

## UNA CAH DE LOS OPERADORES WD DE YVES ESCOUFIER EN UN ACM

FRANCISCO JAVIER DÍAZ-LLANOS SÁINZ-CALLEJA  
JOSÉ LUIS VALENCIA DELFA

### OBJETIVO

El objetivo principal de este artículo es el de mostrar, de la forma más didáctica posible, cómo se construye un **dendrograma** a partir de la **matriz de distancia entre los operadores WD de Yves Escoufier** y un **criterio de agregación**.

Mientras que el cálculo de la **matriz de distancias entre los operadores WD de Yves Escoufier** está suficientemente detallado en (1), el del **criterio de agregación de la varianza** que nos aconseja Pierre Cazes lo está en (2). El algoritmo de clasificación jerárquica ascendente, teniendo en cuenta la distancia entre **operadores** y el **criterio de agregación**, lo desarrollaremos con el suficiente detalle para ser entendido para todos aquellos investigadores en ciencias experimentales que no pertenezcan al área de la Estadística.

Para la resolución de los ejercicios didácticos y el práctico partiremos de **cuatro matrices de distancias** (3, 4) y una **matriz de distancia** (1, 5), respectivamente.

Veremos que la mejor opción para el cálculo de las distancias es la de retener la **distancia entre los operadores WD de Yves Escoufier transformados y normados**.

### PALABRAS CLAVE

Algoritmo de clasificación jerárquica ascendente (**CAH**), análisis de correspondencias múltiples (**ACM**), operador **WD** de Yves Escoufier, distancias entre los operadores **WD** de Yves Escoufier en el **ACM**.

### SUMMARY

Hierarchical-Ascendant-Clustering-based on the variance criterion (2)-Algorithm, distances matrix between-centered (transformed) and standard-operators, has been applied furthermore, and, with the constructed dendrogram, wellknown previous stu-

des, have been -partially-supported-the attainment of homogeneous classes by means of a cut using a horizontal line. All results are discussed.

## KEYWORDS

Hierrarchical clustering analysis; Multiple correspondance analysis; Operator WD Yves Escoufier; Distance betwenn operators WD in multiple correspondande analysis.

## INTRODUCCIÓN

Dado que este tipo de ejercicio no se contempla en ningún libro de **Análisis Estadístico Multidimensional**, hemos tomado la decisión de mostrarlo debido a que presenta un gran interés práctico para los investigadores de ciencias experimentales que no sean de Estadística. De esta manera, podrán utilizarlo y extraer conclusiones de interés de sus datos empíricos.

## MATERIAL Y MÉTODO

Dado que el número de cuestiones contenidas en un cuestionario no es aconsejable que sea elevado, la **matriz de distancias entre los oepradores de Yves Escoufier** será de orden elevado. Por ejemplo, en un cuestionario sobre el **cierre de ventas** diseñado por N. Rackman y contempaldo en castellano por Díaz-Llanos (6), el número de cuestiones es de 15 y cada una de ellas está a cinco modalidades. En este caso la **matriz de distancias** es de orden 15 y, por consiguiente, tendremos que calcular 105 distancias. Ni que decir tiene, en este caso concreto, nos vemos obligados a la realización de un programa informático para el cálculo de las distancias.

En los ejemplos que vemos en este artículo no es necesario la realización de ningún programa informático, ya que el orden de la **matriz de distancias es pequeño**.

Mientras que los resultados obtenidos del **ejercicio didáctico** carecen de interpretación práctica, ya que el orden de la **matriz de datos** es de pequeña, los del **ejercicio práctico** sí se pueden interpretar, ya que el orden de la **matriz de datos** es aceptable aunque sea igual a 5.

El hecho de que el orden de la **matriz de distancias entre los operadores de Yves Escoufier** en el caso de una **encuesta de opinión** no sea mayor que 100, nos inclinamos a aplicar un algoritmo de clasificación jerárquica en lugar de un algoritmo de clasificación no jerárquica. Recordemos que un algoritmo de clasificación no jerárquica tan sólo es recomendado cuando el orden de la matriz de datos sea muy elevado. En caso contrario es preferible el algoritmo de clasificación jerárquica, ya que presenta mejores propiedades.

Así pues, para llegar al **dendrograma** es preciso definir dos conceptos:

1. **La distancia.**
2. **El criterio de agregación.**

## 1. LA DISTANCIA

En cuanto concierne a la **distancia** hemos decidido aplicar para el ejercicio didáctico **cuatro desistancias entre los operadores de Yves Escoufier**, y para el ejercicio práctico la **distancia entre los operadores WD de Yves Escoufier transformados y normados**.

## 2. CRITERIO DE AGREGACIÓN

Entre los distintos  **criterios de agregación** hemos elegido el de  **maximizar el momento central de orden dos de una partición**, ya que presenta propiedades más interesantes que el resto. Una de ellas es que este  **criterio** proporciona simultáneamente  **clases homogéneas y bien separadas las unas de las otras** (7).

Para el desarrollo de dicho algoritmo de clasificación jerárquica hemos tenido en cuenta la publicación de Djillai Maiti y de Yves-F. Thomas (2), en la cual Pierre Cazes, discípulo de J. P. Benzécri (fundador de la revista *Les Cahiers de l'Analyses des Données* publiés par le Laboratoire de Statistique de l'Université Pierre-et-Marie-Curie et par l'Association pour le Développement et la Diffusion de l'Analyse des Données avec le concours du Centre National de la Recherche Scientifique) presenta de forma didáctica en dicha publicación.

Así como en el ejercicio didáctico, carente de interpretación, las cuatro matrices de distancias han sido obtenidas de (3, 4), en el ejercicio práctico la matriz de distancia de orden 5 ha sido calculada a partir de los datos contenidos en (5).

## EJERCICIO DIDÁCTICO

Haciendo uso de las cuatro distancias entre los operadores en el ACM ya deducidas en (1) y del  **criterio de agregación** expuesto por Pierre Cazes en (2) obtendremos los cuatro  **dendrogramas** para el ejercicio didáctico. Ni que decir tiene, debido a que el orden de las  **matrices entre operadores** es pequeña, la interpretación de los datos empíricos de partida carece de valor.

A continuación vamos a mostrar las cuatro  **matrices de distancias entre los operadores** ya obtenidas en (1).

### Primera situación

La  **matriz de distancias entre operadores sin normar**

WD

$$\begin{pmatrix} 0,0000 & 1,5275 & 1,1442 \\ 1,5275 & 0,0000 & 1,0000 \\ 1,4142 & 1,0000 & 0,0000 \end{pmatrix}$$

## Segunda situación

La matriz de distancias entre los operadores normados

$$\frac{WD}{\|WD\|}$$
$$\begin{pmatrix} 0,0000 & 0,9546 & 0,7653 \\ 0,9546 & 0,0000 & 0,5176 \\ 0,7653 & 0,5176 & 0,0000 \end{pmatrix}$$

## Tercera situación

La matriz de distancias entre los operadores transformados

$$W^*D$$
$$\begin{pmatrix} 0,0000 & 1,5275 & 1,4142 \\ 1,5275 & 0,0000 & 1,0000 \\ 1,4142 & 1,0000 & 0,0000 \end{pmatrix}$$

## Cuarta situación

La matriz de distancia entre los operadores transformados y normados

$$\frac{W^*D}{\|W^*D\|}$$
$$\begin{pmatrix} 0,0000 & 1,2364 & 0,9193 \\ 1,2364 & 0,0000 & 0,6058 \\ 0,9193 & 0,6058 & 0,0000 \end{pmatrix}$$

## Construcción de los cuatro dendrogramas. Ejercicio didáctico

Vamos a construir no sólo los cuatro **dendrogramas**, sino también realizar un **corte horizontal** en el **nivel 0,7**. Aunque el resultado obtenido carezca de valor interpretativo, a nivel didáctico sí tiene sentido.

## Primera situación

Matriz de distancias entre operadores sin normar

$$WD$$
$$\begin{pmatrix} 0,0000 & 1,5275 & 1,4142 \\ 1,5275 & 0,0000 & 1,0000 \\ 1,4142 & 1,0000 & 0,0000 \end{pmatrix}$$

En el paso 1, después de haber reunido **2** y **3** en el **nivel 1**, aplicando la fórmula de la **distancia** contenida en (2) y teniendo en cuenta la siguiente notación:

$$m_1 + m_2 = m_{1,2}$$

$$m_1 + m_3 = m_{1,3}$$

$$m_1 + m_2 + m_3 = m_{1,2,3}$$

tenemos:

$$D(1, \{2,3\}) = \frac{m_{1,2} d(2,1) + m_{1,3} d(3,1) - m_1 d(2,3)}{m_{1,2,3}}$$

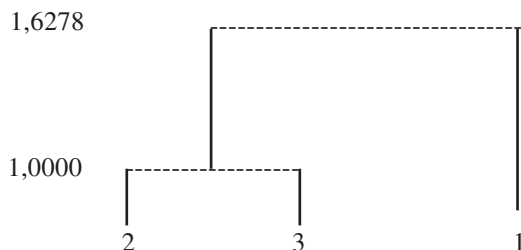
Aplicando esta fórmula a nuestro caso concreto obtenemos los siguientes resultados:

$$D(1, \{2,3\}) = \frac{2(1,5775) + 2(1,4142) - 1}{3} = 1,6278$$

La situación de la nueva **matriz de distancias entre operadores sin normar** es:

X	{2, 3}	1
{2, 3}	0,0000	1,6278
1	1,6278	0,0000

Por tanto, el **dendrograma** obtenido adopta la siguiente forma:



Realizando un **corte horizontal al dendrograma** en el **nivel 0,7000** obtenemos dos clases: la **clase 1** contiene dos operadores (**2** y **3**) y la **clase 2** contiene un operador (**1**).

## Segunda situación

### Matriz de distancias entre operadores normados

$$\frac{WD}{\|WD\|}$$

$$\begin{pmatrix} 0,0000 & 0,9546 & 0,7653 \\ 0,9546 & 0,0000 & 0,5176 \\ 0,7653 & 0,5176 & 0,0000 \end{pmatrix}$$

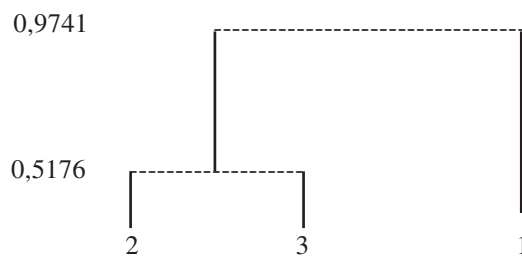
En el paso 1 después de haber reunido **2** y **3** en el **nivel 0,5176**, aplicando la misma fórmula de la **distancia** contenida en (2) y teniendo en cuenta la misma notación que en **primera situación**, tenemos:

$$D(1, \{2,3\}) = \frac{2(0,9546) + 2(0,7653) - 0,5176}{3} = 0,9741$$

La situación de la nueva **matriz de distancias entre operadores normados** es:

X	{2, 3}	1
{2, 3}	0,0000	0,9741
1	0,9741	0,0000

Por tanto, el **dendrograma** obtenido adopta la siguiente forma:



Realizando un **corte horizontal al dendrograma** en el **nivel 0,7000** obtenemos dos clases: la **clase 1** contiene dos operadores (**2** y **3**) y la **clase 2** contiene un operador (**1**).

### Tercera situación

#### Matriz de distancia entre operadores transformados

$$W^*D$$
$$\begin{pmatrix} 0,0000 & 1,5275 & 1,4142 \\ 1,5275 & 0,0000 & 1,0000 \\ 1,4142 & 1,0000 & 0,0000 \end{pmatrix}$$

Dado que la **matriz de distancias entre operadores sin normar y los operadores transformados** es la **misma** los resultados obtenidos son los **mismos** que en la **primera situación**. Por esta razón, para no ser reiterativos los omitimos.

### Cuarta situación

#### Matriz de distancias entre operadores transformados y normados

$$\frac{W^*D}{\|W^*D\|}$$
$$\begin{pmatrix} 0,0000 & 1,2364 & 0,9193 \\ 1,2364 & 0,0000 & 0,6058 \\ 0,9193 & 0,6058 & 0,0000 \end{pmatrix}$$

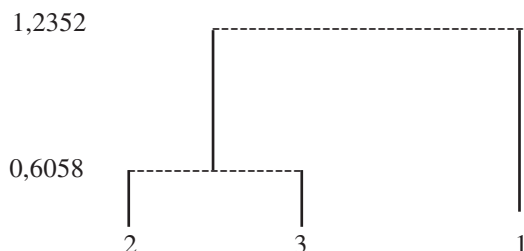
En el paso 1 después de haber reunido **2** y **3** en el **nivel 0,6058**, aplicando la misma fórmula de la **distancia** contenida en (2) teniendo en cuenta la misma notación que en la **primera situación**, tenemos:

$$D(1, \{2,3\}) = \frac{2(1,2364) + 2(0,9193) - 0,6068}{3} = 1,2352$$

La situación de la nueva **matriz de distancias entre operadores transformados y normados** es:

X	{2, 3}	1
{2, 3}	0,0000	1,2352
1	1,2352	0,0000

Por tanto, el **dendrograma** obtenido adopta la siguiente forma:



Realizando un **corte horizontal al dendrograma** en el **nivel 0,7000** obtenemos dos clases: la **clase 1** contiene dos operadores (**2, 3**) y la **clase 2** contiene un operador (**1**).

## CONCLUSIONES

1. Los **dendrogramas** obtenidos mediante la **primera** y **tercera situación** son los mismos, ya que las **distancias entre los operadores** también lo son.
2. En las **cuatro situaciones** observamos que se agregan los operadores **2** y **3** en una **clase**, y se distancian del operador **1**.
3. En la **cuarta situación** la **distancia** entre los operadores **2** y **3** frente a la del operador **1** es ligeramente mayor que en las restantes.

Por consiguiente, la **cuarta situación** es la que consideramos más conveniente a la hora de la construcción de un **dendrograma**.

Esta aseveración sería concluyente en el caso de que el **orden** de la **matriz de partida** fuese de grandes dimensiones, pero en caso contrario no lo es.

## EJERCICIO PRÁCTICO

Haciendo uso de la **distancia entre operadores transformados y normados** en el **ACM** (1) y del **criterio de agregación** indicado por Pierre Cazes (2) obtendremos el **dendrograma** para el **ejercicio práctico**. Dado que el **orden** de la **matriz entre operadores** es aceptable, la interpretación de los **datos empíricos** de partida tendrá cierto valor. Los **datos empíricos** provenientes para la realización del **ejercicio práctico** han sido extraídos del libro de E. Diday, J. Lemaire, J. Pouget y J. Testu (5).

## CONSTRUCCIÓN DEL DENDROGRAMA HACIENDO USO DEL CRITERIO DE AGREGACIÓN DE LA VARIANZA

Vamos a construir no sólo el **dendrograma**, sino también realizaremos un **corte horizontal** en el **nivel 1,4000** con el fin de obtener **clases** lo más **homogéneas** posibles.



## Matriz de distancia entre operadores transformados y normados

$$\frac{W^*D}{\|W^*D\|}$$

$$\begin{pmatrix} 0,0000 & 1,2863 & 1,3976 & 1,3377 & 1,2675 \\ 1,2863 & 0,0000 & 1,2901 & 1,2993 & 0,7310 \\ 1,3976 & 1,2901 & 0,0000 & 1,3369 & 1,2797 \\ 1,3377 & 1,2993 & 1,3369 & 0,0000 & 1,3377 \\ 1,2675 & 0,7310 & 1,2797 & 1,3377 & 0,0000 \end{pmatrix}$$

En el paso 1, después de haber reunido **2** y **5** en el **nivel 0,7310**, aplicando las fórmulas de las **distancias** contenidas en (2) y teniendo en cuenta la siguiente notación:

$$m_1 + m_2 = m_{1,2}$$

$$m_1 + m_3 = m_{1,3}$$

$$m_1 + m_2 + m_5 = m_{1,2,5}$$

$$m_2 + m_3 + m_5 = m_{2,3,5}$$

$$m_2 + m_4 + m_5 = m_{2,4,5}$$

tenemos:

$$D(1, \{2,5\}) = \frac{m_{1,2} d(1,2) + m_{1,5} d(1,5) - m_1 d(2,5)}{m_{1,2,5}}$$

$$D(3, \{2,5\}) = \frac{m_{3,2} d(2,3) + m_{3,5} d(3,5) - m_3 d(2,5)}{m_{2,3,5}}$$

$$D(4, \{2,5\}) = \frac{m_{4,2} d(2,4) + m_{4,5} d(4,5) - m_4 d(2,5)}{m_{2,4,5}}$$

Aplicando estas fórmulas a nuestro caso concreto obtenemos los siguientes resultados:

$$D(1, \{2,5\}) = \frac{2(1,2863) + 2(1,2675) - 0,7310}{3} = 1,4589$$

$$D(3, \{2,5\}) = \frac{2(1,2901) + 2(1,2797) - 0,7310}{3} = 1,4695$$

$$D(4, \{2,5\}) = \frac{2(1,2993) + 2(1,3377) - 0,7310}{3} = 1,5143$$

La situación de la nueva **matriz de distancias entre los operadores transformados y normados** es:

X	{2, 5}	1	3	4
{2, 5}	0,0000	1,4589	1,4695	1,5143
1	1,4589	0,0000	1,3976	1,3377
3	1,4695	1,1376	0,0000	1,3369
4	1,5143	1,3377	1,3369	0,0000

Por tanto, la **primera fase** de la construcción del **dendrograma** es:



En el 2 paso, después de haber reunido **3 y 4** en el **nivel 1,3369**, aplicando las fórmulas de las **distancias** contenidas en (2) y teniendo en cuenta la siguiente notación:

$$m_2 + m_5 = m_{2,5}$$

$$m_2 + m_3 + m_5 = m_{2,3,5}$$

$$m_2 + m_4 + m_5 = m_{2,4,5}$$

tenemos:

$$D(1, \{3,4\}) = \frac{m_{1,3} d(1,3) + m_{1,4} d(1,4) - m_1 d(3,4)}{m_{1,3,4}}$$

$$D(\{2,5\}, \{3,4\}) = \frac{m_{2,5,3} D[\{2,5\}, 3] + m_{2,5,4} D(\{2,5\}, 4) - m_{2,5} d(3,4)}{m_{2,3,4,5}}$$

Aplicando estas las fórmulas a nuestro caso concreto obtenemos los siguientes resultados:

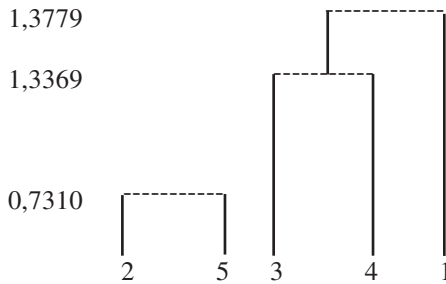
$$D(1, \{3,4\}) = \frac{2(1,3976) + 2(1,3377) - 1,3369}{3} = 1,3778$$

$$D(\{2,5\}, \{3,4\}) = \frac{3(1,4695) + 3(1,5143) - 2(1,3369)}{4} = 1,5694$$

La nueva situación de la **matriz de distancias entre operadores transformados y normados** es:

X	{2, 5}	{3, 4}	1
{2, 5}	0,0000	1,5694	1,4589
{3, 4}	1,5694	0,0000	1,3779
1	1,4589	1,3779	0,0000

Por tanto, la **segunda fase** de la construcción del **dendrograma** es:



En el **tercer paso** después de haber reunido (3 y 4) con 1 en el **nivel 1,3779**, aplicando las fórmulas de las **distancias** contenidas en (2) y teniendo en cuenta los resultados anteriores y la siguiente notación:

$$m_1 + m_2 + m_3 + m_4 + m_5 = m_{1,2,3,4,5}$$

tenemos:

$$D(\{2,5\},\{\{3,4\},1\}) = \frac{m_{2,5,3,4} D(\{2,5\},\{3,4\}) + m_{2,5,1} D(\{2,5\},1) - m_{2,5} d(\{3,4\},1)}{m_{1,2,3,4,5}}$$

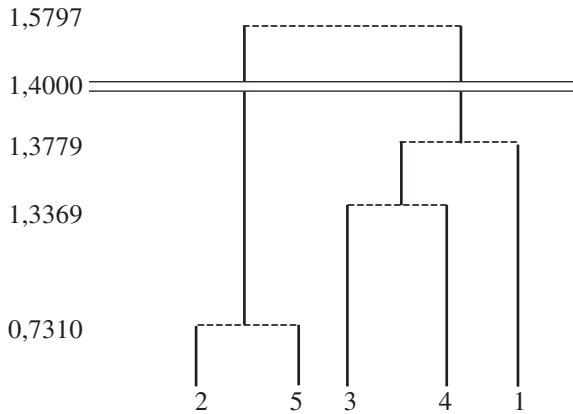
Haciendo uso de los resultados anteriores obtenemos:

$$D(\{2,5\},\{\{3,4\},1\}) = \frac{4(1,5694) + 3(1,4589) - 2(1,3779)}{5} = 1,5797$$

La situación de la nueva **matriz de distancias entre los operadores transformados y normados** es:

X	{2, 5}	{{3, 4},1}
{2, 5}	0,0000	1,5797
{{3, 4},1}	1,5797	0,0000

Por tanto, el **dendrograma** obtenido adopta la siguiente forma:



## CONCLUSIÓN

Tal como indican F. Benzécri, J.-P. Benzécri, Y. L. Cheung y S. Maïza (8), el **corte** que se realice en el **dendrograma**, proveniente de un **algoritmo de clasificación jerárquica ascendente**, puede hacerse mediante una **línea horizontal** o **sinuosa**.

La elección de una u otra estrategia vendrá dada exclusivamente de los criterios del **investigador en ciencias experimentales**. Esta forma de actuar es razonable, ya que ellos son los que aportan los **datos empíricos** con el fin de buscar una **tipología** concreta lo más próxima posible a la realidad.

Por consiguiente, si se desea formar **clases homogéneas** de operadores hay que tener en cuenta tanto la **teoría** como la **práctica**. Esta forma de pensar está en concordancia con la frase del filósofo alemán Immanuel Kant:

La **teoría** sin **práctica** es ciega.

La **práctica** sin **teoría** es absurda.

Como resultado de la realización de un **corte al dendrograma mediante una línea horizontal** en el nivel **1,4000** obtenemos dos clases constituidas por **dos** y **tres operadores transformados y normados**, respectivamente.

La primera **clase** está constituida por:

$$EAU - \frac{W_2^* D}{\|W_2^* D\|} \quad CAL - \frac{W_5^* D}{\|W_5^* D\|}$$

La segunda **clase** está constituida por:

$$GLU - \frac{W_3^* D}{\|W_3^* D\|} \quad PRO - \frac{W_4^* D}{\|W_4^* D\|} \quad PRI - \frac{W_1^* D}{\|W_1^* D\|}$$

De lo que se desprende que existe una gran similitud, por una parte, entre el contenido en agua (EAU) y la cantidad de calorías y, por otra, entre la cantidad de glúcidos (GLU), la cantidad de proteínas (PRO) y el precio (PRI).

Asimismo, existe un cierto grado de **discriminación** entre las **dos clases**.

El hecho de que hayamos presentado este ejercicio de forma didáctica es con el fin de que el **investigador**, no perteneciente a la disciplina de la Estadística, pueda comprender con facilidad los pasos que hay que hacer para llegar a este **dendrograma**. Ni que decir tiene, cuando la **matriz de distancia sea de orden mayor que 5** habrá que recurrir a la aplicación de un programa que construya el **dendrograma**.

## BIBLIOGRAFÍA

- (1) Castilla Plaza, C.; Díaz-Llanos Sáinz-Calleja, Fco. J.; Fernández Cancio, A. (2007): *Distancias entre los operadores WD de Yves Escoufier en el análisis de correspondencias múltiples*. Anales de la R.A.D.E. Vol. 11, n.º 1, pp. 45-64.
- (2) Maiti, D.; Thomas, Yves-F. (1975): «Interactions des plantes et du vent dans les dunes littorales. Memoires du Laboratoire de Geomorphologie de L'ecole pratique des hautes etudes». *F-35800 Dinard*, n.º 28. Con la colaboración de Pierre Cazes, profesor de la Université de Pierre-et-Marie-Curie.
- (3) Díaz-Llanos Sáinz-Calleja, Fco. J. (1995): *El tratamiento estadístico de las encuestas de opinión, pieza clave en la Ingeniería de la Demanda. Un enfoque didáctico y conceptual*. Ediciones CEU (Centro Europeo de Estudios Superiores). ISBN.84-88881-22-3.
- (4) Valencia Delfa, J. L.; Díaz-Llanos Sáinz-Calleja, Fco. J.; Tarazona Lafarga, J. V. (2007): *Efecto y utilidad del coeficiente RV de Yves Escoufier en el análisis de correspondencias múltiples*. Anales de la R.A.D.E. Vol. 11, n.º 1, pp. 9-44.
- (5) Diday, E.; Lemaire, J.; Pouget, J.; Testu, J. (1982): *Éléments d'analyses des données*. Dunod.
- (6) Díaz-Llanos Sáinz-Calleja, Fco. J. (2002): *El análisis de datos en el cierre de ventas*. Editorial La Muralla, S. A. Editorial Hespérides, S. L.
- (7) Benzécri, J.-P. & collaborateurs (1981): *Pratique de l'Analyse des Données. Linguistique et Lexicologie*. Dunod.
- (8) Benzécri, F.; Benzécri, J.-P. & collaborateurs (1986): *Pratique de l'Analyse des Données en Economie*. Dunod.