius de dades $\mathbf{X} = (X_1, ..., X_g)'$, on X_i és la matriu $n_i \times p$ de la població i. Siguin $\overline{\mathrm{x}}'_1, ..., \overline{\mathrm{x}}'_g$ els vectors fila de les mitjanes de cada població. $\mathbf{X}$ és d'ordre $n \times p$, on $n = \sum_{i=1}^{g} n_i$.

La matriu $g \times p$ amb les mitjanes de les g poblacions és $\overline{\mathbf{X}} = (\overline{\mathrm{x}}'_1 - \overline{\mathrm{x}}', ..., \overline{\mathrm{x}}'_g - \overline{\mathrm{x}}')'$.

Les dues formes de quantificar matricialment la dispersió entre les poblacions són:

- La matriu de dispersió no ponderada entre grups:

$$\mathbf{A} = \overline{\mathbf{X}}'\overline{\mathbf{X}} = \sum_{i=1}^{g} (\overline{\mathrm{x}}_i - \overline{\mathrm{x}})(\overline{\mathrm{x}}_i - \overline{\mathrm{x}})'$$

- La matriu de dispersió ponderada entre grups:

$$\mathbf{B} = \sum_{i=1}^{g} n_i (\overline{\mathrm{x}}_i - \overline{\mathrm{x}})(\overline{\mathrm{x}}_i - \overline{\mathrm{x}})'$$

La matriu $\mathbf{A}$ és proporcional a una matriu de covariàncies prenent com a dades només les mitjanes de les poblacions. Aleshores, aquesta matriu serà la matriu de covariàncies entre les poblacions.

La matriu $\mathbf{B}$ participa, juntament amb $\mathbf{W}$ que és la matriu de dispersió dins de grups, en el test de comparació de mitjanes de g poblacions.

D'altra banda, es té la matriu,

$$\Sigma = \frac{1}{n-g} \sum_{i=1}^{g} n_i \Sigma_i$$

que serà la matriu de covariàncies dins de les poblacions.

Continuem definint les variables canòniques.

Siguin $\mathbf{v} = [\mathrm{v}_1, ..., \mathrm{v}_p]$ els vectors propis d'$\mathbf{A}$ respecte de Σ amb valors propis $\lambda_1 > ... > \lambda_p$, és a dir, $\mathbf{A}v_i = \lambda_i \Sigma_i v_i$, normalitzats segons $\mathbf{v}_i' \Sigma_i \mathbf{v}_i = 1$.

Els vectors $\mathrm{v}_1, ..., \mathrm{v}_p$ són els vectors canònics i les variables canòniques són les variables compostes $Z_i = \mathbf{X}\mathrm{v}_i$.

Si $v_i = (v_{1i}, ..., v_{pi})'$ i $\mathbf{X} = (X_1, ..., X_p)$, la variable canònica Z_i és la variable composta

$$Z_i = \mathbf{X}\mathrm{v}_i = v_{1i} X_1 + ... + v_{pi} X_p$$

que té Σ-variància 1 i $\mathbf{A}$-variància λ_i, és a dir:

$$\mathrm{var}_A(Z_i) = \mathrm{v}_i' \mathbf{A} \mathrm{v}_i = \lambda_i, \quad \mathrm{var}_\Sigma(Z_i) = \mathrm{v}_i' \Sigma_i \mathrm{v}_i = 1$$

Siguin $\mathbf{X}$, $\mathbf{Y}$ dues matrius,

$$\mathbf{X} = (X_1, \dots, X_g), \quad \mathbf{Y} = \begin{pmatrix} \mathbb{1}_{n_1} & 0 & \cdots & 0 \\ 0 & \mathbb{1}_{n_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbb{1}_{n_g} \end{pmatrix}$$

aleshores, les variables canòniques de l'anàlisi de correlacions canòniques coincideixen amb les variables canòniques de l'anàlisi canònica de poblacions.

5.3 Anàlisi de correspondències

L'anàlisi de correspondències és una tècnica descriptiva per representar taules de contingència, és a dir, taules on recollim les freqüències de dues o més variables qualitatives.

La informació de partida és ara una matriu de dimensions $I \times J$, que representa les freqüències absolutes observades de dues variables qualitatives A i B obtenint $n = \sum_{ij} f_{ij}$ observacions on f_{ij} és el número de vegades que apareix la intersecció $A_i \cap B_j$ donant lloc a la taula de contingència següent,

$$
\begin{array}{c|cccc|c}
 & B_1 & B_2 & \dots & B_J & \\
\hline
A_1 & f_{11} & f_{12} & \dots & f_{1J} & f_{1.} \\
A_2 & f_{21} & f_{22} & \dots & f_{2J} & f_{2.} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
A_I & f_{I1} & f_{I2} & \dots & f_{IJ} & f_{I.} \\
\hline
 & f_{.1} & f_{.2} & \dots & f_{.J} & n
\end{array}
\tag{5.1}
$$

on $f_i = \sum_j f_{ij}$ és la freqüència absoluta de A_i i $f_j = \sum_j f_{ij}$ és la freqüència absoluta de B_j.

Hem de tenir en compte que, en realitat, la taula (5.1) resumeix la matriu de dades inicials que és de la forma:

$$
\begin{array}{c|cccc|cccc|}
 & A_1 & A_2 & \dots & A_I & B_1 & B_2 & \dots & B_J \\
\hline
1 & 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
i & 0 & 0 & \dots & 1 & 0 & 1 & \dots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
n & 0 & 0 & \dots & 1 & 0 & 0 & \dots & 1 \\
\hline
\end{array}
$$

en la que donem el valor 1 quan es presenta una característica i 0 quan no es presenta.

La matriu de dades $n \times (I + J)$ és doncs,

$$\mathbf{Z} = [\mathbf{X}, \mathbf{Y}].$$

Utilitzarem el nom de variables files i variables columnes a les variables A i B respectivament.

Indiquem per $\mathbf{N} = (f_{ij})$la matriu $I \times J$ amb les freqüències de la taula de contingència i per $\mathbf{1}_k$ el vector de uns de dimensió k.
La matriu,

$$\mathbf{P} = \frac{1}{n}\mathbf{N}$$

és la matriu de correspondències. Sigui $\mathbf{r}$ el vector $I \times 1$ amb els totals absoluts de les files de $\mathbf{P}$ i per $\mathbf{c}$ el vector $J \times 1$ amb els totals absoluts de les columnes de $\mathbf{P}$:

$$\mathbf{r} = \mathbf{P}\mathbf{1}_J \quad \mathbf{c} = \mathbf{P}'\mathbf{1}_I.$$

Tenim doncs,

$$\mathbf{r} = \frac{1}{n}\mathbf{X}'\mathbf{1}_n, \qquad \mathbf{c} = \frac{1}{n}\mathbf{Y}'\mathbf{1}_n$$

que són els vectors de mitjanes de les matrius de dades X, Y.
A més, sigui

$$\mathbf{D}_r = diag(\mathbf{r}), \qquad \mathbf{D}_c = diag(\mathbf{c}),$$

les matrius diagonals que contenen els valors marginals de les files i columnes de $\mathbf{P}$.
Es verifica,

$$\mathbf{X}'\mathbf{X} = n\mathbf{D}_r, \qquad \mathbf{Y}'\mathbf{Y} = n\mathbf{D}_c, \qquad \mathbf{X}'\mathbf{Y} = n\mathbf{P} = \mathbf{N}.$$

Per tant, les matrius de covariàncies entre files, entre columnes i entre files i columnes, són,

$$\Sigma_{XX} = \mathbf{D}_r - \mathbf{r}\mathbf{r}', \qquad \Sigma_{YY} = \mathbf{D}_c - \mathbf{c}\mathbf{c}', \qquad S_{XY} = \mathbf{P} - \mathbf{r}\mathbf{c}'.$$

Com la suma de les variables és igual a 1, les matrius Σ_{XX} i Σ_{YY} són singulars.

El problema de les variables categòriques és que no són quantitatives. La quantificació 0 o 1 anterior és convencional.
Assignem doncs a les categories $A_1, \ldots, A_I$ de la variable fila els valors numèrics $a_1, \ldots, a_I$ i a les categories $B_1, \ldots, B_J$ de la variable columna, els valors numèrics $b_1, \ldots, b_J$, és a dir, indiquem els vectors,

$$\mathbf{a} = (a_1, \ldots, a_I)', \qquad \mathbf{b} = (b_1, \ldots, b_J)',$$

i considerem les variables compostes

$$U = \mathbf{Xa}, \qquad V = \mathbf{Yb}.$$

Si en un individu k observem les categories A_i, B_j, aleshores els valors de U, V sobre k són,

$$U_k = a_i, \, V_k = b_j.$$

Volem trobar a, b tals que les correlacions entre U i V siguin màximes. Estem en un problema de correlacions canòniques amb que les matrius Σ_{XX}, Σ_{YY} són dues matrius singulars.
Una g-inversa de Σ_{XX} és la matriu $\Sigma_{XX}^{-} = \mathbf{D}_r^{-1}$ que verifica,

$$\Sigma_{XX}\Sigma_{XX}^{-}\Sigma_{XX} = \Sigma_{XX}.$$

En efecte,

$$\begin{aligned}
(\mathbf{D}_r - \mathbf{rr}')\mathbf{D}_r^{-1}(\mathbf{D}_r - \mathbf{rr}') &= (\mathbf{D}_r - \mathbf{rr}')(\mathbf{I} - \mathbf{1r}') \\
&= \mathbf{D}_r - \mathbf{D}_r\mathbf{1r}' - \mathbf{rr}' + \mathbf{rr}'\mathbf{1r}' \\
&= \mathbf{D}_r - \mathbf{rr}' - \mathbf{rr}' + \mathbf{rr}' \\
&= \mathbf{D}_r - \mathbf{rr}'.
\end{aligned}$$

Analogament $\Sigma_{YY}^{-} = \mathbf{D}_c^{-1}$ que verifica

$$\Sigma_{YY}\Sigma_{YY}^{-}\Sigma_{YY} = \Sigma_{YY}.$$

En efecte,

$$\begin{aligned}
(\mathbf{D}_c - \mathbf{cc}')\mathbf{D}_c^{-1}(\mathbf{D}_c - \mathbf{cc}') &= (\mathbf{D}_c - \mathbf{cc}')(\mathbf{I} - \mathbf{1c}') \\
&= \mathbf{D}_c - \mathbf{D}_c\mathbf{1c}' - \mathbf{cc}' + \mathbf{cc}'\mathbf{1c}' \\
&= \mathbf{D}_c - \mathbf{cc}' - \mathbf{cc}' + \mathbf{cc}' \\
&= \mathbf{D}_c - \mathbf{cc}'.
\end{aligned}$$

Aplicant descomposició singular,

$$\mathbf{D}_r^{-1/2}(\mathbf{P} - \mathbf{rc}')\mathbf{D}_c^{-1/2} = UD_\lambda V'$$

on $\mathbf{D}_\lambda$ és la matriu diagonal amb els valors singulars en ordre decreixent.

Si u_1, v_1 són els primers vectors canònics, tindrem,

$$\mathbf{a} = \Sigma_{XX}^{-1/2}\mathbf{u}_1, \qquad \mathbf{b} = \Sigma_{YY}^{-1/2}\mathbf{v}_1, \qquad r = \lambda_1,$$

és a dir, el primer valor singular és la màxima correlació entre les variables U i V. Però poden haver-hi més vectors i correlacions canòniques, i per tant la solució general és,

$$\mathbf{a}_i = \mathbf{D}_r^{-1/2}\mathbf{u}_i, \qquad \mathbf{b}_i = \mathbf{D}_c^{-1/2}\mathbf{v}_i, \qquad r_i = \lambda_i, \quad i = 1, \ldots, min\{I, J\}.$$

6 Exemples

Apliquem l'anàlisi de correlacions canòniques a les dades de marques d'automòbils (Härdle 2015, pp. 458). En aquest context estem interessas a relacionar variables de preus amb variables com ara l'esportivitat, la seguretat, etc. En particular, ens agradaria investigar la relació entre les dues variables del valor i el preu del cotxe i totes les altres variables.

Portem a terme l'avaluació de l'anàlisi de correlacions canòniques de les matrius de dades **X** i **Y** que corresponen amb el conjunt de variables {Preu, Valor} i { Economic, Servei, Disseny, Esportiu, Seguretat, Fàcil maneig} respectivament. La matriu de covariàncies estimada Σ ve donada per,

$$\Sigma = \left(\begin{array}{cc|cccccc} 1.41 & -1.11 & 0.78 & -0.71 & -0.90 & -1.04 & -0.95 & 0.18 \\ -1.11 & 1.19 & -0.42 & 0.82 & 0.77 & 0.90 & 1.12 & 0.11 \\ \hline 0.78 & -0.4 & 0.75 & -0.23 & -0.45 & -0.42 & -0.28 & 0.28 \\ -0.71 & 0.82 & -0.23 & 0.66 & 0.52 & 0.57 & 0.85 & 0.14 \\ -0.90 & 0.77 & -0.45 & 0.52 & 0.72 & 0.77 & 0.68 & -0.10 \\ -1.04 & 0.90 & -0.42 & 0.57 & 0.77 & 1.05 & 0.76 & -0.15 \\ -0.95 & 1.12 & -0.28 & 0.85 & 0.68 & 0.76 & 1.26 & 0.22 \\ 0.18 & 0.11 & 0.28 & 0.14 & -0.10 & -0.15 & 0.22 & 0.32 \end{array}\right),$$

on

$$\Sigma_{XX} = \begin{pmatrix} 1.41 & -1.11 \\ -1.11 & 1.19 \end{pmatrix},$$

$$\Sigma_{XY} = \begin{pmatrix} 0.78 & -0.71 & -0.90 & -1.04 & -0.95 & 0.18 \\ -0.42 & 0.82 & 0.77 & 0.90 & 1.12 & 0.11 \end{pmatrix},$$

$$\Sigma_{YY} = \begin{pmatrix} 0.75 & \text{-}0.23 & \text{-}0.45 & \text{-}0.42 & \text{-}0.28 & 0.28 \\ \text{-}0.23 & 0.66 & 0.52 & 0.57 & 0.85 & 0.14 \\ \text{-}0.45 & 0.52 & 0.72 & 0.77 & 0.68 & \text{-}0.10 \\ \text{-}0.42 & 0.57 & 0.77 & 1.05 & 0.76 & \text{-}0.15 \\ \text{-}0.28 & 0.85 & 0.68 & 0.76 & 1.26 & 0.22 \\ 0.28 & 0.14 & \text{-}0.10 & \text{-}0.15 & 0.22 & 0.32 \end{pmatrix}.$$

Les arrels de l'equació quadràtica,

$$|\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX} - \lambda\Sigma_{XX}| = 0,$$

són: $\lambda_1 = 0.9604$, $\lambda_2 = 0,7921$, i per tant les correlacions canòniques són:

$$r_1 = \lambda_1^{1/2} = 0.98, \quad r_2 = \lambda_2^{1/2} = 0.89.$$

Els vectors canònics normalitzats segons $\mathbf{a}'\Sigma_{XX}\mathbf{a} = 1$ i $\mathbf{b}'\Sigma_{YY}\mathbf{b} = 1$, són:

$\mathbf{a_1} = (-0.3379, +0.5817)'$,
$\mathbf{b_1} = (+0.4202, -0.2331, -0.0211, -0.4630, +0.1815, -0.3747)'$.

$\mathbf{a_2} = (+1.5985, +1.6806)'$,
$\mathbf{b_2} = (+0.5593, +0.4202, -0.1411, +0.0067, +0.0826, +0.9044)'$.

Les variables canòniques amb variància 1 són:

$U_1 = \mathbf{a_1'X} = -0.3379x_1 + 0.5817x_2$,
$V_1 = \mathbf{b_1'Y} = +0.4202y_1 - 0.2331y_2 - 0.0211y_3 - 0.4630y_4 + 0.1815y_5 - 0.3747y_6$.

$U_2 = \mathbf{a_2'X} = 1.5985x_1 + 1.6806x_2$,
$V_2 = \mathbf{b_2'Y} = 0.5593y_1 + 0.4202y_2 - 0.1411y_3 + 0.0067y_4 + 0.0826y_5 + 0.9044y_6$.

Podem observar que les variables y_1 (Economic), y_5 (Seguretat) tenen coeficients positius a V_1 i les variables y_2 (Servei), y_3 (Disseny), y_4 (Esportiu) i y_6 (Fàcil maneig) tenen una influència negativa a V_1.

La variable canònica U_1 pot ser interpretada com un índex del preu i del valor del cotxe.

La variable canònica V_1 és formada principalment per variables qualitatives.

Per tant, aquestes variables poden ser interpretades com una apreciació del valor del cotxe. El caràcter esportiu té un efecte negatiu en l'índex de preus i el valor, igual que el disseny, el servei i la facilitat de maneig del cotxe.

7 Conclusions

Des del començament d'aquest treball, s'ha fet un estudi clàssic de les correlacions canòniques. Amb aquest estudi s'ha fet possible tenir una idea teòrica de què són les correlacions canòniques i per a que serveixen.

S'ha observat que és un mètode general de l'anàlisi multivariant que té relacions amb altres tècniques multivariants, en concret amb l'anàlisi de components principals, l'anàlisi canònica de poblacions i l'anàlisi de correspondències.

D'altra banda, hi ha molts altres aspectes de l'anàlisi de correlacions canòniques que queden fora de l'abast del present treball, per exemple, trobar relacions lineals que tinguin un coeficient de correlació màxim entre més de dos conjunts de variables, o trobar una relació no lineal donat que la relació lineal pot ser no adequada per descriure la relació entre dos conjunts de variables.

Referències

[1] Cuadras, C. M. (2014). *Nuevos Métodos de Análisis Multivariante.* Barcelona: CMC Editions.

[2] Härdle, W., Simar, L. (2015).*Applied Multivariate Statistical Analysis* (4ta Ed.) Springer, Berlín.

[3] Hotelling, H. (1936). *Relations between two sets of variables.* Biometrica 28 (3/4), pp. 321-327.

[4] Mardia K. V., Kent J. T., Bibby j. M. (1979). *Multivariate Analysis* Academic Press, London.

[5] Peña, D. (2002). *Análisis de datos multivariantes.* Madrid: McGraw-Hill Interamericana de España, S.A.U.

[6] Takeuchi K., Yanai H., Mukherjee B. N. (1981). *The foundations of Multivariate Analysis.* Wiley El., New Delhi

Annex

A continuació es detallen els codis del programa R que s'han utilitzat per l'elaboració dels càlculs exposats al capítol 6.

```
#
# Cancor.01
#
# Propòsit
#
# Calcular les correlacions canòniques i vectors de coeficients d'una
# matriu de covariàncies
#
#   Entrada:
#
#   S : matriu [p,p] simètrica, definida positiva (a covariance matrix)
#   I1 : un vector d'índexs en l'interval [ 1, p ] , corresponent al primer
#       subconjunt de variables. Suposem I1 està estrictament ordenat .
#   I2 : un vector d'índexs en l'interval [ 1, p ] , corresponent al segon
#       Subconjunt de variables. Suposem I2 està estrictament ordenat .
#       I2 està continguda en  1:p\I1
#
#   Quantitats internes :
#
#   l1  : un nombre sencer positiu. l1 = length(I1)
#   l2  : un nombre sencer positiu. l2 = length(I2)

Cancor.01<-function(S,I1,I2){
  l1<-length(I1)
  l2<-length(I2)
#
#   Aquí hem d'inserir algun codi per verificar si els paràmetres introduïts
```

```
#   satisfan  les condicions establertes. Per exemple:
#
if (!is.matrix(S)) stop("S must be a matrix")
if (nrow(S)!=ncol(S)) stop("S must be a symmetric matrix")
#
#
#
#
    S11<-S[I1,I1]  # An [l1,l1] symmetric, pd matrix
    S12<-S[I1,I2]  # An [l1,l2] matrix
    S22<-S[I2,I2]  # An [l2,l2] symmetric, pd matrix
    S21<-t(S12)
#
#
#  Calculem P1
#
  P1<-S12%*%solve(S22)%*%S21
#
#
# Ara resolem el problema de vectors relativa a P1 respecte a S11
# Per fer això , primer obtenim S11mh , igual a S11 a la potència menys 1/2 i
# S11mg, igual a S11 a la potència 1/2.
#
  E1<-eigen(S11)
  S11mh<-E1$vectors%*%diag(E1$values^(-0.5))%*%t(E1$vectors)
  S11mg<-E1$vectors%*%diag(E1$values^(0.5))%*%t(E1$vectors)
#
#
# Definim N1 com la matriu S11mh * P1 * S11mh
#
  N1<-S11mh%*%P1%*%S11mh
```

```
#
#
#
# Calculem P2
#
  P2<-S21%*%solve(S11)%*%S12
#
#
# Ara resolem el problema de vectors relativa a P2 respecte a S22
# Per fer això , primer obtenim S22mh , igual a S22 a la potència menys 1/2 i
# S22mg, igual a S22 a la potència 1/2.
#
  E2<-eigen(S22)
  S22mh<-E2$vectors%*%diag(E2$values^(-0.5))%*%t(E2$vectors)
  S22mg<-E2$vectors%*%diag(E2$values^(0.5))%*%t(E2$vectors)
#
#
# Definim N2 com la matriu S22mh * P2 * S22mh
#
  N2<-S22mh%*%P2%*%S22mh
#
#
# Calculem les correlacions canòniques agafant l'arrel quadrada
# del's valors propis de N1 o N2.
#

  A1<-eigen(N1)
  vaps<-sqrt(A1$values)
  print("Les correlacions canòniques són:")
  print(vaps)
#
```

```
#
# Calculem els vectors canònics normalitzats. Per fer això calculem els
# vectors propis de les matrius N1 i N2 multiplicats per S11^(-0.5) i
# S22^(-0.5) respectivament.
#
#
  veps1<-S11mh%*%A1$vectors
  print("Els vectors canònics normalitzats són:" )
  print(veps1)



  A2<-eigen(N2)
  veps2<-S22mh%*%A2$vectors
  print(veps2)



}
```