

Combinación del análisis factorial múltiple y el análisis armónico cualitativo en la descripción de datos longitudinales cualitativos

MARTHA LUCÍA CORRALES BOSSIO



UNIVERSIDAD NACIONAL DE COLOMBIA
FACULTAD DE CIENCIAS
DEPARTAMENTO DE ESTADÍSTICA
BOGOTÁ, D.C.
FEBRERO DE 2010

Combinación del análisis factorial múltiple y el análisis armónico cualitativo en la descripción de datos longitudinales cualitativos

MARTHA LUCÍA CORRALES BOSSIO

TRABAJO DE TESIS PARA OPTAR AL TÍTULO DE
MAGISTER EN ESTADÍSTICA

DIRECTOR
CAMPO ELÍAS PARDO, M.SC.
MAGISTER EN ESTADÍSTICA

CODIRECTOR
JUAN RAMOS, PH.D
DOCTOR EN ESTADÍSTICA



UNIVERSIDAD NACIONAL DE COLOMBIA
FACULTAD DE CIENCIAS
DEPARTAMENTO DE ESTADÍSTICA
BOGOTÁ, D.C.
FEBRERO DE 2010

Nota de aceptación

Trabajo de tesis

Jurado

Jurado

Director
Campo Elías Pardo

Codirector
Juan Ramos

Bogotá, D.C., Febrero 12 de 2010

Dedicado a

Dedico este trabajo a mi esposo, Guillermo, y a mis hijos Guillermo Andrés, Diego y Sebastián, quienes con todo su amor supieron entender la ausencia y el tiempo que no les pude dar.

Agradecimientos

Quiero aprovechar esta oportunidad para agradecer a las personas que me han colaborado para poder sacar adelante esta investigación, la cual soporta mi tesis de grado.

De manera muy especial a mi director, Campo Elías Pardo, el cual con sus enseñanzas, exigencia y labor académica me dieron las bases para desarrollar este trabajo; además, por aceptar dirigir mi tesis y por el tiempo empleado en las correcciones y sus respectivas sugerencias.

A mi codirector, Juan Ramos, quien fue la persona que proporcionó los datos para realizar la aplicación del nuevo método propuesto.

Y por último, a mi amigo y jefe, Reinaldo Nuñez, Director de la Escuela de Matemáticas de la Universidad Sergio Arboleda, el cual me facilitó el tiempo necesario para realizar mis estudios apoyándome en todo momento para mejorar mi nivel académico.

CONTENIDO

1. Introducción	1
2. Métodos utilizados	5
2.1. Notación	5
2.2. Análisis armónico cualitativo	7
2.2.1. Análisis de correspondencias difuso	10
2.3. Análisis factorial múltiple	11
2.3.1. AFM y el análisis multicanónico	13
2.3.2. Relación entre los grupos de variables	14
3. Combinación del AAC y el AFM: AAC-AFM	15
3.1. Aspectos generales de la combinación	15

3.2. Propiedades de la combinación del AAC y el AFM	16
3.2.1. Distancias entre individuos y categorías	17
3.2.2. Relaciones de transición	18
3.2.3. Inercias de la combinación	19
3.2.4. Relación de los grupos de las variables	20
4. Aplicación de la combinación del AAC y el AFM	22
4.1. Descripción del rating de TV durante un día en Colombia . . .	22
4.2. Fase de preprocesamiento de los datos	23
4.2.1. Complementos y ajustes de la información	24
4.2.2. Limpieza de los datos	27
4.2.3. Transformación de los datos	28
4.3. Análisis factoriales realizados	30
4.3.1. Análisis armónico cualitativo (AAC)	30
4.3.2. Análisis factorial múltiple sobre el AAC	32
4.4. Clasificación de los televidentes	37
4.5. Descripción de las clases obtenidas	39
4.5.1. Caracterización de las clases	40
4.6. Comparación de las clases obtenidas con ambos métodos . . .	47

5. Librería hqa	50
5.1. Aspectos iniciales	50
5.2. Análisis factoriales	51
5.3. Descripción de la librería hqa	52
6. Conclusiones	53
6.1. Conclusiones de la combinación teórica del AAC y el AFM . .	53
6.1.1. Conclusiones de la descripción del <i>rating</i> de TV	54
6.2. Conclusiones finales del <i>rating</i> de televisión	55
A. Análisis factorial general	60
A.1. Nubes del análisis factorial general	60
A.2. Análisis de la nube de las filas	61
A.3. Análisis de la nube de las columnas	61
A.4. Relaciones de transición	62
B. Canales de televisión	64
C. Clases obtenidas con el AAC	70
D. Clasificación con AAC. Variables ilustrativas.	74
E. Clasificación con AAC. Variable longitudinal activa	78

F. Clasificación con AAC y AFM. Variables ilustrativas.	87
G. Clasificación con AAC y AFM. Variable longitudinal activa	90
H. Planos factoriales AAC	99
I. Ejemplo de uso de la librería <i>hqa</i>	101
I.1. Librería <i>hqa</i>	101

CAPÍTULO 1

Introducción

Con frecuencia en muchas áreas se presentan fenómenos que permiten observar variaciones cualitativas a lo largo del tiempo. Esta variable, a la cual se le denominará, variable longitudinal activa, es de gran importancia al momento de estudiar cualquier evento, ya sea de tipo demográfico, sociológico, psicológico, económico, etc., ya que constituye un punto de partida fundamental en el estudio de una determinada realidad.

En estos casos no es posible utilizar los métodos clásicos de series cronológicas (Lera et al. 2007). Es recomendable, por lo tanto, la utilización de técnicas exploratorias de datos como los llamados métodos factoriales, entre los que se encuentran el análisis armónico cualitativo (AAC) (Deville & Saporta 1980) y el análisis factorial múltiple (AFM) (Escofier & Pagès 1984).

El análisis armónico cualitativo es una herramienta potente en la descripción multivariada de datos de tipo longitudinal categóricos. En la práctica es la realización de un análisis de correspondencias, y por tal motivo se pueden hacer uso de variables ilustrativas tanto transversales como longitudinales, que enriquecen la interpretación del problema tratado. Lo complicado de su aplicación reside en lo difícil que puede resultar la lectura de los factores en algunos casos y las decisiones acerca de los parámetros del análisis que hay

que tomar que dependen del estudio que se esté realizando. Sin embargo, el AAC ha demostrado en muchos estudios prácticos donde se ha utilizado, que es uno de los métodos más efectivos en el tratamiento de este tipo de datos (Barbary 1996).

El análisis factorial múltiple se presenta como una técnica factorial para la descripción de un mismo conjunto de individuos a través de varios grupos de variables. En los estudios de evolución, los grupos se forman según el momento del tiempo en que hayan sido medidas las variables. Se obtiene entonces, una tabla con combinaciones de cortes transversales en el tiempo.

La estructuración en grupos enriquece el estudio, ya que no se limita a la obtención de una tipología de la trayectoria de los individuos sino que se amplía a la búsqueda de relaciones obtenidas al interior de cada uno de los momentos del tiempo y entre estos períodos.

El AFM proporciona una visión gráfica completa de las trayectorias temporales de los individuos, así como los individuos con mayores y menores cambios evolutivos. También muestra las variables responsables de en gran medida de estos cambios.

El objetivo de esta tesis consiste en proporcionar una metodología estadística para el tratamiento de datos longitudinales categóricos, a través de la combinación del análisis armónico cualitativo y del análisis factorial múltiple.

La combinación de estos dos métodos factoriales se realiza para incorporar la ponderación del AFM sobre una tabla codificada de acuerdo a los principios del AAC y de esta manera obtener una metodología que enriquezca el análisis de los datos: 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.

Con el fin de ilustrar esta combinación, se analizan una serie de televidentes colombianos residentes en diferentes zonas del país (centro, antioqueña, pacífica, caribe, cafetera, oriente) que observaron los canales de televisión durante un día específico de la semana. Los cambios de canal que realiza un individuo se detectan minuto a minuto obteniéndose 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 analizada es cualitativa, ya que representa el canal en el cual se encontraba la persona en un tiempo determinado durante el día de estudio y a través de las diferentes franjas de televisión reguladas por el Consejo Nacional de Televisión. Los individuos se encuentran también caracterizados por una serie de variables que ilustran el estudio: el género, el nivel socioeconómico, la zona y la edad.

Para la toma de decisiones respecto a la ejecución del AAC, se tuvieron en cuenta los trabajos realizados por Barbary (1996) y Pinzón (1998). Para el AFM se tomaron como guía los trabajos realizados por Pagés (2004), Landaluce & Valencia (2000) y Abascal & García (2003).

Adicionalmente se desarrolló una librería en **R** (R Development Core Team 2007) denominada **hqa** (la cual se encuentra disponible en el *CRAN*) cuyo objetivo principal es, dada una tabla que contenga un conjunto de individuos, los cuales seleccionan diferentes categorías de una variable longitudinal cualitativa en distintos períodos de tiempo, transformarla en una matriz que se pueda analizar por medio de un análisis armónico cualitativo y/o a través de la combinación del AAC y el AFM.

Este paquete también presenta funciones adicionales para realizar un análisis de clasificación automático y de esta manera obtener tipologías de los individuos de acuerdo a las similitudes entre escogencias de las categorías de la variable estudiada.

En el capítulo 2 se presenta un breve resumen de los métodos utilizados para realizar la combinación y así lograr una comprensión de ella.

En el capítulo 3 se realiza la combinación entre el AAC y el AFM, y se presenta el desarrollo de las fórmulas clásicas de un análisis factorial general, adaptadas a dicha combinación.

El capítulo 4 muestra una aplicación de la combinación, a través del estudio del *rating* de televisión, ya mencionado, para mostrar su funcionamiento.

En el capítulo 5 se puede observar la estructura de la librería *hqa* realizada para la combinación entre el AAC y el AFM, con una pequeña descripción de las funciones que la componen.

En el capítulo 6 se presentan las conclusiones encontradas tanto del desarrollo teórico de la combinación, como de los resultados obtenidos de la parte práctica del método.

CAPÍTULO 2

Métodos utilizados

Los diferentes métodos utilizados en este trabajo se encuentran enmarcadas en el ámbito de los métodos factoriales (ver apéndice A), los cuales buscan resumir la información contenida en una matriz de datos, identificando un reducido número de factores que representan a las variables originales con una pérdida mínima de información. Se presenta a continuación la notación usada y un resumen de los métodos para entender su funcionamiento y lograr una comprensión de la combinación realizada entre ellos.

2.1. Notación

Sean $t = 1, 2, \dots, T$ instantes, $j = 1, \dots, m_t$ las categorías de una variable longitudinal cualitativa, medidas en el período t para $i = 1, \dots, n$ individuos que las seleccionaron.

Se denota:

- \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:

$$p_{ij}^t = \frac{a_{ij}^t}{a_{ij}} \quad (2.1)$$

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

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

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

Se define

$$\bar{p}_j^t = \frac{1}{n} \sum_{i=1}^n p_{ij}^t \quad (2.2)$$

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

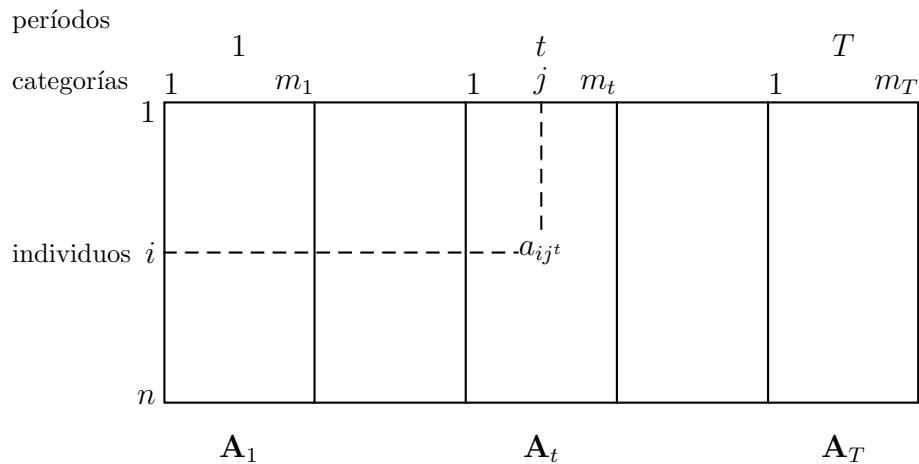


FIGURA 2.1. Tabla de tiempos de permanencia en cada modalidad

2.2. Análisis armónico cualitativo

El análisis armónico cualitativo (AAC) fue presentado por primera vez en un breve artículo teórico por Deville & Saporta (1980) y luego más extensamente por Deville (1982) (en español ver Barbary & Pinzón (1998)).

El AAC es un método exploratorio de datos longitudinales categóricos cuyo objetivo es estructurar la información de los "individuos" de una muestra, durante un período de tiempo determinado. Estos individuos se describen por la cronología de una sucesión de cambios de etapas o estados (cambios de sitios de residencia, cambios de canales de televisión, de estado matrimonial, etc) entre un conjunto de etapas posibles (Saporta 1996). Deville & Saporta (1980) demostraron que este tipo de datos se pueden analizar de forma similar a lo que se haría en un proceso escalar (análisis armónico) y para un conjunto finito de variables cualitativas no temporales (análisis canónico generalizado o análisis de correspondencias múltiples).

El punto de arranque es considerar este tipo de datos como la realización de un proceso cualitativo en un tiempo continuo $X^t(i)$ de m categorías ($X^t(i)$ es el estado entre los m estados posibles, del individuo i en el instante t). Debido a que este proceso es cualitativo, se establece un código escalar del mismo: $Y(t) = f_t(X^t(i))$ (asociar a todas las categorías de cruce de los diferentes

estados con los instantes de observación, un código real). De esta manera se puede realizar el análisis armónico del proceso $Y(t)$, descomponiéndolo en una serie de variables aleatorias independientes del tiempo llamadas armónicas de procesos.

Deville & Saporta (1980) continúan demostrando que este análisis armónico es una generalización de los métodos clásicos de análisis de datos, con las interpretaciones normales de ellos. La forma de realizar este enlace es sencilla: se comienza por discretizar el período de observación del proceso, considerando los intervalos del tiempo en los cuales ningún individuo cambia de estado.

El análisis armónico en tiempo continuo se convierte, en tiempo discreto, en la descomposición de factores resultante del análisis de correspondencias de una tabla disyuntiva particular con n filas (donde n es el número de individuos) y mp columnas (donde p es el número de períodos en los que el proceso permanece estable). Cualquier elemento de la tabla disyuntiva vale 1 si el individuo está en el estado considerado durante el período y 0 si no lo está.

Sin embargo, de acuerdo a los valores de n , m y p la tabla disyuntiva puede ser de un tamaño bastante grande. Debido a esto, Deville & Saporta (1980) proponen una solución cercana al problema global, realizando una recodificación de la información original. Esta consiste en observar a los individuos durante un período de tiempo notado por $[0, T]$, y dividirlo en un número razonable de intervalos, de igual o de diferente amplitud, sin tener en cuenta los cambios de estado individuales. 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 período. Esta tabla se somete entonces a un análisis de correspondencias con las técnicas de interpretación habituales.

Cuando en el AAC se opta por realizar 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 2.1) con las propiedades descritas en la sección de notación (2.1). Esta tabla \mathbf{P} , se puede ver como una tabla de codificación difusa, y por ello se puede analizar, a través de un análisis de correspondencias difuso, el cual se presenta en la subsección 2.2.1.

En la práctica se debe decidir la longitud del tiempo del análisis ($[0, T]$). Esta es una decisión importante, pues determina censura (individuos con trayectorias menores a la longitud del tiempo seleccionado, presentan períodos sin datos) y pérdida de información (individuos con trayectorias mayores a la longitud del tiempo determinado pierden parte de sus datos). Tiempos muy largos implican mucha censura, y tiempos muy cortos, gran pérdida de información (Barbary 1996). La propuesta presentada en Pinzón (1998) conlleva a un compromiso entre censura y pérdida de información. Es decir, tomar diferentes tiempos de análisis y determinar la censura y la pérdida de información para cada uno de ellos. La decisión es tomar el tiempo de análisis que logre un balance entre el tiempo censurado y la pérdida de información.

El número de períodos de tiempo y sus límites (división del período ($[0, T]$) en una cantidad limitada de períodos y la longitud de cada período), conllevan otra decisión empírica a tomar. Por supuesto, se pierde menos información si se toman períodos pequeños. Sin embargo, esto genera un trabajo más dispendioso y no necesariamente implica una ganancia grande en información (Barbary & Pinzón 1998). Por otro lado, se pierde menos información cuando la longitud de los períodos se define de tal manera que se logre un número de etapas similares en todos los períodos.

El análisis se completa con una clasificación realizada a partir de las coordenadas de los primeros ejes obtenidos del análisis de correspondencias. Las clases se caracterizan tanto por la variable longitudinal como por otras variables del estudio, denominadas ilustrativas, las cuales pueden ser transversales o longitudinales.

Para realizar la clasificación se deben tomar nuevamente dos decisiones importantes: el número de ejes y el número de clases a obtener. Un procedimiento propuesto por Barbary (1996) para el número de ejes a utilizar es el de realizar el análisis varias veces cambiando el número de ejes y usando como criterio el número de individuos en la clase mayoritaria. Se toma el número de ejes que origine menos individuos en la clase mayoritaria. El número de clases estará guiado por los saltos del árbol de clasificación jerárquica.

2.2.1. Análisis de correspondencias difuso

El análisis de correspondencias difuso (ACD), propuesto por Chevenet et al. (1994) (en español ver a Gamez (2007)) es una extensión del análisis de correspondencias múltiples, utilizado para el análisis de las posibles tendencias y relaciones existentes entre los rasgos de especies o subpoblaciones.

En muchas áreas, especialmente en ecología, es frecuente encontrar datos que son difusos por su naturaleza, es decir, se encuentra información sobre especies (o individuos) que asumen varias categorías de una misma variable con diferente grado de asociación. De igual manera, las tablas de datos resultantes de la yuxtaposición de diferentes tablas de contingencia se pueden ver como tablas difusas.

La tabla \mathbf{A} , definida en la sección 2.1, se puede considerar como una tabla de codificación difusa, ya que está conformada por la yuxtaposición de las t tablas \mathbf{A}_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.

Como se mencionó anteriormente, el ACD es una extensión del análisis de correspondencias múltiples (ACM). Definiendo las métricas adecuadas en los espacios de los individuos y de las variables, el ACM a su vez, puede definirse como un análisis en componentes principales ponderado (ACP ($\mathbf{X}, \mathbf{M}, \mathbf{D}$)). Por lo tanto, 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 & Pages (1992, Cap.4), Pagés (2004)).

Realizando entonces los ajustes correspondientes el ACD, es el análisis de la tabla de porcentajes \mathbf{P} , definida en la sección de notación, 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 2.2)
 - $\mathbf{M} = \text{Diag} \left(\frac{1}{T} \bar{p}_j^t \right) = \frac{1}{T} \text{diag}(\bar{p}_j^t)$
 - $\mathbf{D} = \text{Diag} \left(\frac{1}{n} \right) = \frac{1}{n} \mathbf{I}_n$

2.3. Análisis factorial múltiple

El análisis factorial múltiple, diseñado 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. La condición impuesta para su aplicación es que las variables que componen cada grupo sean todas del mismo tipo: cualitativas o cuantitativas (Escofier & Pages 1992, Pagés 2004).

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 períodos de tiempo o por todas las variables observadas en el mismo momento del tiempo (Abascal et al. 2001).

Al igual que todos los métodos factoriales, el objetivo principal del AFM, es la búsqueda de un subespacio de dimensiones reducidas en el cual se representen de forma óptima las distancias entre los individuos y la estructura de las correlaciones entre las variables, incluyendo el estudio de los grupos de variables.

El AFM se realiza 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. El papel que desempeña un grupo de variables en un análisis en conjunto intervienen dos elementos: el número de variables que contenga un grupo (a

mayor número de variables, mayor será la influencia del grupo en el análisis conjunto) y la estructura del grupo (mientras más fuerte sea la estructura del grupo, mayor correlación entre las variables que lo conforman y por lo tanto, la influencia sería más determinante en la construcción de los factores de variabilidad) (Escofier & Pages 1992).

Por tal motivo, en el AFM se le da un peso a cada una de las variables de los grupos que equivale al inverso del primer valor propio de los análisis factoriales realizados por separado, lo que 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. Con esta ponderación realizada a las variables de un mismo grupo se efectúa la segunda parte del análisis que es el análisis global de todas las variables con la ponderación mencionada.

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.

Si sólo se realizara la primera parte del AFM, se estaría estudiando la variable en cada momento de tiempo por separado, es decir, se realizaría un estudio transversal de la variable longitudinal, en diferentes momentos.

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.

El AFM proporciona representaciones gráficas y medidas cuantitativas que ayudan a su interpretación (Escofier & Pages 1992, Cap.4):

- Plano factorial de las correlaciones de las variables con los factores, donde es posible darle un significado a los ejes factoriales.
- Nubes parciales de los individuos caracterizados por un grupo de variables.

-
- Representación, en un mismo espacio, de los individuos caracterizados por el conjunto de variables y por cada uno de los grupos: nubes parciales y nube media o global donde se pueden observar los individuos a través de todas las variables (o las trayectorias si es el caso de variables longitudinales).
 - Todas las ayudas a la interpretación que proporcionan los métodos factoriales (cosenos cuadrados, coordenadas de individuos y variables, contribuciones).

2.3.1. AFM y el análisis multicanónico

El AFM se puede ver como un análisis multicanónico (Escofier & Pages 1992) en el sentido de que tiene como uno de sus objetivos la búsqueda de los factores comunes a todos los grupos estudiados. El principio del análisis multicanónico consiste en buscar variables, estandarizadas y no correlacionadas, denominadas generales (variables fuertemente relacionadas con el conjunto de los grupos de variables). Para cada variable general, y para cada conjunto de variables de un grupo, se obtienen las denominadas variables canónicas (combinación lineal de las variables de un grupo más relacionadas con la variable general).

En este sentido, los factores globales del AFM se pueden considerar como las variables generales de un análisis multicanónico y los factores obtenidos del ACP de las nubes parciales se pueden interpretar como las variables canónicas.

Para realizar el análisis multicanónico se debe definir una medida de relación (notada L_g) entre una variable continua, denominada generalmente z (variable general), y las variables del grupo t , las cuales se denominarán $p(t)$.

La medida de relación utilizada por Carroll (1968) es el cuadrado del coeficiente de correlación múltiple. Por definición este coeficiente es el coeficiente de correlación entre la variable z y la combinación de las variables del grupo t más correlacionada con z .

En el análisis factorial múltiple esta medida de relación estaría definida de la siguiente manera:

$$L_g(z, p_t) = \text{inercia de todas las variables del grupo } p_t \text{ proyectadas sobre } z$$

2.3.2. Relación entre los grupos de variables

Además, el AFM proporciona medidas globales de relación entre los grupos. Ellas están basadas en el coeficiente RV , definido para dos grupos cualesquiera t y h :

$$RV(t, h) = \frac{\langle \mathbf{W}_t \mathbf{D}, \mathbf{W}_h \mathbf{D} \rangle}{\|\mathbf{W}_t \mathbf{D}\| \|\mathbf{W}_h \mathbf{D}\|} \quad (2.3)$$

y los coeficientes Lg , ya definidos, que quedarían (Escofier & Pages 1992):

$$Lg(t, h) = \langle \mathbf{W}_t \mathbf{D}, \mathbf{W}_h \mathbf{D} \rangle \quad (2.4)$$

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, ya mencionados.

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). Esta medida se completa con los coeficientes Lg (2.4) que 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.

CAPÍTULO 3

Combinación del AAC y el AFM: AAC-AFM

3.1. Aspectos generales de la combinación

Sea \mathbf{A} una tabla de código difuso, formada por la yuxtaposición de \mathbf{A}_t tablas (definidas en la sección 2.1 del capítulo 2), donde cada una de ellas contiene las modalidades m_t de una variable longitudinal cualitativa en diferentes periodos de tiempo t ($t = 1, \dots, T$), con n filas referidas a los individuos o unidades estadísticas, y m columnas, que corresponden a las categorías de la variable longitudinal tomadas en T períodos de tiempo, donde un individuo i puede tomar, en cada período t , las categorías de la variable con diferente grado de asociación. La forma general de esta tabla se presenta en la sección 2.1, junto con las propiedades que ella posee.

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.

Por lo tanto el AFM, realizado sobre la tabla \mathbf{P} , se ejecuta en dos etapas (Escofier & Pages 1992, Bécue & Pagès 2007):

- 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) . El AC de la tabla \mathbf{P}_t equivale al ACP ponderado, denotado $\text{ACP}(\mathbf{X}_t, \mathbf{M}_t, \mathbf{D})$ donde:

- \mathbf{X}_t es la matriz a analizar de término general: $x_j = \frac{p_j^t}{\bar{p}_j^t} - 1$ (donde \bar{p}_j^t está definido en 2.2)
- $\mathbf{M}_t = \text{diag} \left(\frac{1}{T} \bar{p}_j^t \right) = \frac{1}{T} \text{diag}(\bar{p}_j^t)$
- $\mathbf{D} = \text{diag} \left(\frac{1}{n} \right) = \frac{1}{n} \mathbf{I}_n$

- 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 λ_1^t de la tabla \mathbf{P}_t a la cual pertenece. Esta segunda fase del AFM se logra realizando un $\text{ACP}(\mathbf{X}, \mathbf{M}, \mathbf{D})$ donde:

- \mathbf{X} es la matriz a analizar de término general: $x_{ij} = \frac{p_{ij}^t}{\bar{p}_j^t} - 1$
- $\mathbf{M} = \text{diag} \left(\frac{1}{T} \frac{\bar{p}_j^t}{\lambda_1^t} \right) = \frac{1}{T} \text{Diag} \left(\frac{\bar{p}_j^t}{\lambda_1^t} \right)$
- $\mathbf{D} = \text{diag} \left(\frac{1}{n} \right) = \frac{1}{n} \mathbf{I}_n$

3.2. Propiedades de la combinación del AAC y el AFM

Las siguientes propiedades 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.

3.2.1. Distancias entre individuos y categorías

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:

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

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 períodos 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 períodos de tiempo respectivos t y t' es:

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

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

3.2.2. 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 (ver Anexo A):

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

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 & Pages 1992), 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 período de tiempo t , solamente:

$$F_s^t(i) = \frac{1}{\sqrt{\lambda_s}} \frac{1}{\lambda_1^t} \sum_{j=1}^{m_t} \frac{p_{ij}^t}{T} G_s(p_t) = \frac{1}{\sqrt{\lambda_s}} \frac{1}{T} \frac{1}{\lambda_1^t} \sum_{j=1}^{m_t} p_{ij}^t G_s(p_t) \quad (3.1)$$

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 2007).

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. Analizando 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 períodos de tiempo.

3.2.3. 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 de la 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:

$$I = \frac{1}{nT} \sum_{t=1}^T \frac{1}{\lambda_1^t} \sum_{j=1}^{m_t} \overline{p_j^t} \sum_{i=1}^n \left(\frac{p_{ij}^t}{\overline{p_j^t}} - 1 \right)^2 = \frac{1}{nT} \sum_{t=1}^T \sum_{j=1}^{m_t} \sum_{i=1}^n \frac{1}{\lambda_1^t} \frac{(p_{ij}^t - \overline{p_j^t})^2}{\overline{p_j^t}} \quad (3.2)$$

En (3.2) 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.

Por lo tanto, de acuerdo a lo definido en la sección 3.1 la medida del AAC-AFM estaría dada por:

$$L(z, P_t) = \frac{1}{T} \sum_t \left(\frac{\overline{p_j^t}}{\lambda_1^t} \right) r^2 \left(z, \left(\frac{p_{ij}^t}{\overline{p_j^t}} - 1 \right) \right) \quad (3.3)$$

El primer factor del AFM maximiza la inercia proyectada de todos los grupos de variables, es decir, maximiza $\sum L(z, P_t)$. En este sentido, los factores del AFM (notados como F_s y definidos en el análisis factorial general en el anexo A) se pueden considerar como las variables generales del análisis multicanónico (Escofier & Pages 1992).

De igual forma los factores parciales para cada período t , F_s^t , definidos anteriormente, serán las variables canónicas, de acuerdo a como son definidas éstas en el análisis multicanónico.

3.2.4. 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 , de dimensión $n \times n$ definido en el capítulo anterior es adecuado para representar las variables del período t . Cada una de estas matrices \mathbf{W}_t puede representarse por un punto en R^{n^2} ; es

decir, en este espacio se tiene la nube de los T períodos totales de tiempo, denominada \mathbf{N}_T .

Esta nube se proyecta sobre el subespacio inducido por las variables generales de rango s , z_s . La coordenada del grupo t a lo largo de z_s es igual a la medida de relación definida en 3.3, que es también la contribución del grupo t a la inercia del eje s (Pagés 2002).

De esta manera tiene sentido realizar un producto escalar entre dos grupos. Sean h y t dos grupos cualesquiera. Para el caso del AAC-AFM este producto escalar quedaría:

$$\langle \mathbf{W}_t \mathbf{D}, \mathbf{W}_h \mathbf{D} \rangle = \sum_t \left(\frac{\overline{p_j^t}}{T \lambda_1^t} \right) \sum_h \left(\frac{\overline{p_j^h}}{T \lambda_1^h} \right) \langle \mathbf{P}_t \mathbf{P}_t' \mathbf{D}, \mathbf{P}_h \mathbf{P}_h' \mathbf{D} \rangle \quad (3.4)$$

Este producto escalar (3.4), también conocido como coeficiente L_g , mencionado en el capítulo 2, 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.

CAPÍTULO 4

Aplicación de la combinación del AAC y el AFM

4.1. Descripción del rating de TV durante un día en Colombia

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

Los datos de teleaudiencia (*rating*) fueron suministrados por el profesor Juan Ramos, codirector de la tesis. 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 registra minuto a minuto la sintonía de 584 hogares, conformados en total por 2334 televidentes, información que compilada y procesada permite 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).

Los hogares que conforman el panel, se seleccionan en forma aleatoria y probabilística, por la empresa encargada de esta labor. Una vez seleccionados los hogares, un equipo de representantes en las seis zonas del país, contacta a las personas, con el fin de firmar un acuerdo de colaboración que permite instalar los *People Meter* e iniciar el registro de su audiencia de televisión.

El *People Meter* consiste en una microcomputadora pequeña que registra directamente del televisor, los canales sintonizados durante las 24 horas del día. Este aparato graba automáticamente la audiencia desde el momento en que se enciende el televisor. Para registrar el canal que una persona está viendo o deja de ver, ésta sólo tiene que oprimir un botón en el control remoto que se entrega el día de la instalación. La información de audiencia se transmite a la memoria de la microcomputadora, y luego, por el mismo aparato, al centro de recepción, mediante una comunicación telefónica nocturna y/o vía Internet, con el fin de no interferir con las necesidades de los hogares.

4.2. Fase de preprocesamiento de los datos

La fase de procesamiento de datos comprende las tareas necesarias para adaptar la información, a los métodos que se van a aplicar, de acuerdo a los objetivos propuestos. En esta fase se realizaron, por lo tanto, algunos arreglos a los datos originales.

TABLA 4.1. Cambios de canal de los individuos (dos primeros individuos)

Zona	Individuo	Canal	Hora de inicio	Hora finalización
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	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	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.

4.2.1. Complementos y ajustes de la información

La información proporcionada inicialmente (tabla 4.1) indica el tiempo de permanencia de cada individuo en los diferentes canales de televisión. El tiempo del día considerado inicia a las 6 de la mañana y finaliza a las 2 de la mañana del día siguiente.

Los tiempos en los cuales el televisor se encuentra apagado se considera un nuevo canal de televisión. Como este tiempo no está registrado en la tabla, se incorporó a ella, como una nueva categoría de la variable longitudinal (ver tabla 4.2).

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. 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) como se muestran en la tabla

TABLA 4.2. Cambios de canal de los individuos completos (tres primeros individuos)

Zona	Individuo	Canal	Hora de inicio	Hora 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.
1	110001001	1000	06:00:00 a.m.	06:55:59 a.m.
1	110001001	0031	06:56:00 a.m.	07:01:59 a.m.
1	110001001	1000	07:02:00 a.m.	09:04:59 a.m.
1	110001001	0085	09:05:00 a.m.	01:00:59 p.m.
1	110001001	1000	01:01:00 p.m.	07:04:59 p.m.
1	110001001	0085	07:05:00 p.m.	07:06:59 p.m.
1	110001001	0197	07:07:00 p.m.	07:17:59 p.m.
1	110001001	1000	07:18:00 p.m.	08:55:59 p.m.
1	110001001	0031	08:56:00 p.m.	08:58:59 p.m.
1	110001001	1000	08:59:00 p.m.	09:50:59 p.m.
1	110001001	0031	09:51:00 p.m.	09:54:59 p.m.
1	110001001	1000	09:55:00 p.m.	02:00:00 a.m.

TABLA 4.3. Franjas de audiencias de la televisión colombiana

Franja	Tipo de Audiencia
06:00 -09:30	Familiar
09:30 -12:00	Familiar
12:00-13:00	Familiar
13:00 -14:00	Familiar
14:00 -16:00	Familiar
16:00 -17:00	Infantil
17:00 -18:00	Familiar
18:00 -19:00	Familiar
19:05 -22:05	Familiar
22:10 -22-40	Adultos
22:40 -23:10	Adultos
23:10 -00:10	Adultos

TABLA 4.4. Franjas de audiencia ajustadas

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

4.3. 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 observa en la tabla 4.4 .

Los canales de televisión existentes también se reagrupan. El número de posibles canales elegibles entre los televidentes asciende a un total de 469 (ver Apéndice B). 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 conservan aquellos con alto porcentaje de cambios y se realizan agrupaciones con los demás. Estas agrupaciones fueron realizadas y renombradas teniendo en cuenta la organización realizada por las diferentes empresas prestadoras de servicios de televisión y a las características comunes entre ellos (ver tabla 4.5).

Del total de los individuos de la muestra (2334), sólo 1320 tuvieron encendi-

TABLA 4.5. 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

do 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 1320 individuos y 3986 cambios de canales.

4.2.2. Limpieza de los datos

Adicionalmente, se cuenta con aspectos adicionales de los individuos como la zona en la cual viven, el peso (factor de expansión de la muestra) dado a cada individuo, el nivel socioeconómico, la edad y el género (tabla 4.6). Estas variables (excepto el peso), para efectos del estudio, se consideran ilustrativas, y su labor será la de complementar la caracterización de las clases de televidentes obtenidas.

TABLA 4.6. 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	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
110028201	Hombre	40 a 64	Medio	1	0.10014
110029405	Mujer	12 a 17	Bajo	1	0.08831
110039201	Hombre	25 a 39	Bajo	1	0.10619
220006302	Hombre	18 a 24	Bajo	2	0.04376

TABLA 4.7. Fracción del archivo con los canales vistos minuto a minuto (primeros 6 minutos)

Individuo	Minuto 1	Minuto 2	Minuto 3	Minuto 4	Minuto 5	Minuto 6
110000301	15	15	15	15	15	15
110000701	15	15	15	15	15	15
110001001	15	15	15	15	15	15
110002201	4	4	4	4	4	4
110002801	5	5	5	5	5	5
110002803	15	15	15	15	15	15

4.2.3. Transformación de los datos

El archivo que contiene los cambios de canal realizados por los individuos, fue nuevamente reestructurado, para obtener la tabla 4.7, que indica el canal en el cual se encontraba el individuo en todos los minutos del día (del minuto 1 al minuto 1200).

A partir de la tabla 4.7, se obtuvo la tabla **A**, mencionada en el capítulo 2, la cual contiene el número de minutos que cada individuo permaneció en cada canal de las 6 franjas de tiempo. Una fracción de la tabla **A** puede verse en la tabla 4.8. La notación F1C1 (franja-canal), significa franja 1, canal 1 y así para el total de canales considerados (15) y la cantidad de franjas (6) estructuradas.

TABLA 4.8. Fracción del archivo con el total de minutos vistos por canal y franja

Individuo	F1C1	F1C2	F1C3	F1C4	F1C5	F1C6	F1C7	F1C8	F1C9
11000301	0	0	0	34	1	0	0	0	0
11000701	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
110002501	0	0	0	0	0	0	0	0	0
110002502	0	0	0	0	0	0	0	0	0
110002801	0	6	0	214	30	1	0	1	0
110002803	0	0	0	0	0	0	0	0	0
110002804	0	0	0	349	0	0	0	0	0
110002901	0	6	0	1	3	0	0	0	0
110002903	0	0	0	0	0	1	0	0	0
110003001	0	0	159	0	84	0	0	0	0

4.3. Análisis factoriales realizados

4.3.1. Análisis armónico cualitativo (AAC)

La tabla **A**, de dimensión 1320 (individuos) x 90 (columnas=15 canales x 6 franjas), fue sometida al AAC, realizado mediante un análisis de correspondencias difuso (ACD). Este análisis de correspondencias difuso fue ejecutado utilizando el paquete estadístico *R* (R Development Core Team 2007) por medio de la librería *ade4* (Chessel et al. 2004). En esta fase se tuvo en cuenta el factor de expansión o peso de los individuos de la muestra. Se realizaron dos AAC, 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.

Luego de realizar el AAC de los datos, algunos individuos (14) fueron excluidos del análisis ya que presentaban características especiales (cinco permanecieron todo el tiempo alternando entre los canales 9 y 10; dos de ellos permanecían en el canal 7 todo el día; cuatro permanecían en los canales 4 y 5 gran parte del tiempo; y, tres, veían muy poca televisión, y el tiempo restante permanecían en el canal 2) y por lo tanto formaban ejes factoriales. Después de excluir estos 14 individuos, la tabla final quedó conformada por 1306 filas y 90 columnas. Esta última tabla fue la base de los análisis factoriales realizados.

Los resultados del AAC muestran un gráfico de valores propios (figura 4.1) donde predominan los 6 primeros factores. La interpretación de estos factores puede apreciarse en los planos factoriales (figura 4.2 y Apéndice H. Calidad de representación de los planos: 10%). Estos son determinados básicamente por los canales 4 (RCN), 5 (Caracol) y 14 (otros canales internacionales). El primer plano muestra la estructura evidenciada por estos primeros ejes: los canales son agrupados y ordenados cronológicamente a través de las franjas.

Para lograr la tipología completa se realiza 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), enunciado en el capítulo 2.

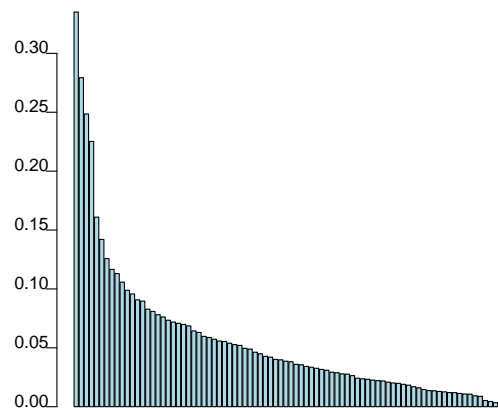


FIGURA 4.1. Primeros valores propios del AAC

Después de realizar varios ensayos, variando el número de ejes desde 4 hasta 20 y de acuerdo al criterio enunciado, se escogieron 6 ejes para la clasificación. El número de clases (8) se obtuvo cortando el árbol de clasificación (figura 4.3). Se consideró, además, una clase adicional conformada por los individuos que no vieron televisión el día observado, mencionados anteriormente.

En el apéndice C se presenta una breve descripción de las clases obtenidas. En los apéndices D y E se encuentran los gráficos con los perfiles de las clases a través de las variables ilustrativas y de la variable longitudinal activa (cambios de canales).

En la figura 4.4 se muestran los individuos con las clases a las cuales pertenecen sobre el primer plano factorial del AAC, lo cual permite tener una síntesis gráfica de las clases obtenidas.

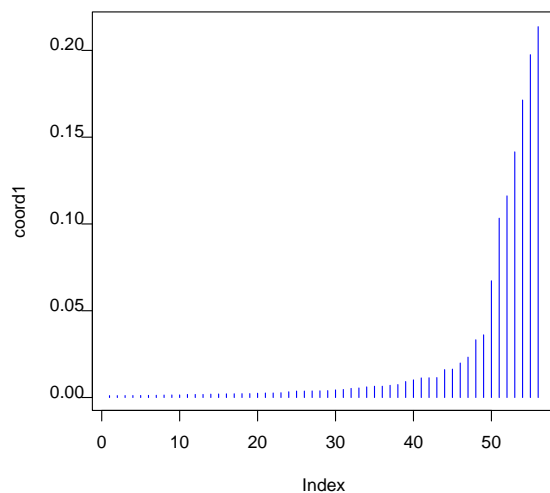


FIGURA 4.3. Histograma de índices de nivel. Clasificación con AAC. Un corte a la altura de 0.5 origina 8 clases

Análisis de cada franja por separado

Los 6 análisis de correspondencias realizados para cada franja dan como resultado diferencias en el primer valor propio (ver tabla 4.9), lo que justifica la ponderación del AFM, que estandariza esos valores propios en 1 para el análisis global.

TABLA 4.9. Primeros cinco valores propios de los análisis separados

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

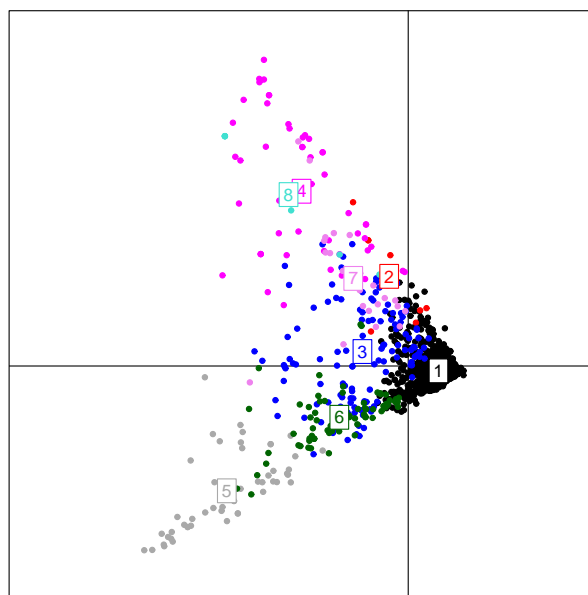


FIGURA 4.4. Nube de los individuos con las clases a las cuales pertenecen. Clasificación AAC.

Análisis global de las franjas

Para analizar la semejanza global entre las franjas en los 6 períodos de tiempo considerados se utilizan coeficientes, ayudas a la interpretación (tabla 4.10) y gráficos propios del AFM (figuras 4.5 y 4.7).

TABLA 4.10. Coordenadas y ayudas a la interpretación de los grupos activos. Ejes 1 a 5

Grupos	Coordenadas					Contribuciones					Cosenos cuadrados				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
F1:Famil	0.24	0.17	0.14	0.18	0.09	08.4	06.9	06.5	09.4	06.1	0.01	0.01	0.00	0.01	0.00
F2:Fami2	0.62	0.44	0.44	0.42	0.28	21.4	18.2	20.6	22.3	19.8	0.09	0.00	0.00	0.04	0.02
F3:Infan	0.69	0.60	0.56	0.51	0.37	24.0	24.6	26.6	26.5	26.2	0.09	0.07	0.06	0.05	0.03
F4:Prime	0.69	0.54	0.54	0.44	0.15	23.8	22.1	25.5	23.3	11.0	0.10	0.06	0.06	0.04	0.01
F5:Fami3	0.46	0.37	0.27	0.28	0.13	15.8	15.3	12.5	14.5	09.2	0.05	0.00	0.00	0.02	0.00
F6:Adult	0.19	0.31	0.18	0.07	0.39	06.5	12.9	08.4	04.0	27.8	0.01	0.00	0.01	0.00	0.03

La proyección de las franjas sobre el primer plano factorial (figura 4.5) 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 si.

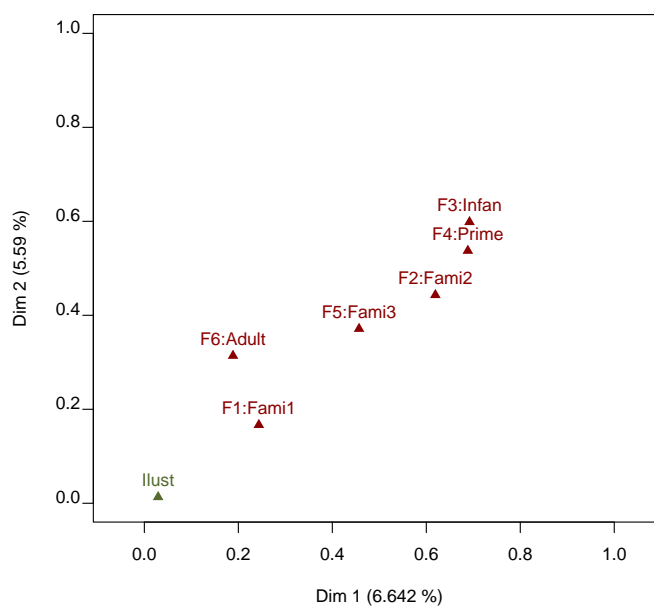


FIGURA 4.5. Franjas en el primer plano factorial

La diagonal de la matriz de coeficientes Lg (tabla 4.11) 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 4.6) se destacan 5 ejes. La matriz de coeficientes RV (tabla 4.12) 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 (ver también figura 4.7).

Comparación de las franjas

Se realiza un análisis detallado de las semejanzas y diferencias entre los elementos, individuos y variables de cada una de las franjas.

En el círculo de correlaciones (franja-canal) de todas las franjas (figura 4.7),

TABLA 4.11. Coeficientes Lg

	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

TABLA 4.12. Coeficientes Rv

	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

se observa la alta correlación entre el canal 15 (televisor apagado) en todas las franjas con el primer factor del AAC-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).

Uno de los principales instrumentos para el análisis de la evolución en el tiempo de la selección de los canales de los individuos es la proyección en los planos factoriales de ellos desde un punto de vista medio (resumen de la información de las 6 franjas) y desde un punto de vista parcial (la posición relativa de cada individuo en cada momento considerado).

Al unir los puntos parciales de un individuo se obtiene su trayectoria la cual permite explicar su evolución (figura 4.8). En esta figura solamente se han

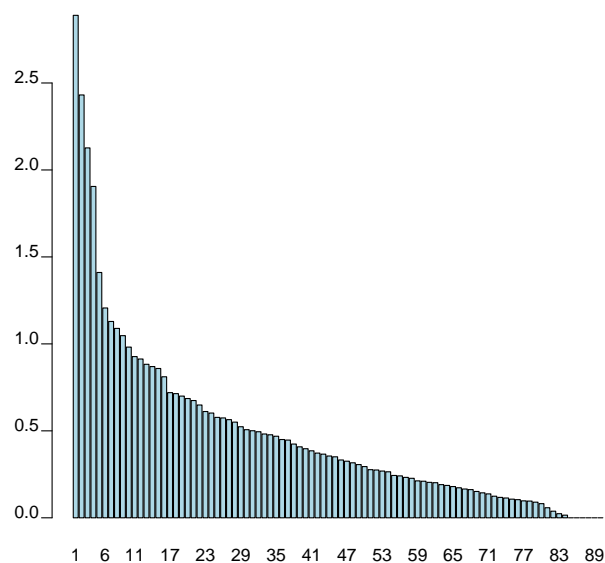


FIGURA 4.6. Primeros valores propios de la combinación del AAC y el AFM

representado las trayectorias de los individuos que presentaban una inercia intra más alta en el primer plano factorial (ver sección 3.2, página 16). Esto es, aquellos individuos cuyos puntos parciales están más separados de su respectivo punto medio, lo cual indica que han experimentado una mayor evolución. En esta figura se muestran también algunas categorías de las variables ilustrativas. Sin embargo, debido a que por medio de la gráfica es complicado realizar la caracterización de todos los individuos, se efectuó de un análisis de clasificación.

4.4. Clasificación de los televidentes

Mediante el AAC-AFM y la clasificación posterior se generaron 8 clases, cada una con características de teleaudiencias diferentes. Adicionalmente, se considera una nueva clase formada por aquellos individuos que no prendieron el televisor en el día de estudio.

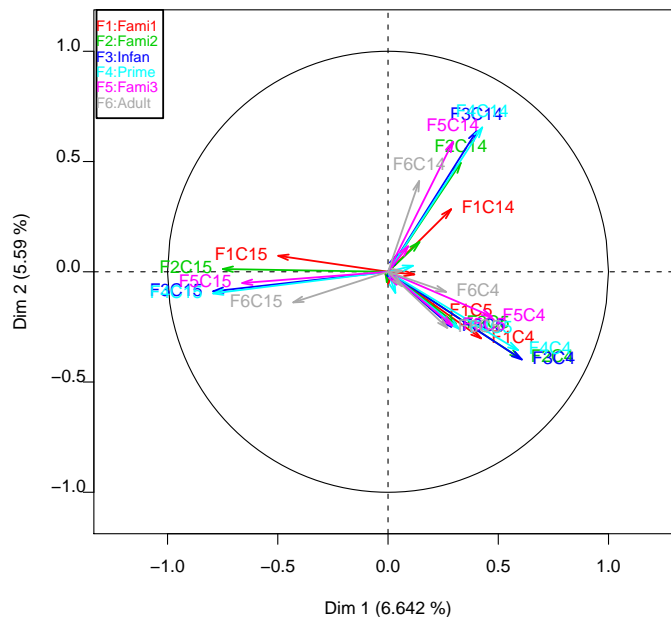


FIGURA 4.7. Círculo de correlación de las franjas-canal del análisis combinado

Para obtener las características especiales de cada una de las clases, como se mencionó anteriormente, 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. Nuevamente se usó el criterio propuesto por Barbary (1996) para decidir el número de ejes.

Se hicieron varios ensayos para el número de factores (tabla 4.13) seleccionando 6 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 vez seleccionado el número de clases (ver por ejemplo Lebart et al. (1995)). El número de clases está guiado por los saltos del histograma de índices de nivel (figura 4.9): se tomaron cuatro particiones en 5, 7, 8 y

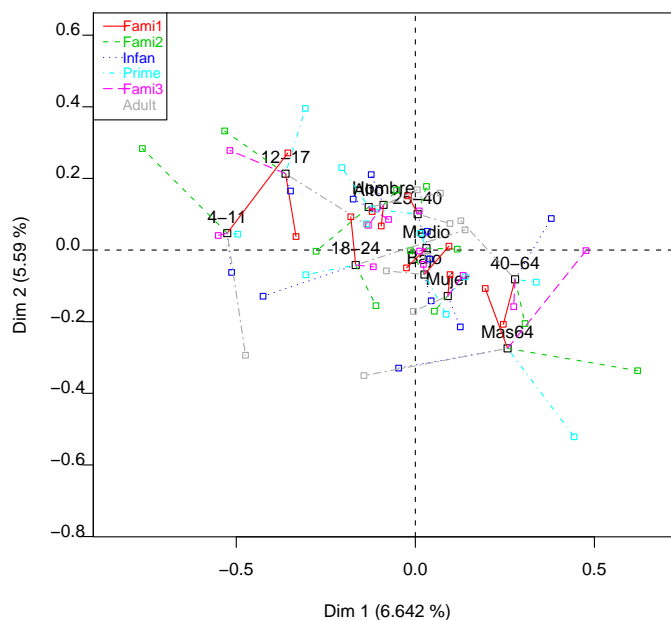


FIGURA 4.8. Evolución de los individuos con inercia intra alta, a través de las franjas de audiencia, con algunas categorías de las variables ilustrativas

11 clases y se decidió describir la partición en 8 clases, ya que es la misma partición seleccionada en el AAC, y de esta forma se puede realizar una comparación entre las clases obtenidas con ambos métodos. Se tiene también la clase formada por los individuos que no vieron televisión el día estudiado.

4.5. Descripción de las clases obtenidas

Los televidentes de la muestra, se repartieron entonces en 8 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 (tabla 4.14), pero se caracterizan también con las variables ilustrativas consideradas: género, estratos socioeconómicos, rangos

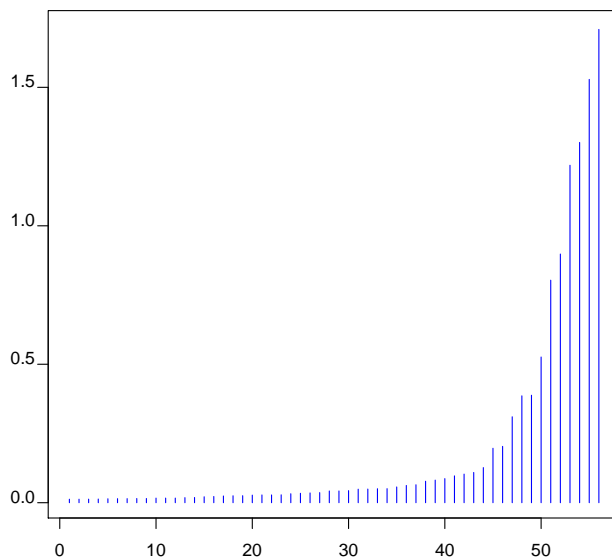


FIGURA 4.9. Histograma de índices de nivel. Nuevamente, un corte a la altura de 0.5 origina 8 clases

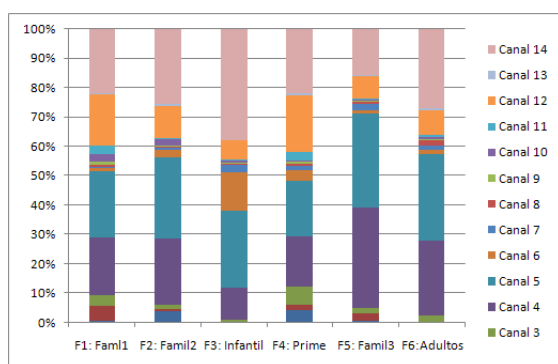
de edad y zonas del país (tablas 4.15, 4.16 y 4.17), teniendo en cuenta que la clase 9 es la conformada por los individuos que no vieron televisión el día del estudio y por lo tanto al no participar en los análisis factoriales, no se reporta información de esta clase en la tabla 4.14.

4.5.1. Caracterización de las clases

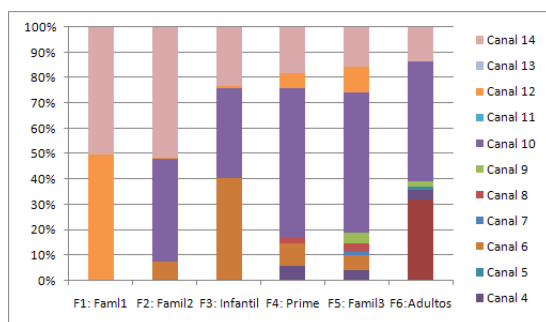
Las tablas 4.14 y 4.17 que ayudan a la descripción de las clases a través de los cambios de canal y variables ilustrativas, con sus respectivos valores test (Lebart et al. (1995), cap. 2) fueron obtenidas con la función *cluster.carac*, de la librería en *R*, *FactoClass*, (Pardo & Del-Campo 2007).

En el gráfico 4.10 se muestran los individuos con las clases a las cuales pertenecen sobre el primer plano factorial del AFM y AAC combinados, lo cual permite tener una síntesis gráfica de las clases.

Los gráficos que acompañan la descripción de las clases se obtuvieron eliminando el canal 15 (televisión apagada) para facilitar la lectura de la selección de los canales en el tiempo que el televisor estuvo encendido. En los apéndices F y G se pueden ver los perfiles gráficos de las clases por las variables ilustrativas así como los gráficos que incluyen el canal 15 en cada una de las clases.



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.



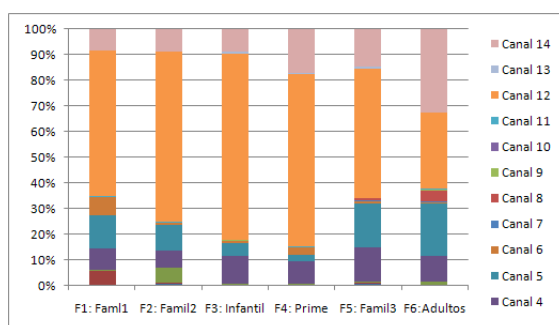
(infantiles internacionales) y 14 (cinemas y otros canales internacionales). En la franjas 2, 3 y 4, se inclinan también por los canales 6 (otros canales

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.

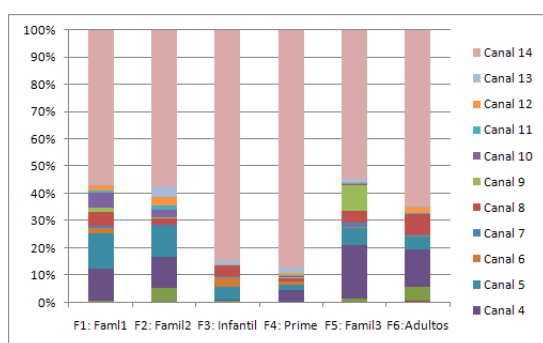
Los canales seleccionados por

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 primera, por el canal 10 (Teleantioquia). En la franja 1 (durante el poco tiempo que permaneció encendido el televisor), prefieren los canales 12

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.



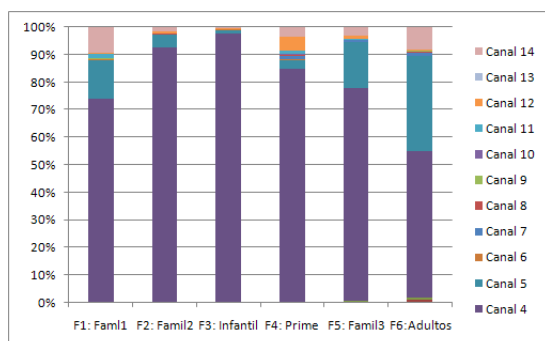
Se encuentra caracterizada también por individuos de nivel socioeconómico bajo, del centro y oriente del país con edades entre los 4 y 11 años, y 25 a 64 años.



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.

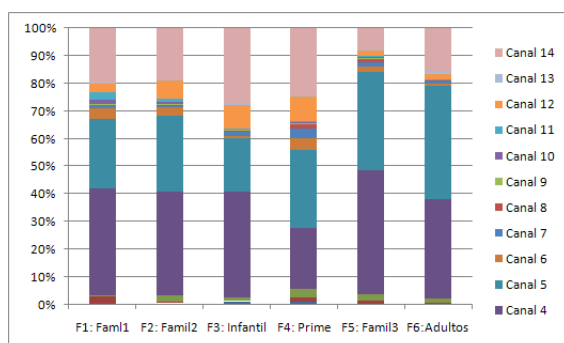
Clase 3 (4.1 %). Formada por televidentes que prefieren principalmente el canal 12 (infantiles internacionales tales como *Cartoon Network*, *Discovery Kids*, *Disney Chanel*, etc). En menor proporción seleccionan los canales 4 (RCN), 5 (Caracol) y 14 (cinemas (*HBO*, *Cinemax*, etc.) y otros canales internacionales). Se

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*,



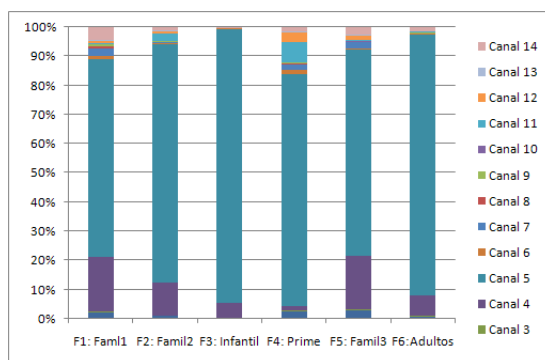
40 a 64 años de niveles bajo y medio pertenecientes en gran proporción a la zona central de Colombia.

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

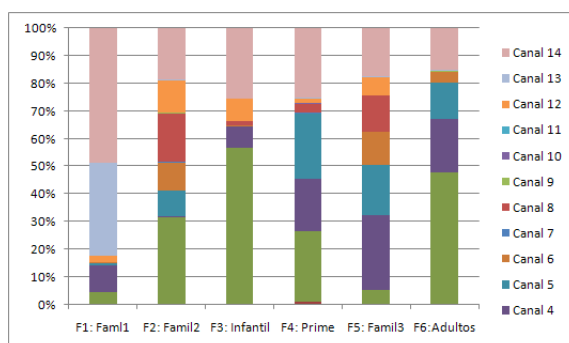


de clase media, con edades entre los 25 y 64 años de las zonas centro, pacífica y oriental principalmente.

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



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 so-



cioeconómicos 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. Esta fue descrita anteriormente en la clasificación realizada con el análisis armónico cualitativo.

TABLA 4.13. Ensayos de ayuda a la clasificación. Se escogieron 6 ejes para realizar la clasificación, ya que con este número se obtiene el menor número de individuos en la clase mayoritaria.

Ejes	Individuos	Escogido
16	931	0
15	1079	0
14	1005	0
13	1067	0
12	1032	0
11	998	0
10	1085	0
09	1037	0
08	919	0
07	863	0
06	750	1
05	892	0
04	796	0
03	907	0
02	755	0

TABLA 4.14. 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	Porcentaje global
F4C15	1	24.9	93.7	76.2
F3C15	1	24.3	98.5	80.4
F5C15	1	22.3	81.4	63.8
F2C15	1	21.0	93.3	80.4
F6C15	1	12.7	90.6	84.9
F4C10	2	25.2	36.2	0.2
F5C10	2	23.7	20.4	0.2
F3C10	2	19.5	19.4	0.1
F6C10	2	13.8	11.6	0.1
F3C6	2	10.2	22.2	0.4
F3C12	3	25.2	48.3	2.5
F4C12	3	21.1	40.6	3.4
F2C12	3	21.1	30.5	2.3
F5C12	3	19.4	30.0	2.5
F1C12	3	11.9	13.2	1.4
F4C14	4	26.3	71.1	6.1
F3C14	4	25.7	67.1	4.8
F5C14	4	20.0	44.7	5.2
F2C14	4	17.3	30.5	3.6
F1C14	4	12.3	16.7	2.5
F6C14	4	10.7	20.2	3.5
F3C4	5	28.3	93.1	5.8
F4C4	5	27.0	68.8	5.2
F2C4	5	26.1	72.0	6.2
F1C4	5	16.4	32.6	3.5
F5C4	5	13.2	60.0	13.1
F6C4	5	6.9	14.5	3.9
F5C4	6	10.7	29.7	13.1
F5C5	6	8.6	23.6	11.6
F4C5	6	4.8	11.0	5.8
F2C4	6	4.4	11.0	6.2
F3C4	6	3.4	10.3	5.8
F3C5	7	28.1	70.7	4.9
F4C5	7	24.4	57.9	5.8
F2C5	7	21.5	45.8	5.4
F5C5	7	14.9	53.0	11.6
F6C5	7	14.7	28.2	5.3
F1C5	7	9.6	14.7	2.8
F3C3	8	28.0	50.6	0.6
F6C3	8	23.1	30.1	0.5
F2C8	8	17.1	11.2	0.2
F2C3	8	15.5	20.1	0.6
F4C3	8	13.1	18.9	0.7
F3C14	8	3.4	22.5	4.8

TABLA 4.15. Perfiles de las clases por género, nivel socio-económico y rango de edad

Clase	Género		Nivel socioeconómico			Rango de edad					
	Hombre	Mujer	Bajo	Medio	Alto	4a11	12a17	18a24	25a39	40a64	mas64
1	51.8	48.2	35.1	46.6	18.3	14.7	12.1	10.4	24.2	32.6	6.0
2	50.0	50.0	0.0	66.7	33.3	0.0	50.0	16.7	0.0	33.3	0.0
3	49.1	50.9	47.2	32.1	20.8	26.4	9.4	3.8	22.6	34.0	3.8
4	57.7	42.3	30.8	46.2	23.1	3.8	11.5	9.6	26.9	42.3	5.8
5	37.5	62.5	39.6	45.8	14.6	4.2	2.1	8.3	16.7	60.4	8.3
6	46.4	53.6	32.7	49.1	18.2	5.9	9.5	9.1	22.3	41.4	11.8
7	47.6	52.4	44.4	46.0	9.5	3.2	6.3	12.7	27.0	39.7	11.1
8	75.0	25.0	16.7	75.0	8.3	0.0	8.3	8.3	41.7	41.7	0.0
9	43.9	56.1	33.5	48.0	18.4	16.7	12.7	20.0	23.0	22.3	5.3

TABLA 4.16. Perfiles de las clases según zonas del país

Clase	-Zonas-					
	1:Centro	2:Antioquia	3:Pacífico	4:Caribe	5:Cafetera	6:Oriental
1	24.53	11.97	18.31	17.49	11.38	16.31
2	33.33	0.00	0.00	33.33	33.33	0.00
3	39.62	7.55	7.55	7.55	15.09	22.64
4	26.92	13.46	13.46	3.85	36.54	5.77
5	43.75	2.08	10.42	14.58	14.58	14.58
6	28.64	9.55	18.18	14.55	10.45	18.64
7	52.38	4.76	4.76	25.40	3.17	9.52
8	0.00	0.00	50.00	0.00	41.67	8.33
9	26.23	12.52	16.07	15.19	13.12	16.86

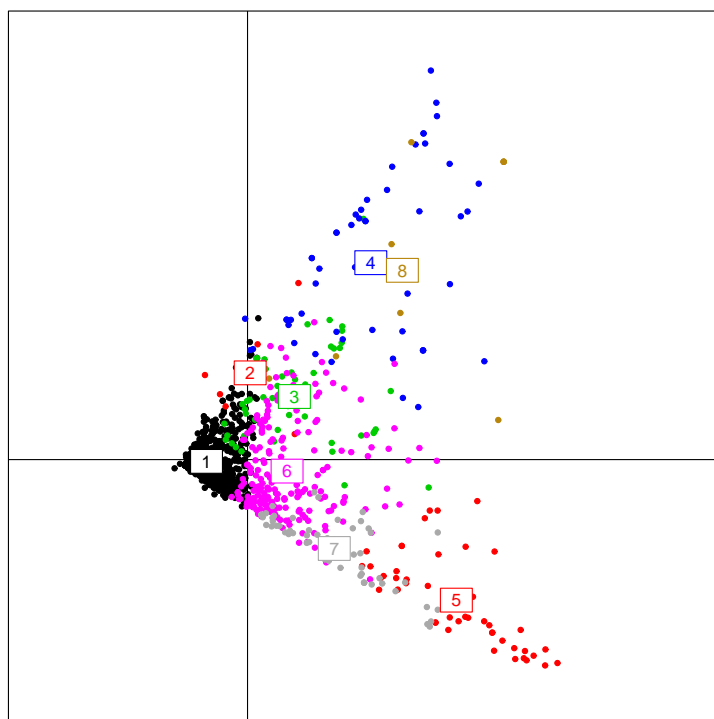


FIGURA 4.10. Nube de los individuos con las clases a las cuales pertenecen. Clasificación combinación AAC y AFM

4.6. Comparación de las clases obtenidas con ambos métodos

En la tabla 4.18 se muestra el número de individuos que pertenecen a cada una de las clases obtenidas con el análisis armónico cualitativo y con la combinación del análisis factorial múltiple y el análisis armónico cualitativo.

De acuerdo a la descripción de las clases halladas con el AAC y con el AAC-AFM, y con lo observado en la tabla 4.18 se notan las siguientes diferencias y similitudes entre las clases obtenidas por ambos métodos:

- Clase 1: ambas clases presentan características similares, con respecto a la selección de los canales. Se diferencian ligeramente en la proporción de

TABLA 4.17. Caracterización de las clases según las variables ilustrativas

Clase	Mod. var. Ilus.	Valor test	Clas.Mod	Mod.clase	Global	Peso
1	Edad: 4 a 11	3.9	79.1	14.7	12.1	158.0
1	Edad: mas 64	-2.0	54.8	6.0	7.1	93.0
1	Edad: 40 a 64	-3.4	59.1	32.6	36.0	470.0
2	Edad: 12 a 17	2.0	2.1	50.0	11.0	144.0
3	Edad: 4 a 11	2.8	8.9	26.4	12.1	158.0
3	Estrato Medio	-2.1	2.8	32.1	46.7	610.0
5	Edad: 40 a 64	3.4	6.2	60.4	36.0	470.0
6	Edad: mas 64	2.7	28.0	11.8	7.1	93.0
6	Edad: 4 a 11	-3.2	8.2	5.9	12.1	158.0
7	Edad: 4 a 11	-2.2	1.3	3.2	12.1	158.0

tiempo que pasan en estos canales.

- Clase 2: las dos clases presentan la misma composición en la franja número 1. En las siguientes franjas la distribución de los canales presentan una selección similar de los canales escogidos. Sin embargo en la combinación es más notorio la preferencia por el canal 10 (Teleantioquia) que en la clasificación con el AAC.
- Clase 3: la clase 3 del AAC es comparable con la clase 6 de la combinación. Los individuos pertenecientes a estas clases prefieren los canales 4 (RCN), 5 (Caracol), 12 (infantiles internacionales) y 14 (cinemas y otros canales internacionales) sobre todos los demás. Sin embargo, se pueden apreciar diferencias en ambas clases con respecto a la proporción de tiempo pasado en estos canales en todas las franjas.
- Clase 4: las clases obtenidas por el AAC y con la combinación presentan una selección de canales casi iguales en todas las franjas, destacándose la preferencia por el canal 14 (otros canales internacionales) sobre los demás canales.
- Clase 5: en ambas clases se observa un favoritismo del canal 4 (RCN) sobre todos los demás. Pero también se destacan diferencias entre la elección de otros canales, con menor proporción, en todas las franjas de las dos clasificaciones. La preferencia del canal 4 es más marcada en la combinación que en el AAC.

TABLA 4.18. Número de individuos en cada clase con los dos métodos

AFM con AAC/AAC	Clase1	Clase2	Clase3	Clase4	Clase5	Clase6	Clase7	Clase8	Total
Clase 1	838	1	13	0	0	0	0	0	852
Clase 2	1	5	0	0	0	0	0	0	6
Clase 3	0	0	28	0	0	1	24	0	53
Clase 4	1	2	2	47	0	0	0	0	52
Clase 5	0	0	0	0	48	0	0	0	48
Clase 6	111	0	89	3	3	14	0	0	220
Clase 7	0	0	0	0	0	63	0	0	63
Clase 8	1	0	0	0	1	0	1	9	12
Total	952	8	132	50	52	78	25	9	1306

- Clase 6: la clase 6 del AAC es comparable con la clase 7 de la combinación. En ellas se observa la inclinación de los televidentes por el canal 5. Se puede apreciar que esta elección es más marcada en la combinación que en el AAC.
- Clase 7: la clase 7 del AAC es comparable con la clase 3 de la combinación, donde sobresale la inclinación por el canal 12 (infantiles internacionales) sobre los demás canales, presentándose pocas diferencias entre las dos clases.
- Clase 8: ambas clases son prácticamente las mismas en ambos métodos, notándose la preferencia por el canal 3 (culturales internacionales).

CAPÍTULO 5

Librería **hqa**

5.1. Aspectos iniciales

Para realizar la combinación entre el análisis armónico cualitativo (AAC) y el análisis factorial múltiple (AFM) utilizando el lenguaje R (R Development Core Team 2007) se programa el paquete denominado **hqa** (disponible en el *CRAN*), el cual utiliza funciones de **ade4** (Chessel et al. 2004) para realizar el análisis armónico cualitativo como un análisis de correspondencias difuso; **FactoMineR** (Husson et al. 2007) para la ejecución del análisis factorial múltiple y la librería **FactoClass** (Pardo & Del-Campo 2007) para caracterizar las clases encontradas por medio de las variables ilustrativas y la variable longitudinal activa; y para realizar gráficos de dicha clasificación. Se programaron también funciones adicionales para la organización de la información longitudinal, y para la realización de la clasificación jerárquica.

5.2. Análisis factoriales

La librería **hqa** utiliza el paquete **ade4** (Chessel et al. 2004) para realizar el análisis armónico cualitativo como un análisis de correspondencias difuso, en particular las funciones:

prep.fuzzy.var: organiza la tabla que se analiza por medio del análisis armónico cualitativo.

dudi.fca: realiza el análisis de armónico cualitativo.

La función **dudi.fca** retorna un objeto tipo **dudi** que contiene los valores y vectores propios del análisis y las coordenadas de las filas y las columnas.

La librería también utiliza el paquete **FactoMineR** (Husson et al. 2007) en especial la función **MFA**, que realiza el análisis factorial múltiple. Esta función retorna un objeto tipo **MFA** que contiene los resultados de los análisis separados de los grupos de variables, los resultados del análisis global, ejes parciales, coordenadas de los grupos, individuos y variables, así como los mapas factoriales de los grupos, los individuos y las variables.

La tabla *dudi\$tab*, generada por la función *dudi.fca*, se convierte en la tabla de entrada para la función *MFA*, y de esta forma se realiza la combinación entre el AAC y el AFM.

El paquete **FactoClass** (Pardo & Del-Campo 2007) se utiliza para realizar la clasificación de los datos, por medio de las funciones **ward.cluster** y **k-meansW**; así como para la caracterización de las clases obtenidas, utilizando la función **cluster.carac**. También se usa la función **planfac** para realizar los gráficos de los planos factoriales del análisis de correspondencias difuso.

5.3. Descripción de la librería `hqa`

La función `hqa` realiza la combinación completa entre el AAC y el AFM, con opciones de realizar únicamente el AAC o de realizar la combinación entre ambos métodos. Sin embargo, algunas funciones se pueden utilizar por separado.

Las funciones del paquete `hqa` se presentan en la tabla 5.1 junto con su descripción.

No es necesario realizar una extensa descripción del paquete, ya que una vez instalado y cargado, se puede obtener la ayuda para las funciones que lo conforman colocando un `?` antes del nombre de la función. Por ejemplo: `?tableclass`.

TABLA 5.1. Funciones y su descripción del paquete `hqa`

Función	Descripción
<code>duration</code>	Calcula la duración pasada en cada modalidad de la variable longitudinal activa
<code>durationtotableA</code>	Transforma la tabla que contiene la duración pasada por cada individuo en cada categoría de la variable longitudinal a la tabla de conteo A descrita en la sección de notación del capítulo 2
<code>fuzzy combination</code>	Realiza el AAC como un análisis de correspondencias difuso Ejecuta la combinación entre el AAC y el AFM
<code>tableclass</code>	Esta función realiza el análisis de clasificación varias veces cambiando el número de ejes y obteniendo el número de individuos en la clase mayoritaria, como ayuda para decidir del número de ejes a considerar en la clasificación.
<code>classif</code>	Realiza la clasificación jerárquica por el método de <i>Ward</i> y la optimiza por medio del método de clasificación <i>kmeans</i> . Para esto se utilizan las funciones <code>ward.cluster</code> y <code>kmeansW</code> del paquete FactoClass .

CAPÍTULO 6

Conclusiones

6.1. Conclusiones de la combinación teórica del AAC y el AFM

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 períodos de tiempo.
- Medidas de asociación que permiten cuantificar las semejanzas entre los diferentes períodos 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.

6.1.1. Conclusiones de la descripción del *rating* de TV

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

Se determinan 8 clases, tanto con el AAC como con la combinación del AAC con el AFM, de televidentes que encienden el televisor en algún momento del día analizado una clase de frecuencia alta (clase 1), que tienen los televisores apagados la mayor parte del día y siete clases diferenciadas por los canales que ven durante el día. Las caracterizaciones obtenidas se presentan en forma detallada en el capítulo 3 y en el apéndice C.

La selección de canales se encuentra asociada al género, estrato socioeconómico, edad y zona del país de los televidentes.

Al tener en cuenta los factores mencionados anteriormente se pueden obtener las siguientes conclusiones:

- La combinación de los dos métodos factoriales logra reducir el número de individuos en la clase que contiene el mayor número de ellos (en este caso la clase 1. Con el AAC esta clase contiene 952 personas. Con la combinación, 852); objetivo importante al realizar la clasificación de los datos (procedimiento propuesto por Barbary (1996), mencionado en el capítulo 2). Esta clase mayoritaria, conserva 852 individuos y los 111 restantes se los pasa a la clase 6, que era la segunda clase mayoritaria. Las clases que

permanecen estables presentan preferencias de canales definidas.

- Las clases de televidentes, en general, se encuentran caracterizadas principalmente por un sólo canal de preferencia. Esta elección es más marcada en la clases obtenidas con la combinación que con las obtenidas a través del AAC.
- Aunque las clases halladas con ambos métodos son comparables, se notan claras diferencias entre ellas. Esto es debido a la ponderación impuesta a las franjas de teleaudiencia en la combinación AAC-AFM.

6.2. Conclusiones finales del *rating* de televisión

Como consideraciones finales cabe resaltar los siguientes aspectos de la selección de canales por parte de los televidentes:

En todas la franjas, seleccionan los canales 4 (RCN), 5 (Caracol), 12 (Infantiles internacionales) y 14 (Otros canales internacionales) con bastante frecuencia y estos canales forman grupos de individuos con características diferentes.

En la franja 5 (*prime time*), de gran interés para las empresas prestadoras de servicios de televisión, y por ende de este trabajo, es donde permanece el televisor más tiempo encendido. En esta franja se nota la preferencia de los individuos por los canales 4 (RCN) y 5 (Caracol), sobre todos los demás, con aproximadamente la misma proporción de tiempo pasada en cada uno de ellos. Los otros canales que se seleccionan en esta franja son: 12 (Infantiles internacionales) y 14 (Otros canales internacionales) con una frecuencia menor que los dos anteriores. Le siguen en elección los canales 3 (culturales internacionales), 8 (series internacionales), 6 (otros canales locales y regionales) y 7 (canal Uno).

La construcción de la librería *hqa* (disponible en el *CRAN*), permite la automatización de la combinación de los dos métodos, y el uso de algunas fun-

ciones adicionales. Además, se convierte en una herramienta de uso futuro para estudios que involucren análisis descriptivos de datos longitudinales de tipo cualitativo.

REFERENCIAS BIBLIOGRAFICAS

- Abascal, E., Aguirre, K. & Landaluce, M. (2001), 'Técnicas factoriales de análisis de tablas múltiples. nuevos desarrollos empíricos', *Universidad del País Vasco* pp. 1–27.
- Abascal, E. & García, I. (2003), 'Una metodología para la evolución de variables latentes. análisis de las infraestructuras de carreteras de las comunidades autónomas (1975-2000)', *Estadística Española* **45**(153), 193–210.
- 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', *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', 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. (2007), 'Multiple Factor Analysis and Clustering of a Mixture of Quantitative, Categorical and Frequency Data', *Computational Statistics and Data Analysis* **10**(1016), 72–79.
- Carroll, J. (1968), Generalization of Canonical Correlation Analysis to Three or more set of Variables, *in* '76th Convention of the American Psychology Association', pp. 227–228. Citado por Escofier-Pages (1992).
- 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. 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).
- Deville, J. & Saporta, G. (1980), 'Analyse harmonic 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. & Pages, J. (1992), *Análisis factoriales simples y múltiples. Objetivos, métodos e interpretación*, Universidad del País Vasco, Bilbao.
- Gamez, N. (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'. url=<http://www.docentes.unal.edu.co/cepardot/docs>.
- Husson, F., Josse, J., Le, S. & Mazet, J. (2007), *FactoMineR: Factor Analysis and Data Mining with R*. R package version 1.07.

-
- Landaluce, M. & Valencia, O. (2000), Estudio comparativo de la metodología STATIS y el análisis factorial múltiple (AFM): una aplicación, *in* Congresos de Economía Regional de Castilla y León, ed., ‘Departamento de Economía. Universidad de Burgos’, pp. 101–110.
- Lebart, L., Morineau, A. & Piron, M. (2006), *Statistique exploratoire multidimensionnelle*, Dunod. 3 Edición, Paris.
- Lebart, L., Piron, M. & Morineau, A. (1995), *Statistique exploratoire multidimensionnelle*, Dunod, Paris.
- Lera, L., Albala, C., Santos, J., Arroyo, P. & García, C. (2007), ‘Estudio multivariado de la evolución de las dimensiones corporales en chilenas mayores de 60 años’, *Archivos Latinoamericanos de Nutrición* **57**, 10–15.
- Pagés, J. (2002), ‘Analyse factorielle multiple appliquée aux variables qualitatives et aux données mixtes’, *Revue de Statistique Appliquée* **50**(4), 5–37.
- Pagés, J. (2004), ‘Multiple Factor Analysis. Main Features and Application to Sensory Data’, *Revista Colombiana de Estadística* **27**(1), 1–26.
- 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 el título de Magister Ciencias-Estadística. Facultad de Ciencias. Departamento de Estadística, Universidad Nacional de Colombia, Bogotá.
- 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.
- Saporta, G. (1996), L’analyse harmonique qualitative, une synthèse de la théorie, *in* ‘Seminario de capacitación e investigación. Recolección y análisis de datos longitudinales’, PRESTA-ORSTOM-Universidad Nacional de Colombia, Bogotá, pp. 111–120.

APÉNDICE A

Análisis factorial general

El análisis factorial general (AFG) describe el proceso común a todos los métodos en ejes principales (Lebart et al. 2006). La notación y los principales aspectos de dicho proceso se pueden resumir en los siguientes puntos:

A.1. Nubes del análisis factorial general

Sea \mathbf{X} la tabla analizada por medio de un análisis factorial, con n filas y k columnas.

Se tiene que:

- El peso p_i de la fila i : estas ponderaciones son agrupadas en una matriz diagonal de dimensión n notada por \mathbf{D} ; $D(i, i) = p_i$; con $\sum_i p_i = 1$
- La ponderación m_k de la columna k : estos pesos son agrupados en una matriz diagonal de dimensión k notada \mathbf{M} ; $M(k, k) = m_k$.

- El conjunto de filas está representada por un conjunto de puntos, notado N_n , en el espacio de k dimensiones, notado \mathfrak{R}^k , donde cada dimensión está asociada a una columna. La matriz \mathbf{M} induce una métrica euclidiana en \mathfrak{R}^k .
- El conjunto de columnas está representado por un conjunto de puntos notados N_k , en el espacio de n dimensiones, notado \mathfrak{R}^n , donde cada dimensión está asociada a una fila. La matriz \mathbf{D} induce una métrica euclidiana en \mathfrak{R}^n .

El análisis factorial general busca los ejes principales de la inercia de las nubes N_n y N_k .

A.2. Análisis de la nube de las filas

En \mathfrak{R}^k , el vector unitario del eje principal de rango s de N_n notado u_s satisface:

$$\mathbf{X}'\mathbf{D}\mathbf{X}\mathbf{M}u_s = \lambda_s u_s \quad \|u_s\|_M^2 = u_s'\mathbf{M}u_s = 1$$

donde λ_s es el valor propio de rango s de $\mathbf{X}'\mathbf{D}\mathbf{X}\mathbf{M}$

Las coordenadas de las filas (puntos de N_n a lo largo de u_s , son reagrupadas en F_s de dimensión $(n,1)$: $F_s = \mathbf{X}\mathbf{M}u_s$; F_s se dice que «es el factor sobre n de rango s ». F_s es un vector de norma:

$$\|F_s\|_D^2 = F_s'\mathbf{D}F_s = \lambda_s$$

A.3. Análisis de la nube de las columnas

En \mathfrak{R}^n , el vector unitario del eje principal de rango s de N_k , notado v_k satisface:

$$v_s = \frac{F_s}{\|F_s\|} = \frac{1}{\sqrt{\lambda_s}} F_s$$

Las coordenadas de las columnas (punto de N_k a lo largo de v_s son agrupadas en G_s , de dimensión $(k,1)$: $G_s = \mathbf{X}'\mathbf{D}v_s$; G_s se dice que «es el factor sobre k de rango s ».

G_s es un vector de \Re^k de norma:

$$\|G_s\|_M^2 = G_s' \mathbf{M} G_s = \lambda_s$$

El vector v_s satisface también:

$$\mathbf{X}\mathbf{M}\mathbf{X}'\mathbf{D}v_s = \lambda_s$$

La matriz $\mathbf{X}\mathbf{M}\mathbf{X}'$ contiene los productos escalares entre los individuos. Ella es usualmente notada como \mathbf{W} .

Además, las direcciones principales de N_n están relacionadas con las proyecciones de N_k con:

$$u_s = \frac{G_s}{\|G_s\|} = \frac{1}{\sqrt{\lambda_s}} G_s$$

A.4. Relaciones de transición

Las coordenadas de las filas (F_s) y de las columnas (G_s) se encuentran relacionadas por las fórmulas de transición:

$$F_s = \frac{1}{\sqrt{\lambda_s}} \mathbf{X}\mathbf{M}G_s$$

Y así para la fila i :

$$F_s(i) = \frac{1}{\sqrt{\lambda_s}} \sum_k x_{ik} m_k G_s(k)$$

Además:

$$G_s = \frac{1}{\sqrt{\lambda_s}} \mathbf{X}' \mathbf{D} F_s$$

Y así para la columna k :

$$G_s(k) = \frac{1}{\sqrt{\lambda_s}} \sum_i x_{ik} p_i F_s(i)$$

Por medio de estas fórmulas es posible interpretar la posición de una fila en función de la posición del conjunto de las columnas, y viceversa.

APÉNDICE B

Canales de televisión

La siguiente tabla muestra los canales de televisión originales, antes de la recodificación, agrupados según la categoría a la que pertenecen una vez efectuado su reagrupamiento (ver tabla 4.5) .

Código	Nombre	Código	Nombre
68	Telecaribe		
21	City TV		
Culturales internacionales			
3	Animal Planet	30	Discovery Home Health
6	Arts	52	History Channel
8	AXN	54	Infinito
1	AyE Mundo	97	Nat geo (National Geographic)
29	Discovery Channel	61	National geographic
4	RCN		
5	Caracol		
Otros canales locales y regionales colombianos			
328	Audio Canal Rock (pereira)	161	Si Colombia
326	Audio Canal Rumba (pereira)	360	Sur tv
319	Ay Hombre!	226	Tamazonas TV
315	Cable Bello TV.	359	Tele 11
361	CALI VISIÓN	347	Tele Atlantico
13	Canal 13	348	Tele Barranquilla
422	Canal 7	428	Tele Caldas

409	Canal 9	277	Tele Medellín
2	Canal Capital	388	Tele Santander
408	Canal Cartagena	312	Tele Soacha (local)
390	Canal del Congreso	411	Teleamazonas
16	Canal Guía	330	Teleboyaca
313	Canal local T Y N (Soacha)	252	Telecafé
17	Canal Noticias	12	Telepacífico
387	Emozónica	393	Todelar
368	Expresiones Del Valle	400	TRT Cucuta
302	Guasca TV	295	TV Agro
407	Info Costavisi3n	275	TV Andina
403	Maria Vision Manizales	336	TV Ciudadania B/manga
137	Nostalgia	337	TV Comunal B/manga
294	Paisa Deportes	305	TV Medellín
292	Paisa Visi3n	339	UNI TV
327	Pentagrama (pereira)	356	Uniaut3noma
401	RC TV Cucuta	340	Univalle
301	Rumba TV	367	Universitario
344	Salsa TV	341	Valle TV
7	Canal Uno		
Series internacionales			
44	Fox	263	Fox Life
47	Hallmark	66	Sony
74	Universal Chanel	42	FX
77	Warner	300	AXN
9	Canal Institucional		
10	Teleantioquia		
11	Señal Colombia		
Infantiles internacionales			
197	Cartoon Network	55	Jetix
31	Discovery Kids	62	Nickelodeon
33	Disney Channel	78	ZAZ
46	Fox sports		
Otros canales internacionales			
187	4 TP	299	Musik
170	6 Noti	289	MV Channel
146	A y M Sports	230	MYN
113	ABN	375	NASA
445	Adsat	374	Naturisimo
503	Advenir Internacional	188	Nice
394	AK	179	Notivision
437	Amis Rael RSS	144	Nova Televisi3n
260	Animax	212	Now TV
89	Antena 3	432	Nt9
253	Antena Latina	245	Nuevo Dia

271	ANTV	514	Ojo 11
194	AO vivo	515	Ojo 12
162	Argentinisima	169	Once
421	Argentinisima	240	Oples
189	Ariran 9	329	OTI
433	Art	149	Panamericana
338	Asaparábolica interno	233	PAT
431	Aster TV	190	Pay Per View
129	ATB	444	Pcn
178	ATD	198	Peliculas
501	Atel	63	People Arts
512	Cine Canal	48	HBO Este
20	Cine Canal 2	49	HBO Oeste
283	Cine Canal Classic	50	HBO plus este
22	Cine Canal Este	51	HBO plus oeste
23	Cine Canal Oeste	272	HBO2
157	Cinema max	57	Movie City
124	ATV	386	Peru magico
371	Azteca 13	64	Playboy
106	Bandeirantes	365	POP TV
14	Barker Channel	502	Portugues
15	BBC	138	Premier Direct TV
99	BBC world	114	Private blue
376	BDP	109	Private gold
93	Bloomberg	276	Promochannel
91	Boca	417	Psn
416	Body	140	Puma
191	Bolivariano	220	Pura Vida
127	Bolivisión	429	Radio Caracas
219	Bolivisión	65	Rai
304	Boomerang	171	Real Madrid Television
195	Brazil Sat	223	Reality TV
244	Buenos Dias Peru	181	Red Global
369	Cable Total	193	Red L
370	Cable Vigia	172	Red Panamericana
101	Cablenoticias	265	Red Sport
309	Cablenovelas	150	Red Uno
286	Cableunión	184	Redes
384	Cadena de Noticias	440	Retro
355	Canaan	227	Retromundo
120	Canal 1	366	RITMO SON
335	Canal 10	96	Ritmo tv
333	Canal 11 Perú	516	Rokola Tv
438	Canal 16	237	SNT
458	Canal 21	358	Solidaria TV

380	Canal 22	291	Solo Tango
442	Canal 23	288	Son Corazón
255	Canal 6	372	Son Latino
363	CANAL C	454	Son Music
508	Canal Comunal de Vigilancia	290	Son Popular
506	Canal Comunal Tele 19 (Litelcoop)	331	Sorpresa
415	Canal Comunitario	284	Speed
131	Canal Cuatro	67	Spice Clips
90	Canal de las estrellas	413	Sport Max
133	Canal de musica	163	Sport TV
174	Canal de Venezuela	156	Sports max
278	Canal EPM TV	130	Starz
102	Canal hipico	79	Super cable channel
430	Canal S	385	Super Show
115	Canal Superview	325	Superland
176	Canal tR	135	SWT
282	Canal U	353	Syria
448	Canal Une	73	T y C Sport
151	Canal Vasco	231	TA
412	Caribe TV	154	TBN
18	Casa club	310	TCM
427	Catolica Luz	402	Tecom
264	CBS	85	Tele Amiga
19	CBS/Worldnet	378	Tele Canarias
504	CDN	364	TELE FAMILIA
318	Cine Latino	121	Tele Futuro
334	Cine Mexicano	242	Tele Globo
225	CineA T.V.	351	Tele Hit
251	Cinema +	303	Tele Humor
141	Cinet	298	Tele Nostalgia
389	Clase	235	Tele Noticias
158	CMB	343	Tele Novelas
123	CMT	180	Tele Ventas
26	CNBC	148	Telecadena 7/4
324	CNC	110	Telefe
27	CNN Español	239	Telegracia
28	CNN Internacional	507	TeleHumos
249	Copan	256	Teleislas
86	Cosmopolitan	261	Telemanía
279	Cosmovisión	192	Telemexicana
455	CTI Movil	405	Telemicro
354	Cuba Visión	69	Telemundo
406	Cubano	316	Telenorte
183	DAP	204	Telepaís
410	DC	143	Telesistema

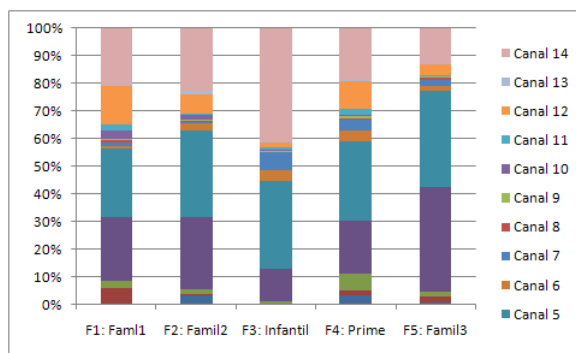
285	De Película	274	Telesistemas
94	Deutsche welle (NO USAR)	159	Telesur
379	Digital 15	182	Teletica 7
452	Digital Kids	404	Televen
418	Discovery Ingles	152	Televicentro
32	Discovery travel Living	280	Televida
202	Dish Network	210	Televisa
320	DJ TV	221	Television Mexiquense
34	DW (Deutsche Welle)	396	The Beat II
81	E! entertainment	70	The film zone
238	Ecuavisión	228	TK
306	Edu Cable	232	TLP
119	Edusat	173	TNP
446	Eduvision	196	TNT
35	Ei	511	Todo Noticias TvCable
126	El Garage TV	84	Torneos y competencias
36	El gourmet. Com	510	Total Help
95	El Kanal	323	TPN
443	El Lider	164	TPR
434	Encore	241	TPS Toot
37	Enlace TBN	436	Tra
270	Epatelevisión	200	TRO
425	Ert	234	Tu TV
38	ESPN	132	Tu Vecino TV
39	ESPN 2	254	TV 1
103	Etb canal vasco	383	TV 15
40	Eurochannel	258	TV 4
287	Europa Europa	71	TV 5
41	EWTN	224	TV 8
297	Exclusive	185	TV Alamos
139	Falabella	311	TV Arte
352	Family Cristiam	98	Tv Azteca
87	Fashion tv	211	Tv Brasil
43	Film Arts	145	TV Cable
346	Formar	435	Tv Centro
321	Formato	88	TV Chile
80	TV guide	216	TV ci
45	Fox News	424	Tv Escola
382	Frecuencia Latina	82	TV España
206	FS TV	205	TV Española
509	FTV	128	TV Galicia
136	Galavision	293	TV Hogar
505	Galicia	373	TV Internacional
160	Globo	177	TV Marli
397	Globo de Brasil (cuadro con ojo)	273	TV Marti

322	Globovisión	168	TV P6
296	Gol TV	453	TV Peru
259	Golden	350	TV Shopping
100	Grandes Documentales TVE	201	TV SS
166	Guatevisión	153	TV Venezuela
447	Guia Une	72	TVE
111	Hispavision	267	TVE 24 Hrs
207	HITN	165	TVT
451	Hot Tv	247	TVT-V
53	HTV	349	U Cinco
332	Humor Chanel	199	UCV
398	Iberoamericano	122	Unitel
208	IMF	414	Musica Latina
395	Informacion	105	Univisión
266	JK México	229	Univisión Puerto Rico
441	JTV	155	Uno
426	K music	112	Utilisima
517	Kodigo	246	UTV
362	LA 14	457	UV
125	La Tele	314	Veintiuno
377	Latitud	134	Venevisión
214	LH TV	439	Venezolana de televisión
56	Locomotion	75	Venus
317	M Total Music	76	VH1
345	Magic Kids	104	Via X
357	Maria Visión	117	Video Plays
268	Market News	342	Video Portero
262	Mas TV	399	Video Rokola
269	Max Prime	459	Video Sol
381	Maya TV	257	View Channel
500	Mega Tv	456	Virus
248	Mexiquese	419	Vision Elite
215	Michoacan	300	VJ Channel
307	Momentum	107	Voice of america (VOA)
186	MP	236	VÓX
60	MTV	217	VTG
449	MTV Hits	218	VTN
450	MTV Jams	147	VTV
118	Multimedios TV	391	World Net
83	Multipremier	243	XHAOX
203	Multivisión	175	Zona Latina
167	Cinemax	513	MGM
24	Cinemax Este	58	Movie city este
25	Cinemax Oeste	59	Movie city oeste
281	Much Music		

APÉNDICE C

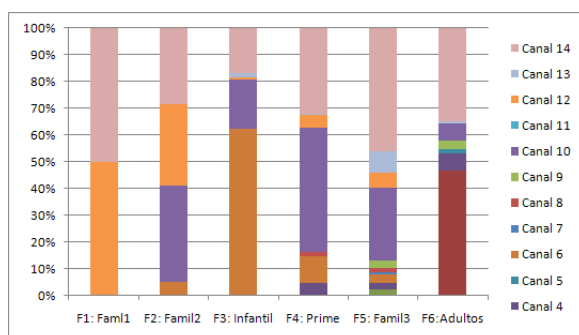
Clases obtenidas con el AAC

A continuación se hace un resumen de la caracterización de las clases obtenidas a través del análisis armónico cualitativo. En el apéndice D se encuentran otros gráficos que ayudaron a caracterizar las clases por medio de las variables ilustrativas. Las figuras que acompañan las clases fueron realizadas eliminando el canal 15 (televisor apagado), para facilitar la lectura de la selección de los otros canales. En el apéndice E se pueden observar los gráficos considerando el canal 15.

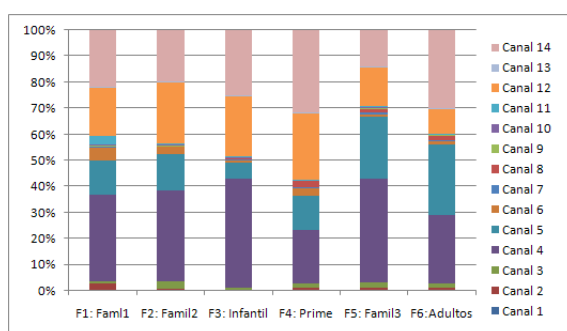


tratatos medio y bajo, con un alto porcentaje de ellos pertenecientes a la zona centro de Colombia.

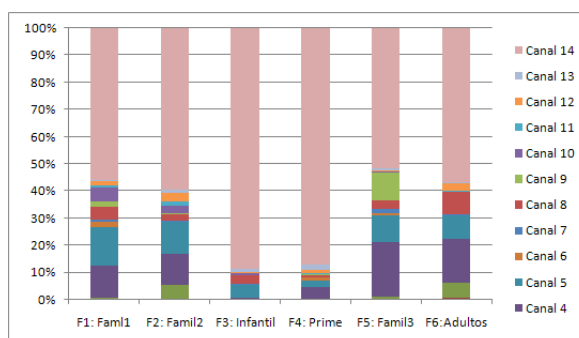
Clase 1 (72.9 %). Los televidentes de este grupo permanecen con el televisor apagado durante gran parte del día. Es un grupo con gustos heterogéneos ya que se inclinan por diversos canales a través de las diferentes franjas de tiempo durante el día. Los preferidos fueron el canal 4 (RCN), el canal 5 (Caracol), el canal 14 (Otros canales internacionales) y en menor grado el canal 12 (Infantiles internacionales). Este grupo está compuesto por personas de edades entre 40 y 64 años, de estratos medio y bajo, con un alto porcentaje de ellos pertenecientes a la zona centro de Colombia.



(otros canales internacionales) principalmente. En la franja 5 se sintonizan diversos canales, destacándose los canales 10 (Teleantioquia) y 14 (otros canales internacionales). En la franja 6 (horario de adultos) eligen principalmente el canal 2 (City TV). El 50% de estos individuos tienen edades entre 12 y 17 años, y el 67% pertenecen al estrato medio.



la zona centro colombiana. El 50% de estos televidentes son mujeres y el 50%, hombres.

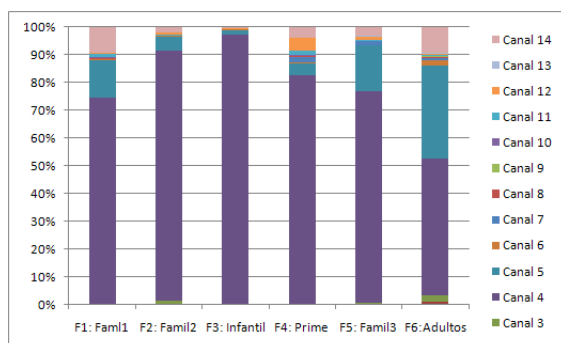


Clase 2 (0.6 %). Este grupo prefiere diferentes canales en cada una de las franjas de tiempo. Se destacan en la franja 2, los canales 14 (cinemas y otros canales internacionales), 12 (canales infantiles internacionales) y 10 (Teleantioquia). En la franja 3, sobresale la elección del canal 6 (canales locales y regionales colombianos). En la franja 4 (*Prime Time*), seleccionan los canales 10 (Teleantioquia) y 14

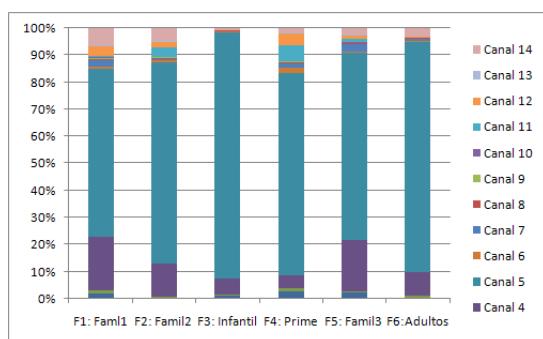
Clase 3 (10.1 %). Este grupo se encuentra conformado por televidentes que prefieren principalmente los canales 4 (RCN), 5 (Caracol), 12 (infantiles internacionales) y 14 (otros canales internacionales) con diferentes proporciones de escogencia durante todas las franjas. Ninguno de los individuos de este grupo pertenece a las zonas 2 y 6. El 41% tienen edades entre 40 y 64 años, de estratos bajos y medio. El 29% de estas personas viven en la

Clase 4 (3.8 %) Este grupo caracteriza a los televidentes que escogen principalmente el canal 14 (otros canales internacionales como: *HBO*, *CineMax*, *MovieCity*, etc.) durante todo el día. Con menor frecuencia seleccionan los canales 4 (RCN) y 5 (Caracol) en las distintas franjas, y otro conjunto de canales en una proporción muy pequeña. El 54% son hombres y el 46% mujeres. El 46% tienen edades entre 40 y 64 años, de las

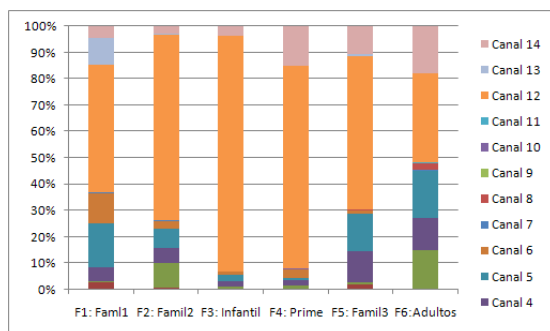
zonas 5 (cafetera) y 1 (centro) principalmente.



años (60%) pertenecientes en gran proporción (52%) a la zona central de Colombia.



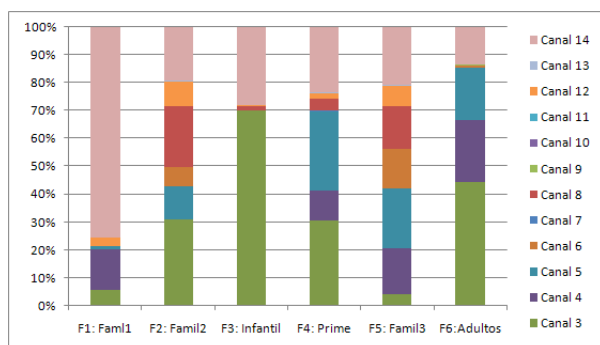
de Colombia.



Clase 5 (4.0%) Este grupo caracteriza a los televidentes que escogen principalmente el canal 4 (RCN) durante todo el día. Con menor frecuencia seleccionan los canales 14 (otros canales internacionales), 12 (infantiles internacionales) y 5 (Caracol) en las distintas franjas, y otra gama de canales en una proporción muy pequeña. El canal 9 (canal institucional) no fue seleccionado, prácticamente, por estos usuarios. En esta clase se encuentran mujeres (62%) entre 40 a 64 años (60%) pertenecientes en gran proporción (52%) a la zona central de Colombia.

Clase 6 (6.0%) Esta clase se encuentra formada por individuos que permanecen casi todo el tiempo de la audiencia en el canal 5 (Caracol). En una proporción muy pequeña, seleccionan también los canales 4 (RCN), 12 (infantiles internacionales) y 14 (cinemas internacionales) y otros canales internacionales. El 64% tienen edades entre 25 y 64 años. El 50% pertenecen al estrato socioeconómico medio, y en igual proporción a la zona centro

Clase 7 (1.9%). Este grupo se caracteriza por preferir el canal 12 (infantiles internacionales) sobre el resto de los canales. También se seleccionan los canales 4 (RCN), 5 (Caracol), 14 (otros canales internacionales) 3 (culturales internacionales) y 13 (*Fox Sport*) en menor proporción. Se encuentra formado principalmente por hombres entre 25 y 64 años, de las zonas 1 (centro) y 4 (caribe) y de niveles socioe-



conómicos bajo y medio.

Clase 8 (0.7%). Esta clase de televidentes prefieren principalmente el canal 3 (culturales internacionales) seguido por el canal 14 (series internacionales). En menor proporción, y en diferentes franjas escogen los canales , 4 (RCN) , 13 (*Fox Sport*), 6 (otros canales locales y regionales colombianos) y 8 (series internacionales). Se encuentra conformado por hombres, entre 18 y 64 años, 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.

APÉNDICE D

Clasificación con AAC. Variables ilustrativas.

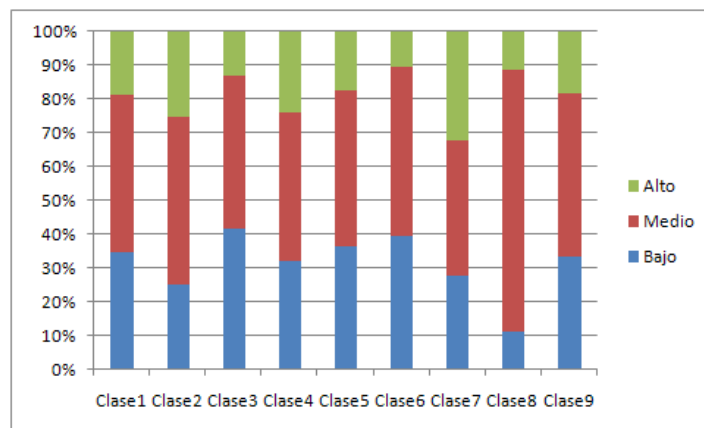


FIGURA D.1. Perfiles de las clases por el estrato socioeconómico

TABLA D.1. Perfiles de las clases AAC por género, nivel socio-económico y rango de edad

Clase	Género		Nivel socioeconómico			Rango de edad					
	Hombre	Mujer	Bajo	Medio	Alto	4a11	12a17	18a24	25a39	40a64	mas64
1	52.1	47.9	34.7	46.6	18.7	13.3	11.7	10.7	24.2	33.4	06.7
2	50.0	50.0	25.0	50.0	25.0	12.5	50.0	00.0	12.5	25.0	00.0
3	40.2	59.8	41.7	45.5	12.9	12.9	09.8	06.1	22.0	40.9	08.3
4	54.0	46.0	32.0	44.0	24.0	02.0	10.0	10.0	26.0	46.0	06.0
5	38.5	61.5	36.5	46.2	17.3	05.8	01.9	07.7	15.4	61.5	07.7
6	48.7	51.3	39.7	50.0	10.3	03.8	09.0	12.8	24.4	38.5	11.5
7	56.0	44.0	28.0	40.0	32.0	24.0	08.0	00.0	32.0	28.0	08.0
8	77.8	22.2	11.1	77.8	11.1	0.00	11.1	11.1	33.3	44.4	00.0
9	43.9	56.1	33.5	48.0	18.4	16.7	12.7	20.0	23.0	22.3	5.3

TABLA D.2. Perfiles de las clases AAC según zonas del país

Clase	-Zonas-					
	1:Centro	2:Antioquia	3:Pacífico	4:Caribe	5:Cafetera	6:Oriental
1	24.9	11.3	18.2	17.6	10.6	17.3
2	25.0	00.0	25.0	25.0	25.0	00.0
3	28.8	12.9	15.9	08.3	15.9	18.2
4	30.0	14.0	10.0	04.0	38.0	04.0
5	42.3	01.9	11.5	15.4	13.5	15.4
6	46.2	05.1	10.3	24.4	05.1	09.0
7	52.0	04.0	08.0	08.0	20.0	08.0
8	00.0	00.0	44.4	00.0	44.4	08.0
9	26.2	12.5	16.1	15.2	13.1	16.9



FIGURA D.2. Perfiles de las clases por la zona



FIGURA D.3. Perfiles de las clases por el género

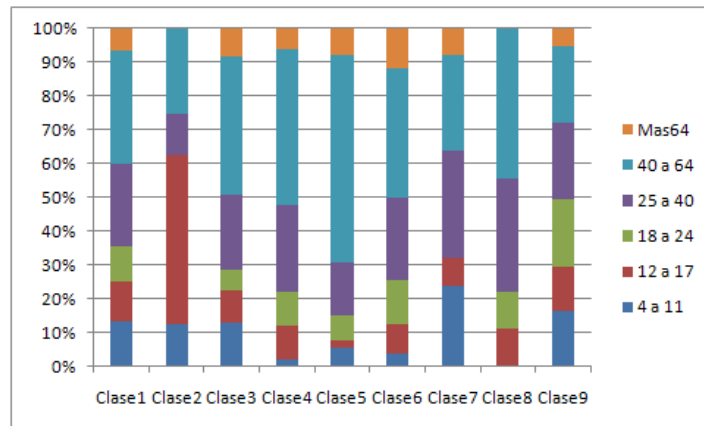


FIGURA D.4. Perfiles de las clases por la edad

APÉNDICE E

Clasificación con AAC. Variable longitudinal activa

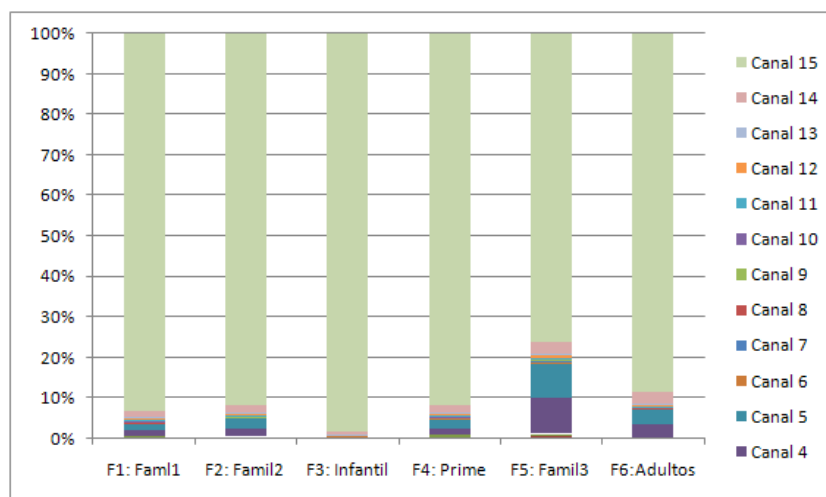


FIGURA E.1. Perfiles de selección de canales por franjas de la clase 1

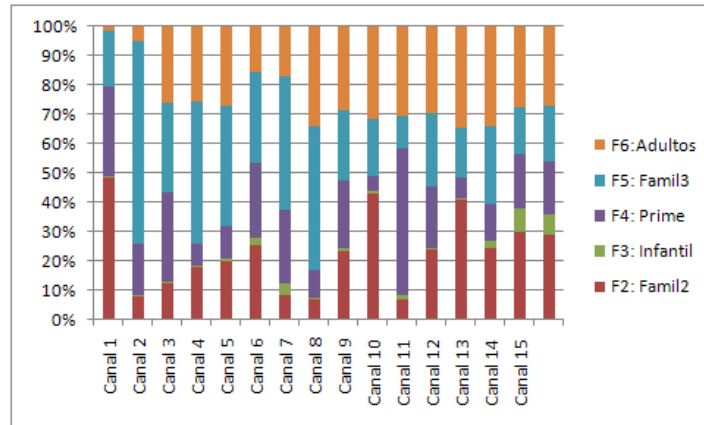


FIGURA E.2. Perfiles de selección de franjas por canales de la clase 1

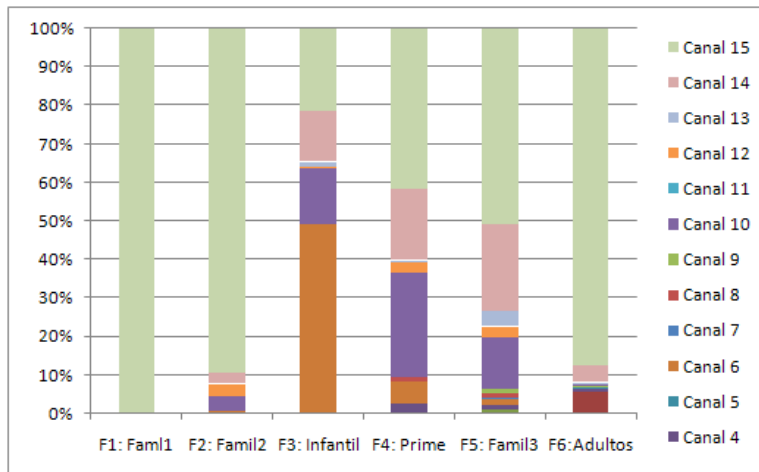


FIGURA E.3. Perfiles de selección de canales por franjas de la clase 2

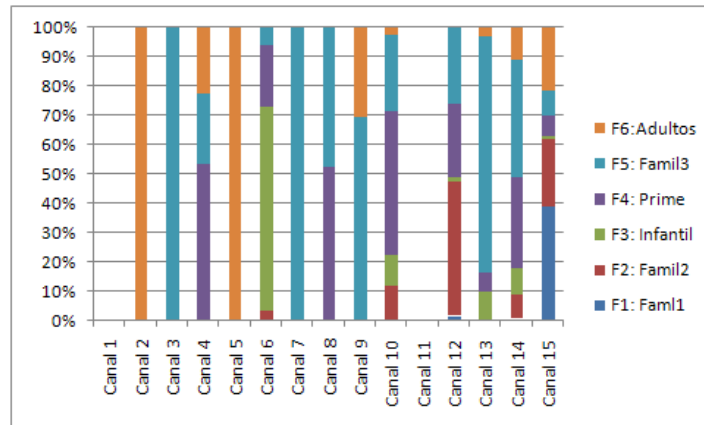


FIGURA E.4. Perfiles de selección de franjas por canales de la clase 2

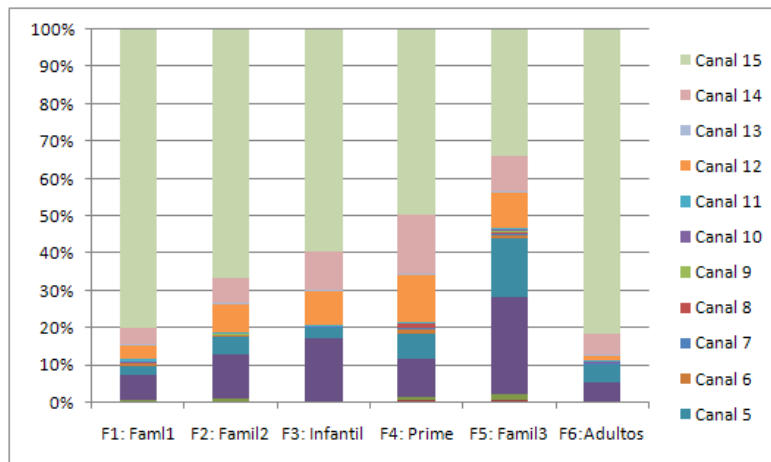


FIGURA E.5. Perfiles de selección de canales por franjas de la clase 3

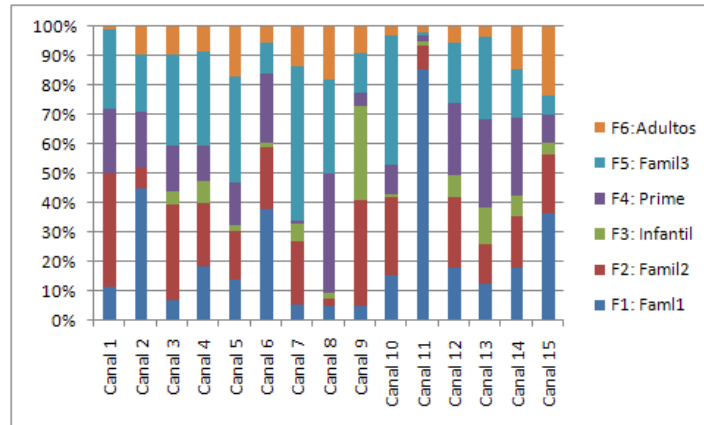


FIGURA E.6. Perfiles de selección de franjas por canales de la clase 3

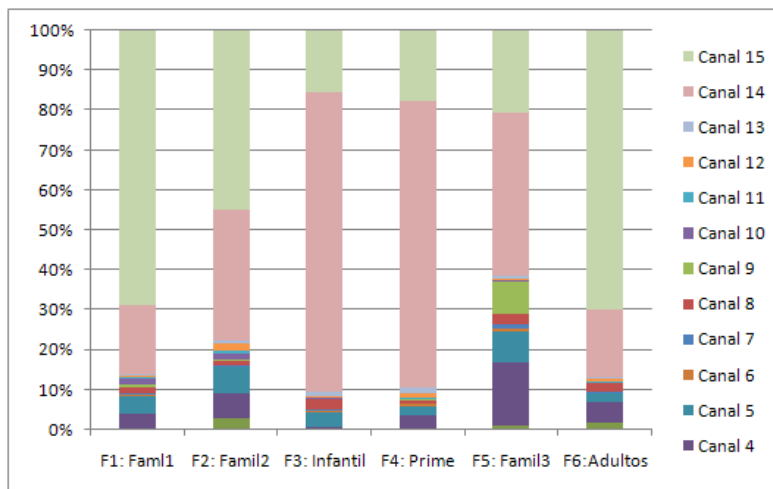


FIGURA E.7. Perfiles de selección de canales por franjas de la clase 4

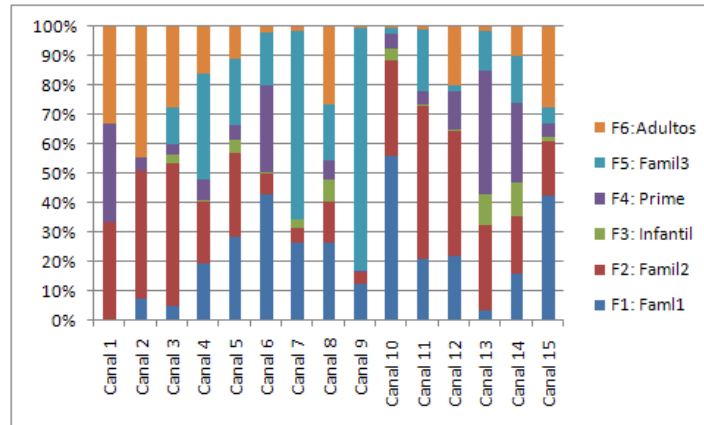


FIGURA E.8. Perfiles de selección de franjas por canales de la clase 4

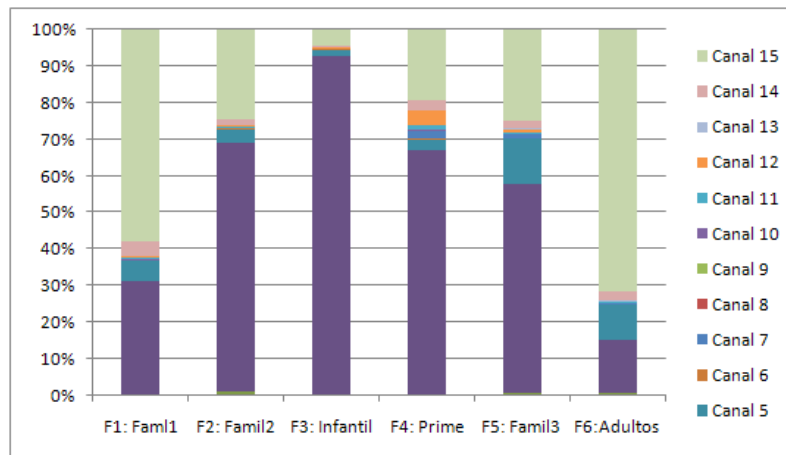


FIGURA E.9. Perfiles de selección de canales por franjas de la clase 5

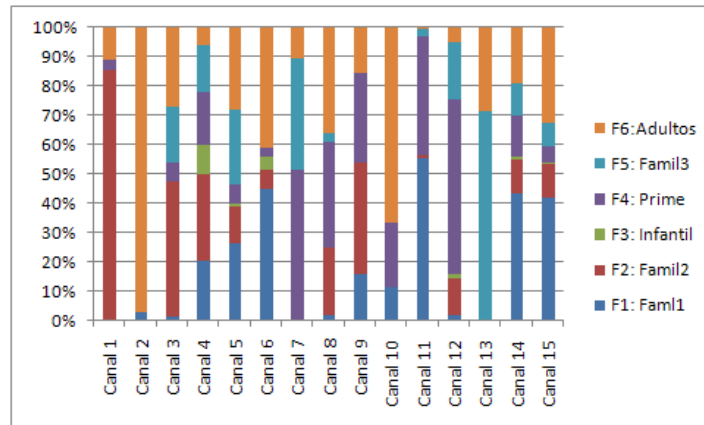


FIGURA E.10. Perfiles de selección de franjas por canales de la clase 5

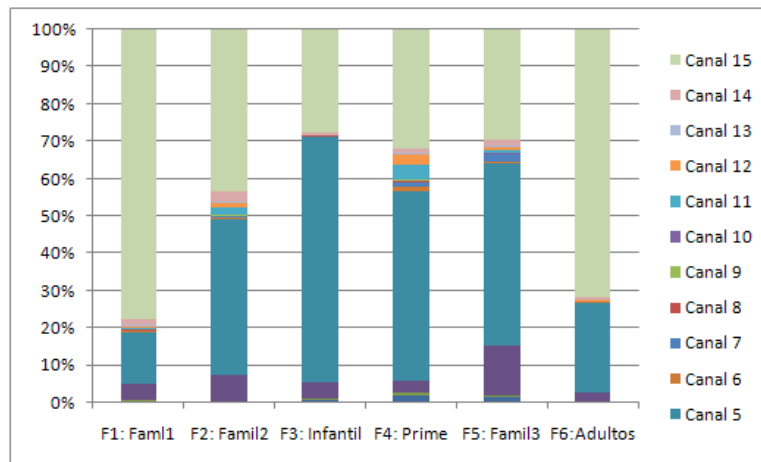


FIGURA E.11. Perfiles de selección de canales por franjas de la clase 6

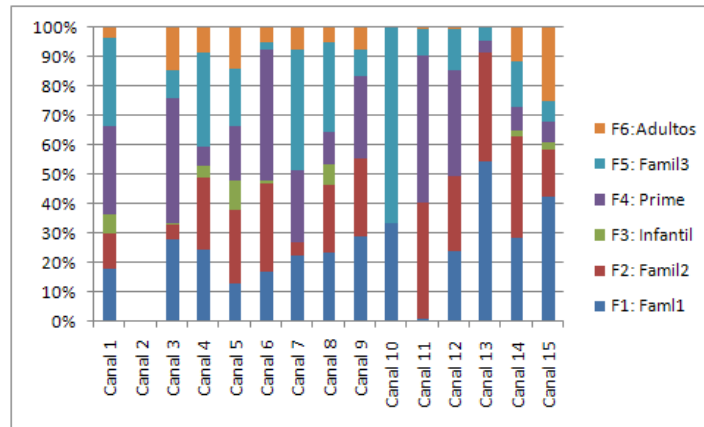


FIGURA E.12. Perfiles de selección de franjas por canales de la clase 6

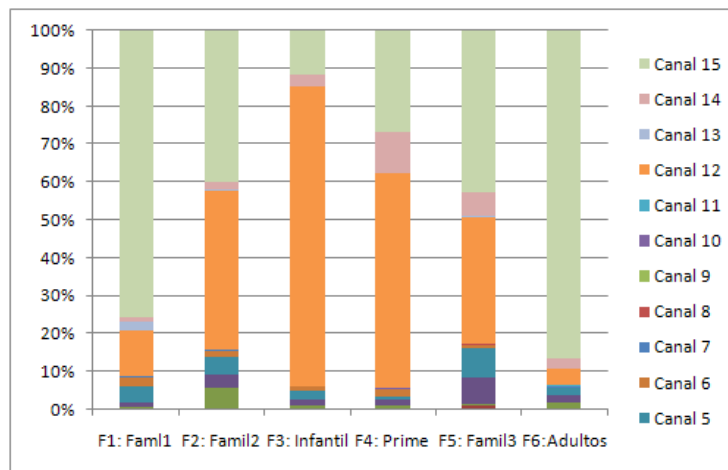


FIGURA E.13. Perfiles de selección de franjas por canales de la clase 7

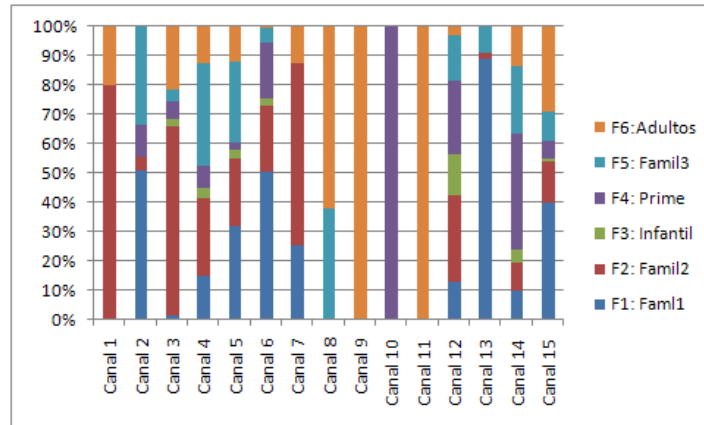


FIGURA E.14. Perfiles de selección de franjas por canales de la clase 7

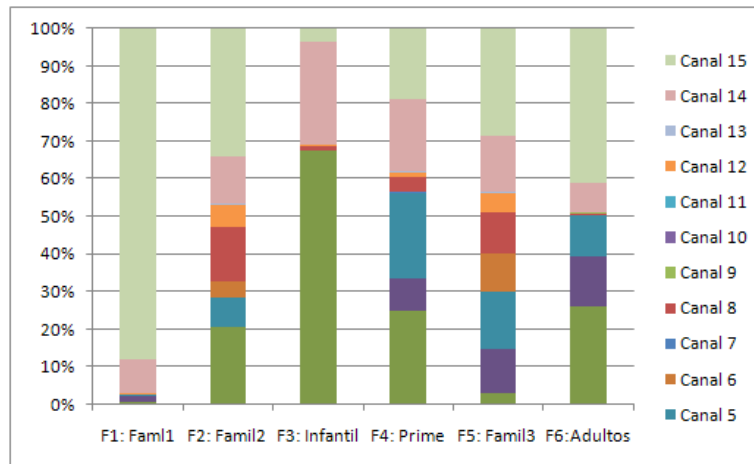


FIGURA E.15. Perfiles de selección de franjas por canales de la clase 8

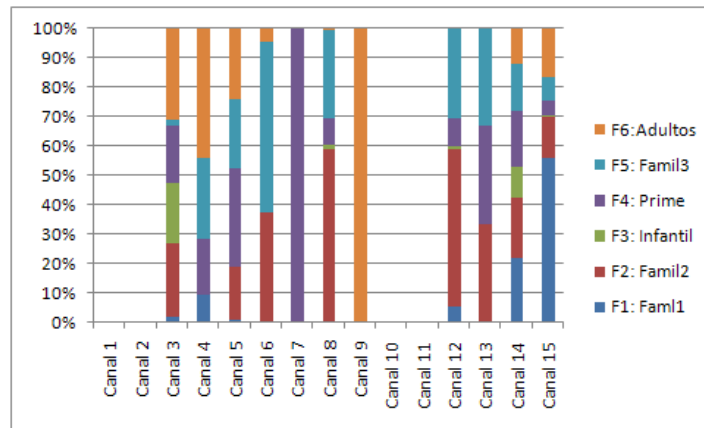


FIGURA E.16. Perfiles de selección de franjas por canales de la clase 8

APÉNDICE F

Clasificación con AAC y AFM. Variables ilustrativas.

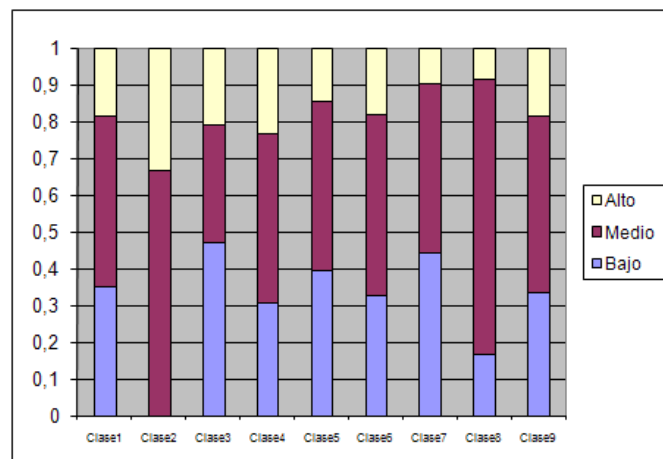


FIGURA F.1. Perfiles de las clases por el estrato socioeconómico

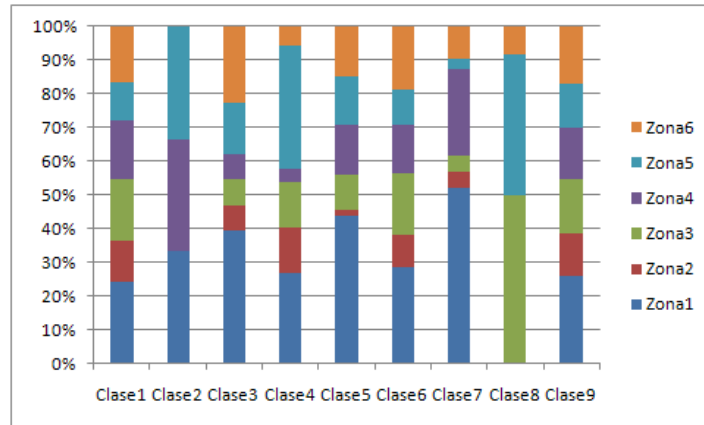


FIGURA F.2. Perfiles de las clases por la zona

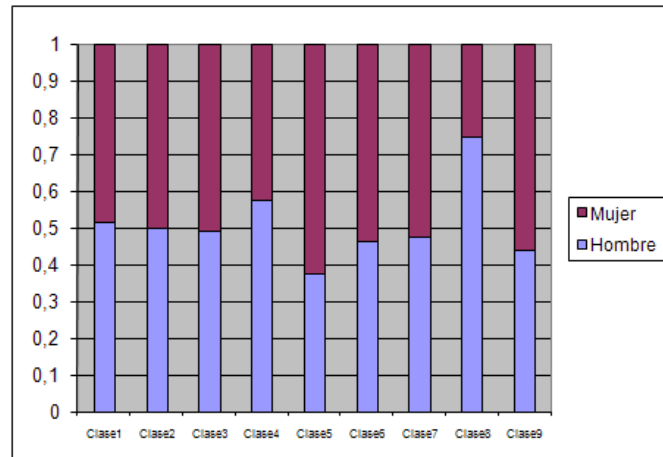


FIGURA F.3. Perfiles de las clases por el género

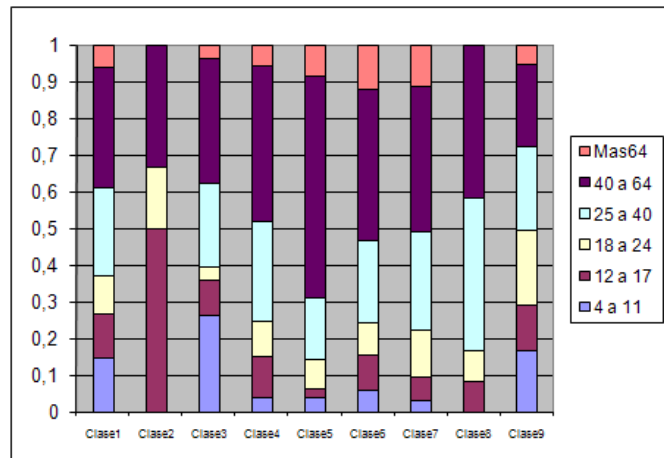


FIGURA F.4. Perfiles de las clases por la edad

APÉNDICE G

Clasificación con AAC y AFM. Variable longitudinal activa

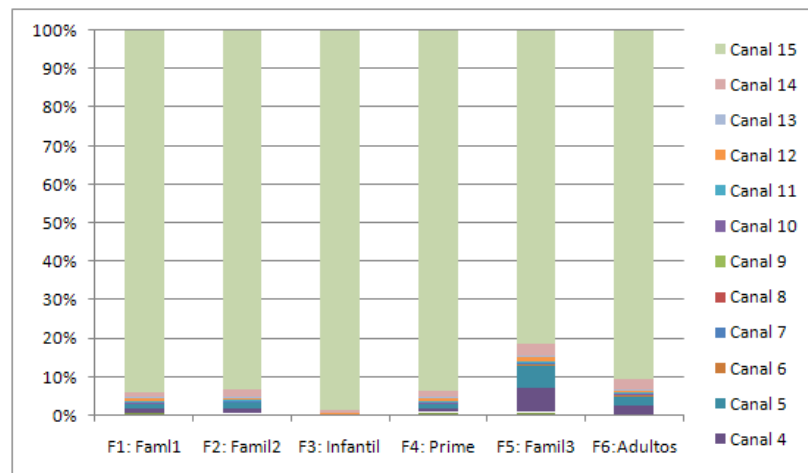


FIGURA G.1. Perfiles de selección de canales por franjas de la clase 1

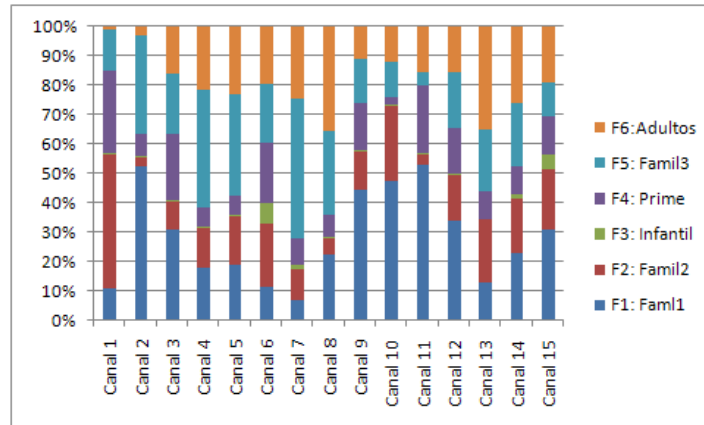


FIGURA G.2. Perfiles de selección de franjas por canales de la clase 1

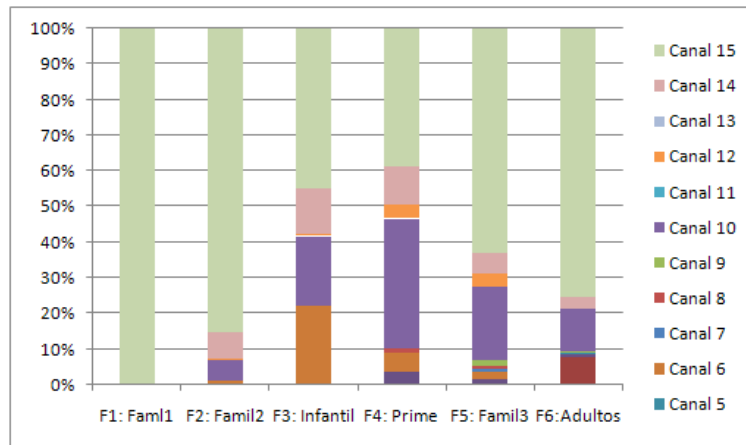


FIGURA G.3. Perfiles de selección de canales por franjas de la clase 2

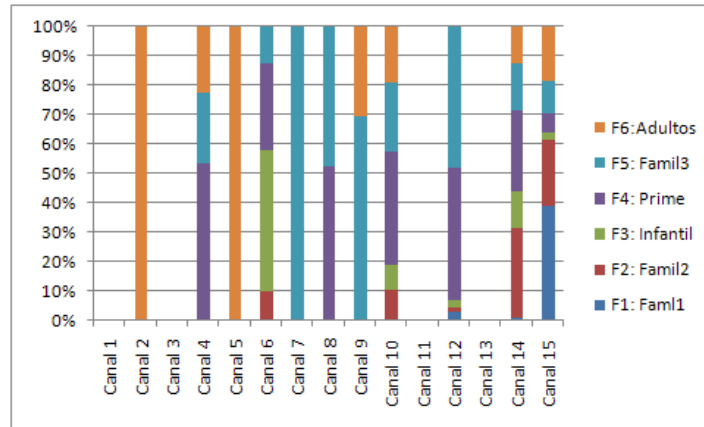


FIGURA G.4. Perfiles de selección de franjas por canales de la clase 2

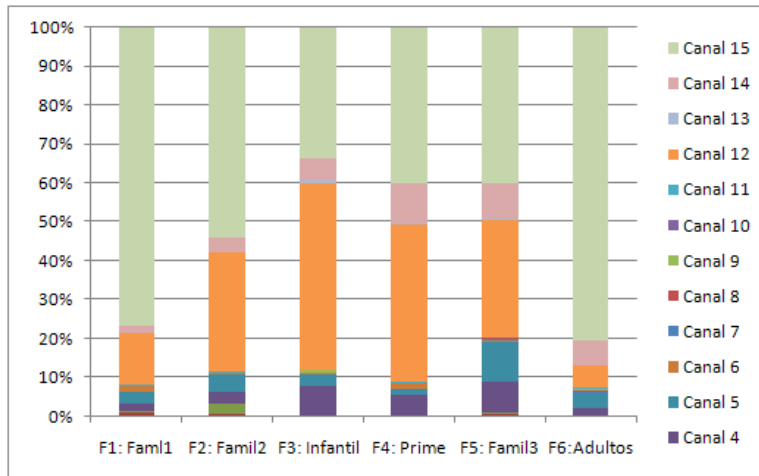


FIGURA G.5. Perfiles de selección de canales por franjas de la clase 3

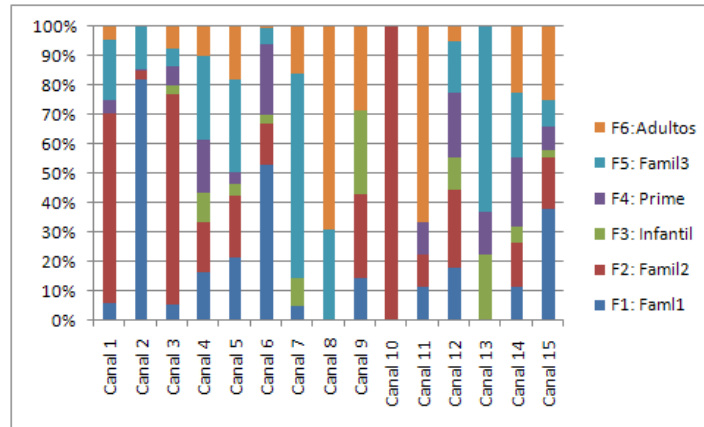


FIGURA G.6. Perfiles de selección de franjas por canales de la clase 3

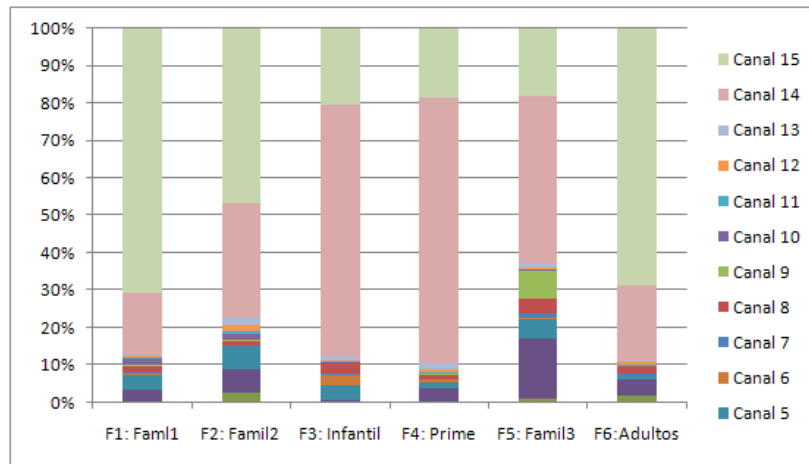


FIGURA G.7. Perfiles de selección de canales por franjas de la clase 4

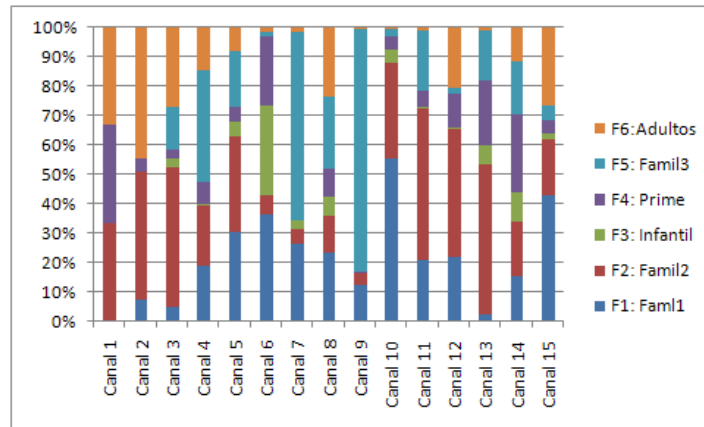


FIGURA G.8. Perfiles de selección de franjas por canales de la clase 4

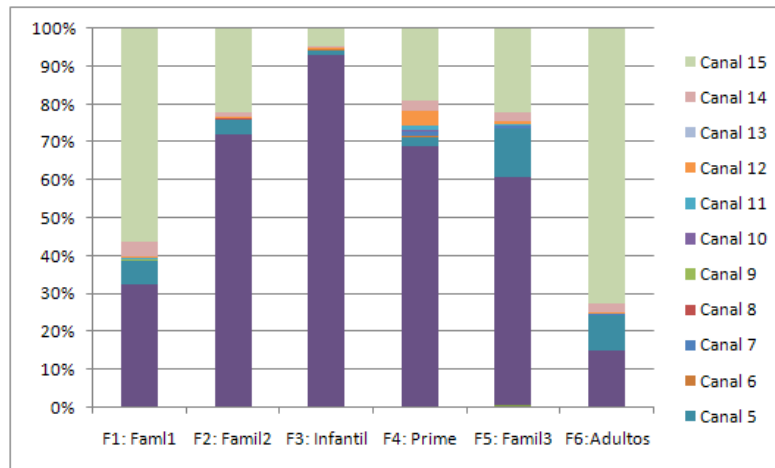


FIGURA G.9. Perfiles de selección de canales por franjas de la clase 5

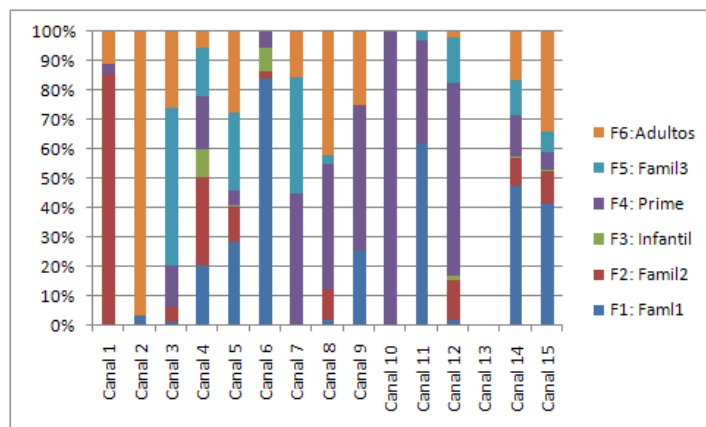


FIGURA G.10. Perfiles de selección de franjas por canales de la clase 5

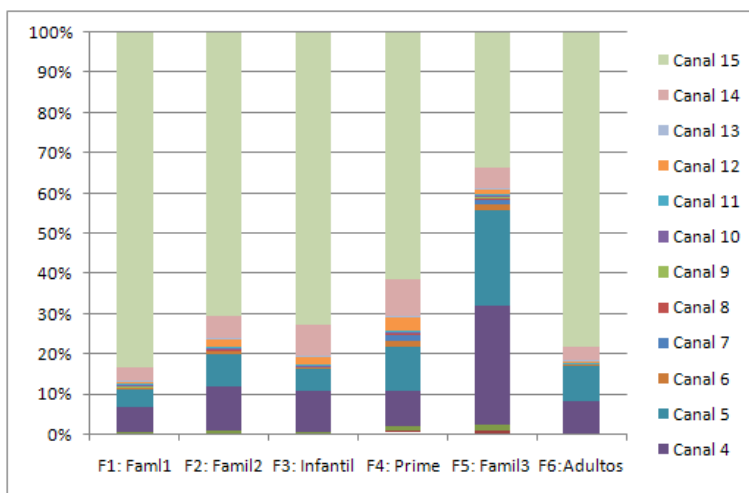


FIGURA G.11. Perfiles de selección de canales por franjas de la clase 6

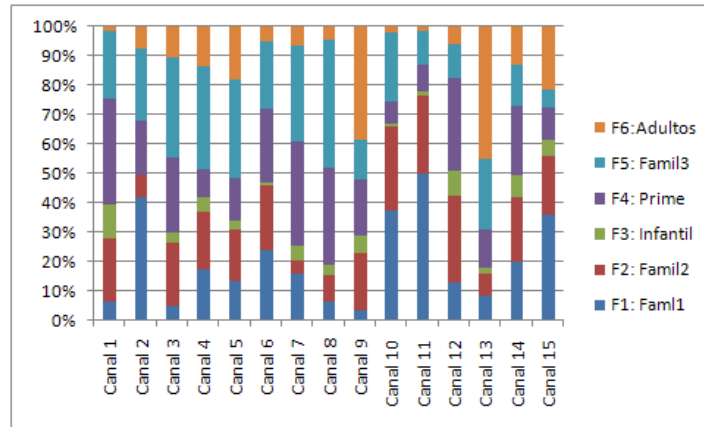


FIGURA G.12. Perfiles de selección de franjas por canales de la clase 6

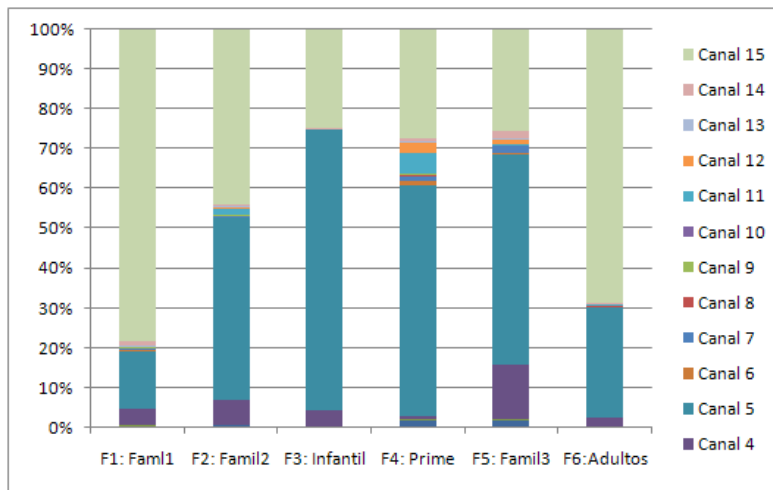


FIGURA G.13. Perfiles de selección de canales por franjas de la clase 7

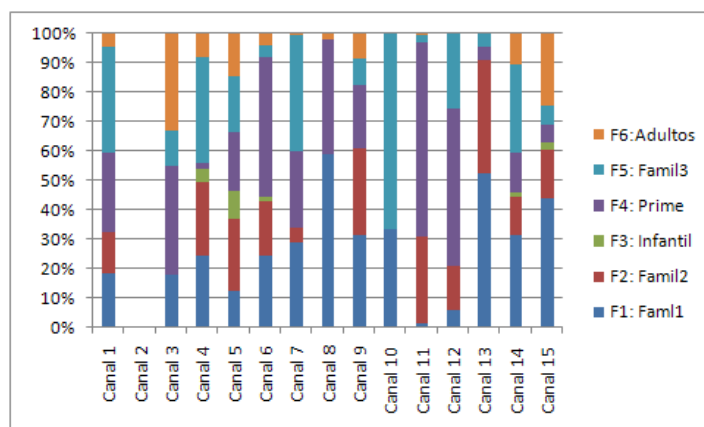


FIGURA G.14. Perfiles de selección de franjas por canales de la clase 7

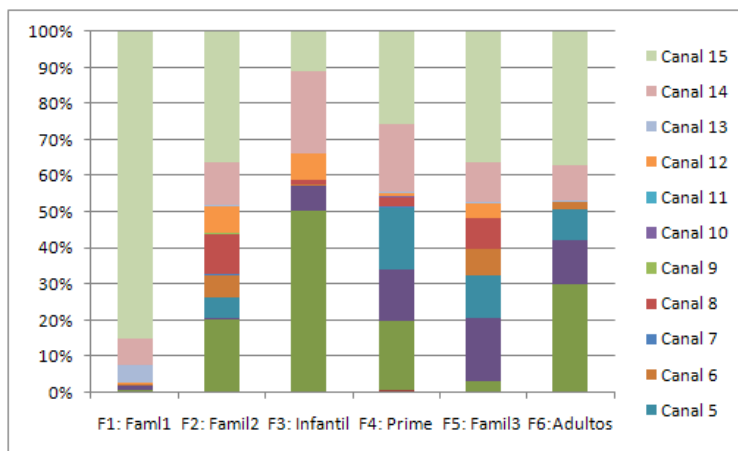


FIGURA G.15. Perfiles de selección de canales por franjas de la clase 8

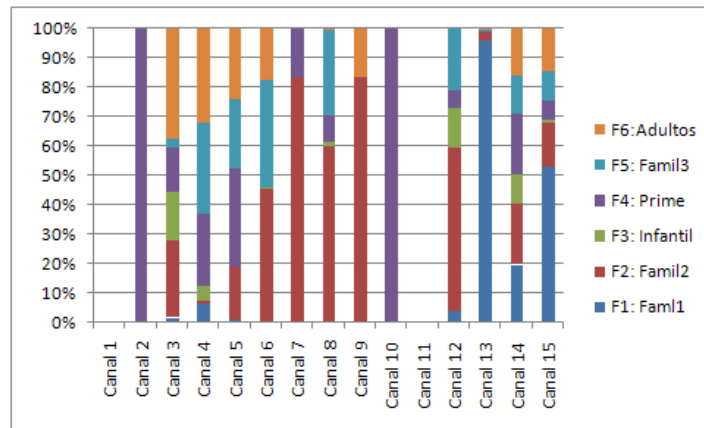


FIGURA G.16. Perfiles de selección de franjas por canales de la clase 8

APÉNDICE H

Planos factoriales AAC

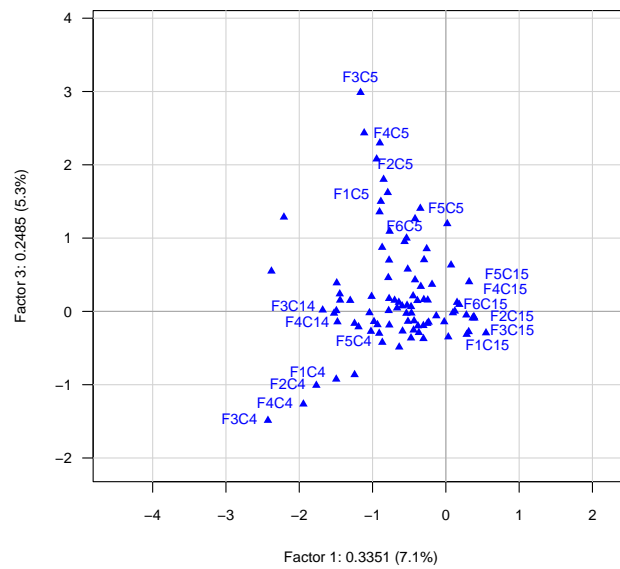


FIGURA H.1. Plano factorial 1-3 AAC

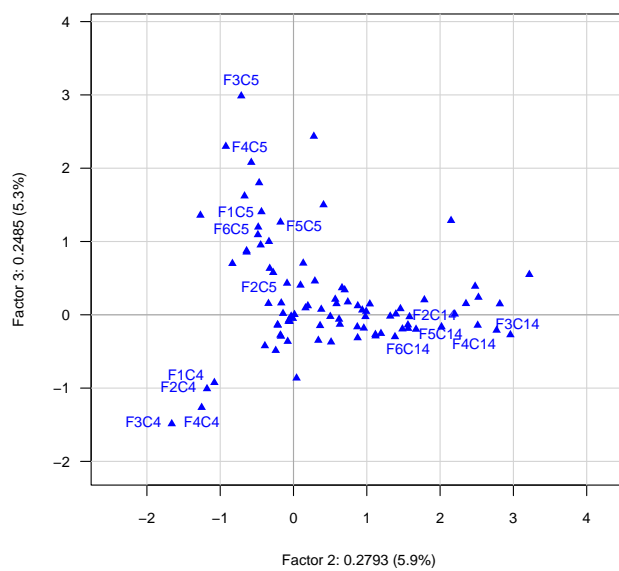


FIGURA H.2. Plano factorial 2-3 AAC

APÉNDICE I

Ejemplo de uso de la librería *hqa*

I.1. Librería *hqa*

Descripción

Realiza la combinación entre el AAC y el AFM para el procesamiento de datos longitudinales categóricos.

Uso

```
hqa(base, conteos=FALSE, units=NULL, durat=FALSE, periodos=NULL, pesos = NULL,
ilustra=NULL, ilustc = NULL, ilust.type = NULL, nfact=5, nfc1=5, k.clust=NULL, com-
binat=TRUE, vector, tableclass=FALSE, clasifica=TRUE)
```

Argumentos

- base : objeto de tipo data frame o matrix
- conteos: TRUE si desea obtener el data frame con los campos ID (identificador),MOD

(modalidad) ,DURATION (duración. Default TRUE

- units time: "secs", "mins", "hours", "days", "weeks", "months", "years". Default = NULL
- durat: TRUE si desea calcular la duración pasada en la modalidad de la variable longitudinal por medio de la función DURATION. Default = FALSE
- periodos: vector que contiene la duración de cada período de tiempo.
- pesos: vector con el peso de las filas
- ilustra: objeto de tipo data frame o matrix con las variables ilustrativas. Default = NULL
- ilustc: vector que contiene el número de variables ilustrativas en cada grupo.
- type: tipo de variable en cada grupo de las variables ilustrativas ("c" para variables cuantitativas, "s" para variables cuantitativas estandarizadas, "n" para variables cualitativas. Por defecto todas las variables son cualitativas
- nfact: número de ejes para usar en el análisis number of axes to use into the factorial analysis . Default nfact=5
- nfcl: number de ejes para usar en la clasificacion. Default nfcl=5
- k.clust: número de clases. Default k.clust= NULL
- combinat: TRUE si quiere realizar la combinación del AAC y del AFM, FALSE si desea realizar el AAC. Default = TRUE
- vector: es el vector que contiene el número de categorías para cada variable difusa.
- tableclass: TRUE si quiere obtener una sugerencia acerca del número de ejes a usar en el clasificación. Default = FALSE
- clasifica: TRUE si desea que la función realice el proceso de clasificación. Default = TRUE

Detalles

Se propone una metodología estadística para analizar datos longitudinales categóricos. Esta metodología consiste en usar la combinación de dos métodos: análisis armónico cualitativo y análisis factorial múltiple. El análisis se complementa con una clasificación de los datos usando las primeras coordenadas factoriales.

Valores devueltos por la función

- HQA: objeto de tipo dudi, si solo se realizó el AAC o de tipo MFA si se efectuó la combinación.
- Clases: Objeto de tipo "kmeans"
- Active: caracterización de las clases considerando la variable longitudinal activa.
- Ilust: caracterización de las clases considerando las variables ilustrativas.

Ejemplo

```
data(ratingTV)
rating = hqa(base=ratingTV$tab, ilustra=ratingTV$ilus, vector=c(15,15,15,15,15,15), ilust=c(4))
90
rating$HQA
rating$Clases
rating$Active
rating$Ilust
```