

Correspondence Analysis of Contingency Tables with Subpartitions on Rows and Columns

Análisis de correspondencias de tablas de contingencia con subparticiones en filas y columnas

CAMPO ELÍAS PARDO^{1,a}, MÓNICA BÉCUE-BERTAUT^{2,b},
JORGE EDUARDO ORTIZ^{3,c}

¹DEPARTAMENTO DE ESTADÍSTICA, FACULTAD DE CIENCIAS, UNIVERSIDAD NACIONAL DE
COLOMBIA, BOGOTÁ, COLOMBIA

²DEPARTAMENTO DE ESTADÍSTICA E INVESTIGACIÓN OPERATIVA, UNIVERSIDAD POLITÉCNICA
DE CATALUÑA, BARCELONA, ESPAÑA

³FACULTAD DE ESTADÍSTICA, UNIVERSIDAD SANTO TOMÁS, BOGOTÁ, COLOMBIA

Abstract

We present Intra-Table Correspondence Analysis using two approaches: Correspondence Analysis with respect to a model and Weighted Principal Component Analysis. In addition, we use the relationship between Correspondence Analysis and the Log-Linear Models to provide a deeper insight into the interactions that each Correspondence Analysis describes. We develop in detail the Internal Correspondence Analysis as an Intra-Table Correspondence Analysis in two dimensions and introduce the Intra-blocks Correspondence Analysis. Moreover, we summarize the superimposed representations and give some aids to interpret the graphics associated to the subpartition structures of the table. Finally, the methods presented in this work are illustrated by their application to the standardized public test data collected from Colombian secondary education students in 2008.

Key words: Multidimensional contingency table, Principal component analysis.

Resumen

Para presentar los análisis de correspondencias intra-tablas, se usan los enfoques del análisis de correspondencias con respecto a un modelo y del análisis en componentes principales ponderado. Adicionalmente, se utiliza la relación de los análisis de correspondencias con los modelos log-lineales para entender mejor las interacciones que cada análisis de correspondencias

^aAssociate professor. E-mail: cepardot@unal.edu.co

^bProfessor. E-mail: monica.becue@upc.es

^cProfessor. E-mail: jorgeortiz@usantotomas.edu.co

describe. Se desarrolla de manera detallada el análisis de correspondencias interno como un análisis de correspondencias intra-tablas en dos dimensiones y se introduce el análisis de correspondencias intrabloques. Por otra parte, se resumen las representaciones superpuestas y las ayudas para la interpretación de las gráficas asociadas a la estructura de subparticiones de la tabla. Finalmente, se ilustran los procedimientos con el análisis de una tabla de contingencia construida a partir de los resultados de las pruebas de estado realizadas a los estudiantes de educación media en Colombia en el año 2008.

Palabras clave: análisis en componentes principales, tabla de contingencias multidimensional.

1. Introduction

Contingency tables (CT) with sub-partitions on rows and columns have row and column categories defined from two nested factors. We use $B(A) \times D(C)$ to denote the table structure. The rows are formed by factors A and B , with B categories nested into A categories. In the same way, C and D factors form the columns, with D categories nested into C categories. Each A category defines a *row band* and each C category defines a *column band*. A sub-table crossing a row band with a column band is called a *block*.

The nesting may occur naturally, for example, in a table crossing subregions and economic sub-sectors, where the subregions are aggregated into regions and the economic sub-sectors are aggregated into sectors. In this case, we say that the CT has a “true” sub-partition structure. In other applications, the researcher will choose the variable defining the coarsest partition according to the objectives of the study. For example, the notation *age-group(sex)* indicates that the categories of the variables *sex* and *age-group* are codified interactively. The *sex* variable defines the partition and the *age-group* categories are nested into the two categories of *sex*. A four-way CT with factors A , B , C and D can be flattened into a two-way table in different manners; for example, into the two-way CT denoted by $B(A) \times D(C)$.

We cite hereafter several examples of CT with row and column sub-partitions extracted from the literature:

Hydrobiological studies: *species(taxonomic groups) × places(dates)*, i.e. phau-nistic tables, with row-species categorized into taxonomic groups and columns *places × dates*, being the same places observed at different dates (Cazes, Chessel & Doledec 1988).

Genomics: *sequences(species) × codons(amino acids)*, a CT crossing sequences aggregated into species and codons aggregated into amino acids (Lobry & Chessel 2003, Lobry & Necsulea 2006).

Genetics: *objects(populations) × aleles(loci)*, a CT with objects split in populations described using alleles clustered into several loci (Laloë, Moazami-Gourdarzi & Chessel 2002).

We aim at presenting different strategies, in the framework of Correspondence Analysis (CA; Lebart, Piron & Morineau 2006, Ramírez & Martínez 2010), to describe contingency tables endowed with sub-partition structures both in rows and columns. Having this objective in mind, we do not discuss inferential methods that might be used to analyze this kind of table.

From a contingency table crossing the row categories $B(A)$ with the column categories $D(C)$, several CA can be performed, depending on the sub-partition structures that are considered. Each CA can be seen as a particular CA with respect to a model, using the generalization of CA proposed by Escofier (1983). This point of view allows us to consider the relationship between Log-Linear Models and Correspondence Analysis applied to the analysis of a two way contingency table. This table is obtained through flattening a four way CT, as described in Van der Heijden (1987).

The structure of the CT, as well as the treatments applied to it, are deduced from the objectives. Dolédec & Chessel (1991) lay out the use of these CA in the environmental sciences.

The first example considers a faunal table in hydrobiology field. The row categories are nested as *species(group)*. The authors apply Intra-group CA (row bands) and argue both that the specialists have different skills to identify species in each taxonomic group, and that, in such a method, the between-groups variability is eliminated. The Intra-date CA (column bands) shows, more clearly, the associations between species and sites. The Internal Correspondence Analysis (ICA) is both Intra-dates and Intra-groups, as proposed by Cazes et al. (1988) to highlight the species-site associations.

Bécue-Bertaut, Pagès & Pardo (2005) present ICA as a double Intra-Table CA and show that it can be computed either as a CA with respect to a model or as a Weighed Principal Component Analysis. Furthermore, they propose to project on the principal planes issued from this ICA, the “partial” rows (“partial” columns), that is, the rows (columns) as seen from the different points of view corresponding to each group of columns (rows). The superimposed representation of the partial rows (partial columns) is obtained following the same rationale that Multiple Factor Analysis (MFA: Escofier & Pagès (1982); Pagès (2004)). These superimposed representations ease the comparison of the different viewpoints and so enrich the interpretation of the results.

In this paper, the theoretical sections presented by Bécue-Bertaut et al. (2005) are extended and Intra-Block Correspondence Analysis (IBCA) is presented. The resulting methodology is applied to a CT built up from the results of the schools standardized test scores answered by last grade Colombian students in secondary education in 2008. The relationship between CA and Log-Linear Models are used to show the interactions described by the different CA.

§2 defines the notation, taking into account the sub-partition structures of the CT. In §3 we present the different CA as specific cases of both CA with respect to a model and Weighted Principal Component Analysis. The superimposed representations are detailed in §4. The interest of the methodology is shown in §5,

by its application to the schools standardized tests scores in Colombia in 2008. In the Appendix, the demonstrations of some formulae are detailed.

2. Notation

The notation adopted in this work is close to this used by Bécue-Bertaut et al. (2005). Let $B(A) \times D(C)$ be a CT with I rows and K columns. The factors A and C have L and J factors, respectively. The L categories from A are sub-partitioned into $I_1, \dots, I_l, \dots, I_L$ categories, respectively; and, similarly, the J categories from C into $K_1, \dots, K_j, \dots, K_J$ categories. We use the same symbols to indicate sets and their cardinality. Thus, I is both the set and the number of rows, that is, the categories of $B(A)$; K is both the set and the number of the columns. The categories of $D(C)$; I_l is both the set and number of categories that are nested into the category l from A . From the CT, the relative frequencies table \mathbf{F} is built up. It is structured as shown in Figure 1.

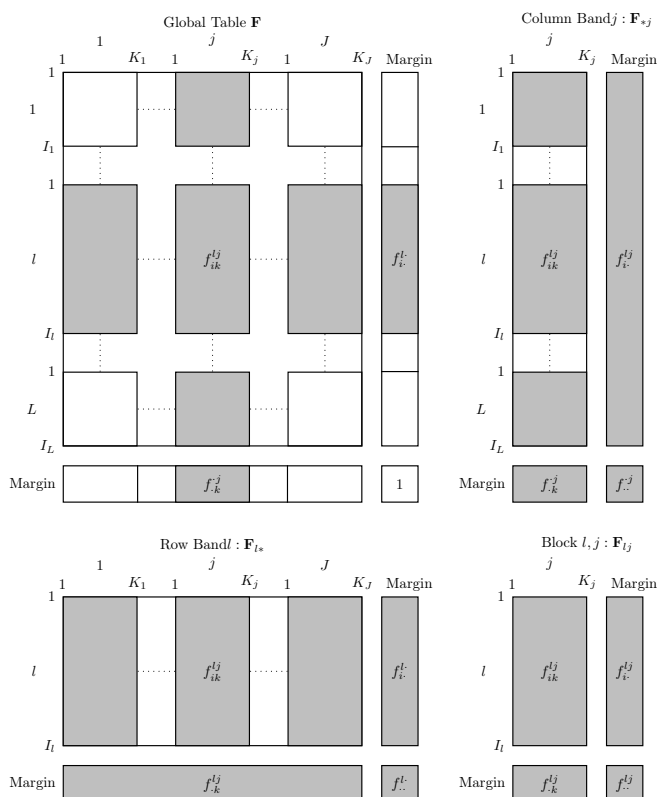


FIGURE 1: Table \mathbf{F} with sub-partition structures in the rows and in the columns.

The general term of \mathbf{F} is noted by f_{ik}^{lj} and its row and column margins by f_i^l and $f_{.k}^j$, respectively. \mathbf{F}_{l*} is the row band l and \mathbf{F}_{*j} the column band j . The

total of the row band \mathbf{F}_{l*} is $f_{..}^l = \sum_{j=1}^J f_{..}^{lj}$ and the total of the column band \mathbf{F}_{*j} is $f_{..}^j = \sum_{l=1}^L f_{..}^{lj}$.

The block (l, j) , noted \mathbf{F}_{lj} , has I_l rows and K_j columns. Its row and column margins are $f_{i.}^{lj} = \sum_{k \in K_j} f_{ik}^{lj}$ and $f_{.k}^{lj} = \sum_{i \in I_l} f_{ik}^{lj}$; and its total is $f^{lj} = \sum_{i \in I_l} \sum_{k \in K_j} f_{ik}^{lj}$.

A cell of \mathbf{F} is identified by the block lj , as superscript, and the specific cell into the block $_{ik}$, as subscript.

\mathbf{F} can be analyzed through the different CA presented in this work: a Simple Correspondence Analysis (SCA); two Intra-Table CA, called here ‘analysis in only one dimension’: the Intra-Column Bands CA and the Intra-row Bands CA; the Internal Correspondence Analysis (ICA) or double Intra-analysis; the Intra-blocks Correspondence Analysis (IBCA).

To avoid misinterpretations, we use the expression ‘Intra-Tables CA’, when the structure only concerns one dimension. When the structure concerns the two dimensions, we use the term ‘Internal Correspondence Analysis’ (ICA) rather than ‘Double Intra-Tables CA’. ICA was proposed, with this denomination, by Cazes et al. (1988). Pagès & Bécue-Bertaut (2006) use the term ICA for referring to Intra-Tables CA only in one dimension because, in this case, the two methods are equivalent.

The clouds of points, associated with CA, are noted by using the letter N and a subscript, referring to both the set of points and its cardinality. For example, N_I is the cloud of the I row points and N_{I_l} is the cloud of the I_l points belonging to the row band l .

3. Correspondence Analysis (CA)

We summarize the use of CA to describe a CT endowed with sub-partitions both in rows and columns. Each CA is presented as a Weighted Principal Component Analysis, denoted $PCA(\mathbf{X}, \mathbf{M}, \mathbf{D})$. \mathbf{X} is the data matrix, issued from the original data possibly conveniently transformed; \mathbf{M} is a diagonal matrix corresponding to both the metric in the row space and the column weights. \mathbf{D} is a diagonal matrix corresponding to both the metric in the column space and the row weights. $PCA(\mathbf{X}, \mathbf{M}, \mathbf{D})$ is also called, in French literature, the *general factor analysis* (Lebart, Morineau & Warwick 1984, Escofier & Pagès 1992, Pagès 2004) or *duality diagram* (Cailliez & Pagès 1976, Tenenhaus & Young 1985). This approach emphasizes the geometric point of view of PCA leading to call several statistical measures as in Physics. For example, the barycentre or centroid corresponds to the vector of means, the inertia corresponds to the generalized variance. *Active* and *illustrative* elements are considered; the former are taken into account to compute the principal axes while the latter, if present, are projected on the principal axes previously computed from the active elements.

3.1. Simple Correspondence Analysis (SCA)

SCA describes the residuals of \mathbf{F} with respect to the independence model. The independence model is defined as the product of the marginal terms. SCA applied to \mathbf{F} is also $PCA(\mathbf{X}, \mathbf{M}, \mathbf{D})$ being \mathbf{X} the matrix with the general term:

$$x_{ik}^{lj} = \frac{f_{ik}^{lj} - f_{i\cdot}^l \cdot f_{\cdot k}^j}{f_{i\cdot}^l \cdot f_{\cdot k}^j} \quad (1)$$

and \mathbf{M} and \mathbf{D} the matrices:

$$\mathbf{D} = \text{diag}(f_{i\cdot}^l) \quad \text{and} \quad \mathbf{M} = \text{diag}(f_{\cdot k}^j) \quad (2)$$

\mathbf{M} (respectively, \mathbf{D}) is the metric matrix (matrix of weights) in row (column) space and the matrix of weights (metric matrix) in the column (row) space.

3.1.1. Centroids of the Subclouds as Illustrative Elements

In the row space induced by CA, the cloud N_I can be considered as the union of the L subclouds N_{I_l} formed, each of them by the points belonging to the row band I_l . The weight of the row point (l, i) within the subcloud N_{I_l} is $f_{i\cdot}^l / f_{\cdot\cdot}^l$; thus, the coordinate (j, k) of the centroid of the subcloud N_{I_l} is:

$$\sum_{i \in I_l} \frac{f_{i\cdot}^l}{f_{\cdot\cdot}^l} \left(\frac{f_{ik}^{lj}}{f_{i\cdot}^l \cdot f_{\cdot k}^j} - 1 \right) = \frac{f_{\cdot k}^j}{f_{\cdot\cdot}^l \cdot f_{\cdot k}^j} - 1 \quad (3)$$

In the same way, the coordinate (l, i) of the centroid of the subcloud N_{K_j} in the column space is:

$$\sum_{k \in K_j} \frac{f_{\cdot k}^j}{f_{\cdot\cdot}^j} \left(\frac{f_{ik}^{lj}}{f_{i\cdot}^l \cdot f_{\cdot k}^j} - 1 \right) = \frac{f_{i\cdot}^l}{f_{\cdot\cdot}^j \cdot f_{i\cdot}^l} - 1 \quad (4)$$

3.1.2. Inertia Decomposition from the SCA

The partition of the cloud N_K into J subclouds N_{K_j} induces the inertia decomposition into *BetweenInertia* + *IntraInertia*:

- Between subclouds N_{K_j} Inertia:

$$\sum_{l,i} f_{i\cdot}^l \sum_j f_{\cdot\cdot}^j \left(\frac{f_{ik}^{lj}}{f_{i\cdot}^l \cdot f_{\cdot k}^j} - 1 \right)^2 = \sum_{l,i} \sum_j \frac{\left(f_{i\cdot}^l - f_{i\cdot}^l \cdot f_{\cdot\cdot}^j \right)^2}{f_{i\cdot}^l \cdot f_{\cdot\cdot}^j} \quad (5)$$

- Intra subclouds N_{K_j} Inertia:

$$\sum_{l,i} f_{i\cdot}^l \sum_j f_{\cdot\cdot}^j \sum_{k \in K_j} \frac{f_{\cdot k}^j}{f_{\cdot\cdot}^j} \left(\frac{f_{ik}^{lj}}{f_{i\cdot}^l \cdot f_{\cdot k}^j} - \frac{f_{i\cdot}^l}{f_{\cdot\cdot}^j \cdot f_{i\cdot}^l} \right)^2 = \sum_{l,i} \sum_j \sum_{k \in K_j} \frac{\left(f_{ik}^{lj} - \frac{f_{i\cdot}^l \cdot f_{\cdot k}^j}{f_{\cdot\cdot}^j} \right)^2}{f_{i\cdot}^l \cdot f_{\cdot k}^j} \quad (6)$$

Through exchanging the subscripts i and j , we obtain the decomposition of the inertia of the cloud N_I into between-clouds N_{I_l} inertia and Intra-clouds N_{I_j} inertia.

3.2. Correspondence Analysis with Respect to a Model

Let \mathbf{A} be the model matrix with general term a_{ik}^{lj} , with the same dimensions and margins as \mathbf{F} . The CA of \mathbf{F} with respect to the model \mathbf{A} , noted $CA(\mathbf{F}, \mathbf{A})$, is equivalent to $PCA(\mathbf{X}, \mathbf{M}, \mathbf{D})$, with \mathbf{M} and \mathbf{D} defined above, in (2), and \mathbf{X} with general term:

$$x_{ik}^{lj} = \frac{f_{ik}^{lj} - a_{ik}^{lj}}{f_{i\cdot}^l \cdot f_{\cdot k}^j} \quad (7)$$

CA with respect to a model keeps almost all of the properties of the classical CA when the model margins are equal to \mathbf{F} margins (Escofier 1984). This is the case for Intra-Tables CA.

The inertia of both clouds N_I and N_K associated to $CA(\mathbf{F}, \mathbf{A})$ is:

$$Inertia(N_I) = Inertia(N_K) = \sum_{l,j} \sum_{i \in I_l, k \in K_j} \frac{(f_{ik}^{lj} - a_{ik}^{lj})^2}{f_{i\cdot}^l \cdot f_{\cdot k}^j} \quad (8)$$

The SCA of \mathbf{F} is obtained if the independence model $\mathbf{H} = (f_{i\cdot}^l \cdot f_{\cdot k}^j)$ is used in the Formula (7).

3.2.1. Decomposition of the Inertia Associated to the SCA when \mathbf{A} Model is Considered

Equation (8) is also the *chi-square* distance centered in \mathbf{H} between the conjoint probability distributions \mathbf{F} and \mathbf{A} , noted $d_{\chi^2_H}^2(\mathbf{F}, \mathbf{A})$ (Cailliez & Pagès 1976, p.449).

It is possible to perform a SCA with respect to model \mathbf{A} , denoted $CA(\mathbf{A}, \mathbf{H})$. The associated clouds N_I and N_K have inertia:

$$Inertia(N_I) = Inertia(N_K) = \sum_{l,j} \sum_{i \in I_l, k \in K_j} \frac{(a_{ik}^{lj} - f_{i\cdot}^l \cdot f_{\cdot k}^j)^2}{f_{i\cdot}^l \cdot f_{\cdot k}^j} \quad (9)$$

The inertia (9) is also the *chi-square* distance, centered in \mathbf{H} , between the conjoint probability distributions \mathbf{A} and \mathbf{H} : $d_{\chi^2_H}^2(\mathbf{A}, \mathbf{H})$.

If \mathbf{A} and \mathbf{F} have the same margins and

$$\sum_{l,i,j,k} \frac{(f_{ik}^{lj} - a_{ik}^{lj}) a_{ik}^{lj}}{f_{i\cdot}^l \cdot f_{\cdot k}^j} = 0 \quad (10)$$

the inertia associated to $CA(\mathbf{F}, \mathbf{H})$ is the sum of the inertias associated to $CA(\mathbf{F}, \mathbf{A})$ and $CA(\mathbf{A}, \mathbf{H})$:

$$d_{\chi^2_H}^2(\mathbf{F}, \mathbf{H}) = d_{\chi^2_H}^2(\mathbf{F}, \mathbf{A}) + d_{\chi^2_H}^2(\mathbf{A}, \mathbf{H}) \quad (11)$$

The demonstration can be found in the Appendix (§Appendix A.1).

In particular, the models associated with CA Intra-bands and ICA, presented hereafter, fulfill the conditions to obtain the inertia decomposition of SCA shown in (11).

3.2.2. Correspondence Analysis and Log-Linear Models

$CA(\mathbf{F}, \mathbf{A})$ describes the residuals with respect to model \mathbf{A} . Hence, it is possible to perform specific CA to analyze the residuals of a log-linear model or to eliminate some interactions in SCA to better describe the non-eliminated ones (Van der Heijden 1987, Van der Heijden, de Falguerolles & de Leeuw 1989).

The saturated log-linear model associated to a four-way table is:

$$\begin{aligned} \ln(\pi_{ijk}^{lj}) = & u + u_{A(i)} + u_{B(j)} + u_{C(k)} + u_{D(l)} + \\ & u_{AB(ij)} + u_{CA(lj)} + u_{AD(lk)} + u_{BC(ij)} + u_{BD(ik)} + u_{CD(jk)} + \\ & u_{ABC(ijk)} + u_{ABD(ijk)} + u_{CAD(ijk)} + u_{BCD(ijk)} + u_{ABCD(ijkl)} \end{aligned} \quad (12)$$

where π_{ijk}^{lj} is the probability of the cell $(\cdot)_{ijk}^{lj}$ and the u terms are the model parameters.

If \mathbf{F} (Figure 1) is the “flattened” $B(A) \times D(C)$ of a four-way CT, the independence model \mathbf{H} corresponds to the log-linear model $[AB][CD]$ ¹ estimation (A and B are jointly independent from C and D). This model is the sum of the four main effects and the first order interactions AB and CD . Then, the CA of \mathbf{F} ($CA(\mathbf{F}, \mathbf{H})$) describes the interactions AC , AD , BC , BD and those of superior order.

From a ‘true’ sub-partition structure, the row factors A and B and the column factors C and D are nested and, therefore, have no interactions between each couple. The saturated model (12) is reduced to:

$$\ln(\pi_{ijk}^{lj}) = u + u_{A(i)} + u_{B(j)} + u_{C(k)} + u_{D(l)} + u_{CA(lj)} + u_{AD(lk)} + u_{BC(ij)} + u_{BD(ik)} \quad (13)$$

In this case, the \mathbf{H} model represents all the main effects and the SCA is the description of all the interactions in (13).

3.3. Intra-Table Analysis

We denominate *Intra-Row Band/Column Analysis*, the two Intra-Table Analysis that are possible to perform on the \mathbf{F} table. We only summarize the Intra-Column Band Analysis, because the other one can be symmetrically deduced.

¹ With this notation, the model includes the whole interactions between the variables that belong to the same square brackets. For example, the $[AB][C]$ model represents the main effects and the interactions between A and B .

\mathbf{F} is considered as the juxtaposition of the J column bands, as shown by Bécue-Bertaut & Pagès (2004) in the Multiple Factor Analysis of Contingency Tables (MFACT):

$$\mathbf{F} = [\mathbf{F}_{*1} \cdots \mathbf{F}_{*j} \cdots \mathbf{F}_{*J}]$$

The Intra-Column Band CA is the CA of \mathbf{F} with respect to the Intra-Bands Independence Model, denoted \mathbf{A}^J , with general term:

$$(a^J)_{ik}^{lj} = \frac{f_{i \cdot}^{lj} f_{\cdot k}^{\cdot j}}{f_{\cdot \cdot}^{\cdot j}} \tag{14}$$

This is the estimation of the log-linear model $[ABC][CD]$ (A and B are jointly independent from D , when C is given). This model includes the interactions AB , AC , BC , CD and ABC ; thus, the $CA(\mathbf{F}, \mathbf{A}^J)$ describes the interactions, AD , BD , ABD , ACD and $ABCD$. If the subpartition structure is ‘true’, the $CA(\mathbf{F}, \mathbf{A}^J)$ describes the interactions AD and BD (see §3.2.2).

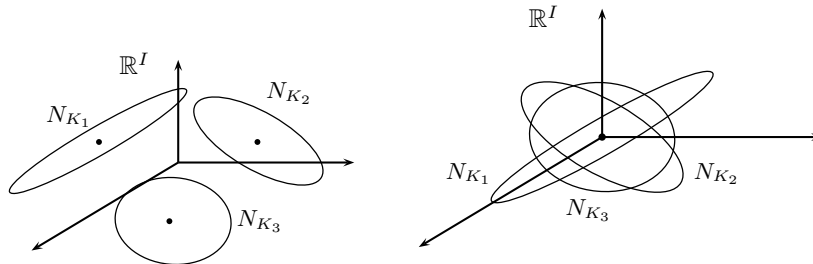
Symmetrically, the Intra-Row Band Independence Model \mathbf{A}^L , $[AB][ACD]$ (C and D are jointly independent from B , given A), includes AB , AC , AD , CD and CAD . Thus, the $CA(\mathbf{F}, \mathbf{A}^L)$ describes the interactions BC , BD , ABD , ABC and $ABCD$.

The Intra-Column Bands Analysis, $CA(\mathbf{F}, \mathbf{A}^J)$, is computed as $PCA(\mathbf{X}, \mathbf{M}, \mathbf{D})$, where \mathbf{X} is the matrix with general term:

$$x_{ik}^{lj} = \frac{f_{ik}^{lj}}{f_{i \cdot}^{\cdot j} f_{\cdot k}^{\cdot j}} - \frac{f_{i \cdot}^{lj}}{f_{i \cdot}^{\cdot j} f_{\cdot \cdot}^{\cdot j}} \tag{15}$$

and \mathbf{M} and \mathbf{D} are metric and weight matrices already defined in (2).

We observe that (15) is equal to (1) - (4): in the Intra-Column Bands CA, the subclouds N_{K_j} in the space \mathbb{R}^I are translated such as their centroids are in the origin. Figure 2a. shows the centroids of the subclouds in SCA and Figure 2b. the same subclouds, but centered in the origin. By centering, the associated inertia to $CA(\mathbf{F}, \mathbf{A}^J)$ is the Intra subclouds N_{K_j} inertia from the SCA of \mathbf{F} .



a. Subclouds associated to SCA b. Centered subclouds (Intra-Column Bands CA)

FIGURE 2: Subclouds in \mathbb{R}^I , associated to the three column bands.

3.3.1. Inertia Decomposition of SCA of \mathbf{F}

In the SCA of \mathbf{F} , the inertia of the N_K cloud in \mathbb{R}^I can be expressed as the sum of the between and intra-inertias subclouds N_{K_j} obtained replacing \mathbf{A} by $\mathbf{A}^{\mathbf{J}}$ in (11):

$$d_{\chi^2}^2(\mathbf{F}, \mathbf{H}) = d_{\chi^2}^2(\mathbf{A}^{\mathbf{J}}, \mathbf{H}) + d_{\chi^2}^2(\mathbf{F}, \mathbf{A}^{\mathbf{J}}) \quad (16)$$

The two right terms in (16) are associated, respectively, to the following CA (see Appendix Appendix A.2):

- $CA(\mathbf{A}^{\mathbf{J}}, \mathbf{H})$, which is also the SCA of the table $\mathbf{T}^{\mathbf{J}}$, with general term f_i^{lj} and dimension $I \times J$.
- $CA(\mathbf{F}, \mathbf{A}^{\mathbf{J}})$, which is the Intra-Column Bands CA of \mathbf{F} .

3.3.2. Subclouds $N_{I_l} \in \mathbb{R}^K$ from the Intra-Column Bands CA

In the Intra-Column Bands CA it is possible to obtain the centroids of the subclouds $N_{I_l} \in \mathbb{R}^K$ and to project them as illustrative elements. The general term of the coordinate (j, k) of the centroid of the sub cloud N_{I_l} is:

$$\sum_{i \in I_l} \frac{f_i^{l \cdot}}{f^{l \cdot}} \left(\frac{f_{ik}^{lj}}{f_i^{l \cdot} f_{\cdot k}^{j \cdot}} - \frac{f_i^{lj}}{f_i^{l \cdot} f^{j \cdot}} \right) = \frac{f_{\cdot k}^{lj}}{f^{l \cdot} f_{\cdot k}^{j \cdot}} - \frac{f^{lj}}{f^{l \cdot} f^{j \cdot}} \quad (17)$$

3.4. Internal Correspondence Analysis (ICA)

The Double Intra Bands CA is obtained by centring the subclouds N_{I_l} of the Intra-Column Bands CA. Then, the general term of \mathbf{X} is equal to (15) - (17):

$$x_{ik}^{lj} = \frac{f_{ik}^{lj}}{f_i^{l \cdot} f_{\cdot k}^{j \cdot}} - \frac{f_{\cdot k}^{lj}}{f_{\cdot k}^{j \cdot} f^{l \cdot}} - \frac{f_i^{lj}}{f_i^{l \cdot} f^{j \cdot}} + \frac{f^{lj}}{f^{l \cdot} f^{j \cdot}} \quad (18)$$

The Formula (18) can also be obtained centering the subclouds N_{K_j} in the Intra-Row Bands CA.

The double Intra CA or Internal Correspondence Analysis (ICA) is the $CA(\mathbf{F}, \mathbf{C})$, where \mathbf{C} is the model with general term:

$$c_{ik}^{lj} = \frac{f_{\cdot k}^{lj} f_i^{l \cdot}}{f^{l \cdot}} + \frac{f_i^{lj} f_{\cdot k}^{j \cdot}}{f^{j \cdot}} - \frac{f_i^{l \cdot} f_{\cdot k}^{j \cdot} f^{lj}}{f^{l \cdot} f^{j \cdot}} \quad (19)$$

We denote \mathbf{E} the matrix with general term $\frac{f_i^{l \cdot} f_{\cdot k}^{j \cdot} f^{lj}}{f^{l \cdot} f^{j \cdot}}$, then \mathbf{C} can be written $\mathbf{A}^{\mathbf{J}} + \mathbf{A}^{\mathbf{L}} - \mathbf{E}$ and expressed as:

$$\mathbf{C} = [\mathbf{A}^{\mathbf{J}} - \mathbf{E}] + [\mathbf{A}^{\mathbf{L}} - \mathbf{E}] + \mathbf{E} \quad (20)$$

The inertia of the SCA of \mathbf{F} can be decomposed as follows:

$$d_{\chi^2}^2(\mathbf{F}, \mathbf{H}) = d_{\chi^2}^2(\mathbf{E}, \mathbf{H}) + d_{\chi^2}^2(\mathbf{A}^{\mathbf{J}}, \mathbf{E}) + d_{\chi^2}^2(\mathbf{A}^{\mathbf{L}}, \mathbf{E}) + d_{\chi^2}^2(\mathbf{F}, \mathbf{C}) \quad (21)$$

Following Sabatier (1987), the right hand terms in (21) are (see §Appendix A.2 in the Appendix):

- SCA of table \mathbf{T} formed by the sum of the blocks (l, j) , with general term f_i^{lj} and dimension $L \times J$. This CA describes the interactions AC , i.e. between the factors defining the row and column bands.
- Intra-Tables CA of $\mathbf{T}^{\mathbf{J}}$, with general term f_i^{lj} and dimension $I \times J$. $\mathbf{T}^{\mathbf{J}}$ is a three-way table, since it is the margin of the column bands of \mathbf{F} , so factor D disappears. The Intra-Tables CA of $\mathbf{T}^{\mathbf{J}}$ corresponds to the residuals with respect to the model $[AB][AC]$ (B is independent of C , given A). The model contains the interactions AB and AC ; thus, the Intra-Tables CA describes the interactions BC and ABC .
- Intra-Tables CA of $\mathbf{T}^{\mathbf{L}}$, with the general term f_k^{lj} and dimension $L \times K$. Table $\mathbf{T}^{\mathbf{L}}$ is the margin of the row bands of \mathbf{F} , hence, it is a three way table. This Intra-Table CA describes the interactions AD and ACD , that are the residuals with respect to the model $[AC][CD]$ (A is an independent from D , when C is given).
- ICA of \mathbf{F} ($CA(\mathbf{F}, \mathbf{C})$). \mathbf{C} is not the estimation of a log-linear model, its structure is additive instead of multiplicative. Because the four CA contain all of the interactions from the $CA(\mathbf{F})$, the ICA describes the interactions that are not in the three former CA, i.e. BD , ABD , BCD and $ABCD$.

In other words, the SCA of \mathbf{F} is a global analysis that can be decomposed into four CA, where the first order interactions present in the SCA of \mathbf{F} , are separated. The inertias associated with the four CA and their relative contributions to the inertia from the SCA are indicators of the importance of these associations.

3.4.1. Intra-Bands CA as Particular Cases of ICA

Intra-Row Bands CA is a particular case of ICA because it can be obtained by considering the L row bands but only one column band with K columns. The Intra-Column Band CA can be obtained considering the J column bands but only one row band with I rows. In the former case, the terms 1 and 3 from (19) cancel one another; in the second case the terms 2 and 3 cancel one another. This justifies the name of “Internal Correspondence Analysis” (ICA) given to one dimension Intra-Tables CA by Pagès & Bécue-Bertaut (2006).

3.4.2. ICA as a Weighted PCA

ICA is the $CA(\mathbf{F}, \mathbf{C})$, i.e. the $PCA(\mathbf{X}, \mathbf{M}, \mathbf{D})$, where \mathbf{X} has the general term given in (18) and \mathbf{M} and \mathbf{D} are defined in (2). In this analysis, the representations in spaces \mathbb{R}^K and \mathbb{R}^I are symmetric: in \mathbb{R}^K the cloud N_I is divided into L

subclouds N_{I_i} ; in \mathbb{R}^I the cloud N_K is divided into J subclouds N_{K_j} . Without loss of generality, the properties are presented below in the space \mathbb{R}^K .

3.4.3. Row Clouds in \mathbb{R}^K

In ICA, the cloud N_I of the I rows is formed by the union of the L subclouds N_{I_i} , each centered in the origin. So, the coordinate of a point (l, i) represents the deviation of the point with respect to the centroid of the subcloud N_{I_i} to which it belongs (Figure 2).

Distances: the square distance between two row points is:

$$d^2[(l, i), (l', i')] = \sum_{j,k} \frac{1}{f_{\cdot k}^{\cdot j}} \left(\frac{f_{ik}^{lj} - c_{ik}^{lj}}{f_i^{\cdot l}} - \frac{f_{i'k}^{l'j} - c_{i'k}^{l'j}}{f_{i'}^{\cdot l'}} \right)^2 \quad (22)$$

Two points (l, i) and (l', i') are close to one another if their deviations to the respective model, weighted with the inverse of $f_{\cdot k}^{\cdot j}$, are similar for every (j, k) . A point (l, i) is located far from the origin if row (l, i) in table **F** differs from the model **C** (Escofier 2003, p. 120).

Transition Formulae: a row coordinate $F_s(l, i)$ on a factorial axis s is a function of the column coordinates $G_s(j, k)$ (see §Appendix A.3):

$$F_s(l, i) = \frac{1}{\sqrt{\lambda_s}} \sum_j \sum_{k \in K_j} \left(\frac{f_{ik}^{lj}}{f_i^{\cdot l}} - \frac{f_{\cdot k}^{lj}}{f_{\cdot \cdot}^{\cdot l}} \right) G_s(j, k) \quad (23)$$

Formula (23) indicates that a row (l, i) lies on the same side that the columns (j, k) whose coordinates are greater than the coordinates of the homologous columns in the l band margin.

Aids to the Interpretation: the contribution to the inertia and the quality of representation on the axes are calculated for each row point. Moreover, aids to the interpretation are defined for each subcloud N_{I_i} :

- Weight subcloud: $f_i^{\cdot l}$.
- Quality of representation on axis s : $Inertia_s(N_{I_i})/Inertia(N_{I_i})$.

$$Inertia_s(l, i) = f_i^{\cdot l} (F_s(l, i))^2$$

Therefore:

$$Inertia_s(N_{I_i}) = \sum_{i \in I_i} f_i^{\cdot l} (F_s(l, i))^2$$

In \mathbb{R}^K the contribution of a row point to the inertia of cloud N_I is:

$$Inertia(l, i) = f_i^{\cdot l} \mathbf{x}'_{li} \mathbf{M} \mathbf{x}_{li}$$

where \mathbf{x}'_{li} is the row (l, i) of **X** and $\mathbf{M} = \text{diag}(f_{\cdot k}^{\cdot j})$.

- Contribution to axis inertia: the sum of the inertia of the points belonging to the subcloud.

3.5. Intra-blocks Correspondence Analysis (IBCA)

Intra-blocks Correspondence Analysis of \mathbf{F} , denoted $IBCA(\mathbf{F})$, is defined as the CA with respect to the Intra-blocks Independence Model \mathbf{B} , using the same metrics as the SCA of \mathbf{F} . The general term of \mathbf{B} is defined by:

$$b_{ik}^{lj} = \frac{f_{i\cdot}^{lj} f_{\cdot k}^{lj}}{f_{\cdot\cdot}^{lj}} \quad (24)$$

\mathbf{B} is the estimation of the log-linear model $[ABC][ACD]$ (B is independent of D , when AC is given). This model includes the interactions AB , AC , BC , CD , AD , ABC and CAD ; thus, the IBCA ($CA(\mathbf{F}, \mathbf{B})$) describes the interactions BD , ABD , BCD and $ABCD$.

If the CT has a 'true' partition structure, the interactions AB , CD and those of superior order including them do not exist. Hence, the model \mathbf{B} contains only the interactions AC , BC and AD and IBCA describes the interactions BD (see §3.2.2).

$IBCA(\mathbf{F})$ is the $PCA(\mathbf{X}, \mathbf{M}, \mathbf{D})$ with:

- $\mathbf{M} = \text{diag}(f_{\cdot k}^j)$
- $\mathbf{D} = \text{diag}(f_{i\cdot}^l)$
- \mathbf{X} with general term given by:

$$x_{ik}^{lj} = \frac{f_{ik}^{lj} - b_{ik}^{lj}}{f_{i\cdot}^l f_{\cdot k}^j} = \frac{f_{ik}^{lj} - \frac{f_{i\cdot}^{lj} f_{\cdot k}^{lj}}{f_{\cdot\cdot}^{lj}}}{f_{i\cdot}^l f_{\cdot k}^j} \quad (25)$$

3.5.1. Centered Clouds and Subclouds

The cloud N_I formed by the I points is centered, because the margins of table \mathbf{F} and model \mathbf{B} are equal.

Each subcloud N_{I_l} formed by the I_l points belonging to the row band l are centered, using the weights $\frac{f_{i\cdot}^l}{f_{\cdot\cdot}^l}$:

$$\frac{1}{f_{\cdot\cdot}^l} \sum_{i \in I_l} f_{i\cdot}^l \frac{f_{ik}^{lj} - \frac{f_{i\cdot}^{lj} f_{\cdot k}^{lj}}{f_{\cdot\cdot}^{lj}}}{f_{i\cdot}^l f_{\cdot k}^j} = \frac{1}{f_{\cdot\cdot}^l} \left(\sum_{i \in I_l} \frac{f_{ik}^{lj}}{f_{\cdot k}^j} - \sum_{i \in I_l} \frac{f_{i\cdot}^{lj} f_{\cdot k}^{lj}}{f_{\cdot\cdot}^{lj} f_{\cdot k}^j} \right) = \frac{f_{\cdot k}^{lj} - f_{\cdot k}^{lj}}{f_{\cdot\cdot}^l f_{\cdot k}^j} = 0$$

3.5.2. Distances

The square distance between two row points is:

$$d^2[(l, i), (l', i')] = \sum_{j,k} \frac{1}{f_{\cdot k}^j} \left(\frac{f_{ik}^{lj} - b_{ik}^{lj}}{f_{i\cdot}^{l\cdot}} - \frac{f_{i'k}^{l'j} - b_{i'k}^{l'j}}{f_{i'\cdot}^{l'\cdot}} \right)^2 \quad (26)$$

Two points (l, i) and (l', i') are close to each other if they similarly differ from the model. Each difference is pondered by $1/f_{\cdot k}^j$. Therefore, a point (l, i) is far from the origin when the row (l, i) of table \mathbf{F} strongly differs from the model \mathbf{B} (Escofier 2003, p.120).

3.5.3. Transition Formulae

The formulae allowing the simultaneous representation of row and column points, as well as their interpretation, are:

$$\begin{aligned} F_s(l, i) &= \frac{1}{\sqrt{\lambda_s}} \sum_{j,k} \left(\frac{f_{ik}^{lj} - b_{ik}^{lj}}{f_{i\cdot}^{l\cdot}} \right) G_s(j, k); \\ G_s(j, k) &= \frac{1}{\sqrt{\lambda_s}} \sum_{l,i} \left(\frac{f_{ik}^{lj} - b_{ik}^{lj}}{f_{\cdot k}^j} \right) F_s(l, i) \end{aligned} \quad (27)$$

Attractions between row and column profiles exist when the observed frequencies are greater than the values in the model.

3.5.4. Aids to the Interpretation

The aids to interpretation used in CA are also available in IBCA, i.e. contribution to the axis inertia and square cosines. Similarly, the aids associated to subclouds N_{K_j} and N_{I_l} are expressed in ICA.

3.5.5. Intra-Blocks CA only in one Dimension

If only one dimension structure is considered, model \mathbf{B} becomes the intra-row bands independence or the intra-column bands model, depending of the case. Thus, the Intra Bands CA can also be considered as Intra-Blocks Analysis in one single dimension.

IBCA has the advantage of being associated with a log-linear model, while ICA allows us to split the inertia of the clouds associated to SCA in four addends, each corresponding to a CA (see 3.4).

4. Superimposed Representation of the Partial and Global Clouds over a Common Referential

In ICA or IBCA, the global representation of the cloud N_I , in the row space, is obtained considering the whole K coordinates for each row point (l, i) . The sub-partition of the columns into J bands also permits to consider each row from the point of view of each J band. Thus, there are J points, denoted $(l, i)^j$ and called *partial points*, considered and projected as illustrative points. The simultaneous projections of global and partial points are denominated superimposed representations.

4.1. Projection of the Partial Clouds

The projections of the partial clouds are defined as done by Pagès (2004) in the frame of multiple factor analysis (MFA).

- Each column j induces the partial cloud $N_I^j \subset \mathbb{R}^{K_j} \subset \mathbb{R}^K = \bigoplus_j \mathbb{R}^{K_j}$, \mathbf{M}_j is the metrics in \mathbb{R}^{K_j} obtained from \mathbf{M} , the coordinates of the points N_I^j are the rows of \mathbf{X}_{*j} and the coordinates of these points in \mathbb{R}^K are the rows of the matrix $\tilde{\mathbf{X}}_{*j}$ defined as:

$$\tilde{\mathbf{X}}_{*j} = [\mathbf{0} \ \cdots \ \mathbf{0} \ \mathbf{X}_{*j} \ \mathbf{0} \ \cdots \ \mathbf{0}]$$

- The union of the J partial clouds form the cloud N_I^J with IJ points, that can also be considered as the union of the I clouds $N_{(l,i)}^J$, each with J partial points $(l, i)^j$ belonging to the same row (l, i) .
- The inertia of the cloud N_I^J can be expressed as WithinInertia + BetweenInertia subclouds $N_{(l,i)}^J$.
- The cloud of the centroids of the I partial clouds $N_{(l,i)}^J$ is $\frac{1}{J} \sum_j \tilde{\mathbf{X}}_j$. To force $F_s(i)$ to lie at the centroid of the J partial points $F_s^j(i)$, the rows of $\tilde{\mathbf{X}}_j$, called partial, are projected as illustrative but dilated by J .

4.2. Restricted Transition Formulae

In (23), each addend j is the restricted formula to the columns K_j belonging to its band. This formula allows us to interpret the position of the partial rows $(l, i)^j$ on the factorial axis s , similarly to the global coordinates:

$$F_s(l, i)^j = \frac{1}{\sqrt{\lambda_s}} \sum_{k \in K_j} \left(\frac{f_{ik}^{lj}}{f_{i\cdot}^l} - \frac{f_{\cdot k}^{lj}}{f_{\cdot\cdot}^l} \right) G_s(j, k) \quad (28)$$

Formula (28) indicates that a row $(l, i)^j$ is placed on the same side that columns $k \in K_j$ whose profile coordinates obtained from the \mathbf{F} table are greater than the profile coordinates obtained from the margin of its band l . The interpretation of the superimposed representations is mainly supported by these formulae. In the graphic representations, the coordinates are amplified by J .

By exchanging the indices, the restricted transition formulae for the partial columns are deduced.

4.3. Aids to the interpretation of the Partial Clouds

In the superimposed representation, for each factorial axis s there are:

- IJ partial coordinates $F_s^j(l, i)$
- I global coordinates $F_s(l, i)$

These points form different projected clouds:

- I partial clouds $N_{(l,i)}^J$, each with centroid $F_s(l, i)$
- J partial clouds N_I^j
- L clouds $N_{I_l}: \{F_s(l, i); i \in I_l\}$.

Since the partial rows $(l, i)^j$ are illustrative they do not contribute to the inertia of the axes. For the partial clouds, the aids to the interpretation are defined as detailed hereafter.

4.3.1. Quality of the Representation of the Partial Clouds

The quality of representation on axis s of each partial cloud N_I^j is computed as the ratio between the projected inertia and the inertia in \mathbb{R}^K .

4.3.2. Similarity Measure between Partial Clouds

The total inertia of N_I^J can be decomposed into within and between inertia of clouds $N_{(l,i)}^J$.

The ratio *BetweenInertia/TotalInertia*, computed for each factorial axis s , is a measure of the proximity of the partial points belonging to the same row and therefore of the global similarity between the J partial clouds projected on axis s . If this ratio is close to 1, the homologous points $\{(l, i)^j; j = 1, \dots, J\}$ are close to each other and the axis s represents a structure common to the different column bands (Pagès 2004, pp.8-9).

4.3.3. Row Contributions to the Within-Inertia

The within-inertia can be decomposed into the contributions of each row, in order to detect differences between the several points of view represented by the column bands. Then, it is possible to identify both the most heterogeneous and homogeneous, in order to interpret the global relations.

It is possible to calculate the contribution to the within-inertia of N_I^J for the cloud associated with a partial row $N_{(l,i)}^J$.

4.4. Zero Partial Points into the Blocks in ICA versus IBCA

Zero Row Inside a Block: in ICA, if the values of the row (l, i) belonging to a column band j of the contingency table are zeros, the partial point does not always lie at the origin. In fact: if $f_{ik}^{lj} = 0, \forall k \in K_j$, then $f_i^{lj} = 0$ but the general term of \mathbf{X} (18) for the cells of row (l, i) in column band j is $x_{ik}^{lj} = \frac{1}{f_{..}^{lj}} \left(\frac{f_{..}^{lj}}{f_{..}^{lj}} - \frac{f_{.k}^{lj}}{f_{.k}^{lj}} \right)$, and this term is not necessarily zero.

In this case, the interpretation of the superimposed representations becomes difficult. Some points belonging to null profiles can lie close to points belonging to non-null profiles.

IBCA solves this problem because the partial point associated with a row of zeros lies at the origin: as $f_{ik}^{lj} = 0; \forall k \in K_j$ then $f_i^{lj} = 0$, thus the cells of \mathbf{X} (25) belonging to the row (l, i) into the column band j are zeros.

Zero Column Inside a Block: in ICA, if the values of a column (j, k) belonging to a row band l of the contingency table are zeros, the partial point is not at the origin, while in IBCA, it is always at the origin. These results can be obtained by exchanging the indices in the former paragraph.

A Block of Zeros: when all the cells inside block \mathbf{F}_{lj} are zeros, the cells of the model (\mathbf{C}_{lj}) inside the block are also zeros; then, the cells of the block \mathbf{X}_{lj} are also zeros. In this block, the cells of model \mathbf{B}_{lj} are not defined, but this problem can be solved defining these cells as zeros.

4.5. Outliers

When few profiles strongly differ from the others, the first axis of the SCA enhance that difference and might hide the differences among the rest of the points. In this case, there are two ways to proceed: 1) to observe the differences on the following axes or 2) to perform the analysis again without the outliers and eventually project them as illustrative elements. These ways to proceed can be used in Intra-Tables CA, ICA and IBCA.

5. Example: Colombia Regional Scores for Secondary Education Standardized Tests

The *Instituto Colombiano para el fomento de la Educación Superior* (ICFES) performs nation-wide secondary education quality assessment based on public standardized tests. Schools are classified into seven levels, according to their scores, ranging from: *very inferior*, *inferior*, *low*, *medium*, *high*, *superior* to *very superior*. The first two categories were joined into one named *inferior* and the last three into another named *high*, leaving four levels. Thus, score is a categorical variable with four levels.

To illustrate the application of the methods proposed in the first sections, the schools classification from their scores in the 2008 tests was used, together with the following information:

1. School attendance shifts: full day, morning and afternoon including evening, Saturdays and Sundays;
2. The Colombian administrative system: Colombia is divided into 33 departments, including Bogotá as capital district. The five departments with less than one hundred thousand inhabitants were collapsed to form a “fictitious department” named *P01*, thus leaving 29 departments.
3. Population size: the departments are grouped into 5 categories depending on their population: *P5* more than two million inhabitants, *P4* between one and two million, *P3* between five hundred thousand and one million and *P2* between one hundred thousand and five hundred thousand. Department *P01* is included into size-group *P2*.

Our prime objective is the comparison of the departments according to their schools standardized tests scores. The departments are grouped according to their population size, since this variable may hide regional differences. The same rationale leads us to consider the school attendance shifts because, generally, students attending full day tuition present advantages over their peers attending partial shifts.

To achieve the main objective, the contingency table (CT) is structured as *department (group) × score (school attendance shift)* (Table 1). According to the notations used in the first sections, four factors are considered: *A* department size-group, *B* department, *C* school attendance shift and *D* score. Since the departments are nested into size-groups, the rows have a “true” sub-partition structure. We have to deal with a CT with $I = 29$ rows and $K = 12$ columns. The 29 rows are the departments divided into $L = 4$ size groups with $I_1 = 7$, $I_2 = 8$, $I_3 = 7$, $I_4 = 7$, according to their population.

The 12 columns correspond to the cross categories of *school attendance shifts × scores*. We consider these 12 columns as divided into $J = 3$ groups according to the three school attendance shifts. Each of the 12 blocks corresponds to a subtable with, in rows, the departments of a given size-group and, in columns, the scores of a given school attendance shift.

The profile for the Choco department is an outlier, not considered as active in the analysis. Being located far from the other departments, it is not projected as illustrative. The table shows the high count of Chocó' schools classified into an inferior score. Therefore, the active table has $I = 28$ departments and $I_4 = 6$.

TABLE 1: Colombian schools classified by departments, school attendance shift and standardized public tests score in 2008.

GRO UP	COD. DEPART MENT	SCHOOL ATTENDANCE SHIFT											
		FULL DAY				MORNING				AFTERNOON			
		infe rior	low	med ium	high	infe rior	low	med ium	high	infe rior	low	med ium	high
P5	BOG Bogotá	5	40	101	309	9	79	219	241	15	171	179	61
	ANT Antioquia	63	180	116	105	38	105	96	90	125	156	84	29
	VAL Valle	35	93	72	81	51	140	118	132	62	113	55	19
	CUN Cundi.	19	80	81	103	11	90	114	50	40	84	33	7
	ATL Atlántico	31	48	22	37	72	62	48	48	106	61	32	15
	SAN Santander	7	27	51	61	10	40	78	79	30	51	24	20
BOL Bolívar	31	31	8	25	90	67	28	30	77	59	14	11	
P4	NAR Nariño	8	15	21	16	31	50	63	56	21	33	26	10
	COR Córdoba	18	32	18	11	35	54	16	9	41	38	13	3
	TOL Tolima	8	19	30	26	28	77	57	28	29	40	21	7
	CAU Cauca	36	56	27	8	24	53	32	24	18	27	13	11
	NSA NorSantander	6	39	20	23	11	40	37	31	31	20	20	7
	BOY Boyacá	6	48	74	40	4	31	52	25	14	39	21	8
	MAG Magdalena	31	19	4	6	58	53	14	12	58	37	6	2
HUI Huila	7	37	42	29	1	16	27	12	24	30	14	10	
P3	CAL Caldas	11	37	26	26	8	38	54	21	10	18	8	2
	CES César	0	16	8	11	9	50	23	15	36	37	19	3
	RIS Risaralda	2	14	14	20	0	25	30	24	12	28	14	5
	MET Meta	4	12	15	6	7	45	24	19	22	21	10	7
	SUC Sucre	9	6	5	3	29	51	19	9	30	28	10	5
	LAG Guajira	7	6	7	8	12	30	6	7	24	10	3	1
QUI Quindío	0	6	7	12	1	17	31	18	10	20	7	1	
P2	CHO Chocó	23	9	6	1	26	10	3	1	20	5	0	0
	CAQ Caquetá	8	10	5	1	2	21	11	7	14	9	5	1
	PUT Putumayo	1	7	10	10	4	6	7	4	4	5	2	0
	CAS Casanare	3	12	10	6	2	16	16	4	8	13	1	1
	ARA Arauca	3	1	5	2	3	7	12	7	7	5	3	1
	GUV Guaviare	0	2	2	0	1	4	1	2	0	2	1	0
	PO1 <100 inh.	6	16	7	2	2	7	3	3	3	4	2	0

Department size groups (inhabitants in millions): P5: more than two,
P4: between one and two, P3: between 0.5 and one, P2: less than 0.5.

5.1. Simple CA on the Global Table

In the factorial planes, each axis inertia and the corresponding percentage in relation to the global inertia are specified.

The simple CA total inertia is equal to 0.2648. The first three axes retain 84.2%: 0.15 (59.9%), 0.5 (18.7%) and 0.02 (8.6%). The first two axes retain an inertia over than the average. According to the sub-partitions structure associated to the CT, the inertia is decomposed into (see §3.4):

- 0.0062 (2.3%) department group - school attendance shift association;
- 0.0442 (16.7%) department group - score and department group - performance - school attendance shift associations;
- 0.0281 (10.6%) department - school attendance shift association;
- 0.1863 (70.4%) department - score and department - performance - school attendance shift associations.

Figure 3a. shows the representation of the departments, and the size-groups as illustrative, on the first factorial plane issued from simple CA applied to the table crossing departments and *scores* \times *school attendance shifts*. Figure 3b. shows the representation of the *scores* \times *school attendance shifts* and the *school attendance shifts* as illustrative, on the same plane. A Guttman (parabola) effect is observed, more tidy in the scores trajectories corresponding to full day and morning shifts. The scores are sorted out on the first axis. The second axis opposes full day shift (on the positive part) to morning and afternoon shifts (on the negative part). Since the sub-clouds are not centered on their own centroid, the second axis opposes departments with a high proportion of full day attendance schools to departments with a high proportion of morning or afternoon attendance shifts. The former are mostly concentrated in the less populated size-groups of departments. This opposition is of no interest in the context of our study.

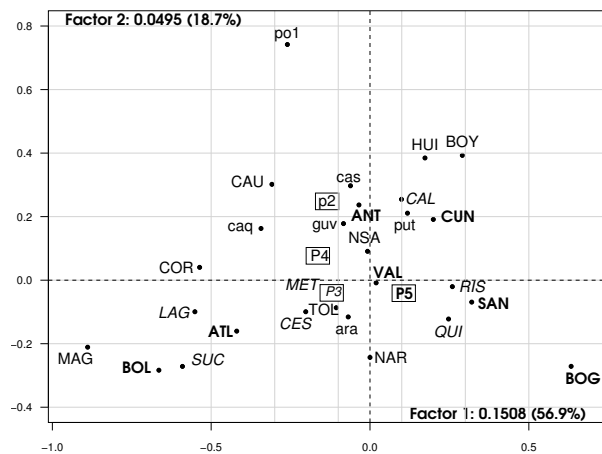
5.2. Intra-School-Attendance-Shift Analysis

The Intra-school-attendance-shift CA allows for removing the variability due to the different profiles of the school-attendance-shifts from one department to another. In this analysis, the inertia is equal to 0.2143 (80.9% of the simple CA's inertia), corresponding basically to the relationship between departments and scores. The first factorial plane retains 83.4% of the inertia and, according to the eigenvalue structure, well synthesizes the results of this analysis.

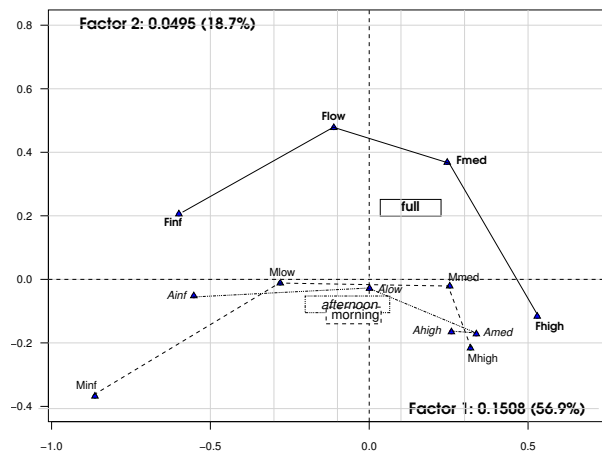
This analysis differs from the simple CA because of the re-centering of the school attendance-shift clouds so that their centroids coincide with the global centroid. In this latter analysis, the departments are more clearly sorted out depending on their schools scores because the interaction between departments and school attendance shifts has been eliminated (Figure 4). The Guttman effect also becomes clearer with this recentering. The departments' scores seem to be linked to their population size. The ICA and the IBCA will allow us compare the scores of the size-groups from an internal point of view.

5.3. ICA and IBCA

The ICA and IBCA results are very similar with total inertias equal to 0.1863 and 0.1856, respectively (70.4 % and 70.1 % of the simple CA's inertia). IBCA main results are described, since they allow for better superimposed representations (§4.4).



a. Departments and centroids of the size-groups: P_5 , P_4 , P_3 y P_2 as illustrative.



b. Scores \times school attendance shifts and the school attendance shifts as illustrative.

FIGURE 3: First factorial plane of the Simple Correspondence Analysis.

Two axes (75.6% of inertia) are retained in the IBCA. Table 2 presents the aids to interpret sub-clouds for both school attendance and department groups. The influence of the sub-clouds in the analysis depends on the weight of each band which is proportional to the percentage of schools that they contain (weight).

Figure 5 shows the simultaneous global representation of rows and columns on the IBCA first factorial plane. This graph synthesizes the CT's analysis. The departments are sorted out by scores issued from the public standardized tests, along a parabola. Bogotá obtained the best scores while Bolivar and Magdalena obtained the worst. Results show that the departments of Bolivar, Magdalena, Atlántico, La Guajira, Sucre and Cordoba, all from the Caribbean Region, obtained inferior results. Cordoba stands out in this region because it has a greater

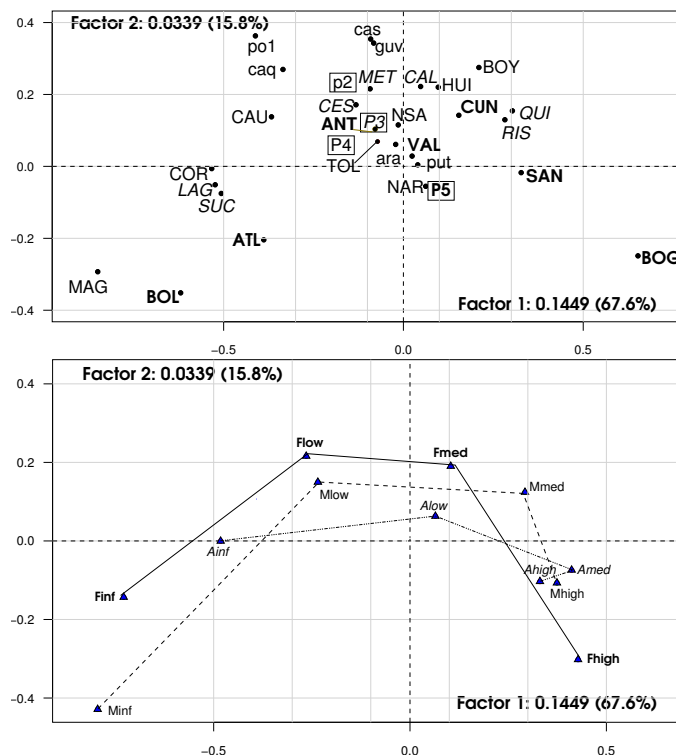


FIGURE 4: First factorial plane of the Intra-school-attendance-shift CA.

TABLE 2: Aids to the interpretation of row and column bands in the IBCA of schools

A. School attendance (column bands)

	Total		Comp1			Comp2			Plane	
	Cont.Inertia x10000	%	Cont.Inertia x10000	%	Cali dad %	Cont.Inertia x10000	%	Cali dad %	Cali dad %	Weight %
School day										
Full	636	34.3	376	29.0	59.2	153	60.3	24.0	83.2	30.5
Morning	806	43.4	611	47.0	75.9	90	35.5	11.2	87.0	40.5
Afternoon	414	22.3	312	24.0	75.3	11	4.2	2.6	77.9	29.0
Total	1856	100.0	1299	100.0		254	100.0			100.0

B. Department groups (row bands)

Group	Total		Comp1			Comp2			Plane	
	Cont.Inertia x10000	%	Cont.Inertia x10000	%	Cali dad %	Cont.Inertia x10000	%	Cali dad %	Cali dad %	Weight %
P5	1162	62.6	872	67.1	75.0	197	77.7	17.0	92.0	58.0
P4	459	24.7	310	23.9	67.5	33	12.9	7.1	74.7	25.2
P3	179	9.6	108	8.3	60.6	14	5.6	7.9	68.4	13.1
P2	56	3.0	9	0.7	16.1	10	3.8	17.4	33.5	3.7
Total	1856	100.0	1299	100.0		254	100.0			100.0

percentage of schools in medium levels. As a rule, the most standing out departments belong to the Andean region. Among the less populated size-group of departments, Arauca (Llanos Orientales) and Putumayo (Amazonia) stand out.

One hypothesis is that these last two departments have succeeded in transferring part of the oil production royalties to the educational system.

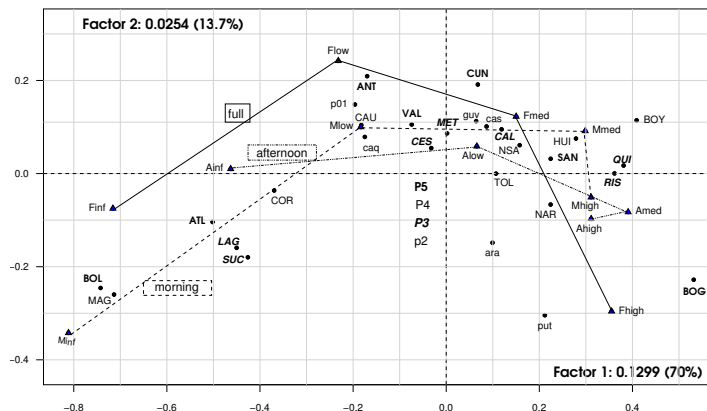


FIGURE 5: First factorial plane of the IBCA for $CT_{department}(group) \times score(shift)$.

5.4. Superimposed Representations in the IBCA

The first axis is a dispersion axis common to the three partial clouds of departments (one for each school attendance shift; ratio BetweenInertia/WithinInertia = 81.1%). Concerning the four score clouds (one for each group of departments, ratio BetweenInertia/WithinInertia = 46.9%), this axis corresponds both to common and specific effects of the clouds. On the second axis, only one third of the inertia corresponds to a common effect to both shift (ratio = 29.9%) and groups of departments (ratio = 28.8%) (see §4.3.2).

5.4.1. Departments

Three partial points correspond to each department in the superimposed representations, one for each shift. The global point is the average of those three points. Figure 6a. shows this representation for the most populated department group.

Bolivar is the partial cloud with the greatest dispersion (0.0028 of within inertia), morning shift has the lowest average score among the departments of the most populated size-group. Table 3 shows that 41.9% of the schools with morning shifts are classified into the inferior category, more than three times the average of the most populated departments (12.6%). In Bolivar, the percentages of these schools classified in medium and high categories are less than half the average. Bogotá has obtained the best results in all three shifts, being the full day shift the most outstanding with 67.9% of its schools in high level, over the average equal to 38.7%. Bogotá and Santander are more alike from the morning shift point of view than for the full day shift. Atlántico and Bolivar's scores are similar in full day

and afternoon shifts but they differ from the morning shift, with better scores for Atlántico.

5.4.2. Columns: $score(shift)$

Each global point represents one shift and one score category. The global point is the centroid of four partial points (one for each group of departments). Figure 6b. shows the superimposed representation of the score categories corresponding to full day shift (see §4). In this shift, the scores differ mostly from the most populated departments ($P5$ and $P4$).

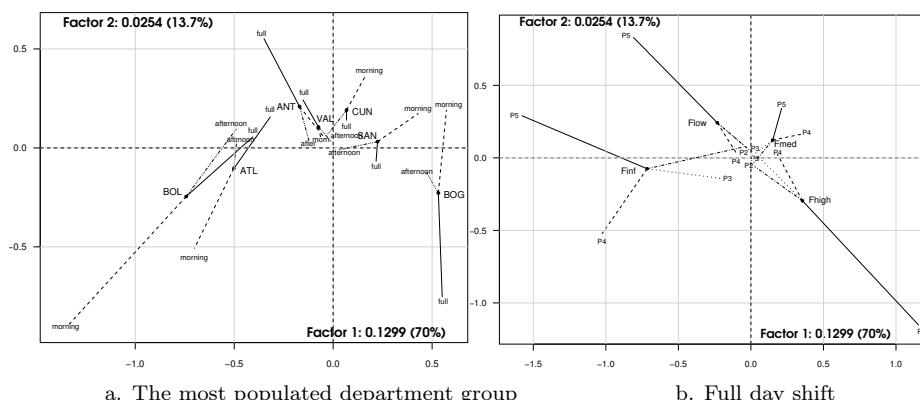


FIGURE 6: Superimposed representation on the IBCA first factorial plane.

5.5. Separate CA for the 12 Blocks

ICBA can be considered as a comparison method for the profiles corresponding to each block. For a given block, the Intra-blocks independence model is the independence model of the table considered separately. The changes in the IBCA, in relation to the separate CA of each block, concerning metrics and weights, are the price that must be paid to have a common reference framework. Figure 7 shows some of the 12 separate correspondence analyses. The axes have been rotated for an easier comparison to one another and to IBCA.

For instance, the planes corresponding to the three shifts in the most populated departments ($P5$) show similar trends that are kept in IBCA.

6. Conclusions

Various correspondence analyses, useful for the description of contingency tables with sub-partition structures both in rows and columns were demonstrated. An extension concerning the theoretical sections presented in Bécue-Bertaut et al. (2005), was made, putting emphasis on the Double Intra Correspondence Analysis

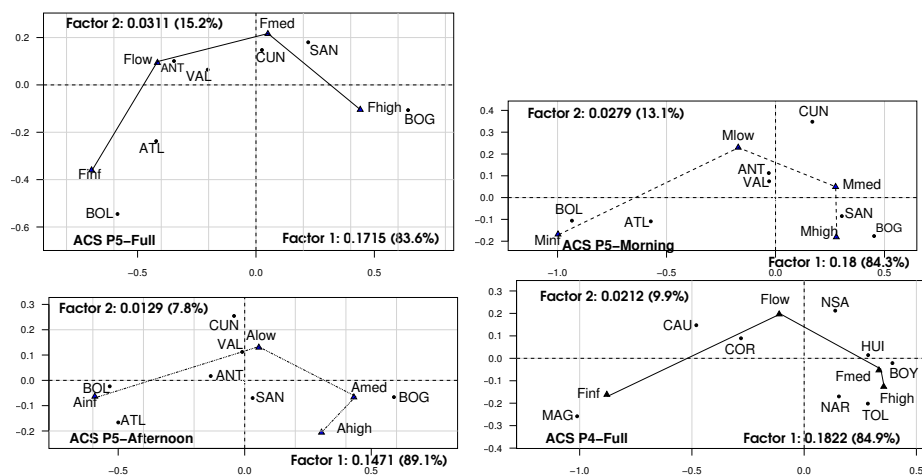


FIGURE 7: First factorial planes of some separate CA.

or Internal Correspondence Analysis (ICA) (§3.4). The Intra-Block Correspondence Analysis (IBCA) was proposed (§3.5). The Intra-Row Bands and Intra-Column-Bands CA are particular cases of ICA and IBCA, whenever the subpartition structure is considered in one single dimension (§3.4.1 and §3.5.5).

The decomposition of a simple CA for a CT with sub-partition structure in rows and columns, presented by Sabatier (1987), was demonstrated, based on the inertia decomposition of the simple CA into four addends. Thus, the four CA to each addend were derived (§3.1.2 and Appendix A.2).

The relation between correspondence analysis and log-linear models applied to multiple ways contingency tables were used to show the interactions described by the ICA and IBCA (§3.2.2).

In the superimposed representation of ICA, some points belonging to null profiles, inside a block, which can lie close to points belonging to non-null profiles. IBCA solves this problem because the partial points belonging to null profiles which always lay at the origin (§4.4).

The methods presented in this work were illustrated by their application to the standardized public test data collected from the Colombian secondary education students in 2008. IBCA provides, in a synthesized way, the regional differences between departments regarding the schools scores. The superimposed representations allow us to compare, on the one hand, the departments scores through the different shifts and, on the other hand, the score-shifts through the groups of departments (§5).

Acknowledgements

We thank both anonymous judges and the Editor in Chief for carefully reading the article and for their valuable feedback which contributed to its improvement;

Professor Jérôme Pagès for his comments about the superimposed representations; D. Castin for her tidy revision of the English version. This document is derived from the doctoral thesis in Statistics presented at the Faculty of Science of the Universidad Nacional de Colombia Sede Bogotá (Pardo 2011), directed and co-directed by the second and third co-authors respectively. The research was funded by the Universidad Nacional de Colombia through a study grant. Computation was done using `pamctdp` developed in R (R Development Core Team 2010) using functions of the `ade4` package (Thioulouse, Chessel, Dolédec & Olivier 1997). `FactoMineR` (Husson, Josse, Le & Mazet 2009) and `FactoClass` (Pardo & DelCampo 2007) packages were used for parts of the graphic functions.

[Recibido: junio de 2011 — Aceptado: mayo de 2013]

References

- Bécue-Bertaut, M. & Pagès, J. (2004), 'A principal axes method for comparing multiple contingency tables: MFACT.', *Computational Statistics & Data Analysis* **45**(3), 481–503.
- Bécue-Bertaut, M., Pagès, J. & Pardo, C. (2005), Contingency table with a double partition on rows and columns. Visualization and comparison of the partial and global structures, *in* J. Janssen & P. Lenca, eds, 'Proceedings ASMDA 2005', Applied Stochastic Models and Data Analysis. Brest, France. May, 17–20, 2005, ENST Bretagne, pp. 355–364.
*<http://conferences.telecom-bretagne.eu/asmda2005/IMG/pdf/proceedings/355.pdf>
- Cailliez, F. & Pagès, J. (1976), *Introduction à l'Analyse des Données*, Smash, Paris.
- Cazes, P., Chessel, D. & Dolédec, S. (1988), 'L'analyse des correspondances internes d'un tableau partitionné. Son usage en hydrobiologie', *Revue de Statistique Appliquée* **36**(1), 39–54.
- Dolédec, S. & Chessel, D. (1991), 'Recent developments in linear ordination methods for environmental sciences', *Advances in Ecology* **1**, 133–155.
- Escofier, B. (1983), Generalisation de l'analyse des correspondances a la comparaison de tableaux de frequence, Rapports de Recherche 207, Institut National de Recherche en Informatique et en Automatique, Centre de Rennes. IRISA.
*<http://hal.inria.fr/inria-00076351>
- Escofier, B. (1984), 'Analyse factorielle en référence à un modèle. Application à l'analyse de tableaux d'échanges', *Revue de Statistique Appliquée* **32**(4), 25–36.
- Escofier, B. (2003), *Analyse des Correspondances. Recherches au Coeur de l'Analyse des Données*, Presses Universitaires de Rennes - Société Française de Statistique, Rennes, France.

- Escofier, B. & Pagès, J. (1982), Comparaison de groupes de variables définies sur le même ensemble d'individus, *Rapports de Recherche* 149, INRIA-IRISA, Rennes, France.
*<http://hal.inria.fr/inria-00076411>
- Escofier, B. & Pagès, J. (1992), *Análisis Factoriales Simples y Múltiples. Objetivos, Métodos e Interpretación*, Universidad del País Vasco, Bilbao.
- Husson, F., Josse, J., Le, S. & Mazet, J. (2009), *FactoMineR: Factor Analysis and Data Mining with R*. R package version 1.12.
*<http://CRAN.R-project.org/package=FactoMineR>
- Laloë, D., Moazami-Gourdarzi, K. & Chessel, D. (2002), Contribution of individual markers to the analysis of the relationships among breeds by correspondence analysis, in '7th World Congress on Genetics Applied to Livestock Production', Montpellier, France.
*<http://pbil.univ-lyon1.fr/R/articles/arti110.pdf>
- Lebart, L., Morineau, A. & Warwick (1984), *Multivariate Descriptive Statistical Analysis*, Wiley, New York.
- Lebart, L., Piron, M. & Morineau, A. (2006), *Statistique exploratoire multidimensionnelle. Visualisation et inférence en fouilles de données*, 4 edn, Dunod, Paris.
- Lobry, J. & Necsulea, A. (2006), 'Synonymous codon usage and its potential link with optimal growth temperature in prokaryotes', *Gene* **385**, 128 – 136.
- Lobry, J. R. & Chessel, D. (2003), 'Internal Correspondence Analysis of Codon and Amino-Acid Usage in Thermophilic Bacteria', *Journal of Applied Genetics* **44**(2), 235–261.
- Pagès, J. (2004), 'Multiple Factor Analysis: Main Features and Application to Sensory Data', *Revista Colombiana de Estadística* **27**(1), 1–26.
- Pagès, J. & Bécue-Bertaut, M. (2006), Multiple Factor Analysis for Contingency Tables, in M. Greenacre & J. Blasius, eds, 'Multiple Correspondence Analysis and Related Methods', Chapman and Hall/CRC, chapter 13, pp. 299–326.
- Pardo, C. & DelCampo, P. (2007), 'Combinacion de metodos factoriales y de analisis de conglomerados en r: el paquete factoclass', *Revista Colombiana de Estadística* **30**(2), 231–245.
*www.matematicas.unal.edu.co/revcoles
- Pardo, C. E. (2011), Métodos en ejes principales para tablas de contingencia con estructuras de partición en filas y columnas, Tesis para optar al título de Doctor en Ciencias-Estadística, Universidad Nacional de Colombia. Facultad de Ciencias. Departamento de Estadística, Bogotá.

R Development Core Team (2010), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

*<http://www.R-project.org/>

Ramírez, J. R. & Martínez, G. (2010), 'Análisis de correspondencia a partir de una muestra peobabilística', *Revista Colombiana de Estadística* **33**, 273–293.

Sabatier, R. (1987), *Methodes factorielles en analyse des données: aproximations et prise en compte de variables concomitantes*, Doctorat d'Etat, Université des Sciences et Techniques du Languedoc, Montpellier.

Sabatier, R., Lebreton, J. & Chessel, D. (1989), Principal Component Analysis with Instrumental Variables as a Tool for Modelling Composition Data, in R. Coppi & S. Bolasco, eds, 'Multiway Data Analysis', Elsevier, Amsterdam, pp. 341–350.

Tenenhous, M. & Young, F. (1985), 'An Analysis and Synthesis of Multiple Correspondence Analysis, Optimal Scaling, Homogeneity Analysis and other Methods for Quantifying Categorical Multivariate Data', *Psychometrika* **50**(1), 91–119.

Thioulouse, J., Chessel, D., Dolédec, S. & Olivier, J. (1997), 'ADE-4: a multivariate analysis and graphical display software', *Statistical and Computing* **7**, 75–83.

*<http://pbil.univ-lyon1.fr/ADE-4/ADE-4F.html>

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 using correspondence analysis and log-linear analysis', *Applied Statistics* pp. 249–292.

Appendix A. Some Proofs

Appendix A.1. Proofs of Formulae (11) and (10), p.122

The CA inertia can be decomposed as:

$$\begin{aligned} \sum_{l,i,j,k} \frac{\left((f_{ik}^{lj} - a_{ik}^{lj}) + (a_{ik}^{lj} - f_{i \cdot}^l f_{\cdot k}^j) \right)^2}{f_{i \cdot}^l f_{\cdot k}^j} &= \sum_{l,i,j,k} \frac{(f_{ik}^{lj} - a_{ik}^{lj})^2}{f_{i \cdot}^l f_{\cdot k}^j} + \sum_{l,i,j,k} \frac{(a_{ik}^{lj} - f_{i \cdot}^l f_{\cdot k}^j)^2}{f_{i \cdot}^l f_{\cdot k}^j} + \\ &+ \sum_{l,i,j,k} \frac{(f_{ik}^{lj} - a_{ik}^{lj})(a_{ik}^{lj} - f_{i \cdot}^l f_{\cdot k}^j)}{f_{i \cdot}^l f_{\cdot k}^j} \end{aligned}$$

Then (11) is true if the last addend is equal to zero. We have:

$$\sum_{l,i,j,k} \frac{(f_{ik}^{lj} - a_{ik}^{lj})a_{ik}^{lj} - (f_{ik}^{lj} - a_{ik}^{lj})f_{i\cdot}^{l\cdot} f_{\cdot k}^{j\cdot}}{f_{i\cdot}^{l\cdot} f_{\cdot k}^{j\cdot}} = \sum_{l,i,j,k} \frac{(f_{ik}^{lj} - a_{ik}^{lj})a_{ik}^{lj}}{f_{i\cdot}^{l\cdot} f_{\cdot k}^{j\cdot}} - \sum_{l,i,j,k} (f_{ik}^{lj} - a_{ik}^{lj})$$

The last term is zero because the totals of \mathbf{F} and \mathbf{A} are equal to one another.

Appendix A.2. Decomposition of the Inertia Associated to the SCA

Sabatier, Lebreton & Chessel (1989) demonstrated the decomposition of the inertia using the correspondence analysis with respect to instrumental variables. We show the inertia decomposition by expressing $\mathbf{F} - \mathbf{H}$ as the sum of differences and using these differences expressed as general terms to calculate the inertia. Then, each inertia term is associated to a CA.

$$\mathbf{F} - \mathbf{H} = (\mathbf{F} - \mathbf{C}) + (\mathbf{A}^J - \mathbf{E}) + (\mathbf{A}^L - \mathbf{E}) + (\mathbf{E} - \mathbf{H})$$

The inertia associated to the SCA of \mathbf{F} is:

$$\begin{aligned} \sum_{l,i,j,k} \frac{(f_{ik}^{lj} - h_{ik}^{lj})^2}{h_{ik}^{lj}} &= \frac{\left((f_{ik}^{lj} - c_{ik}^{lj}) + ((a^J)_{ik}^{lj} - e_{ik}^{lj}) + ((a^L)_{ik}^{lj} - e_{ik}^{lj}) + (e_{ik}^{lj} - h_{ik}^{lj}) \right)^2}{h_{ik}^{lj}} \\ &= \sum_{l,i,j,k} \frac{(f_{ik}^{lj} - c_{ik}^{lj})^2}{h_{ik}^{lj}} + \sum_{l,i,j,k} \frac{((a^J)_{ik}^{lj} - e_{ik}^{lj})^2}{h_{ik}^{lj}} + \sum_{l,i,j,k} \frac{((a^L)_{ik}^{lj} - e_{ik}^{lj})^2}{h_{ik}^{lj}} + \sum_{l,i,j,k} \frac{(e_{ik}^{lj} - h_{ik}^{lj})^2}{h_{ik}^{lj}} \end{aligned} \quad (29)$$

since all the crossed products are equal to zero. In what follows, the equality to zero for the last crossed product is proved:

$$\sum_{l,i,j,k} \frac{((a^L)_{ik}^{lj} - e_{ik}^{lj})(e_{ik}^{lj} - h_{ik}^{lj})}{h_{ik}^{lj}} = \sum_{l,i,j,k} \frac{((a^L)_{ik}^{lj} - e_{ik}^{lj})e_{ik}^{lj}}{h_{ik}^{lj}} - \sum_{l,i,j,k} \frac{((a^L)_{ik}^{lj} - e_{ik}^{lj})h_{ik}^{lj}}{h_{ik}^{lj}}$$

The last term is zero because the totals of \mathbf{A}^L and \mathbf{E} are both zero. The other term is also zero, since:

$$\sum_{l,i,j,k} \frac{((a^L)_{ik}^{lj} - e_{ik}^{lj}) \frac{f_{i\cdot}^{l\cdot} f_{\cdot k}^{j\cdot} f_{\cdot\cdot}^{lj}}{f_{\cdot\cdot}^{j\cdot} f_{\cdot\cdot}^{l\cdot}}}{f_{i\cdot}^{l\cdot} f_{\cdot k}^{j\cdot}} = \sum_{l,j} \frac{f_{\cdot\cdot}^{l\cdot}}{f_{\cdot\cdot}^{j\cdot} f_{\cdot\cdot}^{l\cdot}} \sum_{i,k} ((a^L)_{ik}^{lj} - e_{ik}^{lj}) = 0$$

Now, the CA associated to the inertias contained in the formula can be seen (29):

1. $CA(\mathbf{F}, \mathbf{C})$, i.e., $ICA(\mathbf{F})$.

2. $CA(\mathbf{A}^{\mathbf{J}}, \mathbf{E})$, but the expression can be reduced adding over k :

$$\sum_{l,i,j,k} \frac{\left(\frac{f_i^{lj} f_{\cdot k}^j}{f_{\cdot\cdot}^j} - \frac{f_i^l f_{\cdot k}^j f_{\cdot\cdot}^{lj}}{f_{\cdot\cdot}^l f_{\cdot\cdot}^j} \right)^2}{f_i^l f_{\cdot k}^j} = \sum_{l,i,j} \frac{\left(f_i^{lj} - \frac{f_i^l f_{\cdot\cdot}^{lj}}{f_{\cdot\cdot}^j} \right)^2}{f_i^l f_{\cdot\cdot}^j}$$

We note $\mathbf{T}^{\mathbf{J}}$ the table of dimension $I \times J$ and with general term f_i^{lj} , this inertia is associated to the Intra-Tables CA of $\mathbf{T}^{\mathbf{J}}$.

3. We can obtain an analogous result if we add on the subscript i , i.e. the $CA(\mathbf{A}^{\mathbf{L}}, \mathbf{E})$ is the Intra-Tables CA of $\mathbf{T}^{\mathbf{L}}$, with dimension $L \times K$ and general term $f_{\cdot k}^{lj}$.
4. The last addend is associated to the $CA(\mathbf{E}, \mathbf{H})$. In this case, it is possible to add to both subscripts i and k :

$$\sum_{l,i,j,k} \frac{\left(\frac{f_i^l f_{\cdot k}^j f_{\cdot\cdot}^{lj}}{f_{\cdot\cdot}^l f_{\cdot\cdot}^j} - f_i^l f_{\cdot k}^j \right)^2}{f_i^l f_{\cdot k}^j} = \sum_{l,j} \frac{(f_{\cdot\cdot}^{lj} - f_{\cdot\cdot}^l f_{\cdot\cdot}^j)^2}{f_{\cdot\cdot}^l f_{\cdot\cdot}^j}$$

This inertia is associated to the SCA of \mathbf{T} , with dimension $L \times J$ and general term $f_{\cdot\cdot}^{lj}$, i.e. the table formed by the totals of the blocks (l, j) .

Appendix A.3. Proof of Formula (23), p.126

The coordinate of the row point over the s -axis, as a function of the coordinates of the column points is (Escofier 1984):

$$F_s(l, i) = \frac{1}{\sqrt{\lambda_s}} \sum_{j,k} \left(\frac{f_{ik}^{lj}}{f_i^l} - \frac{c_{ik}^{lj}}{f_{\cdot\cdot}^l} \right) G_s(j, k)$$

Replacing c_{ik}^{lj} (Formula (19), p.124), three sums appear but the two last are zero, because the coordinates $G_s(j, k)$ from each subcloud N_{K_j} are centered with the weights $\frac{f_{\cdot k}^j}{f_{\cdot\cdot}^j}$:

$$\sum_j f_i^{lj} \sum_{k \in K_j} \frac{f_{\cdot k}^j}{f_{\cdot\cdot}^j} G_s(j, k) = 0 \quad \text{and} \quad \sum_j \frac{f_{\cdot\cdot}^{lj}}{f_{\cdot\cdot}^l} \sum_{k \in K_j} \frac{f_{\cdot k}^j}{f_{\cdot\cdot}^j} G_s(j, k) = 0$$

then,

$$F_s(l, i) = \frac{1}{\sqrt{\lambda_s}} \sum_{j,k} \left(\frac{f_{ik}^{lj}}{f_i^l} - \frac{f_{\cdot k}^j}{f_{\cdot\cdot}^l} \right) G_s(j, k)$$