PROPOSAL FOR THE ANALYSIS OF BINARY CORRESPONDENCES MULTIPLE TIMES: AN ADAPTATION OF THE STATIS METHODOLOGY

Eddy Fajardo[1,*] Maura Vasquez** and Guillermo Ramirez**
*Departamento de Ciencias Básicas, Universidad Autónoma de Bucaramanga, Bucaramanga, Colombia.
**Programa Integrado de Postgrado en Estadística y Ciencias Actuariales, Universidad Central de Venezuela, Caracas, Venezuela.

ABSTRACT
In this research, the theoretical and applied foundations of an analytical proposal based on an adaptation of the STATIS methodology is formally presented. It makes use of the binary correspondence analysis to address the statistical treatment of *K* blocks of information generated by characterizing the same individuals, using the same two categorical variables, on different occasions. The proposed procedure, which has been designed and validated in simulated environments and implemented on real data, performs an analysis of the variability of the row or column profiles, of the contingency tables within each block, following the classic approach that defines the BCA, and also, fundamentally, performs a comparative analysis of the changes that occur in the variability of the profiles between the different blocks, throughout the different occasions. The results of these analysis have been summarized on a graphic device, an scatter diagram, in which the points in the plane are determined by the blocks, on which it is possible to visualize the magnitude that describes the measure of variability within each block in terms of the length of the vector that joins the block with the origin of the representation space. It has also been found that the smaller the distance between the information structures corresponding to two different blocks, the stronger the association between the variables that identify those blocks is greater, and vice versa for small distances between blocks. An application is presented on data from the Colombian Institute for the Evaluation of Education (ICFES) in which the educational units of the departments are classified according to the academic results obtained by the students in the Saber 11 test, for the periods 2015, 2016, and 2017.
KEYWORDS: Binary correspondences analysis, STATIS, contingency tables, Saber 11 test.

MSC: 62H25.

RESUMEN
En esta investigación se presentan formalmente los fundamentos teóricos y aplicados de una propuesta analítica basada en una adaptación de la metodología STATIS, que hace uso del análisis de correspondencias binarias (ACB) para abordar el tratamiento estadístico de bloques de información generados al caracterizar los mismos individuos, mediante las mismas dos variables categóricas, en *K* ocasiones diferentes. El procedimiento propuesto, que ha sido diseñado y validado en ambientes simulados e implementado sobre datos reales, realiza un análisis de la variabilidad de los perfiles fila o columna, de las tablas de contingencias a lo interno de cada bloque, siguiendo el enfoque clásico que define el ACB, y además, en lo fundamental, efectúa un análisis comparativo de los cambios que se producen en la variabilidad de los perfiles entre los distintos bloques, a lo largo de las distintas ocasiones. Los resultados de estos análisis han logrado resumirse sobre un dispositivo gráfico tipo diagrama de dispersión en el que los puntos en el plano quedan determinados por los bloques, sobre el que es posible visualizar la magnitud que describe la medida de variabilidad a lo interno de cada bloque en términos de la longitud del vector que une al bloque con el origen del espacio de representación. También se ha encontrado que cuanto menor es la distancia entre las estructuras de información correspondientes a dos bloques diferentes, la fuerza de la asociación entre las variables que identifican a esos bloques es mayor, y viceversa para distancias pequeñas entre bloques. Se presenta una aplicación sobre datos provenientes del Instituto Colombiano para la Evaluación de la Educación (ICFES) en la que se clasifican a las unidades educativas de los departamentos de acuerdo con los resultados académicos obtenidos por los estudiantes en las pruebas Saber 11, para los períodos 2015, 2016 y 2017.
PALABRAS CLAVE: Análisis de correspondencias binarias, STATIS, tablas de contingencia, prueba Saber 11.

1. INTRODUCTION

Binary correspondences analysis (BCA) is a technique that is part of the family of methods of multivariate statistics, which is used in different research areas and has been specially designed to explain the main aspects of the association between two qualitative variables measured on a nominal or ordinal scale (Greenacre and Blasius, [7]), which can also be conceived as a very useful method to understand and explain the variability of the profiles, row or column, of a two-dimensional contingency table (Michailidis [11] and Glynn [6]). If additionally, the data to be analysed is arranged on a set of several tables generated contingencies characterizing the same individuals according to the same set of categorical variables, two particularly, at different times, an approach to analysing these blocks of information, would have the purpose of evaluating the changes that occur in the direction and strength of the association between the variables under consideration throughout the different measurement occasions.

[1] efajardo@unab.edu.co

Considering what was stated in Hair, Anderson, Tatham and Black [9], related to the advances that have occurred in the processes of extraction and storage of data from the same or different sources, it is very common in real applications to come across data sets as described above. In the presence of scenarios like this it cast information useful for understanding the complexity of some phenomena of reality, and it becomes important to develop statistical methods to the analysis of data structures with relative complexity.

In consideration of the previous approach, the problem outlined in this investigation has consisted essentially of making an analytical proposal, based on an adaptation of the STATIS methodology, which makes use of binary correspondence analysis (BCA) to address the statistical treatment of the blocks of information generated by characterizing the same individuals, using the same two categorical variables, at K different times. The procedure that has designed, implemented, and validated these purposes, analyses the variability of the profiles row or column of each of the contingency tables internally in each block, following the classical approach that defines the BCA, but in essence, the proposal makes it possible to carry out a comparative analysis of the changes, major or minor, that occur in the variability of the profiles between the different blocks, throughout the different occasions. They have been defined for this: objects for each block, constructed as matrix arrays containing the essence of the variability of profiles within each block, captured by one indicator strength of the association between categorical variables analysed; and second place, have been established distance measurements between the objects of different defined blocks as a function of the association between the corresponding variables so that the distance between two blocks will tend to be much smaller the larger the association between the variables that identify the compared blocks; and finally one space representation of the objects which can be displayed graphically a measure of variability internal to each block, and the magnitude of the differences and/or similarities between information structures that identify the different blocks.

On the other hand, Greenacre [8] develops a generalization of the BCA to explore the associations in three-way tables, organizing the data according to the different structures described below: First, it groups the information in the three-table (3) tracks on an arrangement of two (2) tracks, each category of a third variable describing the contingency table resulting from the crossing of the first two. In this sense, it proposes a sequence of graphs that allow you to visualize the associations and interactions between variables: A first graph that represents the row profiles of the initial variable, defined by the categories of the second, for each of the categories of the third variable. Subsequently, a second graph obtained from the previous table, but to which the rows of two collapsed two- dimensional tables are obtained by crossing that the first variable with the second one has been added as supplementary categories. In the same way, it is done to the third with the second. Row profiles of these tables, which are essentially the centre of gravity in two-dimensional array in which the information of the three variables are represented as illustrative organized. Besides, a third graph is generated that represents the column profiles of the second variable defined by the categories of the variable resulting from combining the first and third variables. Finally, a joint correspondence analysis (JCA) is developed, to simultaneously scan the inter-associations between pairs of variables. The results are represented on a single graph in a manner analogous to the representation of a matrix of variances and covariances in the principal components analysis (PCA).

Djauhari [2] introduces two mathematical propositions to construct a method to select categorical variables based on the Escoufier vector correlation coefficient (RV) in the framework of the CBA. Based on these propositions, the selection of variables consists of the choice of a set of Escoufier operators whose average maximizes the RV.

For their part, Bécue-Bertaut and Páges [1], propose a method to analyse and group individuals described by a mixed set of variables whose information is organized in a multiple-table that juxtaposes sets of quantitative variables, indicator variables, and frequencies. The authors consider the problem of the weight of the individuals because it is fixed, usually uniform, by the user in the PCA and JCA, but imposed by the marginal in the BCA. The problem is to define a global distance by combining the distances, called separate distances, obtained by the PCA, the multiple correspondence analysis (MCA), and BCA, from an extension of the multiple factor analysis (MFA).

Torres [12] proposes an extension of the STATIS methodology for analysing qualitative data from the measurement of several sets of variables categorical measures on the same individuals (information that can be organized on multiple contingency tables). STATIS method, the analysis, and correlation compound the correspondence analysis binary are combined, generalizing the application and last technique to the treatment of a table of Burt. The distance in the proposed so-called STATIS-C takes into account the inter associations that occur between the respective sets of variables. Furthermore, the proposed strategy fails to describe the structure of spacings between individuals characterized by qualitative variables, being able to evaluate changes or stability in the structures.

Vivien and Sabatier [14] develop a new strategy to analyse two multi-block tables, a method called DO-ACT. This methodology is closely related to STATIS and Tucker's method. The data used are low - dimensional ($n = 7, K = L = 10$) and employs a natural scalar that is the product of operators associated with each array, each of the two multi-blocks. This methodology can be applied to a wide spectrum of knowledge ranging from chemistry to ecology since it allows dividing the study variables among subgroups to choose a selection of these for experimentation.

2. MATERIALS AND METHODS

2.1. STATIS

The STATIS (Structuration des Tableaux a Trois Indices de la Statistique), introduced by Escoufier [3], is an exploratory technique of multivariate data analysis that has been developed for situations in which there is a set of data matrices obtained by characterizing the same individuals using the same set of p quantitative variables on different occasions. In STATIS the data is identified with three subscripts: one for each matrix (occasions), one for the individuals, and another for the variables (Thioulouse et al. [13]). In this way, with this technique, three-way data can be analysed. The main objective of the method is to obtain relevant information contained in the data matrices, which allows analysing the similarity and the differences between them, throughout the occasions. materials and methods section should contain sufficient detail so that all procedures can be repeated. It may be divided into headed subsections if several methods are described.

The basic reference concept in STATIS is called a study. A study defined for an occasion k is a triplet $(\boldsymbol{X_k},\ \boldsymbol{M_k},\ \boldsymbol{D})$ where:

1. $\boldsymbol{X_k}$ is a data table (individuals x quantitative variables).
2. $\boldsymbol{M_k}$ it is a matrix in which weights for the variables are described and are used to define distances between individuals.
3. $\boldsymbol{D}$ is a diagonal matrix of weights that are assigned to individuals.

Each study in STATIS is represented by an object, denoted by $\boldsymbol{W_k}$:

$$\boldsymbol{W_k} = \boldsymbol{X_k M_k X_k}' \tag{1}$$

This object contains the information required to carry out the analysis of distances between individuals. The STATIS method is fundamentally based on the development of the following stages:

Analysis of the inter-structure. Stage in which the objective is to globally compare the information contained in the studies, $(\boldsymbol{X_k},\ \boldsymbol{M_k},\ \boldsymbol{D}), k = 1,2, \ldots K$, which is equivalent to analyze the evolution of the phenomenon under investigation. This analysis is essentially aimed at detecting changes in the structure of distances between individuals, along with different occasions, information that is contained in the objects $\boldsymbol{W_k} = \boldsymbol{X_k M_k X_k}'\ k = 1,2, \ldots K$. Carrying out a comparative analysis of the objects poses the problem of defining a scalar product at this stage that allows the calculation of the distance between them, two to two. The STATIS methodology uses the Hilbert-Schmidt (HS) scalar product to induce the distance between objects in the usual way:

$$\boldsymbol{d}(\boldsymbol{W_k}, \boldsymbol{W_l}) = <\boldsymbol{W_k}, \boldsymbol{W_k}> + <\boldsymbol{W_l}, \boldsymbol{W_l}> \ \boldsymbol{-2}\ <\boldsymbol{W_k}, \boldsymbol{W_l}> \tag{2}$$

Once the scalar product between two-to-two objects is available, an arrangement containing them is constructed, such as by applying on this a PCA, it is possible to obtain approximate representations of the distances between the objects, on the factorial planes determined by the main directions of that analysis.

Search for a compromise. It refers to obtaining an average structure of the objects, that is, the K arrangements $\boldsymbol{W_k}$ are summarized in a single one called commitment, which is defined as a weighted average of the structures corresponding to the representative objects of the information blocks.

Analysis of the intra-structure. In this stage, a detailed representation of the individuals and variables, corresponding to the K occasions, is obtained on the main directions of the compromise space, which allows exploring the differences or similarities that exist between the data tables (in essence it is of the projection of each of the data matrices on the space defined by the commitment object). Thus, representation is obtained in which the position of each of the individuals is projected as an average of their positions along with the different occasions, being possible to illustratively visualize the trajectory of the positions of an individual throughout different occasions. This method can be applied when you have any of the following situations:

- The same individuals measured on different occasions, characterized by the same variables.
- The same individuals measured on different occasions characterized by different variables.
- Different sets of individuals to whom the same variables are measured (Dual-STATIS).

The objectives of the STATIS methodology can be summarized as:

- Globally compare the arrangements of interest, the objects that identify the different tables, using Euclidean representations for which it is possible to visualize similarities and/or differences between them.
- Determine a structure that summarizes, as a weighted average, the corresponding structures of the different tables under consideration.
- Identify the individuals or variables responsible for such similarities or differences.

2.2. Binary correspondence analysis

The technique of multivariate analysis whose objective is to explain the association that exists between two categorical variables is BCA. The aspects of the analysis on which the BCA is essentially based, rest on the identification of the strength of the association between two categorical variables and as a measure of the variability of the row and/or column profiles defined on the contingency table that crosses to these variables. The singular value decomposition (SVD) applied on arrays that contain information corresponding to the association between the variables under study, is the mathematical procedure that allows finding the main directions of variability on which they are projected.

As Michailidis [11] and Glynn [6] refer, these representations, which are constructed as scatter diagrams, can be considered optimal fit planes in the least-squares sense, since their directions capture the maximum possible proportion of the existing association. Among the categorical variables under study, because overall the distance between the row (column) profiles and their projections on the plane is minimized. The representations in reference approximate with a high degree of fidelity the most important aspects of the information contained in the profile tables and similar row (column) profiles appear grouped in projection, which in turn are separated from those of which they differ. Also, graphically it is possible to determine the categories of one of the variables that are most important in the definition of the frequency profile for a fixed category of the other variable. A contingency table is an arrangement $K_{(I\times J)} = (k_{ij})$ whose cells contain the frequencies that describe the distribution of n objects simultaneously characterized by two qualitative variables and categories respectively I and J. Some concepts of interest obtained by making transformations on the contingency table are:

Matrix of relative frequencies concerning the total. Refers to absolute frequencies for total observations. Notation:

$f_{ij} = \frac{1}{n} k_{ij}$ Proportion of individuals who simultaneously possess the modalities X_i and Y_j, concerning the total.

$f_{.j} = \frac{1}{n} k_{.j}$ Proportion of individuals who have the modality Y_j of the column variable, for the total.

$f_{i.} = \frac{1}{n} k_{i.}$ Proportion of individuals who have the modality X_i of the row variable, with respect to the total.

Diagonal matrix of weights for the row categories. This matrix contains weights that are assigned to row categories in terms of the number of objects in each. Notation: $D_I = (diag(f_{i.}))$.

Diagonal matrix of weights for the column categories. This matrix contains weights that are assigned to column categories in terms of the number of objects in each. Notation: $D_I = \left(diag(f_{,j})\right)$.

Matrix of row profiles. The profile of the $i-th$ row of the contingency table K is defined as the frequency distribution of the column variable x, conditioned by the category Y_j of the row variable. The information corresponding to the row profiles is organized about the arrangement: $R_{I\times J} = D_I^{-1}F$.

Column profiles matrix. The profile of the column $j-th$ in the contingency table K is defined as the frequency distribution of the variable x, determined by the category Y_j of the variable column. The information corresponding to profiles for a column is organized on the basis: $C_{n\times p} = FD_I^{-1}$.

Centre of gravity of the row profiles. The center of gravity, or average vector, of the row profiles, is obtained as a weighted average, in the form:

$$g^t = \sum_{i=1}^{I} f_{i.} R_i^t \quad (3)$$

Centre of gravity of the column profiles. The center of gravity, or average vector, of the row profiles, is obtained as a weighted average, in the form:

$$h^t = \sum_{j=1}^{J} f_{.j} C_j^t \quad (4)$$

Point cloud in R^J. It is the set of row profiles R_i^t endowed with the weights f_i and the distance that is used to compare these profiles, using the chi-square metric, the distance defined in the form:

$$d^2(i,s) = \sum_{j=1}^{J} \frac{1}{f_{.j}} \left(\frac{f_{ij}}{f_{i.}} - \frac{f_{sj}}{f_{s.}} \right)^2 = (R_i - R_s)^t D_J^{-1} (R_i - R_s) \quad (5)$$

Point cloud in R^I. It is the set of column profiles C_j^t endowed with the weights c^t the chi-square distance, which is used to compare these profiles, defined in the form:

$$d^2(j,k) = \sum_{i=1}^{I} \frac{1}{f_{i.}} \left(\frac{f_{ij}}{f_{.j}} - \frac{f_{ik}}{f_{.k}} \right)^2 = \left(C_j - C_k\right)^t D_I^{-1} \left(C_j - C_k\right) \quad (6)$$

Total variability of the profiles. The BCA objective is focused on the analysis of the variability of the profile, row, and column clouds, using the following concepts:

The distance of a row profile to its centre of gravity is the chi-square distance between row i and the center of gravity:

$$d^2(R_i, g) = (R_i - g)^t D_p^{-1}(R_i - g) \tag{7}$$

The distance of a column profile to its centre of gravity is the chi-square distance between column j and its center of gravity:

$$d^2(c^j, h) = (c^j - h)^t D_n^{-1}(c^j - h) \tag{8}$$

Total variability: Described as a global measure of the similarity between the row (column) profiles concerning the centre of gravity, which in turn coincide with the Pearson coefficient, which in turn represents a measure of the strength of the association between the variables analysed, $\frac{\chi^2}{n}$, normalized by the number of observations:

$$\text{Total variability} = \sum_{i=1}^{n}\sum_{j=1}^{p}\frac{f_{ij}^2}{f_{i.}f_{.j}} - 1 = \frac{\chi^2}{n} \tag{9}$$

The BCA breaks down the total variability into addends, $\frac{\chi^2}{n} = \sum_{\alpha=2}\lambda_\alpha$, each of which is associated with a principal direction of the analysis, the importance of which is measured by the proportion of the total variability that each of these directions manages to capture.

3. RESULTS AND DISCUSSION

3.1. STATIS-BCA

In this research, an adaptation of the STATIS methodology has been designed, implemented, and tested, which has been designated as STATIS-BCA. This method has - been defined to perform the statistical processing of one data structure generated characterizing the same individuals, by two categorical variables $x_1$and x_2 on K different occasions. A CBA is applied to each of the blocks of information that identify the different occasions and later, using a STATIS-type approach, a comparative analysis is carried out between blocks to evaluate the changes that occur in the association between the two categorical variables considered. The STATIS-BCA applied to each of the contingency tables obtained by crossing the variables within each of the blocks under consideration, allows explaining the most important determinants of the association between those variables and also for the changes in the association structures analysed between occasions.

The analytical bases of the formulated proposal, are located in the perspective of an adaptation of STATIS in which binary correspondences are used to approach the analysis of the information within the blocks, and under the principles that underlie the definition of a distance, a comparison between them is possible. This approach of analysis is essentially based on the following statement (Lebart, Morineau and Warwick, [10]):

"The results of applying the binary correspondence analysis on a contingency table constructed by crossing two qualitative variables x_1 and x_2, with J_1 and J_2 categories respectively, which have been measured on n individuals, are essentially equivalent to those obtained by applying the analysis of multiple correspondences on the complete disjunctive table $Z_{x1.x2}$, resulting from characterizing individuals to the two categorical variables under consideration." (p. 133).

In consideration of this approach, the problem that this research has outlined is to develop an analytical proposal, based on an adaptation of the STATIS methodology, that makes use of BCA, conceived as a particular case of MCA, to approach the statistical treatment of the blocks of information generated by characterizing the same individuals, using the same two categorical variables, on different occasions.

The procedure that has been designed, implemented, and tested for these purposes, analysing the variability of the row or column profiles of each of the contingency tables within the blocks, following the classic approach defined by the CBA, but fundamentally the proposal makes it possible to carry out a comparative analysis of the changes that take place in the variability of the profiles between the different blocks along with the different occasions.

The following arrangements and some measures that are required to be applied to them are defined for this:

An *object* that identifies block k, which is defined as a matrix array that contains the essence of the association between the variables inside the block, or equivalently contains information about the variability of the row or column profiles of the table of associated contingencies.

The Hilbert-Schmidt scalar product of an object with itself, which is defined as an index that accounts for the strength of the association between the categorical variables within a block.

A *measure distance* between objects corresponding to two different blocks, determined as a function of the association between the variables corresponding to the compared blocks, so that the distance between two objects tends to be much smaller the larger the association between variables that identify the blocks in reference, and vice versa when the distance is great.

A *reduced representation space* for objects, on which the length of the vector linking an object to the coordinate origin represents a measure of the variability within the corresponding block, and the distance between two objects is smaller the greater the association between the variables that identify the blocks of the compared objects.

Stages that describe the theoretical design of the proposal

Stage 1

In this stage, the information that constituting the input to apply the STATIS-BCA on the k-th occasion is described, as well as the arrangements and indexes generated are presented, considering that the two qualitative variables are denoted by X_{k1} X_{k2} with J_{1x} and J_{2x} categories respectively.

$$E = (X_k, M_{Xk}, D_{Xk}) \tag{10}$$

where,

$X_k = \frac{1}{2n}\left(X_{k(1)}, X_{k(2)}\right)$ is a partitioned matrix in which disjunctive tables are juxtaposed an it corresponds to the measurement of the variables in the block k, $\boldsymbol{x}_{k(1)}$ y $\boldsymbol{x}_{(k2)}$.

$M_{Xk} = \frac{1}{2n}\begin{pmatrix} D_{xk(1)}^{-1} & 0 \\ 0 & D_{xk(2)}^{-1} \end{pmatrix}$ is the metric entered in the category space that describes the columns of the matrix X_k.

$D_{Xk} = \frac{1}{2n}\boldsymbol{I}$ is the weighting matrix of the individuals. Therefore, the object in block k is described as:

$$W_k = X_k M_{Xk} X_k^t = \left(X_{k(1)}, X_{k(2)}\right)\begin{pmatrix} D_{xk(1)}^{-1} & 0 \\ 0 & D_{xk(2)}^{-1} \end{pmatrix}\begin{pmatrix} X_{k(1)}^t \\ X_{k(2)}^t \end{pmatrix} \tag{11}$$

Stage 2

In this stage, a Hilbert-Schmidt scalar product is defined between objects, which, when applied to the arrangements of the previous stage inside a block, determines an indicator of the strength of the association between the variables on each occasion, and when applied to objects in two different blocks, another indicator is generated that allows comparison between occasions. The scalar product that has been used in this proposal is the Hilbert-Schmidt product, thus maintaining the metric originally used in the STATIS methodology. By applying this product on the object defined for the STATIS-BCA, an expression is obtained that is a direct function of the chi-square statistic that measures the association between the two variables in the block:

$$\langle W_k, W_k\rangle_{HS} = trace(X_k M_{Xk} X_k^t\, X_k M_{Xk} X_k^t) = (J_{1X} + J_{2X}) + \frac{2\chi^2_{xk(1),xk(2)}}{n} \tag{12}$$

That is, the scalar product of an object with itself in the-th block, except for the constant $(J_{1X} + J_{2X})$, is an index of the strength of the association between the corresponding variables, that is, it produces a measure of the variability between row (or column) profiles. The disaggregation of this measure, in terms of the eigenvalues associated with the main directions of variability in the block, allows an BCA-type analysis to be carried out to explain the determinants of the association in this case.

$$x\langle W_k, W_s\rangle_{HS} = traza(W_k W_s) = \frac{(\chi^2_{xs(1),xk(1)} + \chi^2_{xs(1),xk(2)} + \chi^2_{xs(2),xk(1)} + \chi^2_{xs(2),xk(2)})}{n} + 4 \tag{13}$$

Therefore, the dot product between objects in different blocks is a direct function of the strength of the association between the variables of the two blocks.

Taking into account the previous results, it is:

$$d^2(W_{Xk}, W_{Xs}) = \langle W_{Xk}, W_{Xk}\rangle_{HS} + \langle W_{Xs}, W_{Xs}\rangle_{HS} - 2\langle W_{Xk}, W_{Xs}\rangle_{HS} \tag{14}$$

When replaced, the result is:

$$= \left((J_{1Xk} + J_{2Xk}) + \frac{2\chi^2_{xk(1),xk(2)}}{n}\right) + \left((J_{1Xs} + J_{2Xs}) + \frac{2\chi^2_{xs(1),xs(2)}}{n}\right) - 2\frac{\left(\chi^2_{xs(1),xk(1)} + \chi^2_{xs(1),xk(2)} + \chi^2_{xs(2),xk(1)} + \chi^2_{xs(2),xk(2)}\right)}{n} + 4 \tag{15}$$

So, the following can be concluded:

"The distance between the k and s blocks will be the smaller the greater the association between the variables that identify the k and s blocks."

Stage 3

In this phase, a representation of the objects of the different blocks is constructed, on a space whose main directions are defined by the association between the variables on different occasions. When a PCA is carried out on the matrix that in the STATIS-BCA contains the scalar products between the objects of the different blocks, the main addresses of the representation space are obtained:

$$S = (\langle W_k, W_s\rangle_{HS}) \qquad k, s = 1, 2, \dots, K \tag{16}$$

Where S is necessary to construct the space in reference. The eigenvalues and eigenvectors of the matrix S are calculated as follow:

$$Sa^\alpha = \mu_\alpha a^\alpha \tag{17}$$

Finding then that $\sum \mu_\alpha$ is a measure of a global partnership between x_1 and x_2, on K occasions:

$$\sum_{\alpha=1}^{K} \mu_\alpha = K * (\mathrm{J}_{1x} + \mathrm{J}_{2x}) + \frac{2\sum_{k=1}^{K}\chi^2_{xk(1),xk(2)}}{n} \tag{18}$$

and therefore, the strength of the α-th direction in this space of representation is given by the quotient:

$$\frac{\mu_\alpha}{K*(\mathrm{J}_{1x} + \mathrm{J}_{2x}) + \frac{2\sum_{k=1}^{K}\chi^2_{xk(1),xk(2)}}{n}} \tag{19}$$

The projection coordinates of γ^{α} the blocks on the axis α^{α} are in the form:

$$\gamma^{\alpha} = \sqrt{\mu_{\alpha}}a^{\alpha} \qquad \alpha = 1, 2, \dots K \tag{20}$$

In the graphic representation, it is found that for those blocks furthest from the origin of coordinates, the association between x_1 and x_2 is stronger; and on the other hand, the blocks will be closer the greater the association between the variables that identify those blocks, and vice versa for distant blocks.

4. APPLICATION

In this section, an application of the STATIS-BCA is performed on information collected by the Colombian Institute for the Evaluation of Education (ICFES), which corresponds to 5.050 educational institutions (EI) located in different departments in the Andean region in Colombia, and which have been classified according to the academic results obtained by their students in the Saber 11 test for the years 2015, 2016 and 2017. The purpose of the study is to determine if there have been changes in the departments of the region under study, concerning the classification of actual scores of the Saber 11 test assigned by ICFES to EI, for the years indicated.

According to Fajardo, Romero, Plata, and Ramírez [4] and Fajardo, Beleño-Montagut and Romero [5], it can be assured that there is an educational gap in Colombia, which is becoming more acute due to different conditions determined by socioeconomic variables, in addition to the qualifications of teachers, the location of the school according to the department, and the type of settlement of the EI (rural or urban), aspects that may influence in some direction the performance of the Saber 11 test results. These tests are intended to measure the degree of competences acquired by the students who are going to complete the eleventh grade of secondary education in Colombia and specifically address the following aspects of knowledge: mathematics, social, natural sciences, critical reading, and the English language.

The corresponding information is then available for the $n = 5.050$ EI of the Andean Region, which participated in the ICFES tests 2015, 2016, and 2017, classified by the Geographical Department in which the EI is located and by the qualification assigned by the ICFES test, as is described in table 1.

Blocks	Categorical variables	Description	Categories	# of categories
Years 2015, 2016 and 2017	Classification according to the ICFES Test	The categories contemplated by the score in the ICFES test are described according to an ordinal scale, A+, A, B, C, and D, with A+ being the score that indicates much higher results in the test (the schools that obtain the best results) and D it is the lowest rating (the schools with the worst results).	A + TO B C D	5
	Department of the Andean Region where the educational unit is located	The categories that make up this variable, described according to a nominal scale, are made up of the departments of the Andean region of Colombia, a region of particular interest to include the country's capital, as well as some departments of greater social and economic development, such as the of Santander of particular interest to the author for being the seat of her residence and her work environment. It should be noted that Bogotá, even when it does not have a departmental entity, will be considered as a separate department, due to the high concentration of population that congregates.	Antioquia Boyacá Caldas Cundinamarca Huila North Santander Quindio Risaralda Santander Tolima Bogotá	11

Table 1: Database information.

Next, the BCA is presented and the results obtained when applying the STATIS-BCA on this information structure are then shown, using R software, with which a special program was designed for the STATIS-BCA proposal.

Analysis of contingency tables for the blocks studied

The following is an analysis of the evolutionary profile described by the EI classification according to ICFES scores obtained in the Saber 11 test, during the years 2015, 2016, and 2017 for the different departments of the Andean Region in Colombia.

Changes in the profile of ICFES scores (2015 – 2016)

Table 3 shows that almost all (99.83%) of the schools classified as A + (highest category) in 2015 maintain that same rating for 2016. 93% of the schools classified in category A in 2015, they did not change their category for 2016, finding, however, that approximately 5% of these schools improved by moving from one A classification to another with a higher A + level, while a little less than 2% deteriorated, moving from category A to category B.

In schools classified in categories B and C for 2015, approximately 92% maintain the same qualification for 2016. Improvements are observed in the academic qualification of a small number of the EI, both because they rise from category B to A (5%), or because they go from category C to B (5%). The percentage of schools that decreased their rating going from B to C is 3.65%. On the other hand, for the category D EIs in 2015, almost 9% improved their ranking to C in 2016, and 90% kept the same category.

	A+	A	B	C	D
A+	99.83%	0.16%	0.00%	0.00%	0.00%
A	4.89%	93.28%	1.81%	0.00%	0.00%
B	0.00%	4.55%	91.61%	3.65%	0.02%
C	0.00%	0.00%	4.95%	91.86%	3.18%
D	0.00%	0.00%	0.45%	8.71%	90.08%

Table 2: Profile of ICFES scores (2015 vs 2016)

Changes in the profile of ICFES scores (2015 – 2017)

When observing the changes in the rating of the EI between the years 2015 and 2017, all of the EI with an A + rating remain with that rating, also finding that in the remaining rating categories a significant percentage of the EI improve their one-year rating to the other (24% to 29%), as shown in table 4.

	A+	A	B	C	D
A+	100%	0.00%	0.00%	0.00%	0.00%
A	23.91%	76.08%	0.00%	0.00%	0.00%
B	0.059%	20.95%	77.60%	1.19%	0.17%
C	0.00%	0.18%	26.42%	70.15%	3.24%
D	0.00%	0.00%	1.14%	28.66%	70.18%

Table 3: Profile of ICFES scores (2015 vs 2017)

Changes in the profile of ICFES scores (2016 – 2017)

Between 2015 and 2016, as noted above, for all categories only slight improvements were found in the ranking of the EI according to ICFES scores (between 5% and 9%). While between 2016 and 2017, a marked rise was observed in the EI ICFES classification in all scoring categories, highlighting for all categories improvements in the ranking that range between 19 % and 30 %, (see table 4).

	A+	A	B	C	D
A+	99.84%	0.15%	0.00%	0.00%	0.00%
A	18.68%	80.91%	0.40%	0.00%	0.00%
B	0.00%	17.95%	81.11%	0.73%	0.18%
C	0.00%	0.12%	25.35%	71.89%	2.62%
D	0.00%	0.00%	1.10%	28.60%	70.28%

Table 4: Profile of ICFES scores (2016 vs 2017)

The analysis of ICFES scores allows us to conclude that the years 2015 and 2016 are very similar in terms of the ranking of the EI, while 2017 marks a differentiation from previous years. Taking the analysis presented above as a reference, the need naturally arises to ask us about the departments in which there are significant changes in the classification of the corresponding EIs, between the periods under study. For this purpose, BCA is carried out for explanatory purposes, which accounts for the variability of the classification profiles according to ICFES scores of the different departments of the Andean region of Colombia, for each of the periods under study.

6. BINARY CORRESPONDENCE ANALYSIS FOR EACH BLOCK

For each of the years from 2015 to 2017, it was a first analysis performed considering as active variables the classification scoring ICFES, according to five categories, and departments (including Bogotá) of the Andes in Colombia.

Block 1 - 2015

Decomposition *d* and the variability 2015

The total variability of profiles of the departments according to ICFES scores has been decomposed along with four main directions. In table 5 eigenvalues and percentages of variability captured by each direction is. A very important first axis can be seen, which captures 87.15%, and a second axis that captures a portion of significantly less variability, 10.82%, without detracting from its importance for the analysis. These two axes capture almost all (97.98%) of the variability of the phenomenon.

	A+	A	B	C	D
A+	99.84%	0.15%	0.00%	0.00%	0.00%
A	18.68%	80.91%	0.40%	0.00%	0.00%
B	0.00%	17.95%	81.11%	0.73%	0.18%
C	0.00%	0.12%	25.35%	71.89%	2.62%
D	0.00%	0.00%	1.10%	28.60%	70.28%

Table 5: Decomposition of the total variability 2015.

Of the five categories of ICFES scores, three of them: Lower, Low, and Superior define axis 1 in absolute terms, contributing in percentage terms to define the variability captured by the axis, at 28.38%, 26.89%, and 22.82% respectively, for a cumulative of 78.13%. Likewise, the location of these categories is highly determined by this

factor (0.80, 0.91, and 0.92). A negative coordinate of the superior category (A+) is observed on the first axis in contrast to the positive coordinates of the lower (D) and low (C) scores on this axis, (see table 6).
The second axis is defined by the lower score (D) which contributes to explain the variability captured by the axis by 55%. The remaining qualifications were not considered in the explanation of the axis, since, on the one hand, they do not contribute in a substantive way to the definition of the axis, and on the other, they do not have representation quality.

Active frequencies	**Coordinates**		**Contributions**		**Square cosines**	
	Axis 1	**Axis 2**	**Axis 1**	**Axis 2**	**Axis 1**	**Axis 2**
Superior (A+)	-0.5846	-0.1379	22.8680	10.2454	0.9280	0.0516
High (A)	-0.4525	-0.1025	16.3579	6.7597	0.9189	0.0471
Medium (B)	-0.1716	0.0855	5.4923	10.9977	0.7224	0.1797
Low (C)	0.3838	0.1077	26.8999	17.0745	0.9160	0.0722
Inferior (D)	0.7634	-0.3743	28.3816	54.9224	0.8034	0.1931

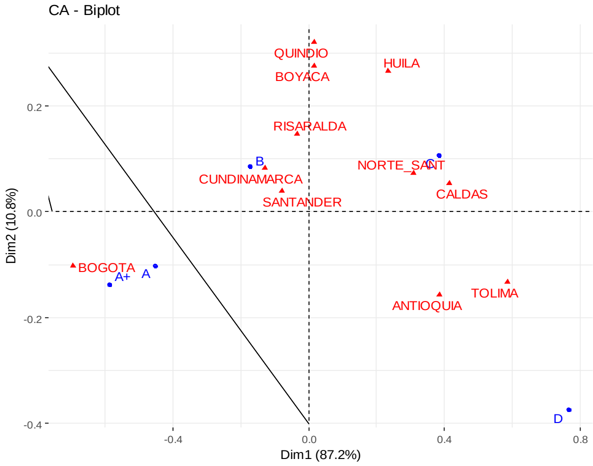
Table 6: Coordinates, Absolute and relative contributions of active frequencies (ICFES results).

Regarding the geographical location of the EI (see table 7), it is observed that Bogotá and the departments of Antioquia and Tolima have an accumulated contribution of approximately 90% (57.56%, 17.94%, and 14.10% respectively) to the variability captured by the first axis. Additionally, it can be seen, through the square cosines (0.97, 0.85, and 0.94), that each of these departments have a high quality of representation on the axis. Likewise, a clear contrast between the departments Antioquia and Tolima (with positive coordinates on the axis) for Bogotá (with negative coordinate) stands out. In particular, it is worth referring to some departments that have a high quality of representation on the first axis, indicating that the Santander EI (0.70), due to their position on the axis, generally obtain medium-level ICFES scores (B), while the UE located in Caldas (0.92) and Norte de Santander (0.86) tend to be rated with lower ICFES scores.

Active frequencies	**Coordinates**		**Contributions**		**Square cosines**	
	Axis 1	**Axis 2**	**Axis 1**	**Axis 2**	**Axis 1**	**Axis 2**
Antioquia	0.3850	-0.1577	17.9463	24.2403	0.8505	0.1427
Bogotá	-0.6927	-0.1019	57.5605	10.0253	0.9782	0.0211
Boyacá	0.0161	0.2763	0.0103	24.3791	0.0027	0.8173
Caldas	0.4139	0.0541	4.4789	0.6166	0.9243	0.0158
Cundinamarca	-0.1282	0.0825	1.2786	4.2599	0.6455	0.2672
Huila	0.2342	0.2660	1.4825	15.3931	0.4099	0.5288
Norte de Santander	0.3089	0.0729	2.7928	1.2528	0.8674	0.0483
Quindío	0.0159	0.3211	0.0030	10.0153	0.0021	0.8848
Risaralda	-0.0337	0.1474	0.0215	3.3043	0.0313	0.5988
Santander	-0.0784	0.0393	0.3174	0.6433	0.7066	0.1778
Tolima	0.5850	-0.1330	14.1077	5.8695	0.9475	0.0489

Table 7: Coordinates, absolute contributions, and relative, square cosines of the cases (departments)

Taking into account the above considerations, the first axis can be conceptualized as a ranking scale of the departments of the Andean Region in Colombia according to the classification defined by the ICFES scores of their EI. The first axis contrasts EI located in Bogotá, which generally receive higher and higher ratings in the ICFES compared to UEs located in the departments of Antioquia and Tolima whose ICFES ratings are in the lower category (D).



In Figure 1 the first factorial map of the BCA is presented cross-tabulated in which simultaneously represent departments of the Andean and the rating will ones of ICFES, for the year 2015. This device allows describing the differences and/or similarities in the ICFES tests qualification level profile. It is noted also that the level is higher s of qualifications ICFES (superior and high) are mostly assigned to EI located in the capital of the country, while that qualifications with the level lowest (inferior) are recorded in EI located in the lower part of the Andean region of the country. In general, the remaining departments tend to be rated with intermediate rating levels.

Figure 1: Andean Region according to classification in Saber 11 in 2015.

Block 2 - 2016

Decomposition *d* and the variability 2016

The total variability of profiles has been decomposed along with four main directions. In Table 8 the values and percentages shown variability captured by each direction, being able to boast the first axis capture 85.03% and the second one 12.42%. These two axis captures approximately 97 % of the total variability of the phenomenon.

Number	Eigenvalue	% variability.	% Accumulated.
1	0.160279850	85.0372396	85.03724
2	0.023415239	12.4230669	97.46031
3	0.003601170	1.9106181	99.37092
4	0.001185694	0.6290754	100,000000

Table 8: Decomposition of the total variability 2016.

Like the behaviour of the phenomenon for 2015, of the five categories of ICFES three of them: Lower, Lower, and Higher contribute highly in absolute terms to the formation of axis 1, contributing 33.32%, 22.99%, and a 23.31% respectively, for a cumulative of 79.62%. Likewise, these categories are highly determined by the said factor (0.81, 0.87, and 0.91). It is found one c o ordinate projection negative for the Superior about the axis and coordinates positive category s Inferior and B garlic on the shaft. The second axis is defined by the Lower (D) rating, which contributes 49.16 %. The other ICFES rating categories were not considered since, on the one hand, they do not contribute substantively to the definition of the axis, and on the other, they do not have representation quality, (see table 9).

Active frequencies	Coordinates		Contributions		Square cosines	
	Axis 1	Axis 2	Axis 1	Axis 2	Axis 1	Axis 2
Superior (A+)	-0.5462	-0.1513	23.3157	12.2564	0.9111	0.0699
High (A)	-0.4179	-0.1123	16.0418	7.9306	0.9032	0.0652
Medium (B)	-0.1466	0.0988	4.3145	13.4220	0.5829	0.2649
Low (C)	0.3411	0.1128	22.9996	17.2265	0.8798	0.0962
Inferior (D)	0.7737	-0.3592	33.3281	49.1641	0.8188	0.1764

Table 9: Coordinates, absolute and relative contributions and square cosines of the active frequencies (ICFES results).

Bogotá and the departments of Antioquia and Tolima give an accumulated contribution of 90.61% (54.19%, 18.67%, and 17.75% respectively) to the first axis. Additionally, it can be seen, through the square cosines (0.96, 0.81, and 0.95) that each of these departments is represented on axis 1 with high quality. Likewise, there is a clear opposition between the departments Antioquia and Tolima (positive coordinates on the axis) concerning Bogotá (negative coordinates on the axis), (see table 10).

Active frequencies	Coordinates		Contributions		Square cosines	
	Axis 1	Axis 2	Axis 1	Axis 2	Axis 1	Axis 2
Antioquia	0.3736	-0.1778	18.6776	28.9615	0.8102	0.1835
Bogotá	-0.6393	-0.1139	54.1904	11.7840	0.9665	0.0307
Boyacá	0.0307	0.1979	0.0415	11.7545	0.0187	0.7750
Caldas	0.4207	0.1051	5.1138	2.1847	0.8740	0.0545
Cundinamarca	-0.1550	0.0828	2.0659	4.0334	0.6887	0.1964
Huila	0.1810	0.3216	0.9789	21.1483	0.2333	0.7364
Norte de Santander	0.1612	0.1542	0.8412	5.2674	0.3569	0.3265
Quindío	0.0287	0.2246	0.0109	4.6029	0.0160	0.9819
Risaralda	-0.0356	0.1451	0.0265	3.0075	0.0352	0.5836
Santander	-0.0721	0.0854	0.2968	2.8495	0.3405	0.4776
Tolima	0.6243	-0.1188	17.7558	4.4059	0.9573	0.0347

Table 10: Coordinates, absolute contributions, and relative, square cosines of the cases (departments) 2016

The first axis contrasts Bogotá with a higher and higher category in the ICFES with the departments Antioquia and Tolima whose classification in the ICFES scores corresponds to the Lower category.

Figure 2 shows the first BCA factorial plan of the contingency table that crosses departments with the ICFES rating. This device allows describing the differences and/or similarities in the profile of the departments according to the level of qualification in the ICFES test. It is observed that the highest levels of ICFES scores (A + and A) are found around the department that corresponds to the country's capital, scores that appear opposed to the lowest level (D) that in frequentist terms defines the departments located towards the lower part of the Andean region of the country.

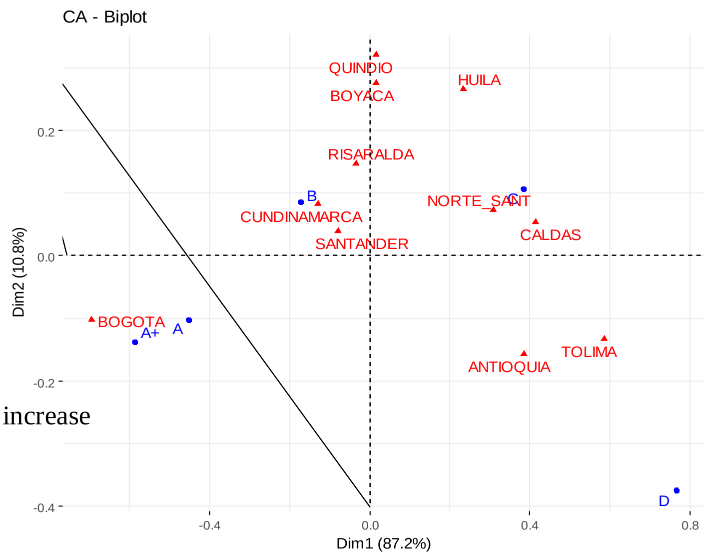


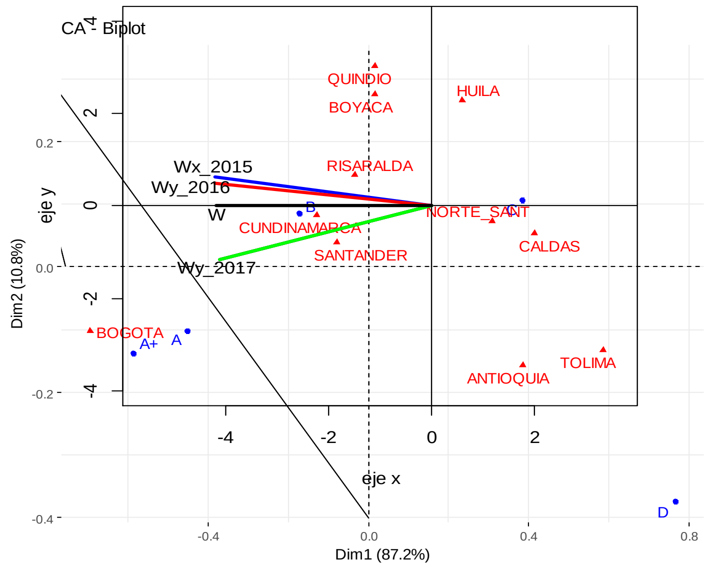
Figure 3 shows that at the intermediate levels of the classification, the department of Norte de Santander improved from being C in the two previous years to being categorized as B in 2017. Besides, it is observed that the departments of Huila and Antioquia also improved concerning the classification of previous years, in the case of Antioquia are closer to classified C and Huila see is closer to be classified as B. It is evident that, by promoting public policies in education, seeking to improve the quality of official establishments, it is possible to have an increase in academic performance. Similarly, as is known the efforts of the national government and of the department of Antioquia in Colombia to the assigned r

Figure 2: Andean Region according to classification in tests Saber 11 in the year 2016.

major portions of spending public investment in education, they have yielded results positive here are observed. Given the change in 2017 with respect to the years 2015 and 2016 on the classification of the EI as scores ICFES, shown in then contingency tables that can describe the direction of changes in departments Antioquia, Huila, and Norte de Santander.

6. STATIS-BCA FOR THE THREE-YEAR STUDY

The following is a synthetic analysis of the changes that have occurred in the profile of ICFES scores corresponding to the Saber 11 scores located in the departments of the Andean region of Colombia throughout the years 2015, 2016, and 2017, all based on the tools provided by the inter-structure.
On the matrix S of the STATIS-BCA shown then scalar products are arranged between objects W_{2015}, W_{2016} and W_{2017}, and, for the three periods under study:



$$S = \begin{bmatrix} 18.40644 & 17.70969 & 16.64381 \\ 17.70969 & 18.37697 & 16.77903 \\ 16.64381 & 16.77903 & 18.34486 \end{bmatrix}$$

As it is clear from the magnitude of the diagonal elements in this arrangement, it is highlighted that the variability of the

Figure 3: Andean Region according to classification in tests Saber 11 in the year 2017.

ICFES scores in the departments of the Andean region of Colombia in the period under study, presents a slight tendency to decrease, which is confirmed in the value that it reaches the variability of the profiles:

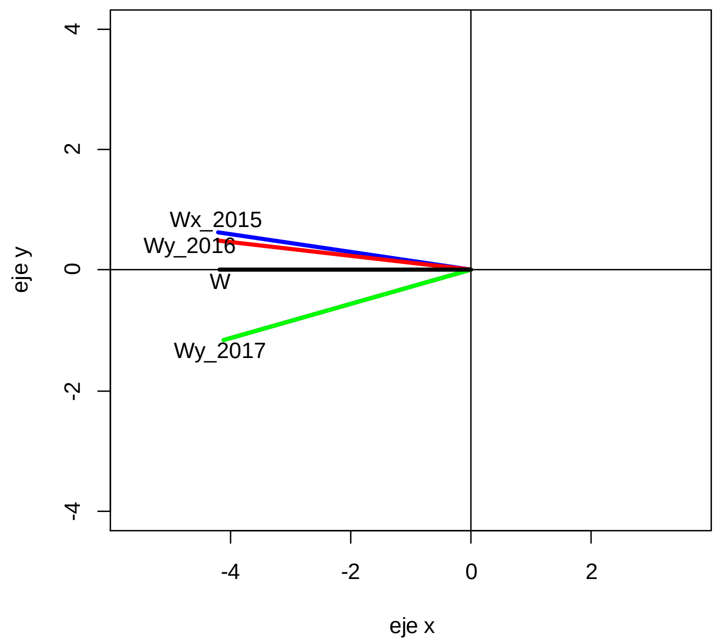


Figure 4: Factorial plane of the inter-structure.

$$\left(\frac{\chi^2_{Depart.Punt.Icfes(2015)}}{5050} = 0.20\,, \quad \frac{\chi^2_{Depart.Punt.Icfes(2016)}}{5050} = 0.19, \frac{\chi^2_{Depart.Punt.Icfes(2017)}}{5050} = 0.17\right)$$

Indeed, the decline in variability is indicative of a slight propensity for one more similarity in the profile scores ICFES between the departments of the Andean region of Colombia, between 2015 and 2017.
Also, it is interesting to note, Table 11, in what regards the distances between objects, are much lower between the blocks W_{2015}, W_{2016} and, and higher of these two compared with W_{2017}. This result is mainly due to the association scores Between $ICFES_{2015-2016}(\chi^2_{Punt.Icfes(2015).Punt.Icfes(2016)} = 16772)$ is higher than the corresponding association Between scores $ICFES_{2015-2017}(\chi^2_{Punt.Icfes(2015).Punt.Icfes(2017)} = 11466)$, to Themselves and Between scores $ICFES_{2016-2017}(\chi^2_{Punt.Icfes(2016).Punt.Icfes(2017)} = 12224)$. These results are in consequence with the construction of the STATIS-BCA distance between blocks in which it is established that the greater the association between the variables that identify the blocks, the smaller the distance between them.

Object comparison	Distance HS
$\mathbf{d}(\boldsymbol{W_{2015}}, \boldsymbol{W_{2016}})$	0.1331403
$\boldsymbol{d}(\boldsymbol{W_{2015}}, \boldsymbol{W_{2017}})$	1.7890067
$\boldsymbol{d}(\boldsymbol{W_{2016}}, \boldsymbol{W_{2017}})$	1.6567194

Table 11: Distances between blocks years 2015, 2016 and 2017.

profiles in the DRA-Col remains more or less constant throughout the three years under study, information that is approximated by the length of the vectors that bind each object to the coordinate origin. Similarly, the graphics device stands near and/or out between objects, in which the great similarity is confirmed in structure profiles scores ICFES of the departments of the Andean region of Colombia for 2015 vs 2016, differing from the corresponding structure to profiles in 2017.
All this is reflected in figure 4, wherein the depicted orthogonal projection of the objects on the first factorial plane inter-structure. Indeed, in this graph, it is possible to visualize that the variability of the ICFES score
The mathematical aspects that allow the previous representation rest on the spectral decomposition of determined by the matrices of eigenvectors $\boldsymbol{a^{\alpha}}$ and eigenvalues $\boldsymbol{\mu_{\alpha}}$:

$$\boldsymbol{a^{\alpha}} = \begin{bmatrix} -0.58072 & 0.44891 & 0.67914 \\ -0.58188 & 0.35454 & -0.73192 \\ -0.56935 & -0.82022 & 0.05531 \end{bmatrix} \qquad \boldsymbol{\mu_{\alpha}} = \begin{bmatrix} 52.46929 & 0 & 0 \\ 0 & 1.98282 & 0 \\ 0 & 0 & 0.6761622 \end{bmatrix}$$

With this information, you obtain the Euclidean representation of objects on the prime plane r factorial, using the first two s factors analysis. The projection coordinates of the objects in the respective blocks are organized on the columns of the following matrix:

$$\boldsymbol{\psi} = \begin{bmatrix} -4.206542 & 0.6321257 \\ -4.214913 & 0.4992488 \\ -4.124173 & -1.1549838 \end{bmatrix}$$

The strength of the first direction in this representation space is given by the quotient:

$$\frac{\mu_1}{\sum_{\alpha=1}^{K} \mu_{\alpha}} = \frac{52.46929}{55.12826} = 0.9517$$

In other words, almost all (95%) of the association between ICFES scores in the departments of the Andean Region in Colombia can be explained on this space, which is equivalent to pointing out that the distances between the objects effectively indicate that the changes in the mobility of the EI from one classification to another in ICFES scores were not significant between 2015 and 2016, with mobility being more evident in 2017, compared to the previous two years.

7. CONCLUSIONS

In this research, the theoretical and applied foundations of an analytical proposal based on an adaptation of the STATIS methodology have been formally presented, which makes use of BCA to address the statistical treatment of blocks of information generated by characterizing them. individuals, using the same two categorical variables, on *K* different occasions. The proposed procedure, which has been validated in real data, allows:
To carry out an analysis of the association between the categorical variables that define the contingency tables within each block, following the classic approach that defines the BCA.
Perform analyses comparative changes that occur in association variables as between blocks along with different occasions.
Measure the separation between two blocks of information as a function of the degree of association between the variables that respectively identify those blocks, highlighting that the distance between two blocks will be smaller the greater the degree of association between the variables that respectively identify those blocks.

Represent the blocks of information on a graphic device in which the variability measurements inside the blocks are displayed, and also account for measures of the distance between the information structures that identify the different blocks. The simulation processes carried out under scenarios differentiated by the degree of association between the variables, within and between blocks, have made it possible to corroborate in practice the postulates derived from the theory on which the proposal rests.

The results of the application on the Saber 11 test in the Andean region, according to the BCA carried out for each year of the study indicated that for both 2015 and 2016, the departments are ordered according to a Guttman effect that shows that, in the capital of the country, Bogotá, mostly schools are rated with A + and A scores. In schools in Cundinamarca and Santander, the rating drops to level B, finding that in schools in the northern departments of Santander and Caldas the ICFES rating it drops to level C. Finally, in Antioquia and Tolima, schools are assigned the lowest grade, D. In 2017 there was an improvement in the department of North Santander having scores in category C in 2015 and 2016, ascending to category B in 2017.

In the graph inter structure it was observed that the axis differentiates two objects achieved, even when the factor levels, it was observed that the structures are substantially similar in the three years of study. These differences are because there has been a migration, although not very marked, of the ICFES scores of some educational institutions.

Acknowledgements: This research is the result of a project financed by the Universidad Autónoma de Bucaramanga (UNAB) under the project code 2021/00003/001/001/036.

RECEIVED: APRIL, 2021.
REVISED: MAY, 2023.

REFERENCES

[1] BÉCUE-BERTAUT, M. and PAGÉS, J. (2008): Multiple factor analysis and clustering of a mixture of quantitative categorical and frequency data. **Computational Statistics & Data Analysis**, 52(6), 3255-3268.

[2] DJAUHARI. M. (1998): Operators averaging and its application in qualitative variables selection. **Journal of Mathematical Science**, 3(1), 41-49.

[3] ESCOUFIER. Y. (2006): Operator related to a data matrix: a survey. In: RIZZI A. and VICHI, M. (eds). **Compstat 2006 - Proceedings in Computational Statistics**, 285-297.

[4] FAJARDO, E., ROMERO, H., PLATA, L. and RAMÍREZ, M. (2018): Determinantes de la calidad de la educación en Colombia a nivel secundario: una aplicación del análisis de correspondencia canónica. **Revista Espacios**, 39, (15), 25.

[5] FAJARDO, E., BELEÑO-MONTAGUT, L. and ROMERO, H. (2021): Incidencia de los factores socioeconómicos en la calidad de la educación media regional en Colombia. **Interciencia**, 46, (3), 118-125.

[6] GLYNN, D. (2014): Correspondence analysis: Exploring data and identifying patterns, in GLYNN, D. and ROBINSON, J. (Editors) **Corpus Methods for Semantics: Quantitative studies in polysemy and synonymy**, John Benjamins Publishing Co, Amsterdam, pp. 443-485.

[7] GREENACRE, M. and BLASIUS, J. (1995): **Correspondence analysis in the social sciences**. Academic Press, San Diego.

[8] GREENACRE, M. (2017): **Correspondence analysis in Practice**. CRC Press, Boca Raton.

[9] HAIR, J., ANDERSON, R., TATHAM, R. and BLACK, W. (2007): **Análisis multivariante**. Prentice Hall, Spain. Madrid.

[10] LEBART, L., MORINEAU, A. and WARWICK. K. (1984): **Multivariate descriptive statistical analysis and related techniques for large matrices**. Wiley, New York.

[11] MICHAILIDIS, G. (2007): Correspondence analysis, in SALKIND, N. (Editor). **Encyclopedia of Measurement and Statistic**, Sage Publications, Thousand Oaks, pp. 191-193.

[12] TORRES. R. (2009): **Una propuesta para el tratamiento de tablas de contingencias múltiples bajo una perspectiva de adaptación de la metodología STATIS.** Postgrado en Estadística, Universidad Central de Venezuela, PhD. Thesis.

[13] THIOULOUSE, J., DRAY, S., DUFOUR, A., SIBERCHICOT, A., JOMBART, T. and PAVOINE, S. (2018): **Multivariate Analysis of Ecological Data with ade4**. Springer, New York.

[14] VIVIEN, M. and SABATIER, R. (2004): A generalization of STATIS-ACT strategy: DO-ACT for two multiblocks tables. **Computational Statistics & Data Analysis**, 46 (1), 151-171.