
Análisis de datos longitudinales cualitativos mediante la combinación de Análisis Armónico Cualitativo y Análisis Factorial Múltiple¹

Longitudinal Data Analysis by combining Harmonic Qualitative
Analysis and Multiple Factor Analysis

Martha Lucía Corrales^a
martha.corrales@usa.edu.co

Campo Elías Pardo^b
cepardot@unal.edu.co

Resumen

Se propone una metodología estadística para el tratamiento de datos longitudinales categóricos, por medio de la combinación del Análisis Armónico Cualitativo (AAC) y del Análisis Factorial Múltiple (AFM). Como ejemplo, se analizan los datos de la medición del *rating* de televisión en Colombia, en un día determinado, y se obtiene una tipología de los televidentes a través de la combinación de los dos métodos factoriales y clasificación

Palabras clave: análisis de correspondencias difuso, análisis de datos longitudinales, análisis factorial múltiple, análisis multivariado de datos, clasificación, rating de televisión.

Abstract

A statistical methodology is proposed in order to analyse longitudinal categorical data. This methodology considers the use of two methods: Qualitative Harmonic and Multiple Factor Analysis. An application for the study of Colombian rating television data for a given day and the resulting typology of television viewers is also shown as an example

Keywords: multivariate data analysis, fuzzy correspondence analysis, longitudinal data analysis, TV rating, cluster analysis.

¹Corrales, M. L., Pardo, C. E. (2015). Análisis de datos longitudinales cualitativos mediante la combinación de Análisis Armónico Cualitativo y Análisis Factorial Múltiple. *Comunicaciones en Estadística*, **8**(1), 11-32.

^aDocente, Escuela de Matemáticas, Universidad Sergio Arboleda. Colombia.

^bProfesor asociado, Departamento de Estadística, Universidad Nacional de Colombia. Colombia.

1. Introducción

El estudio de los cambios de estado de una variable categórica a través del tiempo es de interés en distintas investigaciones. En demografía, por ejemplo, se estudian trayectorias de grupos de individuos: cambios de residencia, carreras laborales, cambios de nivel socioeconómico, entre otras. Este tipo de datos, llamado también datos biográficos, se ha analizado mediante el Análisis Armónico Cualitativo (AAC; Deville & Saporta (1980), Deville (1982)). Barbary & Pinzón (1998) analizan una encuesta sobre las trayectorias residenciales, profesionales y familiares de residentes en el área metropolitana de Bogotá D.C., consiguiendo una tipología de la movilidad intra-urbana y su relación con otros elementos de la encuesta. Corrales (2000) (ver también Dureau & Flórez (2000)) analiza las trayectorias de migración hacia la zona petrolera de Yopal, Aguazal y Tauramena en Colombia.

Los documentos Barbary (1996), Pinzón (1998) y Barbary & Pinzón (1998), son una buena guía metodológica para la aplicación del ACC, que es en la práctica un análisis de correspondencias, el cual se complementa con una clasificación automática, siguiendo el procedimiento descrito en Lebart, Piron & Morineau (2006).

La motivación de este trabajo es el análisis del *rating* de televisión en donde se registra para cada televisor encendido el canal que se sintoniza, cuando la permanencia en el canal es por lo menos de un minuto. Los estados de la variable *canal* son los canales disponibles en la oferta de la región geográfica que se analiza. Si se considera el *rating* de un día es interesante tener en cuenta el tipo de horario, por ejemplo: infantil, familiar, *pruntime*. Se propone combinar el AAC con el Análisis Factorial Múltiple (AFM; Escofier & Pagès (1984); Pagès (2004)), para tener en cuenta esta estructuración de los datos. El AFM se presenta como una técnica factorial para la descripción de un conjunto de individuos a través de varios grupos de variables. Al combinar el AAC con el AFM, cada grupo corresponde a uno o varios periodos de tiempo.

En el AAC, la tabla que se describe en el análisis de correspondencias tiene en filas a los individuos y en columnas los estados de la variable categórica en varios periodos de tiempo. El valor de la celda representa la proporción de tiempo que un individuo permanece en el estado de la variable categórica y el pedido de tiempo de la columna. Cuando la proporción de tiempo se da solo en relación con el respectivo periodo, la tabla se constituye en una codificación difusa de los individuos, dentro de los periodos de tiempo y el AAC es también un análisis de correspondencias difuso (ACD; Chevenet et al. (1994); Gamez (2012)).

En este documento se trata el caso de codificación difusa dentro de los periodos de tiempo, lo que hace que el análisis armónico cualitativo (AAC) sea, en la práctica un análisis de correspondencias difuso (ACD). El AFM se combina con el AAC para equilibrar la influencia de los periodos de tiempo y, también, para tener en cuenta posibles agrupaciones de los periodos de tiempo y enriquecer el análisis de los datos.

En la sección 2 se presenta la notación utilizada en el trabajo. En la sección 3 se muestra un resumen de los métodos involucrados en la combinación del AFM y el AAC. En la sección 4 se realiza la combinación entre el AAC y el AFM, y se presenta el desarrollo de las fórmulas clásicas de un análisis en componentes principales ponderado, adaptadas a dicha combinación. Con el fin de ilustrar esta metodología, en la sección 5, se analiza la medición del *rating* de un día específico de la semana en Colombia, estableciendo una tipología de los televidentes con respecto a la preferencia de los canales vistos. En la sección 6 se presentan algunas conclusiones encontradas en el desarrollo teórico y metodológico de la combinación.

2. Estructura de los datos

Se desarrolla la metodología para el caso en el que el ACC es un ACD, razón por la que se adopta la notación de Chevenet et al. (1994). Sean $t = 1, 2, \dots, T$ los intervalos de tiempo en los que se divide el periodo de estudio $[0, T]$, $j = 1, \dots, m_t$ las categorías de la variable longitudinal que es objeto de análisis, medida en el periodo t para $i = 1, \dots, n$ individuos del estudio. Los intervalos de tiempo en que se divide el periodo de estudio se denominan también periodos.

Se denota:

- \mathbf{A}_t a la tabla de datos formada por n filas y m_t columnas.
- \mathbf{A} a la tabla formada por la yuxtaposición de las T tablas \mathbf{A}_t formada por n filas y $m = \sum_{t=1}^T m_t$ columnas.
- El elemento a_{ij}^t representa el tiempo que el individuo i permanece en la categoría j de la variable longitudinal en el periodo t .

De esta manera, se tiene una misma variable longitudinal cualitativa medida en diferentes periodos de tiempo, ya que cada tabla \mathbf{A}_t contiene las categorías de la variable longitudinal cualitativa para cada una de los periodos de tiempo.

Partiendo de la tabla \mathbf{A} , se conforma una nueva tabla de porcentajes \mathbf{P} , que expresa la proporción del tiempo que un individuo i permanece en una categoría j en el periodo t , de término general:

$$p_{ij}^t = \frac{a_{ij}^t}{a_{i\cdot}^t} \quad (1)$$

donde $a_{i\cdot}^t = \sum_{j=1}^{m_t} a_{ij}^t$, equivale a la duración de cada periodo de tiempo. Es decir, el elemento p_{ij}^t representa la proporción del tiempo pasado por el individuo i en la categoría j de la variable en el periodo t .

La tabla \mathbf{P} posee las siguientes características: $p_{ij} = \sum_{j=1}^{m_t} p_{ij}^t = 1$, $p_{i.} = \sum_{t=1}^T p_{ij} = T$

y $p_{.j} = \sum_{i=1}^n p_{ij} = nT$

Se define

$$\bar{p}_j^t = \frac{1}{n} \sum_{i=1}^n p_{ij}^t \tag{2}$$

como el promedio de las proporciones de tiempo que permanecen los n individuos en la categoría j del periodo t .

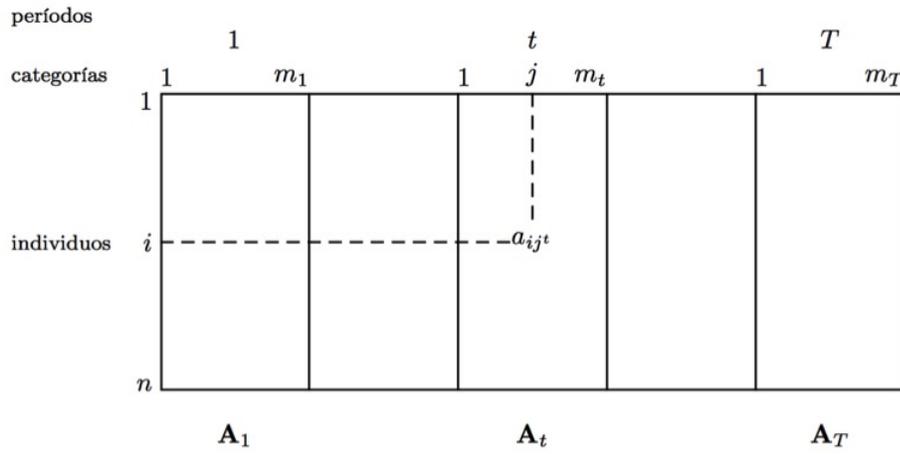


Figura 1: *Tabla de tiempos de permanencia en cada categoría. Fuente: elaboración propia.*

3. Combinación del AAC y el AFM

Se presenta inicialmente un resumen de los métodos factoriales individuales y sus características más importantes, para posteriormente realizar la incorporación del AFM sobre el AAC. El AAC se lleva a cabo, en la práctica, como un análisis de correspondencias difuso (ACD).

3.1. El análisis armónico cualitativo

El AAC es un método exploratorio de datos longitudinales cualitativos con el fin de estructurar la información de los “individuos” de una muestra, durante un periodo de tiempo determinado, los cuales se encuentran caracterizados por la sucesión de cambios de etapas o estados en el tiempo (cambios de sitios de residencia, cambios

de canales de televisión, de estado matrimonial, etc.) entre un conjunto de etapas posibles (Saporta 1996).

El análisis armónico, realizando una serie de transformaciones, se convierte en la descomposición de factores resultante del análisis de correspondencias de una tabla disyuntiva particular. Cualquier elemento de esta tabla vale 1 si el individuo está en el estado considerado durante el periodo y 0 si no lo está.

La tabla disyuntiva puede ser de un tamaño bastante grande y por esto, Deville & Saporta (1980) proponen observar a los individuos durante un periodo de tiempo notado por $[0, T]$, y dividirlo en un número razonable de intervalos, de igual o de diferente amplitud. Se construye luego una tabla que contiene la proporción del tiempo pasado por cada individuo en cada uno de los estados posibles a lo largo de cada periodo. Esta tabla se somete entonces a un análisis de correspondencias con las técnicas de interpretación habituales. También se puede tener la proporción de tiempo en cada categoría y dentro de cada intervalo pero con respecto al periodo total del estudio.

Cuando en el AAC se realiza una recodificación que indique la proporción de tiempo que un individuo i permanece en una categoría j con respecto al periodo t , se obtiene la tabla \mathbf{P} (con término general dado en la fórmula 1) con las propiedades descritas en la sección anterior.

En la práctica se debe decidir la longitud del tiempo del análisis ($[0, T]$), y el número de periodos de tiempo y sus límites (división del periodo ($[0, T]$) decisiones importantes que son tomadas en forma empírica (ver trabajos de Pinzón (1998) y Barbary & Pinzón (1998)).

3.2. El análisis armónico cualitativo como un análisis de correspondencias difuso

El análisis de correspondencias difuso (ACD), propuesto por Chevenet et al. (1994) (en español ver a Gamez (2012)) es una extensión del análisis de correspondencias múltiples, utilizado en ecología, donde es frecuente encontrar datos que son difusos por su naturaleza: especies (o individuos) que asumen varias categorías de una misma variable con diferente grado de asociación, por ejemplo los tipos de alimentos que consume una especie. De igual manera, las tablas de datos resultantes de la yuxtaposición de diferentes tablas de contingencia se pueden ver como tablas difusas.

En el AAC visto como un ACD, la tabla \mathbf{P} , se puede considerar como una tabla de codificación difusa, ya que está conformada por la yuxtaposición de las t tablas \mathbf{P}_t , en donde los individuos toman diferentes categorías de una misma variable longitudinal cualitativa, que pueden verse como diferentes grados de asociación o afinidad con estas categorías.

Debido a que el ACD es una extensión del análisis de correspondencias múltiples (ACM), y este a su vez, definiendo las métricas adecuadas en los espacios de

los individuos y de las variables, puede definirse como un análisis en componentes principales ponderado (ACP $(\mathbf{X}, \mathbf{M}, \mathbf{D})$), el ACD se puede ver como un ACP $(\mathbf{X}, \mathbf{M}, \mathbf{D})$.

La notación ACP $(\mathbf{X}, \mathbf{M}, \mathbf{D})$ indica que la tabla de datos \mathbf{X} (generalmente centrada) será analizada mediante un análisis en componentes principales utilizando como pesos y métricas las matrices \mathbf{M} y \mathbf{D} . De esta forma, \mathbf{M} es la matriz de métrica en el espacio de las filas y de pesos en el espacio de las columnas, y \mathbf{D} , la matriz de métrica en el espacio de las columnas y de pesos en el espacio de las filas (Escofier & Pagès (1984, Cap.4), Pagès (2004)).

Realizando entonces los ajustes correspondientes el ACD, es el análisis de la tabla de porcentajes \mathbf{P} , realizando sobre ella un ACP $(\mathbf{X}, \mathbf{M}, \mathbf{D})$ donde \mathbf{X} , \mathbf{M} y \mathbf{D} se definen como:

- \mathbf{X} es la matriz de término general: $x_{ij} = \frac{p_{ij}^t}{\bar{p}_j^t} - 1$ (\bar{p}_j^t está definido en (2))
- $\mathbf{M} = \text{Diag}\left(\frac{1}{T} \bar{p}_j^t\right) = \frac{1}{T} \text{diag}(\bar{p}_j^t)$
- $\mathbf{D} = \text{Diag}\left(\frac{1}{n}\right) = \frac{1}{n} \mathbf{I}_n$

3.3. El análisis factorial múltiple (AFM)

El análisis factorial múltiple (AFM), propuesto por Escofier & Pagès (1984), es un método que permite el análisis simultáneo de varios grupos de variables sobre el mismo conjunto de individuos, equilibrando la influencia de cada grupo (Escofier & Pagès 1984, Pagès 2004).

El AFM se lleva a cabo en dos etapas: en la fase inicial se realiza un ACP separado de cada una de las tablas que representan las variables de un mismo grupo (si las variables son cualitativas se realizaría un ACM). En la segunda etapa se realiza un ACP de la tabla global con ponderaciones sobre las variables. (Escofier & Pagès 1984).

El peso dado a cada una de las variables de los grupos equivale al inverso del primer valor propio de los análisis factoriales realizados por separado. Esta ponderación conduce a equilibrar el peso de los grupos en la obtención del primer factor, es decir, ningún grupo posee una mayor influencia en el análisis global.

Cuando se dispone de un conjunto de variables recogidas en diferentes momentos de tiempo, un grupo puede estar formado por la misma variable observada en diferentes periodos de tiempo o por todas las variables observadas en el mismo momento del tiempo (Abascal et al. 2001).

Cuando se tiene una sola variable medida en diferentes momentos en el tiempo, la estructura de la matriz que se analiza es similar a la de la tabla \mathbf{A} mencionada

en la sección de notación, y las subtablas \mathbf{A}_t representan los grupos, los cuales contienen las categorías de la variable longitudinal cualitativa en cada uno de los periodos de tiempo.

En la segunda etapa del AFM, se analizan todas las variables en forma simultánea, pero considerando al mismo tiempo la conformación de los grupos, o momentos en el tiempo en que fueron observadas. De esta manera, se puede obtener un marco de referencia común, que permite estudiar la evolución de los individuos.

Adicional a las ayudas a la interpretación que proporcionan los métodos factoriales, el AFM proporciona medidas globales de relación entre los grupos. Ellas están basadas en el coeficiente RV (Escoufier 1973), definido para dos grupos cualesquiera t y h :

$$RV(t,h) = \frac{\langle \mathbf{W}_t \mathbf{D}, \mathbf{W}_h \mathbf{D} \rangle}{\|\mathbf{W}_t \mathbf{D}\| \|\mathbf{W}_h \mathbf{D}\|} \quad (3)$$

y los coeficientes Lg (Escoufier & Pagès 1984):

$$Lg(t,h) = \langle \mathbf{W}_t \mathbf{D}, \mathbf{W}_h \mathbf{D} \rangle \quad (4)$$

donde $\mathbf{W}_t \mathbf{D} = \mathbf{P}_t \mathbf{M}_t \mathbf{P}'_t \mathbf{D}$ siendo \mathbf{P}_t la matriz de datos asociada al grupo t y \mathbf{D} la matriz diagonal que contiene los pesos de los individuos.

El primero se obtiene a partir de los coeficientes de correlación lineal entre dos variables cualesquiera y su valor está comprendido entre 0 (no existe relación entre las variables de los dos grupos considerados) y 1 (las nubes que representan a los grupos son homotéticas) (Abascal & Landaluce 2002). Los coeficientes Lg (4) miden, además, la dimensionalidad (número de factores de inercia considerables) de cada grupo. Estos coeficientes toman el valor 0 cuando no existe relación entre los grupos y no tienen límite superior.

3.4. AFM sobre el AAC: combinación de los dos métodos

La idea de la incorporación del AFM al AAC consiste en realizar una ponderación extra a la tabla de proporciones \mathbf{P} , obtenida de la tabla \mathbf{A} ; es decir, realizar primero un AC a cada tabla \mathbf{P}_t ; y posteriormente realizar el análisis global: un ACP reteniendo de cada análisis de las tablas \mathbf{P}_t , el primer valor propio y ponderando cada columna de la tabla por el inverso de este valor propio. También se pueden considerar varios periodos de tiempo como un grupo de variables para el AFM, en este caso, se realiza un solo ACP previo para la subtabla de todos los periodos involucrados y se ponderan, todos ellos, por el inverso del primer valor propio del análisis.

El AAC-AFM, realizado sobre la tabla \mathbf{P} , se ejecuta en dos etapas (Escoufier & Pagès 1984, Bécue & Pagès 2008):

- T análisis de correspondencias (AC) de las tablas \mathbf{P}_t . De cada uno de estos análisis se retiene el primer valor propio, denotado (λ_1^t) .

- Un ACP de tabla global \mathbf{P} en la que cada columna j se pondera por el inverso del primer valor propio λ_1^t de la tabla \mathbf{P}_t a la cual pertenece. Este es el ACP($\mathbf{X}, \mathbf{M}, \mathbf{D}$) donde \mathbf{X} y \mathbf{D} son las definidas en la subsección 3.2 y

$$\mathbf{M} = \frac{1}{T} \text{Diag} \left(\frac{p_j^t}{\lambda_1^t} \right)$$

4. Aspectos de la combinación del AAC y el AFM

Los siguientes aspectos, propios de un análisis factorial, de la combinación entre el AFM y el AAC (AAC-AFM) se deducen de las definiciones y características de cada uno de los dos métodos e incorporando la ponderación mencionada a cada variable de la tabla analizada.

4.1. Distancias entre individuos y categorías

La distancia entre entre dos filas es la suma ponderada de las distancias separadas producidas por cada conjunto de variables. La distancia al cuadrado entre las filas i y l es, entonces:

$$\begin{aligned} d^2(i, l) &= \sum_{t=1}^T \frac{1}{\lambda_1^t} \sum_{j=1}^{m_t} \frac{nT}{\sum_{i=1}^n p_{ij}^t} \left(\frac{p_{ij}^t}{T} - \frac{p_{lj}^t}{T} \right)^2 \\ &= \frac{1}{T} \sum_{t=1}^T \frac{1}{\lambda_1^t} \sum_{j=1}^{m_t} \frac{1}{p_j^t} (p_{ij}^t - p_{lj}^t)^2 \end{aligned}$$

donde se observa que dos individuos están cercanos si seleccionan aproximadamente en la misma proporción las diferentes categorías de la variable longitudinal en los diferentes periodos de tiempo. Esta distancia se ve afectada por el primer valor propio de los análisis realizados a cada grupo temporal.

La distancia entre dos categorías j y j' de los periodos de tiempo respectivos t y t' es:

$$\begin{aligned} d^2(j^t, j'^{t'}) &= \sum_{i=1}^n n \left(\frac{p_{ij}^t}{\sum_{i=1}^n p_{ij}^t} - \frac{p_{ij'}^{t'}}{\sum_{i=1}^n p_{ij'}^{t'}} \right)^2 \\ &= \sum_{i=1}^n n \left(\frac{p_{ij}^t}{np_j^t} - \frac{p_{ij'}^{t'}}{np_{j'}^{t'}} \right)^2 = \sum_{i=1}^n \frac{1}{n} \left(\frac{p_{ij}^t}{p_j^t} - \frac{p_{ij'}^{t'}}{p_{j'}^{t'}} \right)^2 \end{aligned}$$

donde dos categorías serán cercanas cuando el grupo de individuos que las han seleccionado tengan aproximadamente los mismos perfiles.

4.2. Relaciones de transición

La relación de la coordenada $F_s(i)$ de la unidad estadística i (a través del eje s) con la coordenada de las columnas $G_s(p_t)$ se obtiene aplicando la fórmula general de transición de los análisis factoriales:

$$F_s(i) = \frac{1}{\sqrt{\lambda_s}} \sum_{t=1}^T \frac{1}{\lambda_1^t} \sum_{j=1}^{m_t} \frac{p_{ij}^t}{T}; \quad G_s(p_t) = \frac{1}{\sqrt{\lambda_s}} \frac{1}{T} \sum_{t=1}^T \frac{1}{\lambda_1^t} \sum_{j=1}^{m_t} p_{ij}^t G_s(p_t)$$

De igual forma se obtiene la relación de la coordenada $G_s(p_t)$ con $F_s(i)$

$$G_s(p_t) = \frac{1}{\sqrt{\lambda_s}} \sum_{i=1}^n \frac{1}{\lambda_1^t} \frac{p_{ij}^t}{np_j^t} F_s(i)$$

Estas fórmulas de transición son similares a las de conjuntos categóricos usadas en un análisis de correspondencias múltiples. Ellas permiten la inclusión de unidades suplementarias y/o variables en un análisis factorial. Igualmente, a través de estas fórmulas de transición es posible obtener una representación superpuesta de las T nubes de los individuos, llamadas nubes parciales (Escofier & Pagès 1984), correspondientes a los T análisis separados. La representación de la t -ésima nube parcial se obtiene utilizando la fórmula de transición usual pero restringida a las categorías del periodo de tiempo t , solamente:

$$F_s^t(i) = \frac{1}{\sqrt{\lambda_s}} \frac{1}{\lambda_1^t} \sum_{j=1}^{m_t} \frac{p_{ij}^t}{T}; \quad G_s(p_t) = \frac{1}{\sqrt{\lambda_s}} \frac{1}{T} \frac{1}{\lambda_1^t} \sum_{j=1}^{m_t} p_{ij}^t G_s(p_t) \quad (5)$$

Esta representación superpuesta de los individuos se puede extender a las categorías las cuales se localizan en el centroide de los individuos que las seleccionaron (Bécue & Pagès 2008).

Es precisamente esta representación superpuesta, utilizando las fórmulas de transición, de los individuos y las categorías, la que permite observar en los planos factoriales la evolución, en el tiempo, de los individuos, de acuerdo a la posición de los puntos parciales en el plano; es decir, que por medio de los puntos medios (resumen de la información que contienen los respectivos puntos parciales) es posible obtener una síntesis de toda la información teniendo en cuenta todos los momentos del tiempo. Analizar la proximidad de los puntos parciales de un individuo, con respecto a su punto medio, permite verificar si el individuo ha mantenido la preferencia por la misma categoría a lo largo de todos los periodos de tiempo.

4.3. Inercias de la combinación

En un análisis factorial se pueden observar dos tipos de inercia (Abascal et al. 2006):

- Inercia intra: es la inercia de los puntos parciales con respecto a su punto medio en la nube global, es decir, la inercia de las nubes correspondientes a la imagen de cada individuo desde los diferentes puntos de vista estudiados, respecto a los centros de gravedad de cada individuo
- Inercia inter: es la inercia de los centros de gravedad de cada nube parcial respecto al centro de gravedad global.
- Inercia total: es la variabilidad de la nube global respecto a su centro de gravedad. Se descompone de forma aditiva en inercia intra e inercia inter.

El objetivo del análisis factorial es examinar la existencia de estructuras comunes a todas, o a parte, de las nubes parciales, lo que se reflejaría en una inercia intra mínima o pequeña, o, lo que es lo mismo, en una inercia inter máxima o elevada. Aquellos individuos cuyos puntos parciales se sitúen próximos reflejan una inercia intra débil e ilustran la estructura común que poseen los datos analizados. Por el contrario, aquellos individuos con puntos parciales alejados reflejan una inercia intra alta y constituyen las excepciones a tal estructura común.

La inercia total obtenida del AAC-AFM es:

$$I = \frac{1}{nT} \sum_{t=1}^T \frac{1}{\lambda_1^t} \sum_{j=1}^{m_t} \overline{p_j^t} \sum_{i=1}^n \left(\frac{p_{ij}^t}{\overline{p_j^t}} - 1 \right)^2 = \frac{1}{nT} \sum_{t=1}^T \sum_{j=1}^{m_t} \sum_{i=1}^n \frac{1}{\lambda_1^t} \frac{\left(p_{ij}^t - \overline{p_j^t} \right)^2}{\overline{p_j^t}} \quad (6)$$

En (6) se observa que la inercia no depende del número de variables ni del número de categorías sino de las diferencias entre las proporciones observadas y las teóricas, así como del primer valor propio de los análisis separados.

4.4. Relación de los grupos de las variables

Uno de los objetivos importantes del AAC-AFM es poder realizar comparaciones entre los grupos. El operador $\mathbf{W}_t \mathbf{D}$, de dimensión $n \times n$ es adecuado para representar las variables del periodo t . Cada una de estas matrices $\mathbf{W}_t \mathbf{D}$ puede representarse por un punto en R^{m^2} ; es decir, en este espacio se tiene la nube de los T periodos totales de tiempo, denominada \mathbf{N}_T .

Sobre este espacio se utiliza el producto escalar de Hilbert-Schmidt entre dos grupos, a partir del cual se definen los coeficientes L_g , RV y en la proyección de los grupos sobre los ejes.

La coordenada del grupo t a lo largo de z_s , es igual a la medida de relación:

$$L(z, P_t) = \frac{1}{T} \sum_t \left(\frac{\overline{p_j^t}}{\lambda_1^t} \right) r^2 \left(z, \left(\frac{p_{ij}^t}{\overline{p_j^t}} - 1 \right) \right)$$

que es también la contribución del grupo t a la inercia del eje s , donde r representa el coeficiente de correlación lineal (Pagès 2002).

El producto escalar entre dos grupos h y t , también conocido como coeficiente L_g es:

$$\langle \mathbf{W}_t \mathbf{D}, \mathbf{W}_h \mathbf{D} \rangle = \sum_t \left(\frac{\overline{p_j^t}}{T \lambda_1^t} \right) \sum_h \left(\frac{\overline{p_j^h}}{T \lambda_1^h} \right) \langle \mathbf{P}_t \mathbf{P}_t' \mathbf{D}, \mathbf{P}_h \mathbf{P}_h' \mathbf{D} \rangle \quad (7)$$

L_g será mayor mientras más relacionadas estén las variables de un grupo con las variables del otro grupo y se convierte de esta manera en una buena medida de relación entre los grupos de las variables. La fórmula del coeficiente RV se encuentra en (3).

5. Aplicación: tipología de los televidentes en un día en Colombia

Para ilustrar el AAC-AFM, se analiza un conjunto de televidentes colombianos que observan los canales nacionales e internacionales de televisión durante un día específico de la semana. El objetivo del análisis de este grupo de televidentes es buscar una tipología de ellos en lo que tiene que ver con la preferencia de los canales de televisión vistos, con respecto a distintas franjas de audiencia. Para ver el resultado completo de la aplicación, ver Corrales (2010).

Los datos de teleaudiencia fueron suministrados por una empresa de medición de *rating*, corresponden a un día hábil de la semana. La medición de las audiencias de televisión de canales nacionales e internacionales se realiza utilizando la tecnología electrónica *People Meter*, para conocer el comportamiento de los televidentes. Este *rating* de televisión se estima gracias a un panel de hogares en todo el país, en los que se registraron minuto a minuto la sintonía de 584 hogares, conformados en total por 2334 televidentes.

Para efectos de esta medición el país se encuentra dividido en 6 zonas: Zona 1: Centro del país (Bogotá, D.C. y Soacha). Zona 2: Región Antioqueña (Medellín, Bello, Envigado e Itagüí). Zona 3: Región Pacífica (Cali, Palmira y Yumbo). Zona 4: Región Caribe (Barranquilla, Soledad, Cartagena y Santa Marta). Zona 5: Región Cafetera (Armenia, Pereira, Manizales y Dosquebradas). Zona 6: Región Oriental (Bucaramanga, Floridablanca, Girón, Cúcuta y Los Patios).

La información proporcionada originalmente muestra el tiempo de permanencia de cada individuo en los diferentes canales de televisión. El tiempo del día considerado inicia a las seis de la mañana y finaliza a las dos de la mañana del día siguiente. Para registrar los tiempos en un televisor que se encuentra apagado se introduce una canal ficticio (canal 15) (ver Tabla 1).

Los cambios de canal que realiza un individuo se detectan minuto a minuto obteniéndose, para la información proporcionada durante el día del estudio, un total de 19229 cambios de canal (si en un minuto la persona continúa en el mismo canal, no

Tabla 1: *Cambios de canal de los individuos completos (dos primeros individuos).*
 Fuente: elaboración propia.

Zona	Individuo	Canal	Hora de inicio	Hora finalización
1	110000301	15	06:00:00 a.m.	11:24:59 a.m.
1	110000301	05	11:25:00 a.m.	11:25:59 a.m.
1	110000301	04	11:26:00 a.m.	12:29:59 p.m.
1	110000301	15	12:30:00 p.m.	06:58:59 p.m.
1	110000301	05	06:59:00 p.m.	08:11:59 p.m.
1	110000301	09	08:12:00 p.m.	08:12:59 p.m.
1	110000301	15	08:13:00 p.m.	02:00:00 a.m.
1	110000701	15	06:00:00 a.m.	02:49:59 p.m.
1	110000701	10	02:50:00 p.m.	03:00:59 p.m.
1	110000701	07	03:01:00 p.m.	03:08:59 p.m.
1	110000701	02	03:09:00 p.m.	03:12:59 p.m.
1	110000701	07	03:13:00 p.m.	03:21:59 p.m.
1	110000701	02	03:22:00 p.m.	03:25:59 p.m.
1	110000701	07	03:26:00 p.m.	03:41:59 p.m.
1	110000701	02	03:42:00 p.m.	03:46:59 p.m.
1	110000701	07	03:47:00 p.m.	03:50:59 p.m.
1	110000701	02	03:51:00 p.m.	03:56:59 p.m.
1	110000701	07	03:57:00 p.m.	04:12:59 p.m.
1	110000701	01	04:13:00 p.m.	04:21:59 p.m.
1	110000701	07	04:22:00 p.m.	04:23:59 p.m.
1	110000701	01	04:24:00 p.m.	04:28:59 p.m.
1	110000701	07	04:29:00 p.m.	04:41:59 p.m.
1	110000701	15	04:42:00 p.m.	02:00:00 a.m.

se reporta ningún cambio). La variable a analizar es cualitativa, ya que representa el canal en el cual se encuentra la persona en un tiempo determinado durante ese día.

Los periodos de tiempo considerados son las diferentes franjas de televisión reguladas por el Consejo Nacional de Televisión (Artículo 5 de la Ley 335 de 1996). Sin embargo, debido a que el objetivo de interés consiste en caracterizar la franja de mayor audiencia, *prime time*, y realizar una comparación entre esta y las demás, las franjas de audiencia se ajustaron como se muestra en la Tabla 2.

La cantidad de canales de televisión existentes también se reagruparon. El número de posibles canales elegibles entre los televidentes asciende a un total de 469. Sin embargo, solo 246 de ellos se seleccionaron el día de la medición. Dentro de estos, muchos presentan muy poca frecuencia de tiempo de permanencia en ellos. Se conservaron, por lo tanto, aquellos con alto porcentajes de cambios y se realizaron las agrupaciones de aquellos que no la tuvieron. Estas agrupaciones fueron realizadas y renombradas teniendo en cuenta la estructura que presentan las diferentes

Tabla 2: *Franjas de audiencia ajustadas. Fuente: elaboración propia.*

Franja	Horario	Tipo de Audiencia
F1:Fami1	06:00 -12:00	Familiar
F2:Fami2	12:00 -16:00	Familiar
F3:Infan	16:00 -17:00	Infantil
F4:Prime	17:00 -19:30	<i>Prime Time</i>
F5:Fami3	19:30 -22:10	Familiar
F6:Adult	22:10 -02:00	Adultos

Tabla 3: *Canales de televisión estructurados. Fuente: elaboración propia.*

Canal	Nombre del canal
1	Telecaribe
2	City TV
3	Culturales internacionales
4	Canal RCN
5	Canal Caracol
6	Otros canales locales y regionales colombianos
7	Canal uno
8	Series internacionales
9	Canal Institucional
10	Teleantioquia
11	Señal Colombia
12	Infantiles internacionales
13	Fox Sport
14	Otros canales internacionales
15	Televisión apagada

empresas prestadoras de servicios de televisión y a las características comunes de ellos (Tabla 3).

Adicionalmente, se cuenta con aspectos adicionales de los individuos como la zona a la cual pertenece, el peso (factor de expansión) de la muestra dado para cada individuo, el nivel socioeconómico, la edad y el género (Tabla 4). Las tres últimas variables, para efectos del estudio, se consideran ilustrativas, y su labor será la de complementar la caracterización de las clases de televidentes obtenidas.

Del total de los individuos de la muestra (2334), solo 1320 tuvieron encendido el televisor en el día estudiado y por lo tanto estos fueron los individuos analizados. Los 1014 individuos restantes conforman un grupo aparte con características propias. De esta manera, la tabla analizada, tiene dimensión 1320 x 90, y contiene la información del número de minutos en que permanece cada uno de los 1320 individuos, en los 15 canales de televisión considerados, para cada una de las franjas de audiencia (Tabla 5: tabla A mencionada en la sección de estructura de los datos).

Tabla 4: *Parte del archivo con información adicional de los individuos. Fuente: elaboración propia.*

Individuo	Género	Rango edad	Nivel socio-económico	Zona	Peso
110000701	Mujer	40 a 64	Medio	1	0.19750
110001901	Hombre	40 a 64	Medio	1	0.04248
110004301	Hombre	25 a 39	Medio	1	0.04376
110023101	Hombre	40 a 64	Alto	1	0.09113
110023103	Mujer	25 a 39	Alto	1	0.09113
110027601	Hombre	40 a 64	Alto	1	0.04537

Tabla 5: *Fracción del archivo con el total de minutos vistos por canal y franja. Fuente: elaboración propia.*

Individuo	F1C1	F1C2	F1C3	F1C4	F1C5	F1C6	F1C7	F1C8	F1C9
110000301	0	0	0	34	1	0	0	0	0
110000701	0	0	0	0	0	0	0	0	0
110001001	0	0	0	0	0	0	0	0	0
110001002	0	0	0	0	0	0	0	0	0
110001003	0	0	0	0	0	0	0	0	0
110001901	0	0	0	15	89	0	0	3	0
110001902	0	0	0	84	3	1	0	0	0
110001904	0	0	0	0	0	0	0	0	0
110001905	0	0	0	84	3	1	0	0	0
110002101	0	0	0	0	44	0	0	0	0
110002201	0	0	0	8	30	0	0	0	0
110002202	0	0	0	0	0	0	0	0	0
110002203	0	0	0	0	0	0	0	0	0
110002204	0	0	0	0	0	0	0	0	0
110002205	0	0	0	0	0	0	0	0	0

Inicialmente se realiza un análisis armónico cualitativo (AAC) a la tabla de datos, y el resultado de este análisis es tomado como insumo para el análisis factorial múltiple (AFM), efectuándose así la combinación: AAC-AFM. En la primera fase se tuvo en cuenta el peso o factor de expansión de los individuos de la muestra. Se realizaron dos AAC (vistos como ACD), con y sin el peso de los individuos dando similares resultados. Por tal motivo, los análisis posteriores se realizaron con pesos iguales para los individuos de la muestra. Después de realizar estos dos análisis, se eliminaron catorce individuos, ya que conformaban ejes factoriales. Al final la tabla de trabajo quedó conformada por 1306 individuos (filas) y 90 columnas.

El resultado obtenido del AAC, es incorporado, con las reestructuraciones mencionadas en la sección (3), al AFM. Este AAC-AFM ejecuta seis análisis de correspondencias a cada franja de audiencia, los cuales dan como resultado diferencias en el primer valor propio (Tabla 6), lo que justifica la ponderación del AFM, que estandariza esos valores propios en 1 para el análisis global.

Tabla 6: Los primeros cinco valores propios de los análisis por separado. Fuente: elaboración propia.

Grupo	Valor1	Valor2	Valor3	Valor4	Valor5
F1:Fami1	0.1030	0.0850	0.0780	0.0710	0.0660
F2:Fami2	0.1000	0.0870	0.0755	0.0750	0.0690
F3:Infan	0.1410	0.1350	0.1300	0.1170	0.1050
F4:Prime	0.1130	0.0930	0.0920	0.0870	0.0830
F5:Fami3	0.1180	0.0920	0.0810	0.0760	0.0660
F6:Adult	0.0730	0.0670	0.0600	0.0560	0.0550

La diagonal de la matriz de coeficientes Lg obtenidos del AFM estructurado (Tabla 7) muestra que las franjas tienen dimensionalidades de 4 y 5, las que se recogen en el análisis global pues en el histograma de valores propios se destacan cinco ejes. La matriz de coeficientes RV (Tabla 8) indica que existen grandes diferencias entre las franjas, ya que las relaciones entre ellas son muy bajas. Sin embargo, se observa que las correlaciones más altas se dan entre las franjas vecinas.

Tabla 7: Coeficientes Lg . Fuente: elaboración propia.

	Fami1	Fami2	Infan	Prime	Fami3	Adult	ilust	MFA
F1:Fami1	4.773	0.602	0.250	0.297	0.182	0.135	0.076	2.160
F2:Fami2	0.602	4.248	1.233	0.689	0.344	0.337	0.063	2.580
F3:Infan	0.250	1.233	5.386	1.446	0.414	0.283	0.058	3.120
F4:Prime	0.297	0.689	1.446	4.866	1.016	0.222	0.072	2.955
F5:Fami3	0.182	0.344	0.414	1.016	4.050	0.559	0.102	2.272
F6:Adult	0.135	0.337	0.283	0.222	0.559	5.354	0.063	2.385
ilust	0.076	0.063	0.058	0.072	0.102	0.063	6.391	0.150
AFM	2.160	2.580	3.120	2.955	2.272	2.385	0.150	5.356

Para obtener las características especiales de las clases de televidentes, son insuficientes los planos factoriales, ya que algunas no son fáciles de identificar a través de la interpretación de los factores.

Para conseguir la tipología completa, es necesario, por lo tanto, proceder a la clasificación de los individuos en el espacio vectorial de los primeros factores del análisis. Para este fin se usó el criterio propuesto por Barbary (1996) para decidir el número de ejes (cambiar el número de ejes, tomando aquel que genere menor cantidad de individuos en la clase que contiene mayor cantidad de individuos). Se hicieron varios ensayos para el número de factores seleccionando seis ejes de acuerdo al criterio mencionado.

Se realiza inicialmente un análisis de clasificación jerárquica a través del método de *Ward*. Esta clasificación inicial se optimiza por medio del método *k-means*, una

Tabla 8: *Coficientes Rv. Fuente: elaboración propia.*

	Fami1	Fami2	Infan	Prime	Fami3	Adult	ilust	MFA
F1:Fami1	1.000	0.134	0.049	0.062	0.041	0.027	0.014	0.427
F2:Fami2	0.134	1.000	0.258	0.152	0.083	0.071	0.012	0.541
F3:Infan	0.049	0.258	1.000	0.283	0.089	0.053	0.010	0.581
F4:Prime	0.062	0.152	0.283	1.000	0.229	0.043	0.013	0.579
F5:Fami3	0.041	0.083	0.089	0.229	1.000	0.120	0.020	0.488
F6:Adult	0.027	0.071	0.053	0.043	0.120	1.000	0.011	0.445
ilust	0.014	0.012	0.010	0.013	0.020	0.011	1.000	0.026
AFM	0.427	0.541	0.581	0.579	0.488	0.445	0.026	1.000

vez seleccionado el número de clases (ver por ejemplo Lebart et al. (2006) o Pardo & Del-Campo (2007)).

El número de clases está guiado por los saltos del histograma de índices de nivel (Figura 3): se tomaron cuatro particiones en 5, 7, 8 y 11 clases y se decidió describir la partición en ocho clases (Figura 2), ya que es la que mejor se adapta a los objetivos del estudio. En la figura se observa que la clase (1), con proporción más alta de televidentes, quienes pasan mucho tiempo del día con el televisor apagado y que poseen gustos muy variados de audiencia, se encuentran en el centro del primer plano. Mientras que, en general, las clases más alejadas del centro, poseen menor proporción de individuos y preferencia más marcada de canales de televisión.

Los televidentes de la muestra, se repartieron entonces en ocho clases con los porcentajes de individuos en cada una de ellas: 65.2, 0.5, 4.1, 4.0, 3.7, 16.8, 4.8 y 0.9 respectivamente.

Las clases se construyen con los cambios y permanencias de los televidentes en los diferentes canales, pero se caracterizan también con las variables ilustrativas consideradas: género, estratos socioeconómicos, rangos de edad y zonas del país, teniendo en cuenta que la clase 9 es la conformada por los individuos que no vieron televisión el día del estudio.

Clase 1 (65.2 %). Los televidentes de este grupo permanecen con el televisor apagado durante gran parte del día. Es un grupo con preferencias de audiencia heterogéneas, ya que no se inclinan por ningún canal en especial. Casi todos tienen altas audiencias, en el tiempo que permanece el televisor encendido. Los canales seleccionados por estos individuos son principalmente el canal 4 (RCN), 5 (Caracol), 12 (infantiles internacionales), 6 (otros canales locales y regionales colombianos) y 14 (cinemas internacionales y otros canales de corte internacional de menor audiencia). Son televidentes con edades entre 40 y 64 años, de niveles socioeconómicos medio y bajo de diferentes zonas.

Clase 2 (0.5 %). Conformada por una serie de televidentes con gustos diferentes en cada una de las franjas, pero inclinados en todas las franjas, excepto en la

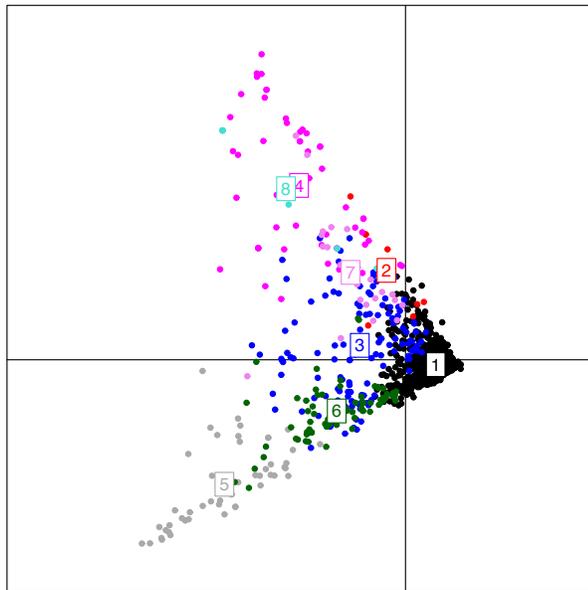


Figura 2: Nube de los individuos sobre el primer plano factorial del AAC-AFM mostrando las clases obtenidas. Se nota una evolución de la tendencia desde tener encendido el TV todo en día, sobre todo en RCN (clase 5), los que alternan sobre todo RCN y Caracol (clase 6), se aumentan las preferencias de canales infantiles (clase 3), luego canales regionales (clase 2), Caracol (clase 7) y canales internacionales (clases 4 y 8). La clase 1 es la más numerosa y heterogénea. Fuente: elaboración propia.

primera, por el canal 10 (Teleantioquia). En la franja 1 (durante el poco tiempo que permaneció encendido el televisor), prefieren los canales 12 (infantiles internacionales) y 14 (cinemas y otros canales internacionales). En la franjas 2, 3 y 4, se inclinan también por los canales 6 (otros canales locales y regionales colombianos) y 14 (otros canales internacionales). En la franja 4 seleccionan además del canal 10 (Teleantioquia), un amplio conjunto de canales. En la franja 6 también se seleccionan el canal 2 (City TV) y el canal 14 (cinemas y otros canales internacionales). Los canales 1 (Telecaribe), 3 (culturales internacionales), 11 (Señal Colombia) y 13 (*Fox Sport*) no se escogieron durante todo el día. Se encuentra caracterizada por televidentes con edades entre 25 y 64 años, principalmente de clase media de las zonas centro, caribe y cafetera.

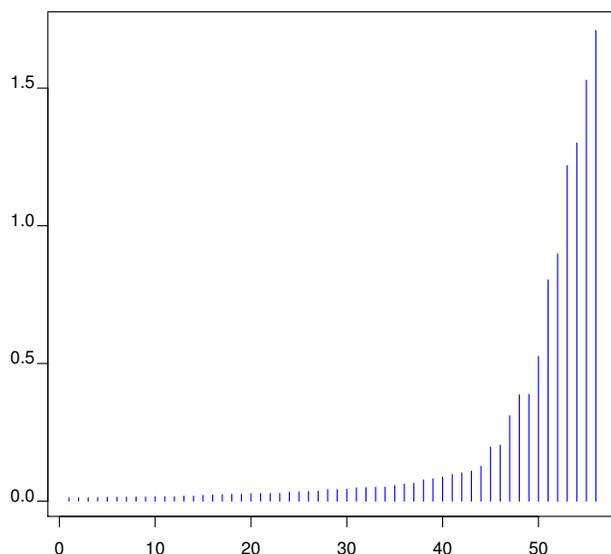


Figura 3: *Histograma de índices de nivel. Un corte a la altura de 0.5 origina ocho clases. Fuente: elaboración propia.*

Clase 3 (4.1 %). Formada por televidentes que prefieren principalmente el canal 12 (infantiles internacionales tales como *Cartoon Network*, *Discovery Kids*, *Disney Chanel*, etc). En menor proporción seleccionan los canales 4 (RCN), 5 (Caracol) y 14 (cinemas (*HBO*, *Cinemax*, etc.) y otros canales internacionales). Se encuentra caracterizada también por individuos de nivel socioeconómico bajo, del centro y oriente del país con edades entre los 4 y 11 años, y 25 a 64 años.

Clase 4 (4.0 %). Este grupo caracteriza a los televidentes que escogen principalmente el canal 14 (otros canales internacionales como *Cinecanal Este y Oeste*, *HBO*, *Cinemax*, *Movie City*, entre otros. En menor proporción seleccionan los canales 4 (RCN), 5 (Caracol), 8 (series internacionales como *Sony*, *Warner Chanel*, *Universal*, entre otros) y otra amplia gama de canales. En esta clase predominan los individuos con edades entre 25 y 40 años, de nivel socioeconómico bajo y medio y de las zonas cafetera y del centro del País.

Clase 5 (3.7 %). Esta clase de televidentes permanece la mayoría de su tiempo en el canal 4 (RCN), con la televisión encendida casi todo el día. En forma poco frecuente seleccionan los canales 5 (Caracol) y 14 (otros canales internacionales). En esta clase se encuentran mujeres entre 40 a 64 años de niveles bajo y medio pertenecientes en gran proporción a la zona central de Colombia.

Clase 6 (16.8 %). Esta clase se encuentra formada por individuos que permanecen casi todo el tiempo de la audiencia alternando entre los canales 4 (RCN) y 5 (Caracol) sin que exista una preferencia notable por ninguno de los dos. En menor

grado también seleccionan el canal 14 (otros canales internacionales). Son personas de clase media, con edades entre los 25 y 64 años de las zonas centro, pacífica y oriental principalmente.

Clase 7 (4.8 %). Este grupo se caracteriza por preferir el canal 5 (Caracol) sobre el resto de los canales. También se selecciona con alguna frecuencia el canal 4 (RCN). No se selecciona el canal 2 (City TV) en ninguna de las franjas. Formado principalmente por individuos de las zonas 1 (centro) y 6 (caribe), de niveles socioeconómicos bajo y medio.

Clase 8 (0.9 %). En esta clase de televidentes, con gustos heterogéneos en las diferentes clases, sobresale la escogencia del canal 3 (culturales internacionales como *National Geographic*, *Discovery Chanel*, etc). Son hombres, de clase media de las zonas pacífica y cafetera colombianas.

Clase 9. (1014 individuos) Esta clase se encuentra formada por los individuos que no vieron televisión. La mayoría son mujeres, de edades entre 18 y 64 años, de clase media y de diferentes zonas del país.

6. Conclusiones

La combinación del análisis armónico cualitativo con el análisis factorial múltiple, dota al AAC de características propias del AFM, como son:

- Síntesis general de toda la información considerando todos los periodos de tiempo.
- Medidas de asociación que permiten cuantificar las semejanzas entre los diferentes periodos de tiempo.
- Inercias intra, que dejan ver los individuos con mayores y menores cambios en el tiempo.

La ponderación utilizada en el AFM equilibra los pesos de los grupos. Esto permite analizar grupos con diferente número de categorías de la variable longitudinal cualitativa. También evita que algún grupo predomine en la formación del primer eje factorial por tener una estructura más fuerte que los demás grupos.

De esta manera, la incorporación del AFM sobre el AAC se presenta como una metodología potente para la descripción de datos cualitativos de tipo longitudinal. Ella recoge las ventajas de ambos métodos lo que permite obtener una mejor herramienta para el tratamiento de este tipo de datos, que el uso de cada metodología en forma separada.

En el análisis del *rating* de televisión en un día entre semana en Colombia, se logra agrupar a los televidentes colombianos, de una muestra, en nueve clases y relacionarlas con algunas de sus características sociodemográficas. Estos resultados ya

no tienen vigencia, pero la metodología y el software asociado, quedan disponibles para realizar análisis con información actual. Esta metodología es aplicable al análisis de variables categóricas a través del tiempo o del espacio en cualquier área.

Agradecimientos

Este artículo es derivado de la tesis de maestría de la primera autora (Corrales 2010), bajo la dirección del segundo autor. El artículo es también resultado del proyecto de investigación “Comparación de métodos multivariados para el análisis de tablas estructuradas”, código Hermes 11319, de la Universidad Nacional de Colombia Sede Bogotá. Agradecemos al los árbitros y la Editora de la Revista, por sus comentarios, que ayudaron a mejorar el documento. Para el ejemplo de aplicación se utilizó el paquete `hqa`, programado en el lenguaje estadístico R (R Development Core Team 2014), desarrollado como parte de la tesis y dentro del proyecto de investigación. El paquete `hqa` utiliza algunas de las funciones de los paquetes: `ade4` (Chessel et al. 2004), `FactoMineR` (Husson et al. 2014) y `FactoClass` (Pardo & Del-Campo 2007).

Recibido: 11 de septiembre de 2014

Aceptado: 27 de junio de 2015

Referencias

- Abascal, E., Aguirre, K. & Landaluce, M. (2001), ‘Técnicas factoriales de análisis de tablas múltiples. Nuevos desarrollos empíricos’, *Universidad del País Vasco*.
- Abascal, E., García, I. & Landaluce, M. (2006), ‘Trayectorias de las comunidades autónomas según sus tasas de paro por categorías de edad. Un análisis de tablas tridimensionales’, *Estadística Española* **48**(163), 225–250.
- Abascal, E. & Landaluce, M. (2002), ‘Análisis factorial múltiple como técnica de estudio de la estabilidad de los resultados de un análisis en componentes principales’, *Questiio* **26**, 109–122.
- Barbary, O. (1996), Una aplicación del análisis cualitativo. La tipología de trayectorias individuales, in ‘Seminario de capacitación e investigación. Recolección y análisis de datos longitudinales’, *PRESTA-ORSTOM-Universidad Nacional de Colombia*, pp. 111–120.
- Barbary, O. & Pinzón, L. (1998), ‘El análisis armónico cualitativo: Teoría y aplicación a la tipología de trayectorias individuales’, *Revista Colombiana de Estadística* **22**(1), 27–51.

- Bécue, M. & Pagès, J. (2008), 'Multiple Factor Analysis and Clustering of a Mixture of Quantitative, Categorical and Frequency Data', *Computational Statistics and Data Analysis* **10**(1016), 72–79.
- Chessel, D., Dufour, A.-B. & Thioulouse, J. (2004), 'The ade4 package-I- One-table methods', *R News* **4**(1), 5–10.
- Chevenet, F., Dolédec, S. & Chessel, D. (1994), 'A Fuzzy Coding Approach For The Analysis of Long-Term Ecological Data', *Freshwater Biology* **31**, 295–309.
- Corrales, M. (2010), Combinación del análisis factorial múltiple y el análisis armónico cualitativo en la descripción de datos longitudinales cualitativos, Tesis de grado para optar el título de Magíster Ciencias-Estadística. Facultad de Ciencias. Departamento de Estadística, Universidad Nacional de Colombia, Bogotá, D.C.
- Corrales, M. L. (2000), Movilidad espacial en ciudades en expansión: Yopal, Aguazul y Tauramena, Trabajo de grado para optar al título de Especialista en Ciencias-Estadística, Universidad Nacional de Colombia. Facultad de Ciencias. Departamento de Estadística, Bogotá, D.C.
- Deville, J.-C. (1982), 'Analyse des données chronologiques qualitatives: comment analyser les calendriers?', *Annales de l'Institut national de la statistique et des études économiques(INSEE)* **45**, 45–104.
- Deville, J. & Saporta, G. (1980), 'Analyse harmonic qualitative', *Data Analysis and Informatics* pp. 375–389.
- Dureau, F. & Flórez, C. (2000), *Aguaitacamino: las transformaciones de las ciudades de Yopal, Aguazul y Tauramena durante la explotación petrolera de Cusiana-Cupiagua*, Ciudades y Ciudadanía, Ed. Bogotá: Uniandes; Tercer Mundo.
*<http://www.documentation.ird.fr/hor/fdi:010022917>
- Escofier, B. & Pagès, J. (1984), *L'analyse factorielle multiple: une méthode de comparaison de groupes de variables*, Data analysis and informatics, III, Diday E. edn, Elsevier Science, Amsterdam, pp. 41–56.
- Escoufier, Y. (1973), 'Le traitement des variables vectorielles', *Biometrics* **29**, 751–760.
- Gamez, N. (2012), 'Fundamentos y aplicaciones del análisis de correspondencias difuso', *Comunicaciones en Estadística* **5**(1), 7–32.
- Husson, F., Josse, J., Le, S. & Mazet, J. (2014), *FactoMineR: Multivariate Exploratory Data Analysis and Data Mining with R*. R package version 1.27.
*<http://CRAN.R-project.org/package=FactoMineR>
- Lebart, L., Piron, M. & Morineau, A. (2006), *Statistique exploratoire multidimensionnelle*, 3 edn, Dunod, Paris.

- Pagès, J. (2002), 'Analyse factorielle multiple appliquee aux variables qualitatives et aux données mixtes', *Revue de Estatistique Appliquée* **50**(4), 5–37.
- Pagès, J. (2004), 'Multiple factor analysis: Main features and application to sensory data', *Revista Colombiana de Estadística* **27**(1), 1–26.
*<http://www.estadistica.unal.edu.co/revista/>
- Pardo, C. & Del-Campo, P. (2007), 'Combinación de métodos factoriales y de análisis de conglomerados en R: el paquete FactoClass', *Revista Colombiana de Estadística* **30**(2), 231–245.
- Pinzón, L. M. (1998), Manejo del tiempo en el análisis armónico cualitativo. Movilidad residencial en la áreas urbanas de Bogotá, Tesis de grado para optar al título de Magíster Ciencias-Estadística. , Universidad Nacional de Colombia. Facultad de Ciencias. Departamento de Estadística, Bogotá, D.C.
- R Development Core Team (2014), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
*<http://www.R-project.org/>
- Saporta, G. (1996), L'analyse harmonique qualitative, une sysnthese de la théorie, in 'Seminario de capacitación e investigación. Recolección y análisis de datos longitudinales', *PRESTA-ORSTOM*-Universidad Nacional de Colombia, Bogotá, D.C., pp. 111–120.