Multiple Factorial Analysis of Symbolic Data

Barnabé Tang Ahanda, Jean Gérard Aghoukeng Jiofack,
Romain Germain Nzangué, Gilbert Hapi Mbiakop
Department de Mathematics, Faculty of Sciences, University of Yaounde 1, Yaounde, Cameroon
Email: tangahanda@yahoo.fr, aghoukeng-maths@yahoo.fr

ABSTRACT

This document presents an extension of the multiple factorial analysis to symbolic data and especially to space data. The analysis makes use of the characteristic coding method to obtain active individuals and the reconstitutive coding method for additional individuals in order to conserve the variability of assertion objects. Traditional analysis methods of the main components are applied to coded objects. Certain interpretation aids are presented after the coding process. This method was applied to poverty data.

Keywords: Symbolic Object; Assertion Object; Multiple Factorial Analysis; Principle Components Analysis; Poverty

1. Introduction

Although the traditional analysis of data is based on a complete theory, it poses restrictive conditions for the resolution of problems: it requires that variables be single-valued (the value of a variable for an object is a unique and specific).

Recent evolutions in the information system have made it possible to find increasing and more complex data which need a more in-depth formalization than the one indicated in the usual rectangular table. The later does not take into account the possible variability of description as well as the uncertainty concerning certain variables or certain objects. A new formalism called symbolic objects has been developed to represent complex data like concepts, skills (see [1,2]). These objects concern complex data and provide new skills at the exit in the form of symbolic objects.

The techniques of factorial analysis (see [3,4]) which are data reduction and representation methods, are well known and well used. Certain stakeholders have extended these methods to symbolic objects by using the coding and decoding techniques (read [5,6]). The issue of multiple factorial analysis (MFA) is of particular interest to us (see [7,8]). A priori, this concerns the case where there is a group structure on variables. Its problem is enhanced by the characterization of groups and the search for a typology of groups of variables. To our knowledge, this method is not yet applied to symbolic objects.

In this article, we extend the MFA to a symbolic data base whose variables are structured into groups. It is a generalization of the MFA basic principle on the one hand, and the principle components analysis (PCA) tech-

niques on the other hand (see [9]). We conclude this document with an application of the method presented to poverty data.

2. Symbolic Data Base

2.1. General Context

A symbolic data base (SDB) is considered as represented in the **Table 1** and made up of triplet $B = \{Y, C, A\}$ with $C = \{C_1, C_2, \cdots, C_I\}$ representing all the I objects, $Y = \{y_1, y_2, \cdots, y_J\}$ and $A = \{a^1, a^2, \cdots, a^I\}$ representing respectively the set of the J variables and the assertion objects used to describe I objects. We suppose that the SDB is associated with a measured space $(\Omega, \mathcal{F}, \mu)$, an σ-algebra $\mathcal{C}$ on C, $O = \{O_1, O_2, \cdots, O_J\}$ where O_j is the space of observation of the variable $y_j, j = 1, 2, \cdots, J$. This space includes an tribe $\mathcal{O}_j$. Let $\left(X_{c_i, j}\right)_{j=1,2,\cdots,J}$ be a set of measurable functions defined on Ω, deriving their value from the measurable spaces $(O_j, \mathcal{O}_j)$ and used to describe all the c_i objects. For i from 1 to I, a^i $a^i = \bigwedge_{j=1}^{J}\left[y_j = X_{a^i, j}\right]$ is a Boolean assertion so that from j from 1 to J, there is a measurable part $V_{a^i, j} \in \mathcal{O}_j$ so that we can have $X_{a^i, j} = X_{c_i, j}$ and $V_{a^i, j} = X_{a^i, j}(\Omega)$. In practice, it is written

$$a^i = \bigwedge_{j=1}^{J}\left[y_j = V_{a^i, j}\right]. \tag{1}$$

Table 1. Symbolic data base.

	y_1		y_j		y_J
c_1	$\left[y_1 = X_{a^1,1} \right]$	$\ldots$	$\left[y_j = X_{a^1,j} \right]$	$\ldots$	$\left[y_J = X_{a^1,J} \right]$
c_i	$\left[y_1 = X_{a^i,1} \right]$	$\ldots$	$\left[y_j = X_{a^i,j} \right]$	$\ldots$	$\left[y_J = X_{a^i,J} \right]$
c_I	$\left[y_1 = X_{a^I,1} \right]$	$\ldots$	$\left[y_j = X_{a^I,j} \right]$	$\ldots$	$\left[y_J = X_{a^I,J} \right]$

Table 2. Symbolic data base of boolean assertions.

	y_1		y_j		y_J
c_1	$\left[\underline{x_{11}}, \overline{x_{11}} \right]$	$\ldots$	$\left[\underline{x_{1j}}, \overline{x_{1j}} \right]$	$\ldots$	$\left[\underline{x_{1J}}, \overline{x_{1J}} \right]$
c_i	$\left[\underline{x_{i1}}, \overline{x_{i1}} \right]$	$\ldots$	$\left[\underline{x_{ij}}, \overline{x_{ij}} \right]$	$\ldots$	$\left[\underline{x_{iJ}}, \overline{x_{iJ}} \right]$
c_I	$\left[\underline{x_{I1}}, \overline{x_{I1}} \right]$	$\ldots$	$\left[\underline{x_{Ij}}, \overline{x_{Ij}} \right]$	$\ldots$	$\left[\underline{x_{IJ}}, \overline{x_{IJ}} \right]$

The observation space is therefore the product space $\Theta_j = \prod_{j=1}^{J} O_j$. This space includes an product tribe $\Xi_j = \prod_{j=1}^{J} \mathcal{O}_j$. To proceed, it is supposed that the descriptors are quantitative, i.e that O_j is a part of $\mathbb{R}$. Thus, it is generalized that $X_{a^i,j}(\Omega) = V_{a^i,j} = \left[\underline{x_{ij}}, \overline{x_{ij}} \right]$ where $\underline{x_{ij}}, \overline{x_{ij}}$ represent respectively the minimum and maximum values per a^i object for the y_j variable. Consequently, we shall treat the Boolean assertions whose SDB is indicated in the following **Table 2**.

2.2. Group Structure

We shall assume that the J variables are divided into K groups with J_k as the number of k group variables; p_i the weight associated with the assertion object a^i and m_ℓ the weight of the y_ℓ variable. We shall designate I, K, J_k as the set and its cardinal. Let us consider $B_k = \{ Y_k, C, A_k \}$ as the symbolic data base derived from B but reduced from the k group variables denoted Y_k. For i from 1 to I, the a^i object for the k group is represented by the assertion $a_k^i = \bigwedge_{j \in J_k} \left[y_j = V_{a_k^i,j} \right]$. Formula (1) may be re-written as follows, while taking into consideration the existence of the following groups:

$$a^i = \bigwedge_{j \in J_1} \left[y_j = V_{a^i,j} \right] \wedge \cdots \bigwedge_{j \in J_k} \left[y_{j + \sum_{r<k} j_r} = V_{a_k^i, j + \sum_{r<k} j_r} \right]$$

$$\wedge \cdots \bigwedge_{j \in J_K} \left[y_{j + \sum_{r<K} j_r} = V_{a_K^i, j + \sum_{r<K} j_r} \right]$$

2.3. Objective

Like in the traditional case, the MFA based on symbolic data corresponds to the following three majors objectives:

1) Drawing up a typology of concepts, especially similar concepts from the point of view of all the variables after balancing the contribution of each group.

2) Comparing the typologies of individuals defined by groups of variables.

3) Exposing a typology of groups of variables and interpreting the proximities between them.

However, taking into consideration the variability of objects adds an interesting element to this problem because coherence warrants us to provide exit elements in keeping with the variability.

3. Analysis and Interpretation

We shall present algorithm for a symbolic digital MFA. This approach essentially repeats the one used by the traditional MFA as well as that of the PCA on symbolic objects. The following stages shall be used:

1) Data coding;

2) PCA of groups;

3) Weighting of the data table;

4) PCA symbolic digital of all data;

5) Traditional MFA representation and interpretation;

6) Superimposed representation of symbolic objects and average cloud;

7) Representation of assertions of variables and symbolic factors of groups

3.1. Data Coding

Either Δ all the assertions in the SDB, the m coding order is any C application to Δ with values in $\left(\mathbb{R}^j \right)^m$. In this work, we shall therefore use the centre and summit methods which is referred to as characteristic coding and reconstitutive coding respectively.

An assertion object is graphically represented by a hypercube. The characteristic coding represents each object by a point which is the centre of gravity of the assertion object. The reconstitutive coding represents an object by 2^J points which constitute the summit of the hypercube. Characteristic coding gives us the active elements of the analysis while reconstitutive coding gives the additional elements of the analysis that enable the reconstitution of the exit symbolic objects.

3.2. Weighting of Variables

Let X be the data table obtained after data coding. A J columns and $I \times \left[1 + 2^J \right]$ lines table is obtained where I individuals are active and the others in addition. For each group, a principle components analysis (PCA) on the $X_k, k = 1 \cdots K$ tables where X_k is the X table reduced to the k group value. A series of

$\lambda_i^k, i = 1, \cdots, k, k = 1, \cdots, K.$ eigenvalues is obtained for each analysis. It is assumed that the real values in each group are classified in descending order. Each y_ℓ variable is weighted by $\dfrac{m_\ell}{\lambda_1^k}$ if the y_ℓ variable is derived from the k group. The weighted table is evaluated $X^{(p)}$.

3.3. Data Processing

Weighting conserves the internal structures of each group but changes the structure of the overall cloud (see [10]). In this case, it is referred to as the average cloud. As in the traditional case, we have represent equivalent objects by respective lines of the same K bases $\tilde{B}_k^{(p)}$ rank, with $B_k^{(p)}$ being the weighted B_k base.

$$\tilde{B}_k^{(p)} = \boxed{0 \mid B_k^{(p)} \mid 0} \Rightarrow \tilde{X}_k^{(p)} = \boxed{0 \mid X_k^{(p)} \mid 0}$$

The N_j^k K clouds may be simultaneously represented in the $\mathbb{R}^J$ space of variables by breaking down $\mathbb{R}^J$ into direct sum spaces isomorphic to $\mathbb{R}^{jk}$ spaces.

The average cloud is obtained by a homothetic transformation of the $B^{(p)}$ $1/K$ relation weighted data base. It represents average individuals and average concepts in $\mathbb{R}^k$. The average cloud is obtained by a homothetic transformation of the $B^{(p)}$ $1/K$ relation weighted data base. It represents average individuals and average concepts in $\mathbb{R}^k$. Let us note B^* the SDB of the average cloud where B_k^* constitutes its restriction in the k group.

Proposition 1. The symbolic digital PCA of the B^* base leads to the same factors as the B base characteristic coding MFA.

Proof. It is immediate. The B^* base is constructed to this effect. □

3.4. Canonic Variable

The re-writing of table X^* in the principle components base $(F_1, \cdots, F_J)$ leads to a representation of objects by a new SDB $B = \{F, O, A_F\}$ whose assertion is written $a_F^i = \bigwedge_{j=1}^{J} \left[F_j = W_{a_F^i, j} \right]$. Assertions of the variables associated with concepts are written $F^j = \bigwedge_{i=1}^{I} \left[F_j = W_{a_F^i, j} \right]$. These factors are called canonic variables.

3.5. Superimposed Representation of Individuals and Concepts

3.5.1. Equivalent Assertions

A $B^{(p)}$ based a^i assertion is represented in each $B_k^{(p)}$ sub-base by a_k^i assertions which are called equivalent assertions associated with a^i. These are the different representatives in the $B^{(p)}$ sub-bases of the same assertion.

3.5.2. Superimposed Representation of Individuals

Here, individuals are described by the characteristic coding. An individual representing an object is represented by the expectation of unpredictable random. variable associated with the skill it represents. This representation no doubt has some shortcomings, but it gives us the first tendencies which are sometimes enough. In concrete terms, each group is projected with additional elements on the PCA of table X^*.

Proposition 2. Each individual (average individual) of table X^* is the centre of gravity of individuals observed from the point of view of all groups.

Proof. This is the property of traditional MFA. □

This proposition explains the notion of average cloud. The study of the attributes of variability of description indicated below will comfort us in this notion.

3.5.3. Superimposed Representation of Concepts

Symbolic digital PCA results will be used to obtain a superimposed representation of objects while taking into consideration variability. It based to the following proposition:

Proposition 3. Either $a^i = \bigwedge_{j=1}^{J} \left[y_j = V_{a^i, j} \right]$ an assertion of BDS B. We note $b^i = \bigwedge_{k=1}^{K} \left[u_k = V_{b^i, k} \right]$ the assertion representing the c_i objects in the space of the PCA principle components of this SDB. Let $G = \left(X_1^G, \cdots, X_J^G \right)$ the centre of gravity of the characteristic coding and u_{jk} the kth component of the jth factorial axis. Thus we have:

$$\min\left(W_{b^i, k} \right) = \sum_{j, u_{kj} < 0} \left[\max\left(V_{a^i, j} \right) - X_j^G \right] u_{kj}$$
$$+ \sum_{j, u_{kj} < 0} \left[\min\left(V_{a^i, j} \right) - X_j^G \right] u_{kj}$$
$$\max\left(W_{b^i, k} \right) = \sum_{j, u_{kj} < 0} \left[\min\left(V_{a^i, j} \right) - X_j^G \right] u_{kj}$$
$$+ \sum_{j, u_{kj} < 0} \left[\max\left(V_{a^i, j} \right) - X_j^G \right] u_{kj}$$

Proof. We note $\left(x_{ij}^G \right)_{j=1, \cdots, J}$ the coordinate of G^i centre of gravity of a^i assertion, and $\left(y_{kj}^G \right)_{j=1, \cdots, J}$ the coordinate of the image of G^i in the new basis. Let u_{jk} be the jth component of the kth factorial axis (denoted u_k), we note $x_r^i = \left(x_{ij}^r \right)_{j=1, \cdots, J}$ the coodinate of an elementary individual of a^i assertion for a point R^i, and $y_r^i = \left(y_{kj}^r \right)_{j=1, \cdots, J}$ his coordinate in the new basis. We have:

$$y_{kj}^r = \overline{GR} \cdot u_k = \sum_{j=1}^{J} \left(x_{ij}^r - x_{ij}^G \right) u_{jk}$$

then

$$\max\left(W_{b^i,k}\right) = \max\left\{y_{ki}^r / x_{ij}^r \in V_{a^i,j}\right\}$$

$$= \max\left\{\left[\sum_{j=1}^J \left(x_{ij}^r - x_{ij}^G\right)u_{jk}\right], x_{ij}^r \in V_{a^i,j}\right\}$$

as y_{kj}^r depend of x_{ij}^r, we have

$$\max\left(W_{b^i,k}\right) = \sum_{j,u_{kj}<0}\left[\left(\min\left\{x_{ij}^r, x_{ij}^r \in V_{a^i,j}\right\}\right) - X_j^G\right]u_{kj}$$

$$+ \sum_{j,u_{kj}>0}\left[\left(\max\left\{x_{ij}^r, x_{ij}^r \in V_{a^i,j}\right\}\right) - X_j^G\right]u_{kj}$$

Therefore

$$\max\left(W_{b^i,k}\right) = \sum_{j,u_{kj}<0}\left[\min\left(V_{a^i,j}\right) - X_j^G\right]u_{kj}$$

$$+ \sum_{j,u_{kj}>0}\left[\max\left(V_{a^i,j}\right) - X_j^G\right]u_{kj}.$$

With the similar methode, we have

$$\min\left(W_{b^i,k}\right) = \sum_{j,u_{kj}<0}\left[\max\left(V_{a^i,j}\right) - X_j^G\right]u_{kj}$$

$$+ \sum_{j,u_{kj}>0}\left[\min\left(V_{a^i,j}\right) - X_j^G\right]u_{kj}.$$

□

This lead that the B_k representing group k will be projected successively with additional elements on the B^* base symbolic digital PCA. The superimposed representation on the same factorial plan enables us to generally observe the position of an object, and the point of view of each group by using variability. We then have the following result :

Proposition 4. Either a^i an assertion object represented in B^* by $\left(a^i\right)^*$. Let us note a_k^i the a^i assertion observed from the k group point of view. In the B^* base, the a^i assertion is written

$$a^i = \bigwedge_{j\in J_1}\left[y_j = \frac{1}{J\sqrt{\lambda_1}}V_{a_1^i,j}\right]$$

$$\wedge\cdots\wedge_{j\in J_k}\left[y_{j+\sum_{r<k}j_r} = \frac{1}{J\sqrt{\lambda_k}}V_{a_k^i,j+\sum_{r<k}j_r}\right]$$

$$\wedge\cdots\wedge_{j\in J_K}\left[y_{j+\sum_{r<K}j_r} = \frac{1}{J\sqrt{\lambda_K}}V_{a_k^i,j+\sum_{r<K}j_r}\right]$$

Either $a^i = \bigwedge_{j=1}^J\left[y_j = V_{a^i,j}^*\right]$ we note

$$a_k^i = \bigwedge_{k=1}^K\left[F_k = W_{a_k^i,k}\right]$$ the assertion representing the c_i object from the k group point of view in the space of the

PCA principle components of this SDB. Either $G = \left(X_1^G,\cdots,X_J^G\right)$ the centre of gravity of X^*, and F_{kj} the kth component of the jth factorial axis. Thus we have:

$$\min\left(W_{a_F^i,k}\right) = \sum_{j\in J_k,F_{kj}<0}\left[\max\left(JV_{a^i,j}^*\right) - X_j^G\right]F_{kj}$$

$$+ \sum_{j\in J_k,F_{kj}>0}\left[\min\left(JV_{a^i,j}^*\right) - X_j^G\right]F_{kj}$$

$$\max\left(W_{a_F^i,k}\right) = \sum_{j\in J_k,F_{kj}<0}\left[\min\left(JV_{a^i,j}^*\right) - X_j^G\right]F_{kj}$$

$$+ \sum_{j\in J_k,F_{kj}>0}\left[\max\left(JV_{a^i,j}^*\right) - X_j^G\right]F_{kj}$$

Proof. This is the consequence of the notations of the previous proposition. □

The representation of superimposed objects in the canonic base (that of factors) therefore respects the principle of centres of gravity. The hypercube representing an object is thus the centre of gravity of the hypercubes representing the same individual from the point of view of different groups. The MFA symbolic digital algorithm helps to preserve this essential property.

3.6. Representation of Variables and Assertions of Variables

The symbolic digital MFA enables us to define many representatives that contribute to the illustration of factors in the general analysis as well as in the different groups. The determination of the main factors of groups and the general analysis lead to many representations of variables and assertions of variables. The superimposed representation of these factors (as assertions of variables) on canonic variables then helps to find and interpret common factors and specific factors.

3.6.1. Representation of Variables and Assertions of Variables

The conduct of the MFA symbolic digital pre-supposes $K+1$ factorial analysis (the K groups and the average cloud). Two representations can be made for each of these analyses: a representation of variables (dual analysis of centres of gravity) and an analysis of assertions of variables taking into account variability. These representations clearly bring out the main variability factors of groups and the factors of general analysis.

3.6.2. Superimposed Representation of Factors (as Assertions of Variables) on Canonic Variables

This representation is the result of a MFA traditional representation which consists in highlighting the main variability factors of groups in the form of additional variables over canonic variables. But in symbolic data analysis, these variability factors are also expressed as

assertions of variables and the representation of these assertions of variables over the canonic variables calls for an original observation. The position and collection of assertions of variables give us precious indications on their importance and the link between the different groups.

3.6.3. Link between Groups: Representation Quality of Groups

This is a traditional indicator in MFA. It is a representation of groups on canonic factors. The coordinates of an axis group is the combined inertia of the group's variables on the corresponding MFA axis of table X^*. For the two groups K_x. and K_y, we have

$\mathcal{L}(K_x, K_y) = \langle W_x^{(p)}D, W_y^{(p)}D \rangle$ where D is the diagonal $I \times I$ matrix of the weight of individuals and $W_j^{(p)}$ is the matrix of scalar products between the individuals of group j for the X_j^p weighted matrix.

A symbolic representation may be proposed for it. Each assertion of variables can be coded into many variables whose projection over canonic variables results in a minimum value and a maximum value. The variability associated to a group will therefore be displayed by two values: the sum of the minimum inertia of the group's variables and the sum of its maximum inertia. The representation of groups is done in the $\mathbb{R}^{I^2}$ space.

3.7. The Notion of Common Factor and Specific Factor

MFA helps to portray the characteristics of groups compared to the canonic variables (also referred to as factors). A common (or specific) factor to the MFA symbolic digital will be the common factor associated to the traditional MFA of centres of gravity.

4. Poverty Data

4.1. Presentation

The data is drawn from the first Cameroonian survey in households (ECAM I) conducted in 1996 by the National Institute of Statistics. The sample is made up of 1800 households with a total of 10230 individuals. The analysis is done on 13 continuous variables out of which 10 are active and 3 illustrative (see **Table 3**).

The average poverty line is used to establish the difference between poor and non poor individuals. We shall consider 10 classes of individuals with 5 poor classes organized in descending order from very poor to less poor individuals and 5 classes of non poor individuals organized in ascending order from less rich to very rich individuals. These classes constitute the concepts that we shall analyze. The groups of variables are regions. For presentation purposes, they shall be presented on a line and not in columns. An overview of the SDB for the Adamawa Region for three variables shows the following **Table 4**.

4.2. Analysis and Interpretation

After the algorithm application, the following results are obtained:

4.2.1. Inertia of the First Six Eigenvalues
Please see **Table 5**.

4.2.2. Relationship between General Variables and Groups
The F_1 variable is the variable which extracts the most important inertia from all groups. The other variables convey specific information to certain regions (see **Table 6**).

Table 3. Variables of ECAM1.

N	Title	Description	Quality
1	DEPEAD	Expenditure per adult equivalent	Additional
2	DEPTET	Expenditure per head	Additional
3	DEPTOT	Total expenditure	Additional
4	DEPALI	Food expenditure	Active
5	DEPHAD	Clothing expenditure	Active
6	DEPMAI	Household expenditure	Active
7	DEPSAN	Health expenditure	Active
8	DEPTRA	Transportation expenditure	Active
9	DEPEDU	Education expenditure	Active
10	DEPSOP	Personal health care expenditure	Active
11	DEPLOI	Leisure expenditure	Active
12	DEPLOY	Rent expenditure	Active
13	DEPLOG	Lodging expenditure	Active

Table 4. An overview of SDB of ECAM1.

Concept	AD		
	DEPALI	DEPHAB	DEPMAI
Poor of class 1	[96,464, 173,375]	[2000, 36,000]	[3950, 36,000]
Poor of class 2	[122,375, 573,832]	[15,000, 31,200]	[28,800, 34,800]
Poor of class 3	[109,500, 573,032]	[5700, 46,800]	[10,800, 28,800]
Poor of class 4	[157,210, 302,428]	[5700, 38,200]	[7200, 15,300]
Poor of class 5	[126,967, 358,482]	[0, 54,500]	[12,000, 53,700]
Non poor of class 1	[102,982, 536,028]	[0, 115,900]	[2450, 55,100]
Non poor of class 2	[220,303, 568,617]	[0, 82,000]	[11,950, 57,600]
Non poor of class 3	[95,421, 1,707,678]	[0, 297,300]	[23,000, 65,100]
Non poor of class 4	[146,782, 967,510]	[3500, 297,300]	[600, 71,300]
Non poor of class 5	[130,617, 813,428]	[8000, 485,200]	[22,800, 199,800]

Table 5. Histogram of eigenvalues.

N	Eigenvalue	Inertia %	Combined %
1	6.80938	68.09	68.09
2	1.08859	10.89	78.98
3	0.60479	6.05	85.03
4	0.56497	5.65	90.68
5	0.31109	3.11	93.79
6	0.29054	2.91	96.69

Table 6. Relationship between general variables and groups.

	F_1	F_2	F_3	F_4	F_5	F_6
AD	52.70	20.95	5.81	3.94	1.15	1.31
CE	77.57	9.29	0.83	0.93	0.35	0.26
ES	58.61	4.60	9.72	8.58	0.96	4.04
EN	62.35	10.50	3.13	5.98	2.81	1.40
LT	82.74	2.18	1.31	1.93	1.09	0.45
NO	51.03	6.12	7.38	8.56	10.40	4.10
NW	67.16	8.38	2.91	5.74	1.36	3.81
OU	46.08	10.14	6.78	9.28	6.78	5.22
SU	57.94	13.26	5.23	2.32	2.02	3.20
SW	57.84	12.42	11.17	3.22	1.13	2.62

4.2.3. Inter-Inertial/Total Inertia Relationship

Table 7 shows that on the first axis, there is a close proximity between equivalent classes. Broadly speaking, these equivalent classes have similar behaviours in their expenditure habits. The same phenomenon is observed in the descending values of *r*, on the 4th, 2nd, 6th and 3rd axes.

4.2.4. Degree of Resemblance between Groups and General Variables: Common and Specific Factors

The first MFA component is a factor common to all the

Table 7. Inter-inertial/total inertia relationship.

Axis	F_1	F_2	F_3	F_4	F_5	F_6
Relationship	0.89	0.70	0.64	0.70	0.30	0.65

Table 8. Degree of resemblance between groups and general variables.

	F_1	F_2	F_3	F_4	F_5	F_6
AD	0.87	0.58	0.65	0.13	0.05	0.09
CE	0.90	0.12	0.03	0.45	0.01	0.02
ES	0.90	0.15	0.31	0.66	0.15	0.80
EN	0.98	0.75	0.45	0.50	0.18	0.38
LT	0.98	0.28	0.11	0.33	0.02	0.08
NO	0.92	0.42	0.44	0.32	0.59	0.32
NW	0.97	0.57	0.12	0.78	0.28	0.35
OU	0.88	0.42	0.43	0.60	0.28	0.62
SU	0.93	0.47	0.60	0.55	0.28	0.70
SW	0.96	0.68	0.72	0.30	0.15	0.16

10 regions. The second factor is common to the Far-North and South-West regions. The 3rd factor is specific to the South-West and Adamawa regions (see **Table 8**).

REFERENCES

[1] E. Diday, "Introduction to the Symbolic Data Analysis," Cahiers du CEREMADE, No. 8823, 1988.

[2] E. Diday, "From the Objects of the Data Analysis to Those of the Analysis of Knowledge, Symbolic and Numeric Induction from Data," Cépadues, 1991.

[3] L. Lebart, A. Morineau and Mr. Piron, "Multidimensional Exploratory Statistics," 3rd Edition, DUNOD, Paris, 2000.

[4] M. Volle, "Data Analysis," Economica, Paris, 1981.

[5] J. F. Martin, "Fuzzy Coding and Its Applications in Statistics," Thesis, University of Pau, Pau, 1980.

[6] F. J. Gallego, "Fuzzy Coding in Correspondence Analysis," *Cahiers de l'Analyse des Données*, Vol. 13, No. 2, 1980, pp. 413-430.

[7] J. Pagès and B. Escofier, "Simple and Multiple Factorial Analysis," DUNOD, Paris, 1990.

[8] B. Escofier and J. Pajes, "Multiple Factorial Analysis, Objectives, Methodology and Interpretation," 2nd Edition, Wiley, Paris, 2005.

[9] P. Cazes, A. Chouakria, E. Diday and Y. Schektman, "Extension of Principal Component Analysis to Interval Data," *Revue de Statistique Appliquée*, Vol. 45, No. 3, 1997, pp. 5-24.

[10] B. Ahanda Tang, "Extension of Factorial Analysis to Symbolic Data," Thesis, University of Paris, Paris, 1998.