

# Combinación del análisis factorial múltiple y del análisis armónico cualitativo en el tratamiento de datos longitudinales categóricos

## Multiple Factor and Qualitative Harmonic Analysis for Longitudinal Categorical Data

MARTHA LUCÍA CORRALES<sup>a</sup>, CAMPO ELÍAS PARDO<sup>b</sup>, JUAN RAMOS<sup>c</sup>

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

### Resumen

Se propone una nueva metodología estadística para el tratamiento de datos longitudinales categóricos, a través 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.

**Palabras clave:** análisis multivariado de datos, análisis de correspondencias difuso, análisis de datos longitudinales, *rating* de televisión.

### Abstract

A new statistical methodology is proposed in order to analyze 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.

**Key words:** Multivariate data analysis, Fuzzy correspondence analysis, Longitudinal data analysis, TV rating.

## 1. Introducción

Cuando se obtienen datos longitudinales cuantitativos y/o categóricos de naturaleza multivariada para analizar fenómenos que evolucionan en períodos cortos de tiempo o cuando se tienen series muy heterogéneas no es posible utilizar los métodos clásicos de series cronológicas. Cuando se observa un fenómeno en el tiempo en el que se consideran múltiples variables, los análisis se complican, siendo recomendable la utilización de técnicas exploratorias como los llamados métodos factoriales, los cuales permiten extraer las relaciones que subyacen entre los elementos de una misma dimensión (datos y variables), así como las relaciones entre las diferentes dimensiones consideradas.

Existen diferentes métodos estadísticos para el estudio simultáneo de varias matrices de datos, la elección de un método u otro depende, por supuesto, de los objetivos que se plantee, del tipo de variable trabajada y del tipo de matriz de datos que se tenga.

El presente trabajo propone la combinación de los métodos AAC (Deville & Saporta 1980) y AFM (Escofier & Pagès 1992) en el manejo de una variable longitudinal categórica, y la aplicación de dicha

<sup>a</sup>Estudiante maestría de estadística. E-mail: mlcorralesb@unal.edu.co

<sup>b</sup>Profesor asociado. E-mail: cepardot@unal.edu.co

<sup>c</sup>Profesor asociado (pensionado). E-mail: juan.ramos@mymcol.com

combinación en el análisis de los datos de *rating* de televisión de un día en Colombia. Para los televidentes de la muestra se observan los cambios de canal, minuto a minuto, en un día específico de una semana.

La incorporación de la ponderación del AFM en el AAC enriquece a este último en varios aspectos: representación de los individuos durante todo el período de tiempo, comparaciones de los diferentes períodos de tiempo y de las trayectorias de los individuos en el tiempo.

## 2. Notación

Sean  $t = 1, 2, \dots, T$  períodos de tiempo,  $j = 1, \dots, m_t$  las modalidades de una variable longitudinal cualitativa, medidas en el período  $t$  para  $i = 1, \dots, n$  individuos que las seleccionaron. Se denotan:

- $\mathbf{A}_t$  a la tabla de datos formada por  $n$  filas y  $m_j$  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 modalidad  $j$  de la variable longitudinal en el período  $t$ .

De esta manera, se tiene una misma variable longitudinal cualitativa medida en diferentes períodos de tiempo. Es decir, cada tabla  $\mathbf{A}_t$  contiene las modalidades 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:

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

donde  $a_{i.}^t = \sum_{j=1}^{m_t} a_{ij}^t$ , equivale a la duración de cada período de tiempo.

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

Se define:

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

como el promedio de las proporciones de tiempo que permanecen los  $n$  individuos en la modalidad  $j$  del período  $t$ .

TABLA 1: Tiempo de permanencia de los individuos en cada categoría durante cada periodo (tabla  $\mathbf{A}$ )

1	1	1	1
	$m_1$	$m_t$	$m_T$
$\mathbf{A}_1$		$\mathbf{A}_t$	$\mathbf{A}_T$

### 3. El análisis armónico cualitativo como un análisis de correspondencias difuso

El análisis armónico cualitativo (AAC), propuesto por Deville & Saporta (1980), es un método exploratorio para el análisis de una variable cualitativa que representa el paso de un conjunto de individuos por diferentes estados (cambios de residencia, estado civil, etc). Su objetivo es estructurar la información de los individuos descritos por la cronología de la sucesión de cambios de estas etapas, identificando aquellos con trayectorias de cambios similares (Saporta 1996).

En la práctica, el período de observación se divide en un número de intervalos razonable  $T$ , de igual o de diferente duración, para luego calcular la proporción de tiempo pasada por cada individuo en cada uno de los estados de la variable longitudinal en el curso de cada intervalo. De esta manera se obtiene una tabla de frecuencias caracterizando las trayectorias de los individuos, la cual se puede someter a un análisis de correspondencias.

Cuando en el AAC se opta por una recodificación que indique la proporción de tiempo que un individuo permanece en una categoría  $j$  con respecto al periodo  $t$ , se obtiene una tabla  $\mathbf{P}$  (término general dado en la fórmula (1)), con las propiedades descritas en la sección de notación. La tabla  $\mathbf{P}$  se puede ver como una tabla de codificación difusa, es decir, los individuos toman diferentes valores de una misma variable con diferentes grados de asociación; y se puede analizar a través de una extensión del análisis de correspondencias múltiples, el análisis de correspondencias difuso (ACD), propuesto por Chevenet et al. (1994) (en español ver Gámez (2007)).

El análisis de correspondencias difuso de la tabla  $\mathbf{P}$  se puede ver como un análisis en componentes principales ponderado,  $ACP(\mathbf{X}, \mathbf{M}, \mathbf{D})$  donde (ver por ejemplo Escofier & Pagès (1992, cap. 4)):

- $\mathbf{X}$  es la matriz de a analizar, con término general:  $x_{ij} = \frac{p_{ij}^t}{p_j^t} - 1$
- $\mathbf{M} = \frac{1}{T} \text{diag}(\overline{p_j^t})$ , matriz de métrica en el espacio de las filas y de pesos en el espacio de las columnas.  $\overline{p_j^t}$  esta dado en la fórmula (2).
- $\mathbf{D} = \text{diag}\left(\frac{1}{n}\right)$ , matriz de pesos en el espacio de las filas y de métrica el espacio de las columnas.

### 4. AFM sobre el AAC

El AFM, desarrollado por Escofier & Pagès (1984), es una técnica factorial para analizar tablas en las cuales un conjunto de individuos está descrito por varios grupos de variables. La condición impuesta para su aplicación es que las variables que componen cada grupo sean todas de la misma naturaleza o del mismo tipo: cualitativas o cuantitativas (Escofier & Pagès 1992, Pagès 2004).

En el AAC se utiliza el AFM para equilibrar los pesos de los periodos de tiempo, es decir que la variable longitudinal cualitativa en un periodo  $t$  forma un grupo. Se tienen, entonces,  $T$  grupos y cada uno tiene asociada la subtabla  $\mathbf{A}_t$ .

El AFM equilibra los pesos de cada periodo de tiempo en el sentido de que ninguno puede influir preponderantemente en el análisis global. El AFM agrega otros elementos que enriquecen la descripción obtenida con el AAC.

El AFM, realizado sobre la tabla  $\mathbf{P}$ , se ejecuta en dos etapas (Escofier & Pagès 1992, Bécue & Pagès 2008):

- $T$  análisis de correspondencias de las tablas  $\mathbf{P}_t$ . De cada uno de estos análisis se retiene el primer valor propio, denotado  $\lambda_1^t$ . El AC de la tabla  $\mathbf{P}_t$  equivale al ACP ponderado, denotado  $ACP(\mathbf{X}_t, \mathbf{M}_t, \mathbf{D})$  donde:

- $\mathbf{X}_t$  es la matriz a analizar, con término general:  $x_j = \frac{p_j^t}{\overline{p_j^t}} - 1$
  - $\mathbf{M}_t = \frac{1}{T} \text{diag} \left( \overline{p_j^t} \right)$ , matriz de pesos de las columnas.  $\overline{p_j^t}$  esta dado en la fórmula (2).
  - $\mathbf{D} = \text{diag} \left( \frac{1}{n} \right)$ , matriz de pesos de las filas del grupo  $t$ .
- Un análisis de correspondencias difuso (ACD) a la tabla global  $\mathbf{P}$  en la que cada columna  $j$  se pondera por el inverso del primer valor propio  $\left( \frac{1}{\lambda_1^t} \right)$  de la tabla  $\mathbf{P}_t$  a la cual pertenece. Esta segunda fase del AFM se logra realizando un  $ACP(\mathbf{X}, \mathbf{M}, \mathbf{D})$  donde:
- $\mathbf{X}$  es la matriz a analizar, con término general:  $x_{ij} = \frac{p_{ij}^t}{\overline{p_j^t}} - 1$  (considerando toda la tabla  $\mathbf{P}$ )
  - $\mathbf{M} = \frac{1}{T} \text{diag} \left( \frac{\overline{p_j^t}}{\lambda_1^t} \right)$ .
  - $\mathbf{D} = \text{diag} \left( \frac{1}{n} \right)$ .

De esta manera, se incorpora el análisis factorial múltiple al análisis armónico cualitativo dotándolo de características propias del AFM como son: una síntesis general de toda la información, teniendo en cuenta todos los momentos del tiempo (cada una de las  $T$  franjas por separado y en conjunto); medidas de asociación (coeficientes  $R_v$  y  $L_g$ , ver por ejemplo Abascal & Landaluce (2002)), que permiten cuantificar la semejanza global existente entre los diferentes períodos de tiempo; y trayectorias parciales de los individuos a través del tiempo.

El AFM induce una distancia entre unidades correspondientes, equivalentes a una suma ponderada de las distancias separadas producidas por cada conjunto de variables. La distancia al cuadrado entre unidades  $i$  y  $l$  es entonces:

$$d^2(i, l) = \sum_{t=1}^T \frac{1}{\lambda_1^t} \sum_{j=1}^{m_t} \frac{1}{T \overline{p_j^t}} (p_{ij}^t - p_{lj}^t)^2$$

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:

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

Esta fórmula de transición permite unidades suplementarias y/o variables a ser consideradas, así como la presentación de los puntos parciales, o trayectorias de los individuos.

## 5. Aplicación

El objetivo de la aplicación es la búsqueda de una tipología de una serie de televidentes colombianos, que observan los canales nacionales e internacionales de televisión durante un día específico de la semana.

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. La información que se compila y procesa para conocer los hábitos de consumo de televisión en los hogares colombianos.

Para efectos de esta medición el país se encuentra dividido en 6 zonas: Zona 1: Centro del país (Bogotá y Soacha). Zona2: Región Antioqueña (Medellín, Bello, Envigado e Itagüí.). Zona3: Región Pacífica (Cali, Palmira y Yumbo). Zona4: Región Caribe (Barranquilla, Soledad, Cartagena y Santa Marta). Zona5: Región Cafetera(Armenia, Pereira, Manizales y Dosquebradas). Zona6: Región Oriental (Bucaramanga, Floridablanca, Girón, Cúcuta y Los Patios).

## 5.1. Ajustes de la información inicial

La información proporcionada originalmente muestra el tiempo de permanencia de cada individuo en los diferentes canales de televisión (ver tabla 2). El tiempo del día considerado inicia a las 6 de la mañana y finaliza a las 2 de la mañana del día siguiente. Para registrar los tiempos en un televisor se encuentra apagado se introduce una canal ficticio (canal 15).

TABLA 2: Cambios de canal de los individuos (dos primeros individuos)

Zona	Individuo	Canal	Hora de inicio	Hora de finalización
1	110000301	1000	06:00:00 a.m.	11:24:59 a.m.
1	110000301	0005	11:25:00 a.m.	11:25:59 a.m.
1	110000301	0004	11:26:00 a.m.	12:29:59 p.m.
1	110000301	1000	12:30:00 p.m.	06:58:59 p.m.
1	110000301	0005	06:59:00 p.m.	08:11:59 p.m.
1	110000301	0009	08:12:00 p.m.	08:12:59 p.m.
1	110000301	1000	08:13:00 p.m.	02:00:00 a.m.
1	110000701	1000	06:00:00 a.m.	02:49:59 p.m.
1	110000701	0010	02:50:00 p.m.	03:00:59 p.m.
1	110000701	0074	03:01:00 p.m.	03:08:59 p.m.
1	110000701	0251	03:09:00 p.m.	03:12:59 p.m.
1	110000701	0074	03:13:00 p.m.	03:21:59 p.m.
1	110000701	0251	03:22:00 p.m.	03:25:59 p.m.
1	110000701	0074	03:26:00 p.m.	03:41:59 p.m.
1	110000701	0251	03:42:00 p.m.	03:46:59 p.m.
1	110000701	0074	03:47:00 p.m.	03:50:59 p.m.
1	110000701	0251	03:51:00 p.m.	03:56:59 p.m.
1	110000701	0074	03:57:00 p.m.	04:12:59 p.m.
1	110000701	0122	04:13:00 p.m.	04:21:59 p.m.
1	110000701	0074	04:22:00 p.m.	04:23:59 p.m.
1	110000701	0013	04:24:00 p.m.	04:28:59 p.m.
1	110000701	0074	04:29:00 p.m.	04:41:59 p.m.
1	110000701	1000	04:42:00 p.m.	02:00:00 p.m.

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 continua en el mismo canal, no se reporta ningún cambio). La variable a analizar es categórica, ya que representa el canal en el cual se encuentra la persona en un tiempo determinado durante ese día. Los períodos 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 interés básico de este ejercicio 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 3.

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, sólo 246 de ellos se seleccionaron el día de la medición. Dentro de estos, muchos presentan muy poca frecuencia de tiempo de permanencia en ellos. Finalmente se tomó la decisión de conservar aquellos con alto porcentajes de cambios y realizar las agrupaciones de aquellos que no la tuvieron. Estas agrupaciones fueron realizadas y renombradas teniendo

TABLA 3: Franjas de audiencia ajustadas

Franja	Tipo de Audiencia
fra1: 06:00 -12:00	Familiar
fra2: 12:00 -16:00	Familiar
fra3: 16:00 -17:00	Infantil
fra4: 17:00 -19:30	<i>Prime Time</i>
fra5: 19:30 -22:10	Familiar
fra6: 22:10 -02:00	Adultos

en cuenta la organización realizada por las diferentes empresas prestadoras de servicios de televisión y a las características comunes de ellos (ver tabla 4).

TABLA 4: Canales de televisión estructurados

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

Del total de los individuos de la muestra (2334), sólo 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, descritos en la tipología. Luego de los ajustes realizados, el archivo final sujeto a análisis quedó conformado por 1306 individuos y 3986 cambios de canales.

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 5). 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.

TABLA 5: Parte del archivo con información adicional de los individuos

Individuo	Género	Rango edad	Nivel socio-económico	Zona	Peso
110000701	Mujer	40 a 64	Medio	1	00.19750
110001901	Hombre	40 a 64	Medio	1	00.04248
110004301	Hombre	25 a 39	Medio	1	00.04376
110023101	Hombre	40 a 64	Alto	1	00.09113
110023103	Mujer	25 a 39	Alto	1	00.09113
110027601	Hombre	40 a 64	Alto	1	00.04537
110028201	Hombre	40 a 64	Medio	1	00.10014
110029405	Mujer	12 a 17	Bajo	1	00.08831
110039201	Hombre	25 a 39	Bajo	1	00.10619
220006302	Hombre	18 a 24	Bajo	2	00.04376

## 5.2. Análisis factorial múltiple sobre el análisis de correspondencias difuso

La tabla sometida al análisis de correspondencias difuso (ACD), de dimensión 1320 x 90, 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. En esta fase se tuvo en cuenta el peso o factor de expansión de los individuos de la muestra. Se realizaron dos 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.

Algunos individuos fueron excluidos del análisis ya que presentaban características especiales (permanencia en un sólo canal durante todo el día, veían muy poca televisión o alternaban el tiempo entre sólo dos canales) y por lo tanto formaban ejes factoriales.

Para la ejecución del AFM se utiliza FactoMineR (Husson et al. 2007), el cual no permite ingresar pesos en las variables, se calcula la matriz que incluye los pesos en las coordenadas: cada columna de la matriz  $\mathbf{X}$  del ACD (*dudi\$tab* de la función *dudi.fca* del *ade4* (Chessel et al. 2004)) se multiplica por la raíz cuadrada de su peso usado en el ACD (*dudi\$cw*) (ver sección 3).

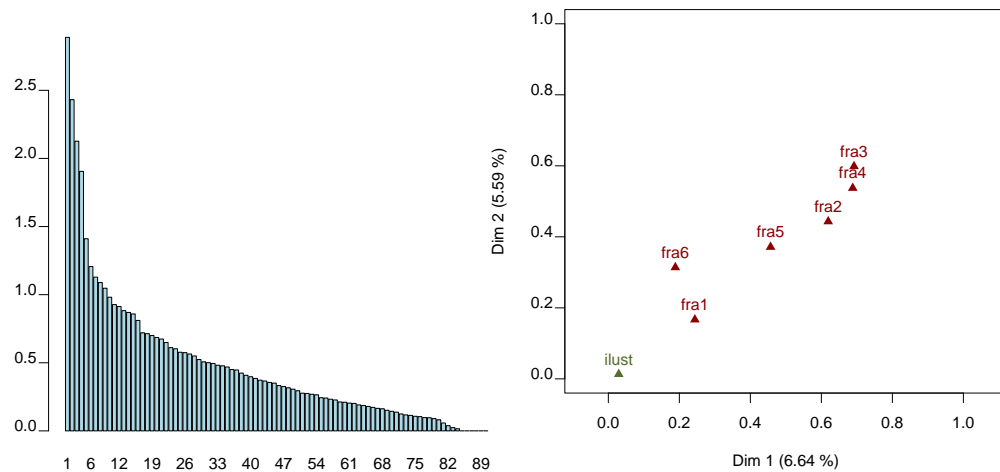


FIGURA 1: Gráfica de los primeros valores propios y del primer plano factorial de las franjas del AFM y del ACD combinados

Los 6 análisis de correspondencias realizados para cada franja dan como resultado diferencias en el primer valor propio (ver tabla 6), lo que justifica la ponderación del AFM, que estandariza esos valores propios en 1 para el análisis global.

TABLA 6: Resumen de valores propios de los análisis separados

Grupo	Valor 1	Valor 2	Valor 3	Valor 4	Valor 5
fra1	0.103	0.085	0.078	0.071	0.066
fra2	0.100	0.087	0.0755	0.075	0.069
fra3	0.141	0.135	0.130	0.117	0.105
fra4	0.113	0.093	0.092	0.087	0.083
fra5	0.118	0.092	0.081	0.076	0.066
fra6	0.073	0.067	0.060	0.056	0.055

La proyección de las franjas sobre el primer plano factorial (ver figura 1) muestra mayor contribución en los dos primeros ejes de las franjas 2,3 y 4, las cuales son además similares. Las franjas 1 y 6 contribuyen menos y son también similares entre sí.

La diagonal de la matriz de coeficientes  $L_g$  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 (figura 1) se destacan 5 ejes.

Las 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  $L_g$ 

	fra1	fra2	fra3	fra4	fra5	fra6	ilust	AFM
fra1	4.773	0.602	0.250	0.297	0.182	0.135	0.076	2.160
fra2	0.602	4.248	1.233	0.689	0.344	0.337	0.063	2.580
fra3	0.250	1.233	5.386	1.446	0.414	0.283	0.058	3.120
fra4	0.297	0.689	1.446	4.866	1.016	0.222	0.072	2.955
fra5	0.182	0.344	0.414	1.016	4.050	0.559	0.102	2.272
fra6	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

TABLA 8: Coeficientes  $RV$ 

	fra1	fra2	fra3	fra4	fra5	fra6	ilust	AFM
fra1	1.000	0.134	0.049	0.062	0.041	0.027	0.014	0.427
fra2	0.134	1.000	0.258	0.152	0.083	0.071	0.012	0.541
fra3	0.049	0.258	1.000	0.283	0.089	0.053	0.010	0.581
fra4	0.062	0.152	0.283	1.000	0.229	0.043	0.013	0.579
fra5	0.041	0.083	0.089	0.229	1.000	0.120	0.020	0.488
fra6	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

En el plano factorial de los indicadores parciales de todos los grupos (ver figura 2), se observa la alta correlación entre el canal 15 (televisor apagado) en todas las franjas con el primer factor del AFM, en oposición con los canales realmente vistos. En el eje dos se destaca la separación de  $C14$ : *otros canales internacionales* (parte superior) con los canales  $C5$ : *Caracol* y  $C4$ : *RCN* (parte inferior).

Para entender las diferencias de teleaudiencia los planos factoriales son insuficientes. Es necesario, por lo tanto, proceder a la clasificación de los individuos en el espacio vectorial de los primeros factores del análisis. Para decidir el número de ejes a utilizar, se usó el procedimiento propuesto por Barbary (1996) el cual consiste en realizar la clasificación varias veces cambiando el número de ejes y usando como criterio el número de individuos de la clase con mayor cantidad de individuos. El análisis a interpretar es aquel que origine el menor número de individuos en la clase mayoritaria.

Se realizaron ensayos con 3, 4, 5, 6, 9 y 10 ejes, seleccionando, de acuerdo al criterio enunciado, la clasificación con 4 ejes.

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 vez seleccionado el número de clases (ver por ejemplo Lebart et al. (1995, cap. 2)). De acuerdo a los saltos observados en el histograma de los índices de nivel (no se incluye), se decide caracterizar a los individuos a través de 7 clases, más una adicional conformada por los individuos no vieron televisión el día observado.

## 6. Descripción de las clases obtenidas

Los televidentes de la muestra que tuvieron encendido en algún momento el televisor durante el día de la medición se repartieron en 7 clases con los porcentajes: 65.0, 3.3, 13.6, 6.0, 3.9, 6.7 y 1.4 respectivamente. A continuación se hace un resumen de caracterización de las clases obtenidas. Las clases se construyen

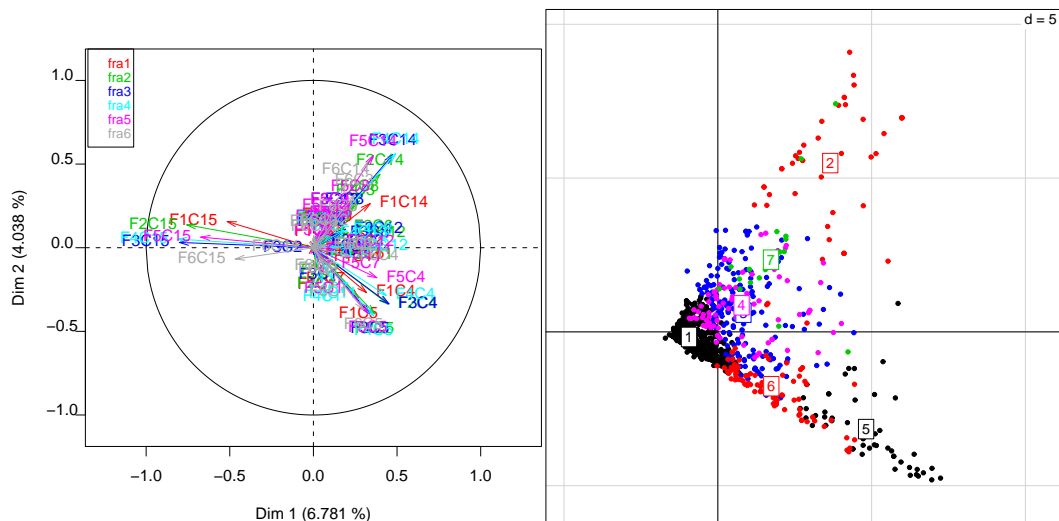


FIGURA 2: Primer plano factorial del AFM-ACD: columnas (franja-canal) e individuos mostrando las clases

con los cambios y permanencias de los televidentes en los diferentes canales (tabla 9), pero se caracterizan también con otras variables: género, estratos socioeconómicos y rangos de edad (tabla 10) y zonas del país (tabla 11).

- Clase 1 (65%). 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. Resalta la escogencia del canal City TV en gran parte del tiempo de la franja 3. Se caracteriza por los televidentes con edades entre 40 y 64 años, niveles socioeconómicos medio y bajo de diferentes zonas del país.
- Clase 2 (3.3%). Este grupo caracteriza a los televidentes que escogen principalmente los canales internacionales como Cinecanal Este, Cinecanal Oeste, CNN en español, HBO, entre otros. En menor proporción se seleccionan los canales *Fox Sport*, series internacionales y el canal institucional. También poseen como característica especial la no elección de los canales TeleCaribe, City TV y Canal Uno durante casi todo el día, y en tener encendido el televisor gran parte del día. En esta clase predominan los individuos con edades entre 25 y 40 años, de nivel socioeconómico medio y pertenecientes a las zonas Cafetera y Centro del País.
- Clase 3 (13.6%). Conformada por individuos que pasan gran parte del día viendo los canales Teleantioquia, otros canales internacionales y Fox Sport. Se encuentra caracterizada por televidentes con edades entre 25 y 64 años, principalmente de clase media de diferentes zonas.
- Clase 4 (6%). Este grupo se encuentra conformado por televidentes que prefieren principalmente los canales infantiles internacionales y los canales *Fox Sport*, series internacionales (*Sony*, *Warner Chanel*, etc.) y algunos canales internacionales como Cinecanal Este, Cinecanal Oeste, CNN en español, *HBO*, entre otros, en casi todas las franjas. Se encuentra formada con mayor proporción de individuos pertenecientes al nivel socioeconómico bajo, del centro del país.
- Clase 5 (3.9%). Esta clase de televidentes permanece la gran parte de su tiempo en el canal RCN, con la televisión encendida la mayor parte del día. En forma poco frecuente selecciona los canales Caracol y series internacionales. En esta clase se encuentran mujeres entre 25 y 40 años pertenecientes en gran proporción a la zona central de Colombia.

- Clase 6 (6.7%). Este grupo se caracteriza por preferir el canal Caracol sobre el resto de los canales. También son seleccionados los canales Telecaribe y Canal Uno. No son seleccionados los canales culturales internacionales (*Discovery Chanel, National Geographic, etc.*) en casi ninguna de las franjas. Formado principalmente por individuos de las zonas Centro y Caribe.
- Clase 7 (1.4%). Esta clase de televidentes prefieren los canales infantiles internacionales en casi todas las franjas. También se caracteriza por no ver prácticamente ningún otro canal. Permanece casi todo el tiempo el televisor encendido. Se encuentra conformada por televidentes de edades entre 4 y 7 años, y de 25 a 40 años. No existen en esta clase individuos entre 12 y 24 años.
- Clase 8. Esta clase se encuentra conformada por los individuos que no vieron televisión. La mayoría son hombres, de edades entre 18 y 64 años, de clases media y baja, de diferentes zonas del país.

En la figura 2 se muestran las clases sobre el primer plano factorial del AFM y ACD combinados, con lo cual se tiene una síntesis gráfica de las clases.

TABLA 9: Extracto de caracterización de las clases según los porcentajes franja-canal (se seleccionan las columnas franja-canal con valor test superior a 2 y porcentaje en la clase mayor de 10)

Franja-canal	Clase	Valor test	Porcentaje clase	Frecuencia	Porcentaje global
F4C15	1	24.8	93.7	849	76.2
F3C15	1	23.8	98.3	849	80.4
F5C15	1	20.8	80.4	849	63.8
F2C15	1	20.3	92.9	849	80.4
F6C15	1	13.2	90.9	849	84.9
F1C15	1	12.0	93.6	849	88.2
F3C14	2	24.9	71.5	43	4.8
F4C14	2	24.9	73.8	43	6.1
F5C14	2	19.6	48.0	43	5.2
F2C14	2	18.2	34.8	43	3.6
F1C14	2	14.2	20.6	43	2.5
F6C14	2	10.2	21.0	43	3.5
F5C4	3	8.8	28.4	178	13.1
F5C14	3	7.2	12.6	178	5.2
F4C14	3	6.1	13.8	178	6.1
F3C14	3	5.6	11.8	178	4.8
F4C12	4	15.4	25.4	79	3.4
F5C12	4	14.7	19.4	79	2.5
F3C12	4	13.0	21.7	79	2.5
F2C12	4	12.7	16.0	79	2.3
F4C14	4	3.5	13.0	79	6.1
F3C4	5	29.1	92.7	51	5.8
F4C4	5	27.3	67.6	51	5.2
F2C4	5	25.6	68.8	51	6.2
F1C4	5	16.6	32.0	51	3.5
F5C4	5	13.2	58.5	51	13.1
F6C4	5	6.9	14.1	51	3.9
F3C5	6	26.5	56.9	88	4.9
F4C5	6	24.7	50.0	88	5.8
F2C5	6	21.1	38.5	88	5.4
F5C5	6	17.4	52.1	88	11.6
F6C5	6	14.6	24.3	88	5.3
F1C5	6	11.2	14.3	88	2.8
F3C12	7	24.9	81.0	18	2.5
F2C12	7	20.6	50.3	18	2.3
F4C12	7	19.7	63.8	18	3.4
F5C12	7	15.1	39.8	18	2.5
F1C12	7	8.6	16.2	18	1.4

TABLA 10: Perfiles de las clases según género, estrato socioeconómico y rangos de edad

Clase	—Género—		Nivel socioeconómico			—Rangos de edad—					
	Mujer	Hombre	Bajo	Medio	Alto	4a11	12a17	18a24	25a39	40a64	mas64
1	48.6	51.4	34.3	47.7	18.0	13.7	12.0	10.5	23.8	33.2	6.8
2	44.2	55.8	30.2	53.5	16.3	2.3	11.6	11.6	27.9	39.5	7.0
3	51.1	48.9	32.0	44.4	23.6	5.1	10.7	09.6	23.0	43.8	7.9
4	51.9	48.1	53.2	32.9	13.9	26.6	11.4	06.3	22.8	27.8	5.1
5	60.8	39.2	37.3	45.1	17.6	5.9	2.0	07.8	15.7	60.8	7.8
6	50.0	50.0	38.6	52.3	9.1	3.4	9.1	11.4	27.3	38.6	10.2
7	44.4	55.6	27.8	44.4	27.8	27.8	0.0	0.00	33.3	33.3	5.6
8	43.9	56.1	33.5	48.0	18.4	16.7	12.7	20.0	23.0	22.3	5.3

TABLA 11: Perfiles de las clases según zonas del país

Clase	—Zonas—					
	1:Centro	2:Antioquia	3:Pacífico	4:Caribe	5:Cafetera	6:Oriental
1	25.1	11.2	16.1	18.7	11.1	17.8
2	25.6	14.0	09.3	2.3	44.2	4.7
3	26.4	12.4	29.8	7.3	12.4	11.8
4	26.6	12.7	16.5	10.1	16.5	17.7
5	42.3	2.0	11.8	15.7	13.7	15.7
6	45.5	3.4	8.0	25.0	5.7	12.5
7	55.6	5.6	5.6	5.6	16.7	11.1
8	26.2	12.5	16.1	15.2	13.1	16.9

## 7. Conclusiones

El AFM permite equilibrar la influencia de los periodos de tiempo y enriquece la interpretación del AAC, ejecutado como un ACD.

La metodología propuesta permite una buena descripción del *rating* de televisión de un día en combinación con la partición obtenida.

Se obtienen 7 clases de televidentes que encienden el televisor en algún momento del día analizado una clase de frecuencia alta (1: 65%), que tienen los televisores apagados la mayor parte del día y seis clases diferenciadas por los canales que ven durante el día (ver sección 6).

La preferencia de canales está asociada al género, estrato socioeconómico, edad y zona del país de los televidentes.

## Software

Para realizar los análisis factoriales mencionados, se utilizó el paquete estadístico *R* (R Development Core Team 2007) y específicamente, las librerías: *ade4* (Chessel et al. 2004) para realizar el análisis de correspondencia difuso, *FactoMineR* (Husson et al. 2007) para el análisis factorial múltiple y *FactoClass* (Pardo & Del-Campo 2007) para la caracterización de las clases.

## Referencias

- Abascal, E. & Landaluce, M. (2002), ‘Análisis factorial múltiple como técnica de estudio de la estabilidad de los resultados de un análisis en componentes principales’, *Questiío* **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’, Universidad Nacional de Colombia. Facultad de Ciencias. Departamento de Estadística; ORSTOM y PRESTA, pp. 121–144.

- Bécue, M. & Pagès, J. (2008), 'Multiple factor analysis and clustering of a mixture of quantitative, categorical and frequency data', *Computational Statistics & Data Analysis* **52**, 3255–3268.
- Chessel, D., Dufour, A.-B. & Thioulouse, J. (2004), 'The ade4 package-I- One-table methods', *R News* **4**, 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.
- Deville, J. & Saporta, G. (1980), 'Analyse harmonique qualitative', *Data Analysis and Informatics* pp. 375–389.
- 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.
- Escofier, B. & Pagès, J. (1992), *Análisis factoriales simples y múltiples. Objetivos, métodos e interpretación*, Universidad del País Vasco, Bilbao.
- Gámez, N. E. (2007), Fundamentos y aplicaciones del análisis de correspondencias difuso (ACD), Trabajo de grado para optar al título de Estadística, Universidad Nacional de Colombia. Facultad de Ciencias. Departamento de Estadística.  
\*<http://www.docentes.unal.edu.co/cepardot/docs/TrabajosGrado/TrabajoGradoNurysGamez.pdf>
- Husson, F., Josse, J., Le, S. & Mazet, J. (2007), *FactoMineR: Factor Analysis and Data Mining with R*. R package version 1.07.  
\*<http://factominer.free.fr>, <http://www.agrocampus-rennes.fr/math/>
- Lebart, L., Morineau, A. & Piron, M. (1995), *Statistique exploratoire multidimensionnelle*, Dunod, Paris.
- Pagès, J. (2004), 'Multiple Factor Analysis: Main Features and Application to Sensory Data', *Revista Colombiana de Estadística* **27**(1), 1–26.  
\*<http://www.matematicas.unal.edu.co/revcoles/>
- Pardo, C. E. & Del-Campo, P. C. (2007), 'Combinación de métodos factoriales y de análisis de conglomerados en R: el paquete FactoClass', *Revista Colombiana de Estadística* **30**(2), 231–245.  
\*<http://www.matematicas.unal.edu.co/revcoles/>
- R Development Core Team (2007), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.  
\*<http://www.R-project.org>
- Saporta, G. (1996), L'analyse harmonique qualitative, une synthèse de la théorie, in 'Seminario de capacitación e investigación. Recolección y análisis de datos longitudinales', pp. 111–120.