Proposal of analyses: DCA, DCIA, DGCA and DGPCA

Gabriel Kissita, Lavie Phanie Moulogho Issayaba, Roger Armand Makany and Dominique Mizére

Institut Supérieur de Gestion
Université Marien Ngouabi, Brazzaville, Congo

Copyright © 2017 Gabriel Kissita, Lavie Phanie Moulogho Issayaba, Roger Armand Makany and Dominique Mizére. This article is distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Several methods of factorial analysis (all the tables have the same individuals) already exist. Of these last, there are only some which has dual versions (all the tables have the same variables). It is thus a question initially of defining a dual matrix of the matrix of inter-covariances between variables of two tables (matrix of the scalar inter-products between individuals of the two tables). This dual matrix will enable us to propose the dual methods of the following analyses: canonical analysis, co-inertia analysis, generalized canonical analysis, generalized principal components analysis.

Keywords: factor analysis, dual factor analysis

1 Introduction

When one has two or several tables of data having the same individuals, of many methods of simultaneous factorial analysis of the tables were proposed, one can quote extensions and alternatives of the canonical analysis (CA) of Hotelling (1936) between two tables such as: co-inertia analyses 1 and 2 (CIA) of Chessel and Mercier (1993) and Tucker (1958) and or of Lafosse and Hanafi (1996), generalized canonical analysis (GCA) of Carroll (1968) and generalized principal component analysis (GPCA) of Casin (1996). Apart from these
Gabriel Kissita et al.

methods, one also has, the ACT (STATIS method) of L'Hermier des Plantes (1976) and multiple factorial analysis (MFA) introduced by Brigitte Escofier and J. Pagés (1984). All these methods are founded on the matrix of inter-covariances between variables of tables taken two to two. In the case of several tables of data (a table by group) having the same variables (vertical multi-table of data), only STATIS and MFA methods have dual versions. The motivations of these last are well known and can be found in Lavit (1988) then in Lé et al. (2010). The dual methods are found in several circumstances, one can quote the following cases: When one measures on the models of cars of various marks the same characteristics; the same macroeconomic variables are measured on the groups of the countries; the same questions are put about several groups of individuals. The dual factor analyses are adapted to this problematical which are that of the analysis of a vertical multi-table.

The idea for these methods is to find axes common of representation some relationships between variables at total level, i.e. measured on the whole some individuals. These axes are also used to represent the relationships at the partial level, i.e. within each group of individuals. To propose the dual methods of those which are quoted before, it is thus a question in this article above all of defining a dual matrix of the matrix of inter-covariances between two tables taken two to two, called matrix of the scalar inter-products between two tables taken two to two. Thus, the principal goal of this article is the proposal for some dual methods some methods of factorial analysis of the tables having the same individuals (direct methods). The dual methods of factorial analysis (methods of factorial analysis having the same variables) which are proposed in this article are those of the following methods: canonical analysis (CA), co-inertia analysis (CIA), generalized canonical analysis (GCA) and generalized principal components analysis (GPCA). Outside co-inertia analysis, for the estimate some factorial axes, all other methods use the metrics inverse matrix of scalar products between individuals which are not invertible such as the matrix of variance-covariances of a table in the case some direct methods (methods which are presented at section 2). This way of proceeding is different from that which is defined in dual STATIS and dual MFA methods. In return, these last are also founded on this matrix some scalar inter-products.

Thus, to section 2 after having briefly recalled the differents base methods and given some notations in the context of the simultaneous analysis of tables having the same individuals, we will propose to section 3, the dual methods some primary methods while giving to passage the notations in this context tables which have the variables in common. The principle of all these methods consists in determining according to Cazes (2004) orthonormed bases of the vector spaces of the individuals and variables in which these last are represented.

To section 4, we compare the various methods. In consideration of to fact
that any dual methods which are proposed here use the metric inverse (cf to subsection 3.2), we will not give an application of these methods to real data. In prospect, we will propose dual versions regularized to make the matrices of scalar products invertible. Under these conditions, applications of these methods will be given. To section 5 in return, we only will give one example of application for the DCIA.

2 Some methods of factorial analysis of tables having the same individuals

In this section, it is question of presenting some methods of factorial analysis of tables having the same lines. But initially, we will define the notations and data which will be used in this section. It will be also made state of the matrix of inter-covariances between tables which found these methods.

2.1 Data and notations

\(A'\) designates the transposed of matrix \(A\).

In this context, \(I\) and \(J_i\) indicates sets respectively of the individuals and of the variables, tables \(X_i\) (\(i = 1, \cdots, M\)) are defined on the products \(I \times J_i\), the whole of the individuals being the same for all the tables.

One indicates by \(n\) the cardinal of \(I\) and by \(p_i\) the cardinal of \(J_i\). Each variable \(l\) of table \(X_i\) is represented by the vector column of \(\mathbb{R}^n\) noted \(x_i^{(l)}\) and each individual \(e\) by the vector column of \(\mathbb{R}^{p_i}\) noted \(x_{i,e}\). One supposes that each individual \(e\) is provided with the weight \(m_e\) such as \(m_e > 0\) and \(\sum (m_e \setminus e \in I) = 1\) and that each variable \(l\) is provided with the weight \(m_l\) such as \(m_l > 0\) and \(\sum (m_l \setminus l \in J_i) = 1\).

One can possibly weighted the tables \(X_i\) by the weights \(\pi_i\) such as \(\pi_i > 0\).

\(\Delta_{\pi} = \text{diag}(\pi_i Id_{p_i} / i = 1, \cdots, M)\) is a diagonal block matrix containing the weights \(\pi_i\) tables \(X_i\).

Let us consider \(M\) triplets statistics \((X_i, Q_i, D)\) where \(X_i\) is a table of dimension \((n, p_i)\).

\(Q_i\) is the metric defined in the space of the individuals \(\mathbb{R}^{p_i}\).

\(Q_{bd} = \text{diag}(Q_i / i = 1, \cdots, M)\) the diagonal block metric defined in \(\mathbb{R}^p\) of the metrics \(Q_i\) with \(p = \sum_{i=1}^{M} p_i\).

\(D\) is the metric of weights of the individuals defined in \(\mathbb{R}^n\).

One Supposes that the tables \(X_i\) are centered and possibly scale compared to the metric \(D\).

\(X = [X_1, \cdots, X_i, \cdots, X_M]\) is a horizontal multi-table of dimension \((n, p)\) with \(p\) the number columns of table \(X\).

Note by \(X_c = X(\Delta_{\pi})^{\frac{1}{2}}\) the horizontal multi-table whose under table \(X_i\) is
affected by the number $\sqrt{\pi_i}$.

$V_{X_i X_j} = X'_i DX_j$ the matrix of inter-covariances between the tables $X_i$ and $X_j$.

$V_{X_i} = X'_i DX_i$ the matrix of variance-covariances of table $X_i$.

$W_{X_i} = X'_i Q_i X_i'$ the matrix of scalar products between individuals of table $X_i$.

$F_i = Im(X_i)$ subspace of $\mathbb{R}^n$ engendered by the columns of $X_i$.

$P_{X_i}$ is the $D$-orthogonal projector on the subspace $F_i$.

$Id_p$ is an identity matrix of order $p$.

$P_a \perp a = Id_p - P_a$ is the $Q$-orthogonal projector on the subspace engendered by the orthogonal of $a$.

$\langle u | v \rangle_Q = u'Qv$ the scalar product between two vectors $u$ and $v$ compared to the metric $Q$.

$\text{cov}(x, y)$ (resp. $\text{cor}(x, y)$) indicates covariance (resp. the correlation) between the variables $x$ and $y$.

$\text{var}(x)$ indicates the variance of the variable $x$.

If $M = 2$, we denote $X_1$ and $X_2$ the tables of the data, in this case the matrix of inter-covariances between $X_1$ and $X_2$ becomes $V_{X_1 X_2} = X'_1 DX_2$ with $V_{X_2 X_1} = V_{X_1 X_2}'$.

$K = V_{X_1 X_2} V_{X_2 X_1}$ and $H = V_{X_2 X_1} V_{X_1 X_2}$ are positive semi-definite symmetric matrices.

All the methods of this section are founded on the matrix of inter-covariances between two tables. This matrix allows to establish the relationships two to two between $M$ tables having the same individuals. It seeks to synthesize the existing between two groups of quantitative variables $X_1$ and $X_2$ having respectively $p_1$ and $p_2$ variables measured on the same $n$ individuals.

2.2 The Canonical Analysis: CA

Canonical analysis (AC) was introduced by Hotelling (1936). Under its general form, canonical analysis has only a limited interest for applications, since it leads to great difficulties of interpretation, but its theoretic border is fundamental. It generalizes several methods of data analysis.

Canonical analysis also has similarities to Principal Component Analysis (PCA) in the determination of axes and interpretation of charts of individuals and variables.

On the geometrical framework, the CA amounts to minimise, dimension by dimension, the angle between the components of these two groups of variables (Cazes, 1980). The CA is the central method of multidimensional descriptive statistics.

It seeks to synthesize the relationships between two groups of quantitative variables $X_1$ and $X_2$ having respectively $p_1$ and $p_2$ variables measured on the same $n$ individuals. Its goal is to summarize, most adequately possible, relationships between these two groups of variables.
$r$ is the rank of the matrix $V_{X_1 X_2}$.
The canonical analysis between two tables $X_1$ and $X_2$ is the research of the linear combinations $c_{X_1} = X_1 d_1$ and $c_{X_2} = X_2 d_2$ ($d_1 \in (\mathbb{R}^p)^*$ and $d_2 \in (\mathbb{R}^p)^*$) maximizing the function

$$f(d_1, d_2) = \text{cor}(X_1 d_1, X_2 d_2)$$

under the constraints

$$\text{var}(c_{X_1}) = \text{var}(c_{X_2}) = 1$$

Once determined the first canonical correlation (the maximum of the square of the linear correlation between $c_{X_1, 1} = X_1 d_{1, 1}$ and $c_{X_2, 1} = X_2 d_{2, 1}$ noted $\rho_1$) as well as the first couple $(c_{X_1, 1}, c_{X_2, 1})$ of the canonical variables, the other canonical correlations of order $s$ strictly higher than one (the maximum of the square of the linear correlation between $c_{X_1, s} = X_1 d_{1, s}$ and $c_{X_2, s} = X_2 d_{2, s}$ noted $\rho_s$) and couple it $(c_{X_1, s}, c_{X_2, s})$ of the canonical variables are determined by recurrence by maximizing the function

$$f(d_{1, s}, d_{2, s}) = \text{cor}(X_1 d_{1, s}, X_2 d_{2, s}) \quad s \in \{2, 3, \ldots, r\}$$

under the constraints

$$\text{var}(c_{X_1, s}) = \text{var}(c_{X_2, s}) = 1$$

and additional constraints

$$\text{cor}(c_{X_1, s}, c_{X_1, h}) = \text{cor}(c_{X_2, s}, c_{X_2, h}) = 0$$

for all $h \in \{1, 2, \ldots, s - 1\}$, $s \neq h$.
The solution of this problem is given by the following relations:

$$V_{X_1}^{-1} V_{X_1 X_2} V_{X_2}^{-1} V_{X_2 X_1} d_{1, s} = \rho_s d_{1, s}$$

$$V_{X_2}^{-1} V_{X_2 X_1} V_{X_1}^{-1} V_{X_1 X_2} d_{2, s} = \rho_s d_{2, s}$$

In the space of the variables, the canonical analysis suggested by Hotelling (1936) between two tables of variables $X_1$ and $X_2$ checks, with the order $s$, the following eigenvalue equations:

$$P_{X_1} P_{X_2} c_{X_1, s} = \rho_s c_{X_1, s}$$

$$P_{X_2} P_{X_1} c_{X_2, s} = \rho_s c_{X_2, s}$$

where $c_{X_1, s} = X_1 d_{1, s}$ and $c_{X_2, s} = X_2 d_{2, s}$ are linear combinations of $\mathbb{R}^n$.
The vectors $d_{i, s}$ and $c_{X_i, s}$ for $(i = 1, 2)$ and $(s = 1, \ldots, r)$ form orthonormed bases.
2.3 The Co-Inertia Analysis: CIA

2.3.1 The Co-Inertia Analysis 1: CIA1

Proposed by Chessel and Mercier (1993), the co-inertia analysis between two tables \(X_1\) and \(X_2\) measured on the same \(n\) individuals is the research of the linear combinations \(c_{X_1} = X_1Q_1a_1\) and \(c_{X_2} = X_2Q_2\) maximizing covariance between these two linear combinations under the constraints of normalisation on the axial vectors \(a_1\) and \(a_2\).

\(Q_1\) and \(Q_2\) are respectively the metric defined ones in spaces of the individuals \(R^{p1}\) and \(R^{p2}\).

Once found the solution of order 1 which corresponds to the first couple \((c_{X_1,1} = X_1Q_1a_{1,1}, c_{X_2,1} = X_2Q_2a_{2,1})\), one continues research until the order \(s\), which corresponds to the couple \((c_{X_1,s} = X_1Q_1a_{1,s}, c_{X_2,s} = X_2Q_2a_{2,s})\) which maximizes covariance between the elements of this couple under the constraints of unit norm with the axial vectors \(a_{1,s}\) and \(a_{2,s}\) and which check besides the additional constraints of orthogonality \(a_{1,s}Q_1a_{1,h} = a_{2,s}Q_2a_{2,h} = 0\) for all \(s \neq h\). One stops the research for the solution when one finds \(r\) solutions where \(r\) is the rank of the matrix \(V_{X_1X_2}\).

The solutions of the co-inertia analysis of order \(s\) in spaces of the individuals \(R^{p1}\) and \(R^{p2}\) given by the eigenvalue equations

\[
V_{X_1X_2}Q_2V_{X_2X_1}Q_1a_{1,s} = \lambda_s a_{1,s}
\]
\[
V_{X_2X_1}Q_1V_{X_1X_2}Q_2a_{2,s} = \lambda_s a_{2,s}
\]

In the space of the variables \(R^n\), the solutions are given by the eigenvalue equations

\[
W_{X_1} DW_{X_2} Dc_{X_1,s} = \lambda_s c_{X_1,s}
\]
\[
W_{X_2} DW_{X_1} Dc_{X_2,s} = \lambda_s c_{X_2,s}
\]

The systems of the vectors \(\{c_{X_1,s}\}_{s=1,\ldots,r}\) and \(\{c_{X_2,s}\}_{s=1,\ldots,r}\) are not \(D\)-orthogonal.

In return, for \(s \neq h\), the vectors \(c_{X_1,s}\) and \(c_{X_2,h}\) are \(D\)-orthogonal.

The co-inertia analysis of two statistical studies \((X_1, Q_1, D)\) and \((X_2, Q_2, D)\) is the principal components analysis of the triplet \((V_{X_2X_1}, Q_1, Q_2)\).

2.3.2 The Co-Inertia Analysis 2: CIA2

In this subsection, we will present an another co-inertia analysis consult Tucker (1958), Lafosse and Hanafi (1997) compared to co-inertia analysis 1 of Chessel and Mercier (1993). It is formulated in the following way: when one has two triplets of the tables \((X_1, Q_1, D)\) and \((X_2, Q_2, D)\), to study the internal structure and the relationship between two tables \(X_1\) and \(X_2\), there are two
ways equivalent to the optimum to achieve these goals. The problem consists to define a couple \((X_1Q_1a_1, X_2Q_2a_2)\), where \(a_1\) and \(a_2\) are \(Q_1\)-normed and \(Q_2\)-normed in \(\mathbb{R}^{p_1}\) and \(\mathbb{R}^{p_2}\) respectively, synthesizing a community of structures of the tables \(X_1\) and \(X_2\) relating to each of the two systems of internal covariations. A simple way to proceed consists to wank simultaneously than the component \(X_1Q_1a_1\) characterizes the system of covariations of the variables \(X_2^{(l)}\) \((l = 1, \cdots, p_2)\) columns of the table \(X_2\), and than the synthetic variable \(X_2Q_2a_2\) characterizes the system of covariations of the variables \(X_1^{(h)}\) \((h = 1, \cdots, p_1)\) columns of table \(X_1\). It’s for that than the first criterion consists to maximize the function:

\[
f(a_1, a_2) = \left(\sum_{l=1}^{p_2} \text{cov}^2(X_1Q_1a_1, x_2^{(l)})\right) \left(\sum_{l=1}^{p_1} \text{cov}^2(X_2Q_2a_2, x_1^{(h)})\right)
\]

or

\[
f(a_1, a_2) = (a_1^t Q_1 K Q_1 a_1)(a_2^t Q_2 H Q_2 a_2)
\]

under the constraints of normalisation

\[
\|a_1\|_{Q_1} = \|a_2\|_{Q_2} = 1
\]

The second criterion amounts to maximize the function

\[
g(a_1, a_2) = (a_1^t Q_1 K Q_1 a_1) + (a_2^t Q_2 H Q_2 a_2)
\]

under the same constraints.

The solution of order \(s\) of the co-inertia analysis 2 is to determine two sets of \(r\) \((r\) is the rank of \(V_{X_1,X_2}\)) synthetic components \(c_{X_1,1} = X_1Q_1a_{1,s}\) and \(c_{X_2,2} = X_2Q_2a_{2,s}\) where \(a_{1,s}\) and \(a_{2,s}\) are axes of co-inertia checking the stationary equations

\[
KQ_1a_{1,s} = r_{a_{1,s}} a_{1,s}
\]

\[
HQ_2a_{2,s} = r_{a_{2,s}} a_{2,s}
\]

with \(a_{1,s} = r_{a_{1,s}} a_{2,s}\) the maximum of the function \(f\) where \(r_{a_{1,s}} = a_{1,s}^t Q_1 K Q_1 a_{1,s}\) and \(r_{a_{2,s}} = a_{2,s}^t Q_2 H Q_2 a_{2,s}\).

The axles of Co-inertia of order \(s\) of the co-inertia analysis for the function \(f\) between two tables \(X_1\) and \(X_2\) check also the following relations:

\[
Q_1^{\frac{1}{2}} K Q_1^{\frac{1}{2}} c_{1,s} = r_{a_{1,s}} c_{1,s}
\]

\[
Q_2^{\frac{1}{2}} H Q_2^{\frac{1}{2}} c_{2,s} = r_{a_{2,s}} c_{2,s}
\]

with \(c_{1,s} = Q_1^{\frac{1}{2}} a_{1,s}\) checking the constraints \(c_{1,s}^t c_{1,s} = a_{1,s}^t Q_1 a_{1,s} = 1\) and \(c_{1,s}^t c_{1,h} = a_{1,s}^t Q_1 a_{1,h} = 0\) for \(s \neq h\) \(c_{2,s} = Q_2^{\frac{1}{2}} a_{2,s}\) checking the constraints
\[ c'_{2,s}c_{2,s} = a'_{2,s}Q_2a_{2,s} = 1 \] and \[ c'_{2,s}c_{2,h} = a'_{2,s}Q_2a_{2,h} = 0 \] for \( s \neq h \), one determines after the axles of co-inertia \( a_{1,s} \) and \( a_{2,s} \) associated spaces \( \mathbb{R}^{p_1} \) and \( \mathbb{R}^{p_2} \).

If one notes by \( c_{X_1,s} = X_1Q_1a_{1,s} \) and \( c_{X_2,s} = X_2Q_2a_{2,s} \) the vectors of \( \mathbb{R}^n \) to the order \( s \), these vectors check the stationary equations

\[
W_{X_1}DX_2'X_2'c_{X_1,s} = r_{a_{1,s}}c_{X_1,s} \\
W_{X_2}DX_1'X_1'c_{X_2,s} = r_{a_{2,s}}c_{X_2,s}
\]

The systems of linear combinations \( \{c_{X_1,s}\}_{s=1,\ldots,r} \) and \( \{c_{X_2,s}\}_{s=1,\ldots,r} \) are not \( D \)-orthogonal. It is the same for the linear combinations \( c_{X_1,s} \) and \( c_{X_2,h} \) for \( s \neq h \).

### 2.4 The GCA method

The purpose of the generalized canonical analysis (GCA) within the sense of Carroll (1968), is to measure, characterize linear relationships between \( M \) groups of variables \( X_i \).

Let us consider \( M \) groups of variables \( X_i \) each table is of format \( (n, p_i) \) for all \( i = 1, \cdots, M \). One notes by \( X \) the horizontal multi-table of format \( (n, p) \) with \( p = \sum_{i=1}^{M} p_i \) the total number of variables of the horizontal multi-table \( X \).

The tables \( X_i \) can be weighted by weights \( \pi_i \) and are centered and possibly scale.

The generalized canonical analysis Carroll search \( r \) (\( r = \min(\text{rank}(X_i)/i = 1, \cdots, M) \)) compromise variables \( z_s \) and linear combinations \( c_{X_i,s} = X_id_{i,s} \) where \( d_{i,s} \in (\mathbb{R}^{p_i})^* \) maximizing the function

\[
\sum_{i=1}^{M} \pi_i cor^2(X_id_{i,s}, z_s)
\]

under the constraints \( \text{var}(z_s) = 1 \) and \( \text{cor}(z_s, z_h) = 0 \) for \( s \neq h \) for all \( s = 1, \cdots, r \) and \( i = 1, \cdots, M \).

The solution of this problem leads to determine vectors \( z_s \) and \( c_{X_i,s} \) checking the relations

\[
(\sum_{i=1}^{M} \pi_i P_{X_i})z_s = \lambda_s z_s
\]

\[ c_{X_i,s} = P_{X_i}z_s \]

where \( P_{X_i} \) is the \( D \)-orthogonal projector on \( F_i \) and \( \lambda_s \) the greatest eigenvalue associated at the eigenvector \( z_s \) of the matrix \( \sum_{i=1}^{M} \pi_i P_{X_i} \). By noting to \( V_{bd}^{-1} = \text{diag}(V_{X_i}^{-1}/i = 1, \cdots, M) \) the block-diagonal metric defined in \( \mathbb{R}^p \) some metrics
$V_{X_i}^{-1}$ of Mahalanobis and $X_c = X(\Delta_x)^{\frac{1}{2}}$, the eigenvalue and eigenvector relation preceding can be still written in the space of the variables.

$$X_c V_{bd}^{-1} X'_c D z_s = \lambda_s z_s$$

If one poses $z_s = X_c d_s$ in the last relation where $d_s \in (\mathbb{R}^p)^*$ the block column vector whose the $i$th block is the vector $d_{i,s}$, and by pre-multiplying this last by $X'_c D$ afterwards while simplifying, it results the relation

$$V_{bd}^{-1} V_{X_c} d_s = \lambda_s d_s$$

$V_{X_c} = X'_c DX_c$ the matrix of variance-covariances of $X_c$. Consequently, the GCA of Carroll (1968) is the PCA of the triplet $(X_c, V_{bd}^{-1}, D)$. The vectors $z_s$ and $\frac{d_s}{|d_s|_{V_{bd}}}$ forms orthonormed bases for all $s = 1, \cdots, r$.

### 2.5 The GPCA method

Let us consider $M$ statistical triplets $(X_i, Q_i, D)$ possibly weighted by the weights $\pi_i$. $Q_i$ is the diagonal metric of the weights in $\mathbb{R}^{p_i}$ whose the principal diagonal is constituted of the weights $m_l$. $Q_{bd}$ the block diagonal metric whose the principal diagonal is composed of the metrics $Q_1$.

The tables $X_i$ are centered for each variable. The generalized principal components analysis (GPCA) of Casin (1996) consists in at step 1 to determine vectors $z_1$ of $\mathbb{R}^n$ and then $z_{i,1} = p_{X_i} z_1$ of $F_i$ such that

$$f(z_1) = \sum_{i=1}^{M} \sum_{l=1}^{p_i} (\pi_i m_l \text{cov}^2(z_1, x_i^{(l)}))$$

either maximum under the constraint $z'_1 D z_1 = 1$.

With the step $s$, it is a question of determine $z_s$ to $\mathbb{R}^n$ and then $z_{i,s} = p_{X_{i,s-1}} z_s$ of $F_i$ such that

$$f(z_s) = \sum_{i=1}^{M} \sum_{l=1}^{p_i} (\pi_i m_l \text{cov}^2(z_s, x_i^{(l)}))$$

either maximum under the constraints $z'_s D z_s = 1$ and the constraints of orthogonalisation $z'_{i,s} D z_{i,h} = 0$ for all $h < s$ and $i = 1, \cdots, M$.

The solution of order 1 of the GPCA checks the equation with the values and eigenvectors

$$\left(\sum_{i=1}^{M} \pi_i W_{X_i}\right) D z_1 = \lambda_1 z_1$$
3 The dual of preceding methods

In this section, it is a question of to propose some dual methods of that which are presented at subsection 2.1. The context of the dual methods is that where the same variables are measured on several groups of individuals.

3.1 Data and notations

In this context, $I_i$ and $J$ indicates sets respectively of the individuals and of the variables, $X_i$ ($i = 1, \cdots, M$) are definite on the products $I_i \times J$, the set of the variables being the same for all the tables. One indicates by $n_i$ the cardinal of $I_i$ and by $p$ the cardinal of $J$.

Each variable $l$ of table $X_i$ is represented by the vector row of $\mathbb{R}^{n_i}$ noted $x_i^{(l)}$ and each individual $e$ by the vector row of $\mathbb{R}^p$ noted $x_{i,e}$. If for all $i$ ($i = 1, \cdots, M$), where $i \in I_i$, one has $X_i = \sum_{e=1}^{n_i} x_{i,e} x_{i,e}'$.

Let us suppose that each individual $e$ is provided with the weight $m_e$ such as $m_e > 0$ and $\sum_{e} (m_e \setminus e \in I_i = 1)$ and that each variable is provided with the weight $m_l$ such as $m_l > 0$ and $\sum_{l \in J} (m_l \setminus l \in J) = 1$.

One considers $M$ triplets statistics $(X_i, Q, D_i)$ where $X_i$ is a table of dimension $(n_i, p)$.

$Q$ is the metric defined in the space of the individuals $\mathbb{R}^p$.

$D_i = diag(m_e \setminus e \in I_i)$ is the metric weights of the individuals defined in $\mathbb{R}^{n_i}$.

$D_{bd} = diag(D_i / i = 1, \cdots, M)$ the diagonal block metric of individuals defined in $\mathbb{R}^n$ some metrics $D_i$ with $n = \sum_{i=1}^{M} n_i$.

Vertical table $X$ is the superposition of tables $X_i$ which are centered and possibly scale compared at the metric $D$.

We can also possibly weighted tables $X_i$ by the weights $\pi_i$ such as $\pi_i > 0$.

$\Delta_{\pi} = diag(\pi_i D_i / i = 1, \cdots, M)$ is a diagonal block matrix containing the
weights \( \pi_i \) tables \( X_i \) and in this case \( X \) becomes \( X_c = (\Delta_w)^{\frac{1}{2}}X \), the vertical multi-table whose under table \( X_i \) is affected by the number \( \sqrt{\pi_i} \).

\( X = [X'_1, \cdots , X'_i, \cdots , X'_m]^\prime \) being the vertical multi-table of dimension \((n, p)\) with \( n \) the number columns of table \( X \).

\( V_{X_i} = X'_iD_iX_i \) the matrix of variance-covariances of tables \( X_i \).

\( W_{X_i} = X'_iQX'_i \) the matrix of scalar products between individuals of table \( X_i \).

\( W_{X_iX_j} = X'_iQX'_j \) the matrix of inter-scalar products between individuals of \( X_i \) and that of \( X_j \). This matrix is dual matrix of inter-covariances \( V_{X_iX_j} = X'_iDX_j \) between tables \( X_i \) and \( X_j \) defined at 2.1. It is at the base of all the dual methods in factorial analysis of data in particular that which are proposed in this article.

\( E_i = Im(X'_i) \) subspace of \( \mathbb{R}^p \) engendered by the columns of \( X'_i \).

\( P_{X'_i} = X'_iW_{X'_i}X_iQ \) is the \( Q \)-orthogonal projector with \( E_i \).

\( Id_n \) is an identity matrix of order \( n \).

\( P_{u_i} = Id_n - P_{u_i} \) is the \( D_i \)-orthogonal projector with the subspace engendered by the vector \( u_i \) of \( \mathbb{R}^{n_i} \) which is \( D_i \)-normed, where \( P_{u_i} = u_iu'_iD_i \).

Let us note by \( a_{i,s} = X'_i v_{i,s} \) the vector of \( \mathbb{R}^p \) which is a linear combination of the columns of table \( X'_i \) associated at the solution of order \( s \) where \( v_{i,s} \in (\mathbb{R}^{n_i})^* \).

### 3.2 The Dual Canonical Analysis: DCA

We will propose in this subsection the dual canonical analysis between two tables of the canonical analysis of Hotelling (1936).

Let us consider two statistical study \((X_1, Q, D_1)\) and \((X_2, Q, D_2)\), or, one measures on two groups of individuals \( I_1 \) and \( I_2 \) having respectively \( n_1 \) and \( n_2 \) individuals the same \( p \) variables.

\( Q \) is the metric defined in the space of individuals \( \mathbb{R}^p \).

\( D_1 \) and \( D_2 \) are of the metrics weights defined in \( \mathbb{R}^{n_1} \) and \( \mathbb{R}^{n_2} \) respectively.

The tables \( X_1 \) and \( X_2 \) are supposed to be centered.

\( W_{X_1X_2} = X_1QX_2' \) the matrix of the scalar inter-products between the individuals of the table \( X_1 \) and those of \( X_2 \) with \( W_{X_2X_1} = W_{X_1X_2}' \). \( r \) is the rank of Dual matrix \( W_{X_1X_2} \).

The dual canonical analysis (DCA) between two tables is a method which allows to measure the proximity between the individuals of table \( X_1 \) and those of table \( X_2 \). These tables are respectively of format \((n_1, p)\) and \((n_2, p)\). The dual canonical analysis between these two tables \( X_1 \) and \( X_2 \) is the research of the linear combinations \( X'_1 v_1 \) and \( X'_2 v_2 \) \((v_1 = D_1 u_1 \in (\mathbb{R}^{n_1})^* \) and \( v_2 = D_2 u_2 \in (\mathbb{R}^{n_2})^* \) maximizing the function

\[
    f(v_1, v_2) = \langle X'_1 v_1 | X'_2 v_2 \rangle_Q = v'_1 W_{X_1X_2} v_2
\]

under the contraints

\[
    v'_1 W_{X_1} v_1 = v'_2 W_{X_2} v_2 = 1
\]
with \( u_1 \in \mathbb{R}^{n_1} \) and \( u_2 \in \mathbb{R}^{n_2} \).

Once given the first canonical proximity (the maximum of the square of the scalar product between \( X'_1 v_{1,1} \) and \( X'_2 v_{2,1} \) noted \( \lambda_1 \)) thus which the initialy couple \((X'_1 v_{1,1}, X'_2 v_{2,1})\) canonical vectors, the other canonical proximities of order \( s \) greater than one (the maximum of the square of the scalar product between \( X'_1 v_{1,s} \) et \( X'_2 v_{2,s} \) noted \( \lambda_s \), where, \( v_{1,s} = D_1 u_{1,s} \) and \( v_{2,s} = D_2 u_{2,s} \)) and the couple \((X'_1 v_{1,s}, X'_2 v_{2,s})\) canonical vectors are determined by recurrence by maximizing the function

\[
f(v_{1,s}, v_{2,s}) = \langle X'_1 v_{1,s} | X'_2 v_{2,s} \rangle_Q = v'_{1,s} W X_1 X_2 v_{2,s}
\]

under the contraints

\[
v'_{1,s} W X_1 v_{1,s} = v'_{2,s} W X_2 v_{2,s} = 1
\]

and additional constraints

\[
\langle X'_1 v_{1,s} | X'_1 v_{1,h} \rangle_Q = \langle X'_2 v_{2,s} | X'_2 v_{2,h} \rangle_Q = 0
\]

for all \( h \in \{1, 2, \cdots, (s - 1)\} \), \( s \neq h \) and \( s = 2, \cdots, r \).

The solution of this problem is given by the following relations:

\[
W^{-1} X_1 W X_1 v_{1,s} = \lambda_s v_{1,s}
\]

\[
W^{-1} X_2 W X_2 v_{2,s} = \lambda_s v_{2,s}
\]

\( W^{-1}_X \) indicates the generalized inverse of the matrix \( W_X = X Q X' \) with \( X \) a matrix of dimension \((n, p)\), because \( W_X \) is according to Saporta (1990, page 184) of rank strictly lower than \( n - 1 \) thus not invertible.

If one indicates by \( P_X' = X'_i W^{-1}_X X_i Q \) the \( Q \)-orthogonal projector on \( E_i \) for \( i = 1, 2 \), in the space of individuals, the dual canonical analysis between two tables of variables \( X_1 \) and \( X_2 \) checks with the order \( s \), the following eigenvalues equations:

\[
P_{X'_1} P_{X'_2} a_{1,s} = \lambda_s a_{1,s}
\]

\[
P_{X'_2} P_{X'_1} a_{2,s} = \lambda_s a_{2,s}
\]

where \( a_{1,s} = X'_1 v_{1,s} \) and \( a_{2,s} = X'_2 v_{2,s} \) two vectors of \( \mathbb{R}^p \).

Vectors \( v_{i,s} \) and \( a_{i,s} \) for \((i = 1, 2)\) and \((s = 1, \cdots, r)\) form orthonormed bases.
3.3 The Dual Co-Inertia Analysis: DCIA

3.3.1 The Dual Co-Inertia Analysis 1: DCIA1

We will propose in this subsection the dual co-inertia analysis between two tables of the co-inertia analysis of Chessel and Mercier (1993). Let us consider two statistical study \((X_1, Q, D_1)\) and \((X_2, Q, D_2)\), or, one measure on two groups of individuals \(I_1\) and \(I_2\) having respectively \(n_1\) and \(n_2\) individuals the same \(p\) variables.

The context is the same one as that of the dual canonical analysis. Thus, the DCIA1 consists to find two linear combinations \(X_1' D_1 u_1 \in \mathbb{R}^{n_1}\) and \(X_2' D_2 u_2 \in \mathbb{R}^{n_2}\) maximizing the function

\[
f(u_1, u_2) = \langle X_1' D_1 u_1 | X_2' D_2 u_2 \rangle_Q = u_1' D_1 W_{X_1, X_2} D_2 u_2
\]

under contraints of normalization

\[
u_1' D_1 u_1 = u_2' D_2 u_2 = 1
\]

where \(u_1\) and \(u_2\) are respectively two vectors of \(\mathbb{R}^{n_1}\) and \(\mathbb{R}^{n_2}\).

The solution of order \(s\) in spaces of the variables \(\mathbb{R}^{n_1}\) and \(\mathbb{R}^{n_2}\) is given by the stationary equations

\[
(D_1)^{\frac{1}{2}} W_{X_1, X_2} D_2 W_{X_2, X_1} (D_1)^{\frac{1}{2}} c_{1,s} = \lambda_s c_{1,s}
\]

\[
(D_2)^{\frac{1}{2}} W_{X_2, X_1} D_1 W_{X_1, X_2} (D_2)^{\frac{1}{2}} c_{2,s} = \lambda_s c_{2,s}
\]

where \(u_{1,s} = (D_1)^{\frac{1}{2}} c_{1,s}\) and \(u_{2,s} = (D_2)^{\frac{1}{2}} c_{2,s}\).

If on notes by \(a_{1,s} = X_1' D_1 u_{1,s}\) and \(a_{2,s} = X_2' D_2 u_{2,s}\) the vectors of \(\mathbb{R}^p\) in order \(s\). Thus, these vectors checks the stationary equations

\[
V_{X_1} Q V_{X_2} Q a_{1,s} = \lambda_s a_{1,s}
\]

\[
V_{X_2} Q V_{X_1} Q a_{2,s} = \lambda_s a_{2,s}
\]

The systems of vectors \(\{a_{1,s}\}_{s=1,\ldots,r}\) and \(\{a_{2,s}\}_{s=1,\ldots,r}\) are respectively \(D_1\)-orthogonal and \(D_2\)-orthogonal. On the other hand, the systems of vectors \(\{a_{1,s}\}_{s=1,\ldots,r}\) and \(\{a_{2,s}\}_{s=1,\ldots,r}\) are not \(Q\)-orthogonal.

The dual co-inertia analysis 1 of co-inertia analysis 1 amounts making the PCA of the triplet \((W_{X_2, X_1}, D_1, D_2)\). This PCA is dual of PCA of the triplet \((V_{X_2, X_1}, Q_1, Q_2)\) of CIA of Chessel and Mercier (1993).
3.3.2 The Dual Co-Inertia Analysis 2: DCIA2

In this subsection, it is a question of proposing of the dual co-inertia analysis of the co-inertia analysis of Lafosse and Hanafi (1997).

The context is identical to that of the DCIA1 and that of the DCA. One poses simply here that

\[ X'_1 X_1 = \sum_{e=1}^{n_1} x'_{1,(e)} x_{1,(e)} \quad \text{and} \quad X'_2 X_2 = \sum_{e=1}^{n_2} x'_{2,(e)} x_{2,(e)}. \]

The dual co-inertia analysis 2 enables to study the proximity between two tables \( X_1 \) and \( X_2 \) by simultaneously intending that the vector \( a_1 = X'_1 D_1 u_1 \) describes the system of proximities of individuals columns of \( X'_2 \), and that the vector \( a_2 = X'_2 D_2 u_2 \) describes the system of proximities of individuals columns of \( X'_1 \).

The solution of this method is given by the maximization of one of the criterion equivalent to the optimum following:

The first criterion consists to maximize the function

\[
 f(u_1, u_2) = \left( \sum_{e=1}^{n_2} (X'_1 D_1 u_1 | x_{2,(e)})^2 \right) \left( \sum_{e=1}^{n_1} (X'_2 D_2 u_2 | x_{1,(e)})^2 \right)
\]

or

\[
 f(u_1, u_2) = (u'_1 D_1 W_{X_1 X_2} W_{X_2 X_1} D_1 u_1)(u'_2 D_2 W_{X_2 X_1} W_{X_1 X_2} D_2 u_2)
\]

under the contraints

\[
 u'_1 D_1 u_1 = u'_2 D_2 u_2 = 1
\]

where \( u_1 \) and \( u_2 \) are two vectors of \( \mathbb{R}^{n_1} \) and \( \mathbb{R}^{n_2} \) respectively.

The second criterion returns to maximise

\[
 f(u_1, u_2) = (u'_1 D_1 W_{X_1 X_2} W_{X_2 X_1} D_1 u_1) + (u'_2 D_2 W_{X_2 X_1} W_{X_1 X_2} D_2 u_2)
\]

under same contraints that the first criterion.

The solutions of order \( s \) are given by stationnary equations

\[
 (D_1)^{\frac{1}{2}} W_{X_1 X_2} W_{X_2 X_1} (D_1)^{\frac{1}{2}} c_{1,s} = r_{u_1,s} c_{1,s}
\]

\[
 (D_2)^{\frac{1}{2}} W_{X_2 X_1} W_{X_1 X_2} (D_2)^{\frac{1}{2}} c_{2,s} = r_{u_2,s} c_{2,s}
\]

where \( u_{1,s} = (D_1)^{-\frac{1}{2}} c_{1,s} \) and \( u_{2,s} = (D_2)^{-\frac{1}{2}} c_{2,s} \) checks stationnary equations

\[
 W_{X_1 X_2} W_{X_2 X_1} D_1 u_{1,s} = r_{u_1,s} u_{1,s}
\]

\[
 W_{X_2 X_1} W_{X_1 X_2} D_2 u_{2,s} = r_{u_2,s} u_{2,s}
\]
whith \( r_{u_1,s} = u_{1,s}'D_1W_{X_1X_2}X_1D_1u_{1,s} \) and \( r_{u_2,s} = u_{2,s}'D_2W_{X_2X_1}X_2D_2u_{2,s}, \) 
\( \lambda_s = r_{u_1,s}r_{u_2,s} \) the maximum of the function \( f. \)

If one notes by \( a_{1,s} = X_1'D_1u_{1,s} \) and \( a_{2,s} = X_2'D_2u_{2,s} \) vectors of \( \mathbb{R}^p \) in order \( s, \) then, these vectors checks stationnary equations

\[
V_{X1}QX_2'X_2Qa_{1,s} = r_{u_1,s}a_{1,s}
\]

\[
V_{X2}QX_1'X_1Qa_{2,s} = r_{u_2,s}a_{2,s}
\]

If one notes by \( r \) the rank of the matrix \( W_{X_1X_2}, \) the systems of vectors \( \{u_{1,s}\}_{s=1,...,r} \) and \( \{u_{2,s}\}_{s=1,...,r} \) are \( D_1 \)-orthogonal respectively \( D_2 \)-orthogonal.

The systems of vectors \( \{a_{1,s}\}_{s=1,...,r} \) and \( \{a_{2,s}\}_{s=1,...,r} \) are not \( Q \)-orthogonal. Moreover, for \( s \neq h, \) vectors \( a_{1,s} \) and \( a_{2,h} \) are not \( Q \)-orthogonal contrary to the DCIA1 where these vectors are \( Q \)-orthogonal.

### 3.4 The Dual Generalized Canonical Analysis: DGCA

It is a question to propose in this subsection the dual generalized canonical analysis of the generalized canonical analysis of carroll (1968). The data and notations are those which were defined in the 3.1.

Let us consider \( M \) statistical triplets \((X_i, Q, D_i)\) where \( X_i \) is a table of dimension \((n_i, p)\).

\( Q \) is metric defined in the space of individuals \( \mathbb{R}^p. \)

\( D_i \) is the metric of the weights of the individuals defined in \( \mathbb{R}^{n_i}. \)

The vertical table \( X \) is the superposition of tables \( X_i \) which are centered and possibly scale. One can also weighted the tables \( X_i \) by weights \( \pi_i. \)

\( \Pi = diag(\pi_i/i = 1, \cdots, M) \) is the diagonal matrix of weights of tables \( X_i. \)

\( W_{X_iX_j} = X_iQX_j' \) the matrix of the scalar inter-products between individuals of \( X_i \) and those of \( X_j. \)

The goal of the dual generalized canonical analysis is to measure proximities between \( M \) groups of individuals to \( X_i. \)

The dual generalized canonical analysis (DGCA) is the research of the \( r \) \((r = \min(rank(X_i)/i = 1, \cdots, M)) \) vectors common \( a_s \) to all tables and linear combinations \( a_{i,s} = X_i'v_{i,s} \) of \( \mathbb{R}^p \) maximizing the function

\[
\sum_{i=1}^{M} \pi_i\cos^2(X_i'v_{i,s}, a_s)
\]

under the contraints \( \|a_s\|_Q = 1 \) and \( \langle a_s|a_h\rangle Q = 0 \) for \( s \neq h \) where \( v_{i,s} \in (\mathbb{R}^{n_i})^* \) for all \( i = 1, \cdots, M \) and \( s = 1, \cdots, r, \) where \( \cos(a,b) \) indicates the cosine of the angle between two vectors \( a \) and \( b. \)
The solution of this problem driven of to determine for all \( i = 1, \cdots, M \) and \( s = 1, \cdots, r \) vectors \( a_s \) and \( a_{i,s} \) checking relations

\[
\left( \sum_{i=1}^{M} \pi_i P_{X'_i} \right) a_s = \gamma_s a_s
\]

\[
a_{i,s} = P_{X'_i} a_s
\]

where \( P_{X'_i} = X'_i W_{X'_i} X'_i Q \) is the \( Q \)-orthogonal projector in \( E_i \) and \( \gamma_s \) the greatest eigenvalue associated at the eigenvector \( a_s \) of the matrix \( \sum_{i=1}^{M} \pi_i P_{X'_i} \). By noting to \( W_{bd} = \text{diag}(W_{X_i}/i = 1, \cdots, M) \) the block-diagonal matrix defined in \( \mathbb{R}^n \) of the matrices \( W_{X_i} \) of scalar products between individuals of the table \( X_i \) and \( X_c = (\Delta_n)^{1/2} X \), the relation with eigenvalues and eigenvectors preceding one can be still written in the space of individuals

\[
X'_c W_{bd} X_c Q a_s = \gamma_s a_s
\]

If one notes by \( a_s = X'_c v_s \) in this relation where \( v_s \in (\mathbb{R}^n)^* \) the block column vector whose the \( i \)th block is the vector \( v_{i,s} \), and while pre-multiplying by \( X_c Q \) this relation then while simplifying afterwards, one finds the relation in the space of the variables

\[
W_{bd} W_{X_c} v_s = \gamma_s v_s
\]

\( W_{X_c} = X_c Q X'_c \) being the matrix of the scalar products between individuals of table \( X_c \).

Consequently, the DGCA is the PCA of triplet \((X_c, Q, W_{bd})\). The vectors \( a_s \) and \( \frac{v_s}{\|v_s\|_{W_{bd}}} \) form orthonormed bases for all \( s = 1, \cdots, r \).

### 3.5 The Dual Generalized Principal Components Analysis: DGPCA

In this subsection, it is a question to propose the dual generalized principal components analysis. The data and notations are identical to those which are defined in 3.1.

One notes by \( P_{X'_i} = X'_i W_{X'_i} X'_i Q \) the \( Q \)-orthogonal projector on \( E_i \) let us recall that \( X_{i,0} = X_i \) for all \( i = 1, \cdots, M \), the dual generalized principal components analysis enables to determine with order 1 the vector \( a_1 \) which is \( Q \)-normed of \( \mathbb{R}^p \) and the \( a_{i,1} = p_{X'_{i,0}} a_1 \) of \( E_i \) maximizing the function

\[
f(a_1) = \sum_{i=1}^{M} \pi_i \left( \sum_{e=1}^{n_e} m_e (x_{i,e} | a_1)^2 \right)_{Q}
\]
taking into account what \( \sum_{i=1}^{n_i} m_e x_{i(e)} x'_{i(e)} = V_{X_i} \) for all \( i = 1, \cdots, M \), the preceding function can be still written

\[
f(a_1) = a_1'Q\left(\sum_{i=1}^{M} \pi_i V_{X_i}\right)Qa_1
\]

Once found a first \( M + 1 \)-uple \( (a_{i,1}, a_1) \) of vectors, one continuous the research.

The solution of order \( s \) consist to find a \( s \)th \( M + 1 \)-uple \( (a_{i,s}, a_s) \) of vectors for all \( i = 1, \cdots, M \) and \( h = 1, \cdots, s - 1 \) such that \( a_{i,s}'Qa_{i,h} = 0 \) and \( a_s'Qa_h = 0 \).

The solution of order 1 to DGPCA for all \( i = 1, \cdots, M \) checks relations below

\[
\left(\sum_{i=1}^{M} \pi_i V_{X_i}\right)Qa_1 = \lambda_1 a_1
\]

\[
a_{i,1} = P_{X_{i,0}}' a_1
\]

\( a_1 \) is the first eigenvector \( Q \)-normed associated in the greatest eigenvalue \( \lambda_1 \) of the matrix \( \left(\sum_{i=1}^{M} \pi_i V_{X_i}\right)Q \). The eigenvalues equations and eigenvectors can be still written

\[
V_{X_i}Qa_1 = \lambda_1 a_1
\]

where \( X_c = (\Delta_{\pi})^{\frac{1}{2}}X \) is the vertical multi-table and \( V_{X_c} = X_c'D_{bd}X_c \) the matrix of variance-covariances of \( X_c \). The relation with the eigen values and vectors preceding one is equivalent to

\[
W_{X_i}D_{bd}c_1 = \lambda_1 c_1
\]

where \( c_1 = X_cQa_1 \) is the principal component of the vertical multi-table \( X_c \) with \( W_{X_c} = X_cQX_c' \) the matrix of scalar products between individuals of \( X_c \) general elements \( \sqrt{\pi_i\pi_j}W_{X_iX_j} \) for all \( i = 1, \cdots, M \) and \( j = 1, \cdots, M \). The matrix \( W_{X_iX_j} = X_iQX_j' \) is the matrix of scalar inter-products between individuals of \( X_i \) and that of \( X_j \) let us point out it. The solution of order \( s \) of the DGPCA is determined by recurrence by relations

\[
\left(\sum_{i=1}^{M} \pi_i V_{X_{i,s-1}}\right)Qa_s = \lambda_s a_s
\]

\[
a_{i,s} = P_{X_{i,s-1}}' a_s
\]

where \( X_{i,s-1} = X_{i,s-2}P_{a_{i,s-2}}^\perp \) the residue to regression of \( X_{i,s-2} \) in the vector \( a_{i,s-1} \) in order \( s \) with \( X_{i,0} = X_i \) for all \( i = 1, \cdots, M \) and all \( s = 1, \cdots, \dim(E_i) \). Such, vectors \( \frac{a_{i,s}}{\|a_{i,s}\|_Q} \) for all \( s = 1, \cdots, \dim(E_i) \) and \( i = 1, \cdots, M \) constitute \( Q \)-orthonormed bases. By making the same proof that Casin (1996), one shows that the vectors \( a_s \) are \( Q \)-orthonormed, which enables to have a total representation of individuals in this system of vectors.
4 Comparison of methods

The comparison of those will be done on the basis of three elements: the criterion, the principle of the methods and the solution.

On the level of the criterion, methods of the first group (the tables have the same individuals). These methods in general maximize functions under constraints by determining initially a compromise (variable auxiliaries) and afterwards, one finds the linear combinations associated with each group of variables. These methods are founded on the matrix of inter-covariances between variables of two tables taken two by two. For the methods of the second group (the tables have the same variables), it is a question in general of maximizing under constraint a criterion containing a common vector of representation of all the tables, and afterwards, one determines the linear combinations associated with each table. These methods are based on the matrix of the scalar inter-products between individuals of two tables taken two by two.

On the level of the principle of these methods (all two groups of methods confused), it is a question of establishing orthonormed bases in vector spaces and or proper subspaces. In the plan of the solution, the methods of the first group in general determines the auxiliary variables $z_s$ by eigenanalysing the weighted sum of the operators $W_{X_s}D$ between individuals of tables $X_i$. For all these methods the solutions of order 1 are in general the same ones, the difference is at level of the solution of an order equal to or higher than two. The methods of the second group in general determines the vectors common to all the tables $a_s$ by diagonalising the weighted sum of the operators $V_{X_s}Q$ between variables of the tables $X_i$. For these methods also, the solutions of order 1 are the same ones, the difference is with the solution of an order equal to or higher than two. Ones noticed that, the solutions of the dual methods can be obtained by a simple adaptation of the formulas (it is enough to modify the context).

5 Example of application

For reasons which were evoked at the end of section 1, we chose to apply the dual co-inertia analysis (DCIA) to data analyzing the variability of the performance of the models of cars: CITROEN and BMW. On these models partitioned in two groups (CITROEN and BMW marks), the same variables (cylindered, power, length, breadth, weight, speed) were measured. The data can be found in Internet site ”www.motorlegend.com”. The models and their abbreviations are consigned in table 1. The objective of this example is to identify inside each mark the model of car which shows good characteristics and which can induce an interest for the users. The analysis will be articulated in two points: global analysis and partial analysis. We will make the
principal components analysis of the table $X$ regrouping the two marks of car to have highlights of the proximities of models of two marks of car and the relationships between the various variables. $X$ is the vertical multi-table made up of the superposition of under tables $X_1$ and $X_2$ respectively associated with the two marks. It is supposed that table $X$ is centered and scale. The two axles of the PCA of table $X$ (global analysis) have a very high percentage of inertia explained 90 percent, which makes it possible to retain only these two axles for the representation of the individuals and the variables (Figure 1) and (Figure 2). It comes out from this representation that all the variables are strongly correlated between them in general (Table 2) and with the first axle (effect of size), or, the models are organized on this axle by order of importance, of most powerful with the least powerful. We apply to this play of the data (Figures 3 to 10), the two dual co-inertia analyses (DCIA1 and DCIA2). We notice that the results which are provided by the PCA find also rather well at the partial level for the two co-inertia analyses. The difference is simply a problem of sign on the level of the DCIA2, but restored information is the same one for two methods (DCIA1 and DCIA2). It spring from the matrix of the correlations of table $X_1$ associated with CITROËN (Table 3) with correlations strong between variables: breadth and length, breadth and cylindered, weight and cylindered, weight and length, weight and breadth. One notes also a correlation strong between the speed and the power. For the matrix of correlations of table $X_2$ associated with BMW (Table 4), one in general notes correlations strong between variables. At the partial level of table $X_1$ and methods (DCIA1 and DCIA2) (cf figures (3, 4, 7 and 8)), apart from the variables power and speed, the other variables are strongly correlated with axle 1. Let us constate that model CDB is characterized in general by all variables. The axle 2, is an axe of power and speed, models CC1 and CDV are most powerful and fastest, contrary to models CDC and CDR. Concerning the table $X_2$ and for two methods (DCIA1 and DCIA2) to refer to figures (5, 6, 9 and 10). All variables are correlated with axle 1 (effect of size). It opposes the good models of bad models. Models BMX5 and BMX4 for mark BMW are most interesting. On the other hand models BMS3 and BMS4 are not good. The axle 2 is characterized by variables speed and length, thus, the model BMS5 is very fast, but, model BMM is very long. The analysis ”inter-groupe” (between the marks) reveals that, models CDB and BMX5 are very close from the point of view of the characteristics; it is the same for models CDV, CC1, BMS3 and BMS5.
Table 1: List models and abbreviations

<table>
<thead>
<tr>
<th>Models</th>
<th>Abbreviations</th>
</tr>
</thead>
<tbody>
<tr>
<td>CITROEN DS3 1.6THP</td>
<td>CD16</td>
</tr>
<tr>
<td>CITROEN DS5 BLUEHDI</td>
<td>CDB</td>
</tr>
<tr>
<td>CITROEN DS3 CABRIO 1.6THP</td>
<td>CDC</td>
</tr>
<tr>
<td>CITROEN DS3 CABRIO 1.6THP</td>
<td>CDC16</td>
</tr>
<tr>
<td>CITROEN CS3 1.6THP</td>
<td>CC1</td>
</tr>
<tr>
<td>CITROEN DS3 RACING 1.6</td>
<td>CDR</td>
</tr>
<tr>
<td>CITROEN DS3 1.6 VTRI</td>
<td>CDV</td>
</tr>
<tr>
<td>BMW X4</td>
<td>BMX4</td>
</tr>
<tr>
<td>BMW Serie 4</td>
<td>BMS4</td>
</tr>
<tr>
<td>BMW M2</td>
<td>BMM</td>
</tr>
<tr>
<td>BMW X5</td>
<td>BMX5</td>
</tr>
<tr>
<td>BMW Serie 3</td>
<td>BMS3</td>
</tr>
<tr>
<td>BMW Serie 5</td>
<td>BMS5</td>
</tr>
</tbody>
</table>

Table 2: Matrix of correlations of table X.

<table>
<thead>
<tr>
<th></th>
<th>cylindered</th>
<th>power</th>
<th>length</th>
<th>breadth</th>
<th>weigth</th>
<th>speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>cylindered</td>
<td>1.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>power</td>
<td>0.966</td>
<td>1.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>length</td>
<td>0.663</td>
<td>0.567</td>
<td>1.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>breadth</td>
<td>0.733</td>
<td>0.601</td>
<td>0.838</td>
<td>1.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>weigth</td>
<td>0.928</td>
<td>0.859</td>
<td>0.851</td>
<td>0.860</td>
<td>1.</td>
<td></td>
</tr>
<tr>
<td>speed</td>
<td>0.625</td>
<td>0.743</td>
<td>0.563</td>
<td>0.428</td>
<td>0.612</td>
<td>1.</td>
</tr>
</tbody>
</table>
Table 3: Matrix of correlations of table $X_1$.

<table>
<thead>
<tr>
<th></th>
<th>cylindered</th>
<th>power</th>
<th>length</th>
<th>breadth</th>
<th>weight</th>
<th>speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>cylindered</td>
<td>1.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>power</td>
<td>-0.007</td>
<td>1.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>length</td>
<td>0.866</td>
<td>-0.268</td>
<td>1.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>breadth</td>
<td>0.940</td>
<td>-0.190</td>
<td>0.984</td>
<td>1.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>weight</td>
<td>0.974</td>
<td>-0.142</td>
<td>0.925</td>
<td>0.971</td>
<td>1.</td>
<td></td>
</tr>
<tr>
<td>speed</td>
<td>0.129</td>
<td>0.985</td>
<td>0.139</td>
<td>-0.054</td>
<td>0.0017</td>
<td>1.</td>
</tr>
</tbody>
</table>

Table 4: Matrix of correlations of table $X_2$.

<table>
<thead>
<tr>
<th></th>
<th>cylindered</th>
<th>power</th>
<th>length</th>
<th>breadth</th>
<th>weight</th>
<th>speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>cylindered</td>
<td>1.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>power</td>
<td>0.987</td>
<td>1.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>length</td>
<td>0.222</td>
<td>0.156</td>
<td>1.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>breadth</td>
<td>0.920</td>
<td>0.887</td>
<td>0.548</td>
<td>1.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>weight</td>
<td>0.897</td>
<td>0.870</td>
<td>0.595</td>
<td>0.983</td>
<td>1.</td>
<td></td>
</tr>
<tr>
<td>speed</td>
<td>0.591</td>
<td>0.698</td>
<td>-0.271</td>
<td>0.455</td>
<td>0.417</td>
<td>1.</td>
</tr>
</tbody>
</table>
Figure 1: Representation of variables of $X$

Figure 2: Representation of individuals of $X$
Proposal of analyses: DCA, DCIA, DGCA and DGPCA

Figure 3: Representation of variables of $X_1$ with DCIA1

Figure 4: Representation of individuals of $X_1$ with DCIA1
Figure 5: Representation of variables of $X_2$ with DCIA1

Figure 6: Representation of individuals of $X_2$ with DCIA1
Proposal of analyses: DCA, DCIA, DGCA and DGPCA

Figure 7: Representation of variables of $X_1$ with DCIA2

Figure 8: Representation of individuals of $X_1$ with DCIA2
6 Conclusion

We have just proposed in this article the dual ones of some factorial methods of analysis having the same lines. These duals are based on a common matrix, the matrix of the scalar inter-products between tables making it possible to describe the proximities between individuals of tables taken two to two if the
tables have the same variables. This matrix is dual matrix of inter-covariances between two tables taken two to two if the tables have the same rows. It is at the base of all the theory of the duality in factorial analysis of the data. But the use of this metric in certain methods limits their use. From where need for adopting the regularized methods. It is noted that, the solutions of the dual methods can be obtained by a simple adaptation of the solutions of the direct methods. If the tables have the same rows and the same variables (cubic tables) both approaches are applicable.

References


https://doi.org/10.1080/03610920903140114


https://doi.org/10.1007/bf02294050

Received: January 30, 2017; Published: March 24, 2017