Available online at www.sciencedirect.com
SCIENCE DIRECT
Computational Statistics & Data Analysis 45 (2004) 481 – 503
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

A principal axes method for comparing contingency tables: MFACT☆

Mónica Bécue-Bertaut[a,*], Jérôme Pagès[b]

[a] Universitat Politècnica de Catalunya, Edif. FME, Pau Gargallo, 5 Estadistica i Inv. Operativa, 08028 Barcelona, Spain
[b] ENSA/INFSA, Rennes, France

Received 1 April 2001; received in revised form 12 December 2002

Abstract

A new methodology is introduced for comparing the structures of several contingency tables. The latter, built up from different samples or populations, present the same rows and different columns (or vice versa). This methodology combines some aspects of principal axes methods (global maximum dispersion axes), canonical correlation techniques (canonical dispersion axes) and Procrustes analysis (superimposed representations) but takes into account the particularities of contingency tables in order to extend correspondence analysis to multiple contingency tables. Two main problems arise: the differences between the margins of the common dimension and the need for balancing the influence of the different tables in global processing. A study of the four structures induced on Spanish regions by mortality causes (by gender) and by age distribution (by gender), in conjunction, will illustrate the methodology.

Keywords: Principal axes methods; Multiple contingency tables; Correspondence analysis; Principal component analysis; Generalised canonical analysis; Procrustes methods

1. Introduction

The methodology presented here is for comparing the structures of several contingency tables with the same rows and different columns (or vice versa), generally built up from different samples or populations. By structure of a table, we mean the relationship between the rows and the columns as expressed in correspondence analysis (CA)

☆ Research was partially supported by the Catalan DGR, grant 1998BEAI400100.
* Corresponding author. Tel.: +34-934017031; fax: +34-934015855.
E-mail address: monica.becue@upc.es (M. Bécue-Bertaut).

0167-9473/$ - see front matter
doi:10.1016/S0167-9473(03)00003-3

through graphical displays (Benzécri, 1973; Escofier and Pagès, 1988–1998; Gower and Hand, 1996; Greenacre, 1984; Lebart et al., 1984; Lebart et al., 1998). The comparison has to deal with both rows and columns, that is to say, the structure induced by the rows over the different sets of columns, and also the structure induced over the rows by the different sets of columns.

If we adopt a CA-like approach, several methods are available for analysing a set of contingency tables. However, most of them deal with three-way contingency tables (Carlier and Ewing, 1992; Carlier and Kroonenberg, 1996, 1998; Kiers, 1989) and thus are not applicable to the case in which the columns are not the same throughout the tables.

The CA of all the tables juxtaposed row wise (Benzécri, 1982; Cazes, 1980; VanderHeijden, 1987) allows a comparison, in a single framework, of the different sets of columns but not of the rows, described only by the columns as a whole. Furthermore, when the tables have no proportional column margins, CA does not actually compare the internal structures of the tables, because of the different centroids.

Benzécri (1983) and Escofier and Drouet (1983) proposed a generalisation of CA, itself generalised by Cazes and Moreau (1991, 2000) which they called internal correspondence analysis (ICA) and which solves the problem by centring the subtables on their own margins. However, this method only represents the rows from a global point of view, that is to say the set of tables as a whole.

The methods based on canonical correlation analysis tackle a similar problem that should be solved, when dealing with different groups of variables observed in one set of individuals. A common structure is highlighted through a set of correlated canonical variables, one per group, which is a linear combination of the variables of its group. The canonical correlation technique (Hotelling, 1936), concerned with only two groups of variables, has been generalised to various groups from different points of view: Horst's generalised canonical correlation analysis (Horst, 1961), Carroll's generalised canonical correlation analysis (Carroll, 1968), STATIS (Escoufier, 1985; Lavit, 1988), multiple factor analysis (MFA) (Escofier and Pagès, 1988–1998, 1994), PLS path modelling (Lohmöller, 1989) and GPCA (Casin, 2001). Along similar lines, generalised Procrustes analysis (GPA, Gower (1975, 1984)) deals with the comparison of systems of distances between individuals as induced by various groups of variables using superimposed representations of the individuals as described by each group.

We adopt the point of view of MFA (Escofier and Pagès, 1988–1998, 1994; Pagès and Tenenhaus, 2001; Pagès and Husson, 2001), which combines the principal component analysis approach (global maximum dispersion axes), the canonical correlation analysis approach (canonical dispersion axes)—in such a way that the canonical variables take into account the correlation structure of the group that they summarise—and the Procrustes methods characteristics (superimposed representations) in order to introduce an extension of correspondence analysis combined with MFA characteristics and thus compare several contingency tables. For a comparison between MFA and Caroll's generalised canonical correlation analysis, see Escofier and Pagès (1988–1998, pp. 162–164), Escofier and Pagès (1994) and Pagès and Tenenhaus (2001). After the notation is introduced (Section 2), we briefly recall some CA properties (Section 3) and present the methodology that we call MFA for contingency tables or MFACT

(Section 4). The properties of MFACT are developed (Section 5) and an application is presented (Section 6).

2. Notation

2.1. A single contingency table X

Let us consider a two-way contingency table X of order $I \times J$. Its general term f_{ij} is the relative frequency or proportion with which row i ($i=1,\dots,I$) is associated with column j ($j=1,\dots,J$) such that $\sum_{ij} f_{ij}=1$. We denote the ith row margin term as $f_{i.}=\sum_j f_{ij}$. D_I is the diagonal matrix with general term $f_{i.}$. The column margin is $f_{.j}=\sum_i f_{ij}$ ($j=1,\dots,J$), and D_J is the diagonal matrix with general term $f_{.j}$.

When inspecting a contingency table, data are considered through conditional proportions, usually called *profiles*. The row-i profile is $\{f_{ij}/f_{i.}; j=1,\dots,J\}$ and the column-j profile is $\{f_{ij}/f_{.j}; i=1,\dots,I\}$. The mean row profile, as calculated by weighting the rows using their margin sum $f_{i.}$, is $\{f_{.j}; j=1,\dots,J\}$. Symmetrically, the mean column profile is $\{f_{i.}; i=1,\dots,I\}$.

2.2. Juxtaposed contingency tables

T tables $X_1,\dots,X_t,\dots,X_T$, are juxtaposed into a two-way table (Fig. 1). Columns are not the same throughout the tables. Table X_G is this "global" table, of dimension $I \times J$, and (*sub*)*table* X_t, of dimension $I \times J_t$, the tth (sub)table. We denote f_{ijt} the relative frequency or proportion, in table t ($t=1,\dots,T$), with which row i (=category $i; i=1,\dots,I$) is associated with column j (=category $j; j=1,\dots,J_t$; $\bigcup_t J_t=J$). $\sum_{ijt} f_{ijt}=1$.

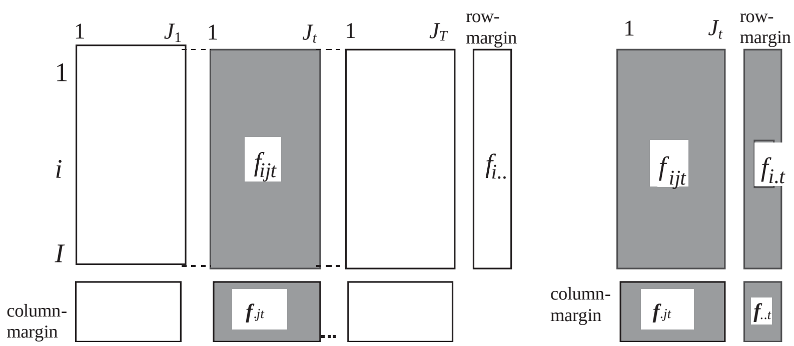

Fig. 1. Multiple contingency table: notations. T contingency tables, having the same rows, are juxtaposed. On the left, the global table and margins. On the right, the table t and margins. In the example, Section 6, $T=4$; $I=17$; $J_1=14$, $J_2=14$, $J_3=7$, $J_4=7$.

We denote the row margin of table X_G as $f_{i..} = \sum_{jt} f_{ijt}$. The column margin of table X_G is $f_{.jt} = \sum_i f_{ijt}$. The row margin of table t, as a subtable of table X_G, is $f_{i.t} = \sum_j f_{ijt}$, and the sum of the terms of table t inside table X_G is $f_{..t} = \sum_{ij} f_{ijt}$.

3. Correspondence analysis concepts

In the following sections, we look at the main concepts and properties of classical CA and the wide-ranging possibilities opened up by its generalisation to models other than the independence model.

3.1. Correspondence analysis

Here we summarise classical CA, that is to say the variant of CA that leads to superimposed displays of approximations to χ^2 distance between rows and between columns. Other variants of CA are presented and compared with the classical one in Israëls (1987) and Gower and Hand (1996). For non-symmetric correspondence analysis (NSCA), in which the variables do not play a symmetric role, see D'Ambra and Lauro (1989).

3.1.1. Geometric approach and χ^2 distance

Correspondence analysis is concerned with two-way contingency tables through a geometric approach to the set of the rows (and symmetrically, to the set of columns). For a detailed presentation and proofs, see Benzécri (1973), Escofier and Pagès (1988–1998), Greenacre (1984), Gower and Hand (1996), Lebart et al. (1984, 1998), Nishisato (1980).

Each row i of matrix X, of order $I \times J$, is considered as a point in R^J (without any loss of generality, we can assume $J < I$) with coordinates $\{f_{ij}/f_{i.}, j = 1, \dots, J\}$ and weight $\{f_{i.}, i = 1, \dots, I\}$; the centroid of the rows set is the point $\{f_{.j}, j = 1, \dots, J\}$. The proximities between rows are measured using the χ^2 distance. So, the square distance between rows i and l is

$$d^2(i,l) = \sum_{j \in J} \frac{1}{f_{.j}} \left(\frac{f_{ij}}{f_{i.}} - \frac{f_{lj}}{f_{l.}} \right)^2.$$

Symmetrically, each column j is considered as a point in R^I, the coordinates of the columns and the distances between columns are obtained by exchanging the role of the index i and j in the former paragraph.

Consistently with the geometric approach, the dispersion of the set of rows (and symmetrically, of the set of columns) around its centroid is measured through the inertia, also called, in a statistical context, Pearson's mean square contingency coefficient:

$$\Phi^2 = \sum_i f_{i.} d^2(i, Centroid) = \sum_j f_{.j} d^2(j, Centroid) = \frac{\chi^2}{n}$$

being n the total effective of the counted up units.

CA inspects the dispersion structures of both rows and columns sets, that is to say the distances between row profiles as a whole (and symmetrically, between columns profiles), or equivalently, the distances between each profile and the mean profile. Consequently, CA describes the discrepancy from the independence model, mainly by displaying approximations to the distances between rows (and symmetrically, between columns) on the axes of maximum dispersion, also called principal axes.

3.1.2. CA as a particular PCA

The principal axes search can be obtained by performing a principal component analysis (PCA) on the table Y whose general term is the residual with respect to the independence model (weighted by the inverse of the cross product of the row and column marginal sums):

$$y_{ij} = \frac{f_{ij} - f_{i.} \cdot f_{.j}}{f_{i.} \cdot f_{.j}}.$$

This PCA uses the diagonal matrix D_I as row weights and metric in the column space and the diagonal matrix D_J as column weights and metric in the row space (Escofier and Pagès, 1988–1998, pp. 95–97).

By applying the usual formulae of PCA, the formulae and properties of CA are easily deduced. We recall some of them below.

3.1.3. Principal coordinates of rows and columns

In the row space, the inertia axis with rank s corresponds to the eigenvectors u_s, ($\|u_s\|_{D_J} = 1$) of the matrix $Y'D_IYD_J$, associated with the eigenvalues λ_s (in decreasing order). In the column space, the inertia axis with rank s corresponds to the eigenvectors v_s, ($\|v_s\|_{D_I} = 1$) of the matrix $YD_JY'D_I$, associated with the same eigenvalues λ_s (in decreasing order). The vectors of the row scores are $F_s = YD_Ju_s = \sqrt{\lambda_s}v_s$ and the vectors of column scores are $G_S = Y'D_Iv_s = \sqrt{\lambda_s}u_s$. These scores, or principal coordinates (Greenacre, 1984), lead to both sets of distances (those between rows and those between columns) corresponding to the χ^2 distances defined above. In the following, we denote the vector of the principal coordinates on axis s either of the rows (denoted by F_s) or of the columns (denoted by G_s) as the principal component with rank s, in accordance with our PCA-like approach to correspondence analysis.

The same result is obtained by performing the generalised SVD of matrix Y in the metrics mentioned above:

$$Y = V\Lambda^{1/2}U'$$

with $V'D_IV = Id$, $U'D_JU = Id$, where Id is the unity matrix.

The general term of the diagonal matrix $\Lambda^{1/2}$ is $\sqrt{\lambda_s}$; the columns of V are the vectors v_s and the columns of U are the vectors u_s as defined above.

With this notation, the score matrices for the rows and for the columns are, respectively, $F = V\Lambda^{1/2}$ and $G = U\Lambda^{1/2}$.

3.1.4. Transition formulae and supplementary elements

Eqs. (1) (usually known as transition formulae but also as quasi-barycentric coordinates) allow the transition from the set of row coordinates to the set of column coordinates (and vice-versa).

$$F_s(i) = \frac{1}{\sqrt{\lambda_s}} \sum_j \frac{f_{ij}}{f_{i.}} \cdot G_s(j), \quad G_s(j) = \frac{1}{\sqrt{\lambda_s}} \sum_i \frac{f_{ij}}{f_{.j}} \cdot F_s(i). \tag{1}$$

An important application of the transition formulae is that of projecting new (supplementary) rows or columns on CA displays.

3.1.5. Data reconstitution

Matrices Y and X can be reconstituted from coordinates and eigenvalues:

$$\frac{f_{ij} - f_{i.} \cdot f_{.j}}{f_{i.} \cdot f_{.j}} = \sum_s \frac{F_s(i) G_s(j)}{\sqrt{\lambda_s}},$$

$$Y = \sum_s \frac{1}{\sqrt{\lambda_s}} F_s G_s' = \sum_s \sqrt{\lambda_s} v_s u_s', \tag{2}$$

$$\frac{x_{ij}}{n} = f_{ij} = f_{i.} \cdot f_{.j} \left(1 + \sum_s \frac{F_s(i) G_s(j)}{\sqrt{\lambda_s}} \right),$$

$$\frac{X}{n} = D_I \left(1_{IJ} + \sum_s \frac{F_s G_s'}{\sqrt{\lambda_s}} \right) D_J \tag{3}$$

being 1_{IJ} the matrix $I \times J$ filled with 1.

3.1.6. CA as a minimisation problem

The restriction of formula (2) to its first S terms corresponds to an approximation of Y by a matrix with rank S, through a weighted least squares criterion, giving the cell i, j the weight $f_{i.} f_{.j}$ (Escofier and Pagès, 1988–1998, pp. 104–106). The squared norm of the difference between Y and its rank S approximation is equal to the sum of the eigenvalues with rank superior to S.

3.2. CA relative to a general model with imposed metrics

Escofier (1983, 1984) (see also Van der Heijden et al., 1989) introduces a generalisation of CA to any model M in any diagonal metrics P (whose general term is p_i) and Q (whose general term is q_j). The general term m_{ij} of M is the expected value of f_{ij} under the model considered.

Performing the CA of X with respect to the model M in metrics P and Q is equivalent to performing a PCA, in metrics P and Q, on the matrix whose general term is

$$\frac{f_{ij} - m_{ij}}{p_i q_j}.$$

This generalisation leads to a large range of applications allowing residuals with respect to very varied models to be decomposed, while retaining CA properties. Specifically, Van der Heijden et al. (1989) show how generalised CA can be seen as a flexible technique to decompose residuals from very diverse log-linear models and allow to analyse the differences between two log-linear models. They propose a combine approach using both CA and log-linear analysis (LLA), as complementary methods. Firstly, LLA detects relationship between variables by the way of significant interactions. Secondly, generalised CA is applied either to explore the residuals of the chosen model, or to visualise interactions intentionally not included in the model so as to make easier their interpretation in case of large multiway contingency tables.

4. MFA for contingency tables (MFACT)

We aim to extend CA to a set of contingency tables in such a way that:

- As in MFA, the dispersion structures common to all the tables are highlighted by means of global and canonical axes; the global axes correspond to directions of maximum inertia of the global table and the canonical axes correspond, as far as possible, to directions of subtables both of high variance and highly related to global axes (Pagès and Tenenhaus, 2001).
- As in CA, the structure of each subtable corresponds to the discrepancy of the independence model as given by its own margins.
- As in CA, the metrics and weights are induced by the row and column margins.

These considerations determine the model with which the global table has to be compared, and the metrics/weights for the rows and the columns.

4.1. Intra-tables independence model

When performing extended CA on the global table, the model must juxtapose the models used in the separate CA, that is to say the intra-tables independence model M of order $I \times J$ whose general term is

$$m_{ijt} = \left(\frac{f_{i.t}}{f_{..t}}\right) \cdot \left(\frac{f_{.jt}}{f_{..t}}\right).$$

4.2. Metrics in row and column spaces

4.2.1. Row weights (and metric in column space)

The row weights (and metric D_{I_T} in column space) are the row margins as calculated for the whole table $\{f_{i..}, i = 1, \dots, I\}$, which are also the mean of the row weights in the separate CA, weighted by the subtable counts. The differences between the row margins of the different tables induces a difficulty since each row has not the same weight in the separate CA of each table. The choice of $f_{i..}$, as the weight for row i, is a compromise which leads to several good properties described in the following. An other method for the simultaneous analysis of several contingency tables has recently been proposed (Zarraga and Goitisolo, 2001) which solves this difficulty in another way. The data are transformed so as to take into account, in each table t, weights and metrics of the separate CA of table t (in others words, the row weights are "included" in the coordinates). Thus, in this approach, the global analysis of the juxtaposed tables is "closer" from separate analysis than MFACT is. Nevertheless, this good feature induces the loss of some important properties: e.g. the representation of the set of rows is not centred. Furthermore, the weight of the rows being different from one table to another, the introduction of supplementary variables is problematic. Due to the importance of supplementary variables, the interest of a unique set of row weights (necessarily a compromise) is obvious.

4.2.2. Column weights (and metric in row space)

The column weights have to be chosen in such a way that the influence of each column group could be comparable in a global analysis, and so enhance the links between the tables. We must avoid that a single table t could contribute on its own to the construction of the first axis; nothing can be required for further axes because a multidimensional table will always influence more axes than a unidimensional one. In the case of contingency tables, the predominant role of table t can have two origins:

- a strong structure of table t, that is to say a strong relationship between rows and columns, which leads to high eigenvalues;
- a high global proportion ($f_{..t}$), which could be due, for example, to a high sample size, which would lead to high weights for its columns.

In a general study, these two elements are of major importance, but they must be removed in a global analysis.

To balance the influence of the subtables, we choose, as in MFA, to standardise to 1 the highest axial inertia of each subtable. Therefore, the weight of the jth column belonging to set t is the margin sum $f_{.jt}$, divided by λ_1^t, the first eigenvalue of CA performed or subtable t, but in metrics D_{I_T} (instead of D_I, because of the need for a common weight for the rows) and D_{J_t}, restriction of D_J to J_t.

So, the metric in the row space is the diagonal matrix D_{J_T} whose general term is $f_{.jt}/\lambda_1^t$. This over weighting of columns by $1/\lambda_1^t$ has several properties (Escofier and Pagès, 1988–1998, p. 151, pp. 157–158); in particular, the intra-tables structures are not modified and, except for very special cases, the first axis of the global analysis cannot be generated by a single table.

4.3. MFACT as a weighted PCA of the global table

MFA for contingency tables consists of PCA in metrics D_{I_T} and D_{J_T} of table Z whose general term is the weighted residual with respect to the intra-tables independence model:

$$\frac{f_{ijt}-(f_{i.t}/f_{..t})\cdot f_{.jt}}{f_{i..}f_{.jt}}=\frac{1}{f_{i..}}\left[\frac{f_{ijt}}{f_{.jt}}-\frac{f_{i.t}}{f_{..t}}\right]. \tag{4}$$

Note that Z is divided into T subtables, Z_t.

It is easy to verify that the weighted rows are centred, the weighted columns are centred and the weighted columns of each set t are also centred.

The subtables Z_t differ slightly from the subtables Y_t (Section 3.1.2) as they would have been built up to perform the separate CA on each table. In both cases, the general term is the residual with respect to the independence model, but weighted differently. In fact, the need for a common weight for the same row through all the subtables leads us to compare not the tables Y_t but the tables Z_t. The distortion inherent in the substitution of Y_t by Z_t, which is small if the row margins differ only slightly through the subtables, results from the necessary compromise in order to compare contingency tables with different margins. In what follows, we use the term "separate analysis" of table t to refer to that based on table Z_t.

4.3.1. *Global principal components*

From this PCA, also called global analysis, the global principal components are obtained in the usual way for columns and rows, F_s and G_s respectively; $s = 1, \ldots, \min(I-1, J-1)$.

4.3.2. *Canonical components*

F_s can be written as

$$F_s = \frac{1}{\lambda_s} \sum_t Z_t D_{J_t} Z_t' D_{I_T} F_s.$$

In MFACT, the canonical components, associated with the global principal components F_s in each group t, are defined in the following way:

$$F_s = \sum_t F_s^t,$$

$$F_s^t = \frac{1}{\lambda_s} Z_t D_{J_t} Z_t' D_{I_T} F_s. \qquad (5)$$

4.3.3. *Data reconstitution*

The classical PCA formulae lead to the following decomposition expressions:

$$Z = \sum_s \frac{1}{\sqrt{\lambda_s}} F_s G_s' = \sum_s \sqrt{\lambda_s} v_s u_s',$$

$$f_{ijt} = \left(\frac{f_{i.t}}{f_{..t}}\right) \cdot f_{.jt} + f_{i..} \cdot f_{.jt} \left(\sum_s \frac{F_s(i) G_s(j,t)}{\sqrt{\lambda_s}} \right). \qquad (6)$$

4.3.4. *MFACT as a minimisation problem*

The restriction of formula (6) to its first S terms corresponds to an approximation of Z by a matrix with rank S, through a weighted least-squares criterion, giving to the cell i, j, t the weight $f_{i..} f_{.jt} / \lambda_1^t$. As in CA, this property follows from a general PCA property (Escofier and Pagès, 1988–1998, pp. 104–106). The squared norm of the difference between Z and its rank S approximation is equal to the sum of the eigenvalues with rank superior to S.

5. Properties of MFACT

5.1. Distances between rows and between columns

5.1.1. Distances between rows

The squared distance between rows i and l, calculated from coordinates given in (4) using the weight $f_{.jt}/\lambda_1^t$ for the column j of set t, is

$$d^2(i,l) = \left[\sum_t \frac{1}{\lambda_1^t} \sum_{j\in J_t} \left(\frac{f_{ijt}}{f_{i..}} - \frac{f_{ljt}}{f_{l..}}\right)^2 \cdot \frac{1}{f_{.jt}}\right] - \left[\sum_t \frac{1}{\lambda_1^t \cdot f_{..t}} \left(\frac{f_{i.t}}{f_{i..}} - \frac{f_{l.t}}{f_{l..}}\right)^2\right]. \tag{7}$$

In expression (7), disregarding weighting by the reverse of the first eigenvalue:

- the first term corresponds to the distance (between profiles i and l) in the CA of the juxtaposed tables;
- the second term corresponds to the distance (between profiles i and l) in the CA of the table containing the sums by row and by subtable. The general term $i.t$ in this table is the sum of row i in table t. We see here how this last table is neutralized by recentring each subtable on its own margins.

5.1.2. Proximities between columns

The squared distance between column j (belonging to table t) and column k (belonging to table r), calculated from coordinates given in (4), is

$$d^2(j\in t,\ k\in r) = \sum_i \frac{1}{f_{i..}} \left[\left(\frac{f_{ijt}}{f_{.jt}} - \frac{f_{i.t}}{f_{..t}}\right) - \left(\frac{f_{ikr}}{f_{.kr}} - \frac{f_{i.r}}{f_{..r}}\right)\right]^2, \tag{8}$$

$$d^2(j\in t,\ k\in r) = \sum_i \frac{1}{f_{i..}} \left[\left(\frac{f_{ijt}}{f_{.jt}} - \frac{f_{ikr}}{f_{.kr}}\right) - \left(\frac{f_{i.t}}{f_{..t}} - \frac{f_{i.r}}{f_{..r}}\right)\right]^2. \tag{9}$$

The proximities between columns can be interpreted as a similar association with rows.

Case 1: *the columns belong to the same table* ($t = r$).

The proximity between two columns is interpreted in terms of the similarity between profiles, exactly as in the usual CA.

Case 2: *the columns belong to different tables* ($t \neq r$).

The column profiles are relativised by the average profiles, as shown in the two expressions of the squared distance. Expression (8) shows that the profile of a column intervenes by its deviation from the mean profile of the corresponding table. Expression (9) shows how the differences between column profiles are relativised by the differences between mean profiles.

5.2. Global representation of rows and columns

The PCA performed in Section 4.3 provides a representation of the rows and of the columns that must be analysed together. Moreover, these representations can be superimposed in a biplot, as in correspondence analysis. This is the consequence of the transition formulae which relate coordinates of rows and of columns, as examined hereafter.

5.2.1. Transition formulae and interpretation rules

Rows among columns: The relation giving (along the s-axis) the coordinate $F_s(i)$ of row i from the coordinates $\{G_s(j,t); j=1,\dots,J_t; t=1,\dots,T\}$ of columns is

$$F_s(i)=\frac{1}{\sqrt{\lambda_s}}\sum_t\sum_{j\in J_t}\frac{f_{.jt}}{\lambda_1^t}\left[\frac{1}{f_{i..}}\left[\frac{f_{ijt}}{f_{.jt}}-\frac{f_{i.t}}{f_{..t}}\right]\right]G_s(j,t).$$

The centroid of the column profiles belonging to J_t being 0 (see Section 4.3), we have

$$\sum_{i\in J_t} f_{.jt}G_s(j,t)=0.$$

Using this property, the transition formula can be written as follows

$$F_s(i)=\frac{1}{\sqrt{\lambda_s}}\sum_t\frac{1}{\lambda_1^t}\frac{f_{i.t}}{f_{i..}}\left[\sum_{j\in J_t}\frac{f_{ijt}}{f_{i.t}}G_s(j,t)\right]. \tag{10}$$

Except for a constant, each row lies in the centroid of the columns associated with this row.

Globally, a row is attracted by the columns with which it is associated.

Columns among rows: The expression (along the s-axis) for the coordinate $G_s(j,t)$ of column j,t from the coordinates $\{F_s(i),\ i=1,\dots,I\}$ of rows is

$$G_s(j,t)=\frac{1}{\sqrt{\lambda_s}}\left[\sum_i\left(\frac{f_{ijt}}{f_{.jt}}-\frac{f_{i.t}}{f_{..t}}\right)F_s(i)\right]. \tag{11}$$

As the coefficient of $F_s(i)$ can be negative, the columns are not in the centroid of the rows, except when the row weights are the same in all the tables.

This coefficient measures the discrepancy between the profile of columns j,t and the column margin of table t. A column is attracted (or repelled) by the rows that are more (or less) associated with it than if there were independence between rows and columns in table t.

5.3. Superimposed representations of rows through canonical components

In order to compare the structures of the rows induced by the different tables, it is possible to use the canonical components F_s^t, as given by (5), to obtain a superimposed representation of these structures, called partial structures.

In practice, the canonical components of the rows, also called partial components, can be calculated very easily. Therefore, the rows of the separate tables t are completed by zeroes, in such a way that they have the same number of columns as the global table, and projected, as supplementary rows, on the global axes. In this way, we superimpose the representations of the rows associated with each table t (*called partial rows*) and the representation associated with the global table X_G (*called global or mean rows*).

The superimposed representation benefits from CA properties. In particular, these partial representations can be related to column representation by means of a "restricted" transition formula:

$$F_s^t(i) = \frac{1}{\sqrt{\lambda_s}} \cdot \frac{f_{i.t}}{f_{i..}} \left[\sum_{j \in J_t} \frac{f_{ijt}}{f_{i.t}} \cdot \frac{G_s(j,t)}{\lambda_1^t} \right]. \qquad (12)$$

This expression is derived from (10) restricted to the columns of table t.

In the graph superimposing partial representations, these partial representations are dilated by the coefficient T (number of tables). Thus, a (mean) row point is located in the centroid of the corresponding partial row points.

5.4. Representation of the separate principal axis on the global axis

The separate principal components, that is to say the principal components calculated in separate CA but in metrics D_{I_T} and D_J, can be represented on the global axis by way of their correlations with the first two global principal components.

6. Application

Within the framework of a study requested by an insurance company, a first objective is to summarise and visualise the geographical distribution of the mortality causes among the autonomous regions of Spain in reference to the age structure variability, distinguishing by gender. It is also wished to link the mortality causes variability with social, economic and health indicators.

Spain is divided into 19 autonomous regions showing a great diversity in climatic, geographic and socioeconomic conditions, as well as dietetic habits.

The inequality of the regions faced with mortality is nowadays discussed from recent epidemiologic studies (Benach and Yasui, 1999; Benach et al., 2001b; Borrell and Pasarin, 1999), incipient domain in Spain. In particular, Benach et al. (2001a) build up an "Atlas of mortality in small areas in Spain (1987–1995)" from the standardised mortality ratios, specifically the global ratio concerning all the causes of mortality, and also the ratios concerning the ten principal mortality causes, different for men and women. These studies tackle the mortality causes one by one, looking for the associated risk factors very precisely.

Although the work is done at small area level, the authors notice that it makes sense to aggregate the results and summarise them at the region level and assess that the mortality risk is significantly great in South of Spain, as compared with North. Isolated northern areas of high mortality, as Galicia or Asturias are pointed out. To try to explain the differences between regions, these researchers emit the hypothesis of a

Table 1
Mortality causes as classified in the ICD

Cause	Description	ICD codes
1	Infectious and parasitic diseases	(001–139)
2	Neoplasms	(140–239)
3	Endocrine, nutritional and metabolic diseases, immunity disorders	(240–279)
4	Diseases of the blood and blood-forming organs	(280–289)
5	Mental disorders	(290–319)
6	Diseases of the nervous system and sense organs	(320–389)
7	Diseases of the circulatory system	(390–459)
8	Diseases of the respiratory system	(460–519)
9	Diseases of the digestive system	(520–579)
10	Diseases of the genitourinary system	(580–629)
11	Complications of pregnancy, childbirth, and the puerperium	(630–676)
12	Diseases of the skin and subcutaneous tissue	(680–709)
13	Diseases of the musculoskeletal system and connective tissue	(710–739)
14	Congenital anomalies	(740–759)
15	Certain conditions originating in the perinatal period	(760–779)
16	Symptoms, signs, and ill-defined conditions	(780–799)
17	Injury and poisoning	(800–999)

combination of social, work conditions and environment factors. Specifically, they note that increasing deprivation is associated with mortality risk, differently by both cause and gender.

Our approach is very different as it does not tackle the over mortality at all, but only the mortality distribution among the regions in order to highlight, for each region, the predominant causes and, for each cause, regions having high proportions among its own deaths. Our aim is an exploratory one, mainly to obtain a good description of the mortality causes profiles by region in reference to age structure.

6.1. Data

The most recent complete data, as compiled by Statistical Office of Spain (INE: Instituto Nacional de Estadística), correspond to year 1995. Only individuals over 20 are taken into account. Ceuta and Melilla, very small Spanish regions situated in North-Africa coast, are not considered.

Mortality data: the death counts are classified by cause, gender and main region of residence. The mortality causes are classified using the international classification of diseases (ICD). Here we use the more general level including 17 causes (Table 1). We neglect causes 14 (*congenital anomalies*) and 15 (*certain conditions originating in the perinatal period*) almost including children (given that we are studying only adult mortality), and also cause 11 (*complications of pregnancy*, *childbirth*, *and the puerperum*), which is very seldom (fewer than three cases per region are observed).

Table 2
χ^2 and eigenvalues of the four separate correspondence analyses

	Male mortality	Female mortality	male age structure	Fem. age structure
χ^2	1484.6	1858.9	99473.2	106990.0
λ_1	0.0038 (31%)	0.0058 (39%)	0.0058 (81%)	0.0059 (82%)
λ_2	0.0026 (21%)	0.0026 (18%)	0.0010 (14%)	0.0010 (14%)

Table 3
Correlations between principal components derived from the four separate correspondence analyses

		Male mortality Axis 1	Male mortality Axis 2	Fem. mortality Axis 1	Fem. mortality Axis 2	Male age struct. Axis 1	Male age struct. Axis 2
Fem. mortality	*Axis 1*	−0.26	−0.57				
	Axis 2	**−0.82**	0.46				
Male age struct.	*Axis 1*	**−0.87**	0.18	−0.09	**0.80**		
	Axis 2	0.29	0.28	**−0.83**	−0.10		
Fem. age struct.	*Axis 1*	**−0.85**	0.21	−0.13	**0.81**	**0.99**	0.10
	Axis 2	−0.38	−0.36	**0.86**	0.14	0.09	**−0.97**

The mortality data lead to two contingency tables, one for each gender. The row-classification variable is the region and the column-classification is the mortality cause.

Age structure data: we also know the distribution of the residents by age (seven categories: in 10 years intervals from 20 to 79, and 80 and over), gender and region. The counts are official estimates, calculated by projection since the previous census performed in 1991. The age data lead to two contingency tables, one for men and another for women. In each table, the row-classification variable is the region and the column-classification variable is the age interval.

Supplementary information: we also have information measuring economic and social development for the regions.

Finally, the overall table has 17 rows (regions) and (14 + 14 + 7 + 7) columns (mortality causes and age intervals, by gender).

6.2. Results

6.2.1. Results from separate analysis

The structure of regions as induced by the age distribution, for men and women separately, is two dimensional (Table 2), with a very dominant first dispersion axis. Male and female structures are almost identical as highlighted by the high correlations (0.99; −0.97) between their first two axes (having the same rank) (Table 3).

The structure of regions as induced by the mortality causes is more complex (five and four dispersion axes for men and women, respectively) although in both cases

Table 4
Decomposition by group of the inertia of the five first principal components

	Axis 1	Axis 2	Axis 3	Axis 4	Axis 5
Eigenvalue	3.23	1.82	0.94	0.68	0.47
Groups					
Male mortality	0.90 (27.8%)	0.56 (30.8%)	0.57 (60.4%)	0.31 (45.9%)	0.33 (71.4%)
Female mortality	0.40 (12.4%)	0.98 (54.1%)	0.34 (36.3%)	0.32 (47.9%)	0.11 (22.7%)
Male ages	0.97 (30.0%)	0.13 (7.1%)	0.02 (2%)	0.03 (3.5%)	0.02 (3.8%)
Female ages	0.96 (29.9%)	0.15 (7.9%)	0.01 (1.3%)	0.02 (2.7%)	0.01 (2.2%)

the first two axes clearly retain the largest part of the dispersion (52% and 56.7%) (Table 2). According to the difference between these sequences of eigenvalues, the whole mortality causes variability cannot be explained by age distribution variability. Furthermore, male and female structures present common elements but their comparison is difficult because the same rank axes do not coincide (see Table 3).

6.2.2. Results from global analysis

The sequence of eigenvalues suggests five interpretable axes, with two main directions (see Table 4).

For a given global axis, the amount of inertia due to the set of columns in a group (Table 4) is a measurement of the importance of this direction in this group. In this case, the first two principal axes are common to all mortality causes and age groups, although the second one is a dispersion direction of major importance only in mortality causes groups. Axes with rank 3 or higher are specific to mortality groups.

This result is essential as it already shows that a part of the variability of mortality causes follows a regional pattern similar to the age structure; the other part is clearly not linked to age (mainly with reference to the principal axes with rank higher than two).

The columns of the two mortality causes groups have an explained variance equal to 45.2% and 54.1% for men and women, respectively, on the first principal plane. The comparison with those obtained in the separate analysis (respectively, 52% and 57%, see Table 2) shows that the visualisation of all the groups in a same reference space is obtained with only a low loss in explained variance.

6.2.3. Structures displayed on the first principal plane

As in classical CA, the displays of the rows and of the columns (Figs. 3 and 4) can be superposed and must be interpreted in conjunction.

First axis: The first axis sorts the age intervals into their natural order. It opposes regions where young adults are in excess ("young regions": Canary Islands, Andalusia, Madrid) with regions where there are more elderly people than in the whole Spain ("old regions": Asturias, Aragon, Castile-Leon). For a given age, men and women have a similar distribution among the regions. The most important difference is relative to the two older categories: the oldest age intervals for men better characterise the old regions than the older age intervals for women.

On the other hand, this first axis opposes the most frequent mortality causes in the "young regions" (causes 3 "*endocrine disorders*" and 9 "*digestive system*" for men and women, causes 10 "*genitourinary system*" and 12 "*skin diseases*" for women only) with the most frequent mortality causes in the "old regions" (causes 5 "*mental disorders*", 13 "*musculoskeletal system*" and 16 "*ill-defined conditions*" for men and women, and cause 4 "*blood diseases*" for men only).

We would like to underline that the first principal component does not sort mortality causes depending on their association with age. For example, causes 10 and 12 concern elderly women but living in "young regions". Cause 17 "*Injury and poisoning*", most linked to young people, lies in a central position on the graphic because this cause is associated with regions which are very different as to their age distribution.

The first principal axis reflects features different of the age structure. "Young regions" retain (or attract) the young adults, that probably indicates a certain recent economic dynamism.

As shown in the following, these regions can be more or less developed, depending on their position on the second axis.

Second axis: The second axis opposes the regions with an excess of middle-aged (40–59 years old) people (mainly: the Basque country, Catalonia, Madrid, Navarre and Asturias) to the regions having a deficit of these categories (mainly: Extremadura, Andalusia and Castile-La Mancha).

This opposition goes hand in hand with the increase of certain causes in the first group of regions (causes 2 "*neoplasms*", 5 "*mental disorders*" and 6 "*nervous system*" for men and women, causes 4 "*blood diseases*" and 17 "*injury and poisoning*" for women, cause 3 "*endocrine disorders*" for men) and of other causes in the second group of regions (cause 7 "*circulatory system*" for men and women, cause 12 "*skin diseases*" for men, cause 3 "*endocrine disorders*" for women).

The second axis is linked to development and industrialisation variables, as shown by the high correlation between this axis and various socioeconomic indicators (Figs. 2–4). Female mortality variability is very linked this axis: thus, for women, causes 3 (which includes diabetes) and 7 (*circulatory systems*) decrease with development, and causes 17 (*injury and poisoning*), 2 (*neoplasms*) and other causes frequent in elderly women increase.

In fact, the first bisector approximately reflects the opposition "North-East/South-West" commented by Benach and Yasui (1999) as a result concerning global mortality.

Third, fourth and fifth axes: The following axes highlight more specific associations between causes and regions such as, the high incidence of cause 8 ("*respiratory system diseases*") for men and women in Galicia, and the high incidence of the same cause but only for men in Aragon and Asturias.

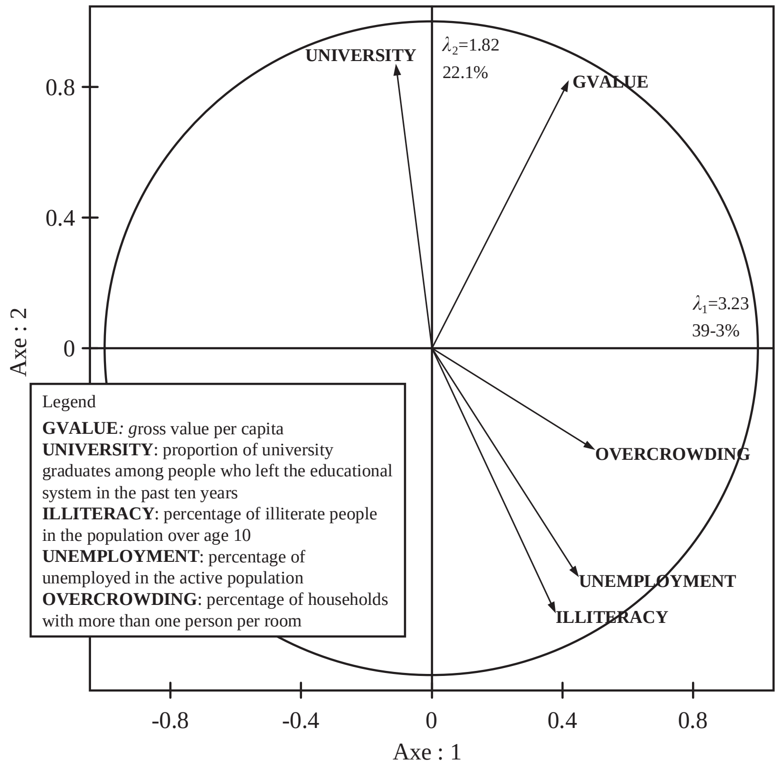

Fig. 2. Representation of supplementary variables through their correlations with the first two global principal components.

6.2.4. *Superimposition of partial structures*

In order to compare the structure of regions induced by mortality causes on one hand and by age distribution on the other hand, MFACT provides a superimposed representation of rows (see Section 5.3). Each region is represented by five points: two points per gender (one for mortality causes and the other for age distribution) and the global point, which is the centroid of the previous points (named partial points) (Fig. 5).

For each region, along the first axis, the points representing male and female age categories are very close to one another. This goes hand in hand with the fact that the first axis opposes regions both from mortality causes and age structure points of view. Some discrepancies can be observed from this global proximity between the four points referring to one region being highlighted. For example, Murcia and Andalusia are geographically close and also as regards to their age structure (partial points corresponding to groups 3 and 4). However, these two regions differ notably regarding the structure of men's mortality: only Andalusia is well characterised by male mortality causes associated with "young regions". Causes 3 "*endocrine and immunity disorders*"

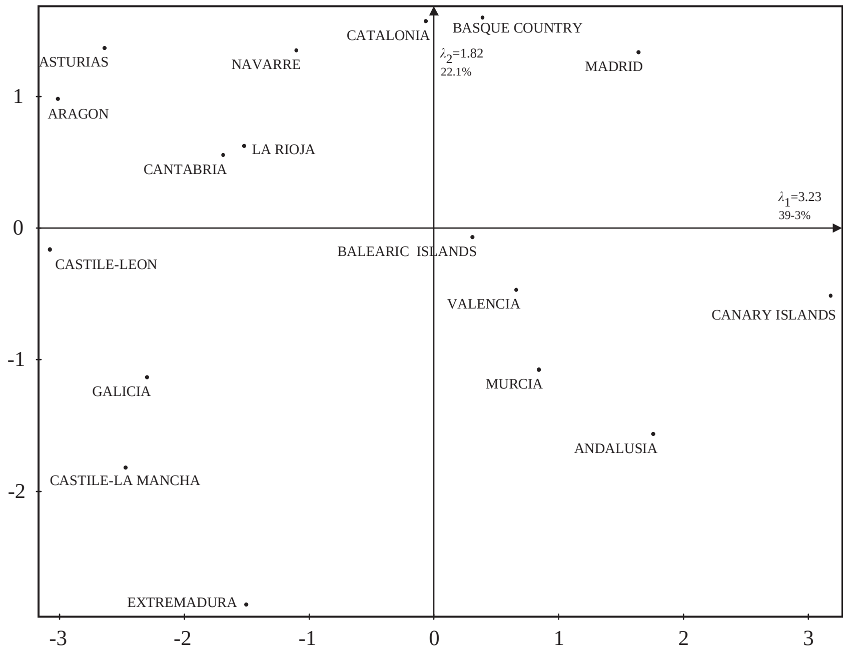

Fig. 3. Principal plan provided by MFACT: global representation of the regions.

and 9 "*digestive system*" sum up 11,75% of deaths in Andalusia, only 10.07% in Murcia (9.29% in all Spain).

At the opposite, Asturias is characterised by an excess of old men and women and by mortality causes linked to "old regions". This agreement exists between age structure and mortality causes structure.

An other interest of this representation is that the proximities and/or oppositions among the points representing the regions for only one given group can also be interpreted. For example, Extremadura, as described by female mortality, lies in a very extreme position on axis 2. In fact, causes 3 "*endocrine and immunity disorders*", 7 "*circulatory system*" and 16 "*ill-defined conditions*" for women, strongly associated with the less-developed regions, present a high frequency in this region: altogether they represent 59% of female mortality, whereas only 52% in Spain. The extreme position of this region is reinforced by the great deficit in female mortality causes associated with developed regions as causes 17 "*injury and poisoning*", 5 "*mental disorders*" and 6 "*nervous system*".

The high specificity of female mortality in Extremadura is highlighted because the whole of the causes intervene in the analysis, which allows to detect their cumulate effect, although each cause, taken one by one, can seem to have little importance. This specificity, which can be found at a lower level in Andalusia, was not underlined in the other mentioned studies.

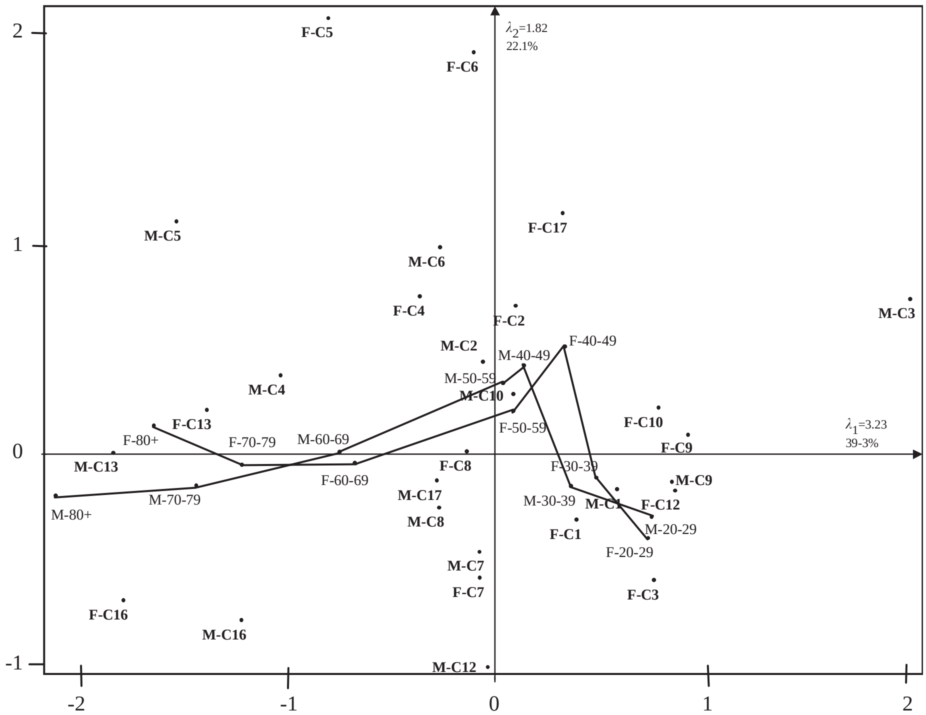

Fig. 4. Principal plan provided by MFACT: representation of the columns. Mortality causes in bold, age intervals in capital letters, M for males and F for females. Segments join the male and female age intervals in natural order.

6.3. Conclusions about the data analysis

The MFACT approach and the classical epidemiologic studies like the one by Benach et al. (2001a) are not competitive: they can be used as complementary methods as they bring very different information and ways to synthesise it.

The epidemiologic studies offer a synthesis using a geographic support. Mapping from MFACT is another kind of support which can more easily reflect the resemblance (both from mortality causes and age structure points of view) between regions not geographically close to one another. In this way, the information contained in the map is not used in the construction of the regions representation but only during its interpretation.

MFACT approach presents the classical advantages of multidimensional exploratory methods. So, it constitutes a flexible tool in order to obtain a very complete description of the variability of the causes mortality distribution among the regions in reference to age structure. It is possible to use other region characteristics (economic indicators, etc.) instead of age structure. Therefore, MFACT helps to detect both mortality causes linked to region characteristics and causes profiles not expected on the basis of the

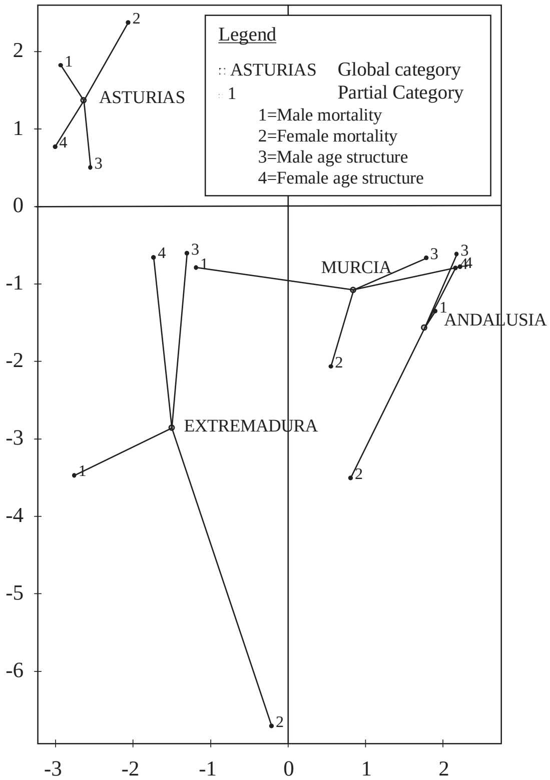

Fig. 5. Excerpt of the superimposed representation of global and partial structures on the first plan of MFACT. Here Fig. 3 is completed by the representation of global and partial points corresponding to Asturias, Andalusia, Murcia and Extremadura.

regions features. These descriptions are useful in the scope of planification insurance and sanitary policy.

7. Conclusion

The methodology described provides an original and operational point of view for the simultaneous analysis of several contingency tables having a common dimension (which includes, of course, those tables having two common dimensions).

Centering each table with respect to its row margin solves the problem of the distinct margins (corresponding to the common dimension) by way of a compromise between two contradictory requirements: to respect the intra-table structures in the sense it is given in correspondence analysis, and to give the same weight to each row in a simultaneous analysis of all the tables.

The "MFA" aspect of MFACT solves the problem of balancing the influence of the various subtables in a global analysis by way of a solution that is already well tested in the case of T tables individuals variables (quantitative and/or qualitative). This solution is now extended to the case of contingency tables having different row margins.

The MFA framework opens up the way for various extensions. A particularly interesting one is the possibility of comparing the structure of subtables of different natures in the same analysis. Thus, it is possible to add various groups of quantitative and/or qualitative variables describing the same rows to a set of contingency tables. The example presented here was not selected to emphasise this possibility, given that the development level indicators are only used a posteriori, as supplementary variables. However, since regions could also be described by different sets of economic or social quantitative indicators, this information can also be used.

These possibilities together constitute a complete methodology for exploratory analysis of a set of individuals, or clusters of individuals, described by data of different types.

8. Software

Note MFACT as described in this work was programmed by the second author in Fortran within ADDAD software and can be obtained from the authors as an independent (DOS) program upon request. It can also be performed using SPAD-V software (CISIA, Paris), provided that matrix and weights are suitably adapted.

References

Benach, J., Yasui, Y., 1999. Geographical patterns of excess mortality in Spain explained by two indices of deprivation. J. Epidemiol. Comm. Health 53, 423–431.

Benach, J., Yasui, Y., Borrell, C., Rosa, E., Pasarín, M., Benach, N., Nol, E.E., Martínez, J., Daponte, A., 2001a. Atlas of mortality in small areas in Spain (1987–1995). Universitat Pompeu Fabra, Barcelona.

Benach, J., Yasui, Y., Borrell, C., Sáez, M., Pasarin, M., 2001b. Material deprivation and leading causes of death by gender: evidence from a nationwide small area study. J. Epidemiol. Comm. Health 55, 239–245.

Benzécri, J., 1973. Analyse des Données, Vol. 1: Analyse des correspondances (Data analysis, vol. 1: Correspondence analysis). Dunod, Paris.

Benzécri, J., 1982. Sur la généralisation du tableau de burt et son analyse par bandes (about burt table generalisation and its analysis by strips). Les Cahiers de l'Analyse des Données 7 (1), 33–43.

Benzécri, J., 1983. Analyse de l'inertie intraclasse par l'analyse d'un tableau de contingence (intra-class inertia analysis through the analysis of a contingency table). Les Cahiers de l'Analyse des Données 8 (3), 351–358.

Borrell, C., Pasarin, M., 1999. The study of social inequalities in health in Spain: where are we?. J. Epidemiol. Comm. Health 53, 388–389.

Carlier, A., Ewing, G., 1992. Longitudinal analysis of contingency tables with application to demographic data. Comput. Statist. 7, 329–353.

Carlier, A., Kroonenberg, P., 1996. Biplots and decompositions in two-way and three-way correspondence analysis. Psychometrika 61 (2), 355–373.

Carlier, A., Kroonenberg, P., 1998. Three-way correspondence analysis, the case of the French cantons. In: Blasius, J., Greenacre, M.J. (Eds.), Visualization of Categorical Data. Academic Press, London, pp. 253–276.

Carroll, J., 1968. Generalization of canonical correlation analysis to three or more set of variables. In: Proceedings of the 76th Convention of the American Psychology Association, Vol. 3. pp. 227–228.

Casin, P., 2001. A generalisation of principal component analysis to *k* sets of variables. Comput. Statist. Data Anal. 35, 417–428.

Cazes, P., 1980. Analyse de certains tableaux rectangulaires décomposés en blocs (analysis of some rectangular tables decomposed into blocks). Les Cahiers de l'Analyse des Données 5 (2, 4), 145–161, 387–403.

Cazes, P., Moreau, J., 1991. Analysis of a contingency table in which the rows and the columns have a graph structure. In: Diday, E., Lechevallier, Y. (Eds.), Symbolic-numeric Data Analysis and Learning. Nova Science Publishers, New York, pp. 271–280.

Cazes, P., Moreau, J., 2000. Analyse des correspondances d'un tableau de contingence dont les lignes et les colonnes sont munies d'une structure de graphe bistochastique (correspondence analysis of a contingency table whose files and columns have a bistochastic graph structure). In: Moreau, J., Doudin, P.A., Cazes, P. (Eds.), L'analyse des correspondances et les techniques connexes. Approches nouvelles pour l'analyse statistique des données (Correspondence analysis and close techniques. New approaches for data statistical analysis). Springer, Berlin, Heidelberg, pp. 87–103.

D'Ambra, L., Lauro, N., 1989. Non symmetrical analysis of three-way contingency tables. In: Coppi, R., Bolasco, S. (Eds.), Multiway Data Analysis. North-Holland, Amsterdam, pp. 301–315.

Escofier, B., 1983. Analyse de la différence entre deux mesures sur le produit de deux mêmes ensembles (analysis of the difference between two measures on the two same sets product). Les Cahiers de l'Analyse des Données 8 (3), 325–329.

Escofier, B., 1984. Analyse factorielle en référence à un modèle: application à l'analyse d'un tableau d'échanges (principal axes method in reference to a model: application to the analysis of an exchanges table). Rev. Statist. Appl. 32 (4), 25–36.

Escofier, B., Drouet, D., 1983. Analyse des différences entre plusieurs tableaux de fréquence (analysis of the differences among several frequency tables). Les Cahiers de l'Analyse des Données 8 (4), 491–499.

Escofier, B., Pagès, J., 1988–1998. Analyses factorielles simples et multiples; objectifs, méthodes et interprétation (Multiple and simple principal axes methods; objectives, methods and interpretation). Dunod, Paris.

Escofier, B., Pagès, J., 1994. Multiple factor analysis: afmult package. Comput. Stat. Data Anal. 18, 121–140.

Escoufier, Y., 1985. Objectifs et procédures de l'analyse conjointe de plusieurs tableaux (objectives and procedures of several tables joint analysis). Statist. et Anal. des Données 10 (1), 1–8.

Gower, J., 1975. Generalized procrustes analysis. Psychometrica 40, 33–51.

Gower, J., 1984. Procrustes analysis. In: Lloyd, E.H. (Ed.), Handbook of Applicable Mathematics, Vol. 6. Wiley, Chichester, pp. 397–405.

Gower, J., Hand, D., 1996. Biplots. Chapman & Hall, London.

Greenacre, M., 1984. Theory and Applications of Correspondence Analysis. Academic Press, New York.

Horst, P., 1961. Relation among *m* sets of measures. Psychometrika 26, 129–149.

Hotelling, H., 1936. Relation between two sets of variables. Biometrika 28, 129–149.

Israëls, A., 1987. Eigenvalue Techniques for Qualitative Data. DSWO Press, Leiden.

Kiers, H., 1989. Three-way Methods for the Analysis of Quantitative and Qualitative Two-way Data. DSWO Press, Leiden.

Lavit, C., 1988. Analyse conjointe de tableaux quantitatifs (Joint analysis of quantitative tables). Masson, Paris.

Lebart, L., Morineau, A., Warwick, K., 1984. Multivariate Descriptive Statistical Analysis. Wiley, New York.

Lebart, L., Morineau, A., Piron, M., 1998. Statistique exploratoire multidimensionnelle (Multidimensional exploratory statistics). Dunod, Paris.

Lohmöller, J., 1989. Latent Variables Path Modelling with Partial Least Squares. Physica Verlag, Heidelberg.

Nishisato, S., 1980. Analysis of Categorical Data: Dual Scaling and Its Applications. University of Toronto Press, Toronto.

Pagès, J., Husson, F., 2001. Inter-laboratory comparison of sensory profiles: methodology and results. Food Qual. Preference 12, 297–309.

Pagès, J., Tenenhaus, M., 2001. Multiple factor analysis combined with pls path modelling. Application to the analysis of relationships between physicochemical variables, sensory profiles and hedonic judgements. Chemometric Intell. Lab. Syst. 58, 261–273.

Van der Heijden, P., 1987. Correspondence Analysis of Longitudinal Categorical Data. DSWO Press, Leiden.

Van der Heijden, P., de Falguerolles, A., de Leeuw, J., 1989. A combined approach to contingency table analysis and log-linear analysis (with discussion). Appl. Statist. 38, 249–292.

Zarraga, A., Goitisolo, B., 2001. Méthode factorielle pour l'analyse simultanée de tableaux de contingence (principal axes method for simultaneous analysis of contingency tables). Rev. Stat. Appl. 50 (2), 47–70.