

Complementariedad entre el modelo log-lineal y el análisis de correspondencias en el análisis de tablas de contingencia

DIANA CAMILA BUITRAGO^a
AUTOR

CAMPO ELÍAS PARDO^b
DIRECTOR

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

Resumen

Las relaciones entre los modelos log-lineales y el análisis de correspondencias permiten desarrollar una metodología para el análisis de tablas de contingencia que combina las dos técnicas. En este enfoque, los modelos log-lineales se usan para evaluar la importancia de los efectos de interacción y luego se emplea el análisis de correspondencias para evaluar los residuales del modelo. También es posible visualizar en el análisis de correspondencias interacciones intencionalmente no incluidas en el modelo para hacer más fácil su interpretación. En este trabajo se resume esta metodología de análisis y se presenta un ejemplo de aplicación en una tabla de tasas de mortalidad estandarizadas para algunas comunidades autónomas de España.

Palabras clave: Análisis de correspondencias, modelos log-lineales, análisis exploratorio de datos..

Abstract

The relationship between log-linear models and analysis of correspondence can be used to develop a combined approach for the analysis of contingency tables. In this approach, log-linear models are used to evaluate the importance of interaction effects. Afterward, analysis of correspondence is used to evaluate the residuals. This approach leads to the visualization of interactions not included in order to get an easier interpretation. In this paper the method of analysis is summarized and an application is presented for a table containing mortality rates for some regions of Spain.

Key words: Analysis of correspondence, log-linear models, exploratory data analysis..

1. Introducción

El análisis de correspondencias y los modelos log-lineales son dos métodos estadísticos muy utilizados en el análisis de variables cualitativas (Van der Heijden et al. 1989). La información de este tipo se puede presentar en tablas de contingencia, que agrupan los individuos según las categorías que asumen para cada variable. Como resultado de esta agrupación se obtienen tablas de dimensión dos o mayor, según el número de variables en estudio.

El objetivo principal del análisis de correspondencias es encontrar la mejor representación simultánea de dos conjuntos, constituidos por las filas y columnas de una tabla de contingencia, a través de una

^aEstudiante de estadística. E-mail: dcbuitragoo@unal.edu.co

^bProfesor Asociado. E-mail: cepardot@unal.edu.co

reducción de dimensión que permita aislar el ruido para examinar las relaciones existentes entre las variables.

En el análisis de correspondencias no es necesario asumir una distribución de la cual provienen los datos ni un modelo hipotético que estos siguen. Por el contrario, los modelos log-lineales sí requieren estos dos supuestos. Las estimaciones de los parámetros del modelo se hacen por el método de máxima verosimilitud y se utilizan para calcular las frecuencias esperadas. Éstas se comparan con las frecuencias observadas para verificar la bondad de ajuste del modelo.

A partir de la propuesta de generalización de Escofier (1984) en cuanto al análisis factorial con referencia a un modelo se han estudiado las posibilidades de análisis que surgen al utilizar las dos metodologías en forma complementaria, lo cual permite mayor claridad en la interpretación de las relaciones existentes entre las variables.

En este trabajo se hace una revisión de algunos resultados que muestran las relaciones existentes entre los dos métodos y se resume una metodología de análisis combinado presentada en Van der Heijden et al. (1989), donde se interpreta el análisis de correspondencias como una herramienta para el análisis de los residuales de un modelo log-lineal.

Se muestran las ventajas que presenta esta metodología de análisis combinado, mediante su aplicación a los datos sobre mortalidad en las diferentes comunidades autónomas de España. Se presenta la aplicación del método para una tabla de contingencia donde algunas variables presentan anidamientos y se busca el modelo log-lineal asociado al análisis de correspondencias intra tablas donde hay una doble partición de filas y columnas.

2. Análisis de correspondencias y modelos log-lineales

2.1. Análisis de correspondencias simples (ACS)

En esta sección se hace una breve descripción del ACS siguiendo los resultados de Lebart et al. (1995).

Se parte de una tabla de contingencia \mathbf{K} , de dimensión $n \times p$, cuya entrada k_{ij} es el número de individuos que asumen la categoría i de la variable fila y j de la variable columna. El total de la tabla es $k = \sum_{i=1}^n \sum_{j=1}^p k_{ij}$. A partir de ésta, se construye una tabla (\mathbf{F}) de frecuencias relativas $f_{ij} = \frac{k_{ij}}{k}$.

Los perfiles fila y columna son $\left\{ \frac{f_{ij}}{f_{i.}}, j = 1, \dots, p \right\}$ y $\left\{ \frac{f_{ij}}{f_{.j}}, i = 1, \dots, n \right\}$ respectivamente, donde $f_{i.} = \sum_{j=1}^p f_{ij}$ y $f_{.j} = \sum_{i=1}^n f_{ij}$ son las frecuencias marginales de las filas y columnas.

Los n perfiles fila se ubican en un espacio euclidiano de dimensión $p - 1$. La distancia χ^2 permite interpretar las proximidades entre dos puntos fila o entre dos puntos columna. Entre dos categorías de la variable fila se define la distancia al cuadrado como

$$d^2(i, i') = \sum_{j=1}^p \frac{1}{f_{.j}} \left(\frac{f_{ij}}{f_{i.}} - \frac{f_{i'j}}{f_{i'.}} \right)^2 \quad (1)$$

e indica en qué medida las dos modalidades se relacionan en forma similar o no con las modalidades de la variable columna, dando más peso a las columnas de baja frecuencia. Cuando dos puntos fila están cerca, sus perfiles son similares. Cuando están muy separados, las dos filas se relacionan en forma diferente con las columnas. Cuando un punto fila está cerca al origen, su perfil es similar al perfil medio. En forma similar se calcula e interpreta la distancia χ^2 entre dos puntos columna, ubicados en un espacio de dimensión $n - 1$.

El objetivo del ACS es encontrar una representación simultánea de las filas y las columnas, en un espacio de menor dimensión, a través de una rotación de la nube de puntos, que retenga la mayor variabilidad posible.

La fórmula de reconstitución de la tabla de datos a partir de las coordenadas factoriales está dada por:

$$f_{ij} = f_{i.}f_{.j} \left(1 + \sum_{\alpha=1}^{p-1} \sqrt{\lambda_{\alpha}} \varphi_{\alpha j} \psi_{\alpha i} \right) \quad (2)$$

que en forma matricial puede verse como

$$\mathbf{F} = \mathbf{E} + \mathbf{D}_n \boldsymbol{\psi} \boldsymbol{\Lambda}^{1/2} \boldsymbol{\varphi}' \mathbf{D}_p \quad (3)$$

donde $\boldsymbol{\Lambda} = \text{diag}\{\lambda_1, \dots, \lambda_{p-1}\}$ es la matriz que contiene los valores propios del ACS, \mathbf{E} es la matriz de frecuencias esperadas bajo el modelo de independencia, cuyo término general es $f_{i.}f_{.j}$, $\boldsymbol{\psi}$ y $\boldsymbol{\varphi}$ son matrices que contienen las coordenadas factoriales de las filas y columnas, respectivamente y $\mathbf{D}_n = \text{diag}(f_{i.})$, $\mathbf{D}_p = \text{diag}(f_{.j})$ contienen los pesos de las filas y las columnas.

Esta fórmula muestra una descomposición de la matriz formada por las frecuencias observadas, f_{ij} , en una matriz de valores esperados bajo el modelo de independencia y unos residuales bajo dicho modelo.

El valor propio asociado a cada eje factorial es la inercia de la nube de puntos proyectada sobre éste. Por lo tanto, la inercia total es la suma de todos los valores propios, es decir, la traza de $\boldsymbol{\Lambda}$. Por otra parte, la estadística χ^2 de Pearson permite verificar si los desvíos entre las frecuencias observadas y las esperadas bajo el modelo de independencia son significativos. La inercia total de la nube de puntos (medida de su dispersión) es una partición de esta estadística, como se observa en la fórmula (4).

$$\text{traza}(\boldsymbol{\Lambda}) = \chi^2/n \quad (4)$$

Con lo anterior se observa que se puede obtener una visión de la interacción entre las variables fila y columna a partir del análisis de correspondencias, el cual se puede ver como el análisis de los residuales con respecto al modelo de independencia.

2.2. Generalización del análisis de correspondencias

La generalización del análisis de correspondencias presentada en Escofier (1984) es útil cuando el interés se centra en descomponer los residuales de otros modelos diferentes al de independencia. También permite asignar a los puntos fila y columna pesos diferentes a las frecuencias relativas marginales.

Sea \mathbf{A} la tabla de frecuencias relativas esperadas bajo algún modelo de interés y sean \mathbf{Q} y \mathbf{S} dos matrices de pesos de tamaños n y p respectivamente (diagonales, de traza uno y cuyas entradas diagonales q_i y s_j son números entre cero y uno).

Se construye la nube de n puntos fila, en el espacio de dimensión p , utilizando las siguientes coordenadas, para $i = 1, \dots, n$:

$$\left(\frac{f_{i1}}{q_i} - \frac{a_{i1}}{q_i}, \dots, \frac{f_{ip}}{q_i} - \frac{a_{ip}}{q_i} \right)' \quad (5)$$

La distancia al cuadrado entre dos filas está dada por

$$d^2(i, i') = \sum_j \frac{1}{s_j} \left(\frac{f_{ij} - a_{ij}}{q_i} - \frac{f_{i'j} - a_{i'j}}{q_{i'}} \right)^2 \quad (6)$$

Dos puntos fila están cerca si los desvíos entre sus filas correspondientes entre las matrices \mathbf{F} y \mathbf{A} son similares para todas las columnas. En forma simétrica se construye la nube de p puntos columna y se interpretan las distancias entre éstos.

Los términos generales de los centros de gravedad de las nubes de puntos son:

$$w_n(j) = \sum_i q_i \frac{f_{ij} - a_{ij}}{q_i} = f_{.j} - a_{.j}$$

$$w_p(i) = \sum_j s_j \frac{f_{ij} - a_{ij}}{s_j} = f_{i.} - a_{i.}$$

Lo anterior indica que la nube de puntos está centrada sólo cuando las frecuencias marginales son iguales en las matrices \mathbf{F} y \mathbf{A} . Por lo tanto, para que las distancias entre puntos estén bien representadas en los planos factoriales, estas marginales no deben ser muy diferentes.

La fórmula de reconstitución de la tabla de frecuencias observadas es: (Van der Heijden et al. 1989)

$$\mathbf{F} = \mathbf{A} + \mathbf{Q}\psi\mathbf{\Lambda}^{1/2}\varphi'\mathbf{S} \quad (7)$$

donde $\mathbf{\Lambda} = \text{diag}\{\lambda_1, \dots, \lambda_{p-1}\}$.

Comparando con la fórmula (3) se observa que se generaliza tanto la matriz de la cual se miden los desvíos como los pesos asignados a los puntos fila y columna.

Cuando $a_{i.} = f_{i.} = q_i$ y $a_{.j} = f_{.j} = s_j$, los resultados del análisis de correspondencias generalizado se pueden encontrar realizando ACS de la tabla de término general $f_{ij} - a_{ij} + f_{i.}f_{.j}$ (Escofier 1984).

2.3. Modelos log-lineales

Otra metodología que permite examinar las relaciones existentes entre variables categóricas que forman una tabla de contingencia es el análisis log-lineal. El modelo especifica la forma en que las frecuencias esperadas dependen de los niveles de las variables categóricas para cada celda, así como de sus interacciones.

En el caso de una tabla de contingencia bivariada, bajo la hipótesis de independencia, la frecuencia relativa esperada para cada celda (i, j) es el producto de las frecuencias observadas marginales, es decir $a_{ij} = f_{i.}f_{.j}$. A partir de esta expresión se obtiene el modelo log-lineal de independencia, presentado en (8), que sólo involucra los efectos principales de cada variable.

$$\ln a_{ij} = u + u_{1(i)} + u_{2(j)} \quad (8)$$

Por otra parte, el modelo saturado, que incluye todas las interacciones posibles es

$$\ln a_{ij} = u + u_{1(i)} + u_{2(j)} + u_{12(ij)} \quad (9)$$

Este modelo tiene en total np parámetros, con los que se consigue construir toda la tabla. Cuando las variables tienen muchas categorías el número de parámetros es grande y su interpretación puede ser difícil.

En tablas de mayor dimensión se presentan diferentes tipos de asociación entre las variables, como señala Agresti (2002). Para el caso de una tabla a tres vías, las tres variables son mutuamente independientes cuando la frecuencia esperada es $a_{ijk} = f_{i..}f_{.j.k}$, para todo i, j, k . El modelo log-lineal correspondiente es

$$\ln a_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} \quad (10)$$

En el segundo tipo de asociación, la variable 2 es conjuntamente independiente con las variables 1 y 3 si $a_{ijk} = f_{i.k}f_{.j.}$, para todo i, j, k . El modelo log-lineal está dado por:

$$\ln a_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{13(ik)} \quad (11)$$

Finalmente, las variables 1 y 2 son condicionalmente independientes dada la variable 3 si $a_{ij|k} = f_{i.|k}f_{.j|k}$, para todo i, j, k , o en forma equivalente

$$a_{ijk} = \frac{f_{i.k}f_{.jk}}{f_{..k}}$$

El modelo log-lineal para este caso, presentado en (12), indica que hay relación entre las variables uno y tres y entre dos y tres, pero no entre las variables uno y dos.

$$\ln a_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{13(ik)} + u_{23(jk)} \quad (12)$$

Los modelos log-lineales se pueden validar utilizando la estadística χ^2 o la estadística de razón de verosimilitud. Una vez se ha validado el modelo, es posible interpretar los parámetros y hacer inferencia sobre estos.

La interpretación de los parámetros algunas veces es difícil, especialmente cuando el número de categorías es grande, con lo cual habrá muchos parámetros de interacción; también cuando hay muchas variables y las interacciones son de orden alto. Por esta razón, puede ser útil buscar herramientas complementarias para el análisis de datos categóricos que faciliten la interpretación de los parámetros.

2.4. Complementariedad entre el AC y los modelos log-lineales

El AC puede ser útil para restringir los parámetros de interacción en un modelo log-lineal. En Van der Heijden et al. (1989) se presentan las siguientes relaciones entre los dos métodos.

Tomando el logaritmo natural de la fórmula de reconstitución (2), y teniendo en cuenta que si un número positivo x es cercano a cero se tiene la aproximación $\ln(1+x) \approx x$, entonces

$$\ln f_{ij} \approx u + u_{1(i)} + u_{2(j)} + \sum_{\alpha=1}^k \sqrt{\lambda_{\alpha}} \varphi_{\alpha j} \psi_{\alpha i} \quad (13)$$

donde $u = 0$, $u_{1(i)} = \ln f_{i.}$, $u_{2(j)} = \ln f_{.j}$. Si $k = 1$, se tiene

$$\ln f_{ij} \approx u + u_{1(i)} + u_{2(j)} + \phi u_{1(i)}^* u_{2(j)}^* \quad (14)$$

Entonces este último término puede interpretarse como la interacción entre las dos variables, que se descompone aproximadamente en forma log-multiplicativa en el AC, dando una visión de los parámetros que deben restringirse.

Cuando se estudian más de dos variables, se pueden construir tablas a dos vías realizando aplanamientos de la tabla multidimensional. Este procedimiento se explica en la sección 2.5.

En algunas situaciones es de interés estudiar modelos que sólo permiten independencia condicional o ausencia de interacciones de orden superior. Luego de plantear el modelo correspondiente, la generalización del AC, presentada en la sección 2.2, descompone sus residuales, mostrando sólo las interacciones que no se estudian en el modelo.

Van der Heijden & de Leeuw (1985) proponen usar el AC y los modelos log-lineales en forma complementaria, empleando cada uno de los métodos en etapas diferentes.

- En el primer paso el modelo log-lineal muestra qué variables están relacionadas, detectando interacciones importantes.
- En el segundo paso, se puede usar el AC en dos formas: primero cuando un modelo de interés no ajusta bien, el AC permite explorar los residuales. Segundo, al escoger un modelo que ajusta mal, se puede emplear el AC en lugar de la interpretación de los parámetros de interacción. En los dos contextos la complementariedad es útil cuando la tabla en estudio es grande, ya que una representación gráfica puede simplificar la interpretación.

2.5. Análisis de correspondencias de tablas de contingencia estructuradas

Para aplicar análisis de correspondencias simples a una tabla de contingencia con más de dos variables es necesario realizar combinaciones de sus categorías que lleven a una tabla bivariada. Estas combinaciones en adelante recibirán el nombre de aplanamientos de la tabla multidimensional.

Se definen dos grupos disyuntos de variables A y B. En cada uno se combinan las categorías de las variables, obteniendo dos variables cruzadas que forman una tabla a dos vías. Por ejemplo, en el caso de una tabla de tamaño $2 \times 2 \times 2$, formada por las variables I, J, K , se pueden combinar las categorías de

las variables I y J , obteniendo la variable cruzada que toma los valores $i_1j_1, i_1j_2, i_2j_1, i_2j_2$. Esto lleva a un aplanamiento de la tabla multivariada a una tabla de dimensión 4×2

Como indican Van der Heijden et al. (1989), se puede interpretar el AC como la diferencia entre los modelos $[AB]$ y $[A][B]$, y sirve para analizar sus residuales¹. Se debe tener en cuenta que si se pretende analizar la interacción entre dos variables en particular, no se deben realizar los cruces entre ellas.

2.5.1. Estructura de partición a una vía

Cuando el interés está en determinar las relaciones existentes entre tres variables categóricas I, J, K , hay varias formas de combinar las variables para aplanar la tabla de datos.

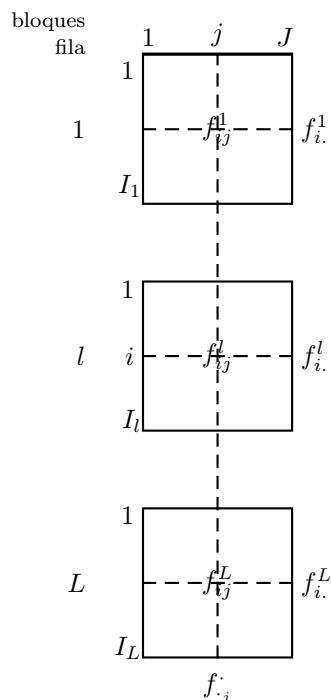


FIGURA 1: Esquema de la tabla de contingencia a tres vías, al aplanar por bloques fila

En primer lugar, se puede formar una tabla con estructura de partición en bloques fila, como la que se presenta en la figura 1. En este caso, las frecuencias marginales fila son $f_{i.}^l$ y marginales columna son $f_{.j}$, con las cuales se construye el modelo de independencia

$$a_{ij}^l = f_{i.}^l \cdot f_{.j} \quad (15)$$

Éste tiene asociado el modelo log-lineal $[IL][J]$, el cual tiene la forma

$$\ln a_{ij}^l = u + u_{I(i)} + u_{L(l)} + u_{J(j)} + u_{IL(il)} \quad (16)$$

En este caso el análisis de correspondencias se interpreta como la diferencia entre el modelo saturado $[IJL]$, dado por

$$\ln a_{ij}^l = u + u_{I(i)} + u_{L(l)} + u_{J(j)} + u_{IL(il)} + u_{IJ(ij)} + u_{LJ(lj)} + u_{IJL(ijl)} \quad (17)$$

¹Esta notación indica que el modelo incluye todas las interacciones entre las variables que se encuentran en el mismo paréntesis cuadrado. Por ejemplo, el modelo $[IJ][K]$ incluye sólo la interacción entre las variables I y J , además de los efectos principales

y el modelo $[IL][J]$, es decir, muestra las interacciones de primer orden restantes (IJ , LJ) y la interacción de segundo orden IJL .

Si la estructura de los datos incluye el anidamiento de las variables L e I , es decir, las posibles categorías de la variable I dependen de la categoría asumida en la variable L , entonces algunos parámetros $u_{IL(il)}$ y $u_{IJL(ijl)}$ toman el valor cero. Esto ocurre debido a que la tabla de datos contiene ceros estructurales, ya que algunas combinaciones entre las categorías de estas dos variables no son posibles en la población.

Otro modelo que se puede estudiar a partir de la tabla presentada en la figura 1 es el modelo de independencia intra bloques, el cual, como indican Escofier & Pagès (1992, p. 218) suprime la dispersión entre bloques dejando sólo la inercia al interior de cada bloque. Para cada bloque fila l , las marginales fila y columna son f_i^l y $f_{.j}^l$, y el total de la tabla es $f_{..}^l$. El modelo de independencia intra bloques fila es

$$a_{ij}^l = \frac{f_i^l f_{.j}^l}{f_{..}^l} \quad (18)$$

Esta expresión corresponde al modelo de independencia condicional de la fórmula (12), por lo tanto el modelo log-lineal asociado es $[IL][JL]$, es decir

$$\ln a_{ij}^l = u + u_{I(i)} + u_{J(j)} + u_{L(l)} + u_{IL(il)} + u_{JL(jl)} \quad (19)$$

El AC con respecto a este modelo descompone sus residuales, estudiando la diferencia entre el modelo saturado (17) y el modelo (19), es decir, muestra las interacciones restantes IJ e IJL .

2.5.2. Estructura de partición a dos vías

A continuación se considera el caso de una tabla a cuatro vías, como se muestra en la figura 2. Cada bloque F_{lj} es una tabla de contingencia cuya frecuencia total es $f_{..}^{lj}$ y sus frecuencias marginales son $f_{i.}^{lj}$ y $f_{.k}^{lj}$.

Como indican Escofier & Pagès (1992, p. 182), a partir de estos bloques se puede responder varios interrogantes. En primer lugar, se examina si existe una estructura común a las LJ tablas. Esto equivale a determinar la interacción entre las variables I y K condicional a las variables L y J . En segundo lugar, se pueden analizar las desviaciones entre estos bloques, es decir, la interacción entre las variables L y J . Por último, se puede hallar cuál de las dos interacciones anteriores tiene mayor presencia en los datos.

La inercia de la nube de puntos se descompone en inercia inter bloques e inercia intra bloques. Estas cantidades permiten resolver el último punto planteado, hallando el porcentaje de inercia de cada tipo.

El modelo log-lineal saturado para la tabla F a cuatro vías es

$$\begin{aligned} \ln a_{ik}^{lj} = & u + u_{L(l)} + u_{I(i)} + u_{J(j)} + u_{K(k)} \\ & + u_{LI(li)} + u_{LJ(lj)} + u_{LK(lk)} + u_{IJ(ij)} + u_{IK(ik)} + u_{JK(jk)} \\ & + u_{LIJ(lij)} + u_{LIK(lik)} + u_{LKJ(lkj)} + u_{IJK(ijk)} + u_{LIJK(lijk)} \end{aligned} \quad (20)$$

Por otra parte, el modelo de independencia asociado al análisis de correspondencias de la tabla bivariada (formada por los cruces LI y JK) es:

$$\ln a_{ik}^{lj} = u + u_{L(l)} + u_{I(i)} + u_{J(j)} + u_{K(k)} + u_{LI(li)} + u_{JK(jk)} \quad (21)$$

El ACS de la tabla formada por estos cruces muestra entonces las interacciones LJ , LK , IJ , IK y las de orden superior, que corresponden a la diferencia entre el modelo saturado (20) y el modelo (21).

También se puede trabajar el modelo de independencia intra bloques columna, en el cual las frecuencias relativas esperadas a_{ik}^{lj} dependen de las marginales fila y columna de cada bloque, $f_{i.}^{lj}$ y $f_{.k}^{lj}$ y del total del bloque, $f_{..}^{lj}$. En este modelo las frecuencias relativas esperadas son

$$a_{ik}^{lj} = \frac{f_{i.}^{lj} f_{.k}^{lj}}{f_{..}^{lj}} \quad (22)$$

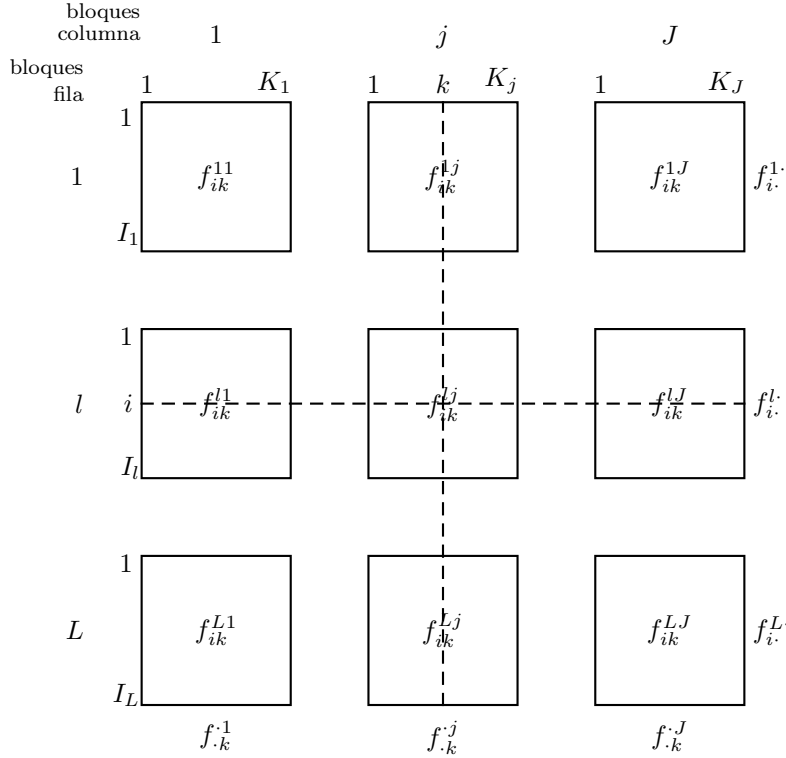


FIGURA 2: Esquema de la tabla de contingencia a cuatro vías, al aplanar en una estructura de doble partición

y el modelo log-lineal asociado es $[ILLJ][JK]$. El análisis de correspondencias se puede interpretar como la diferencia entre este modelo y el modelo saturado (20), es decir, muestra las interacciones de primer y segundo orden restantes IK , LK , ILK , IJK , LJK , y la interacción de tercer orden $IJKL$.

Por último, si se tiene en cuenta la partición tanto de las filas como de las columnas, se puede definir el modelo intra bloques, para examinar si existe una estructura común a las LJ tablas, es decir, si hay interacción entre las variables I y K condicional a las variables L y J . En este modelo las frecuencias relativas esperadas se expresan como

$$a_{ik}^{lj} = \frac{f_{i.}^{lj} f_{.k}^{lj}}{f_{..}^{lj}} \quad (23)$$

y el modelo log-lineal correspondiente es $[ILLJ]$, $[KLLJ]$, es decir

$$\begin{aligned} \ln a_{ik}^{lj} = & u + u_{L(l)} + u_{I(i)} + u_{J(j)} + u_{K(k)} + u_{LI(li)} + u_{LJ(lj)} \\ & + u_{LK(lk)} + u_{IJ(ij)} + u_{JK(jk)} + u_{LIJ(lij)} + u_{LKJ(lkj)} \end{aligned} \quad (24)$$

Con el análisis de correspondencias intra bloques se analizan las interacciones IK , ILK , IJK , $IJKL$, es decir, se estudia la diferencia entre el modelo (24) y el modelo saturado (20).

En la práctica algunas veces se encuentran tablas que involucran variables anidadas, es decir, donde las categorías de una variable no son siempre las mismas para todas las categorías de otra variable. Por ejemplo, en datos sobre mortalidad, como los presentados en la sección 3, hay causas de mortalidad específicas para algunos grupos poblacionales según el grupo de edad y el género. Por lo tanto se presenta un anidamiento entre las variables causa de mortalidad y grupo poblacional.

A continuación se considera el caso en el que la estructura de los datos es la que se presenta en la figura 2 y las categorías de la variable K dependen de la categoría asumida por J , es decir, están anidadas. Algunos parámetros de interacción que involucran estas dos variables simultáneamente toman el valor cero.

El modelo log-lineal (21) asociado al ACS de la tabla presentada en la figura 2 incluye la interacción JK , entre los grupos y las causas de mortalidad. Algunos de estos parámetros son cero, debido al anidamiento de las dos variables, y hay que tener esto en cuenta para realizar la interpretación. Así mismo, los parámetros que quedan para la interpretación en el ACS que involucran simultáneamente las variables J y K sólo muestran las interacciones posibles entre estas dos variables.

Lo anterior se extiende al caso del modelo de independencia intra bloques (23), donde también se tiene que algunos de los parámetros que incluyen estas dos variables toman el valor cero, debido al anidamiento.

3. Aplicación

En esta sección se utiliza información sobre las tasas de mortalidad estandarizada para las comunidades de España, calculadas a partir de la información sobre las tasas brutas de mortalidad suministrada por la Eurostat. Esta organización ha recogido las tasas de mortalidad para diferentes países de la Unión Europea, a nivel nacional y regional desde 1994 (Niederlaender 2006). La descripción de los cálculos para la estandarización se encuentra en el apéndice Apéndice A.

La clasificación de las causas de mortalidad que sigue Eurostat tiene 65 grupos que se basan en la clasificación ICD (International Statistical Classification of Diseases and Related Health Problems) desarrollada y mantenida por la Organización Mundial de la Salud .

La información que se utiliza se encuentra organizada de la siguiente forma. Se consideran las tasas de mortalidad por género y por dos grupos de edad, muertes prematuras (entre 35 y 64 años) y no prematuras (65 años o más). Con esto se construyen cuatro grupos poblacionales, masculino prematuro (MP) , masculino no prematuro (MNP), femenino prematuro (FP) y femenino no prematuro (FNP). Para cada uno se consideran las cinco causas de mortalidad más frecuentes en dicha población. Con esto se presenta anidamiento entre los grupos poblacionales y las causas de mortalidad.

Con el propósito de ilustrar los principales resultados expuestos en este trabajo, se consideran sólo las cinco comunidades autónomas españolas con mayor número de habitantes (Instituto Nacional de Estadística 2008) las cuales son Andalucía, Cataluña, Madrid, Valencia y Galicia. Además, para determinar la evolución en el tiempo de las diferentes causas de mortalidad se consideran dos periodos, de 1994 a 1996, y de 2004 a 2006. En cada periodo se calcula la tasa promedio para los tres años.

Entonces se tiene una tabla de datos a cuatro vías (tabla 1), conformada por las variables grupo poblacional (género-edad, identificada con J), causas de mortalidad (K), año (L) y comunidad (I).

3.1. Análisis de correspondencias simples

El modelo log-lineal asociado al ACS de la tabla 1, que presenta anidamiento entre las causas y los grupos poblacionales, está dado por la fórmula (21). Este modelo incluye las interacción entre el año y la comunidad y entre grupo y causa de mortalidad, teniendo en cuenta que algunos de estos últimos son cero debido al anidamiento.

Las estimaciones se realizaron en el programa SAS (SAS Institute 2002) utilizando el procedimiento CATMOD, que permite estimar modelos log-lineales para tablas con ceros estructurales. En la tabla 3 del apéndice se presenta la estadística de razón de verosimilitud, que indica que el modelo es significativo al 1 %.

En la tabla 5 se presentan las estimaciones de los parámetros y su significancia. Para los parámetros de interacción *año* \times *comunidad* se observa que la mortalidad por las causas consideradas en este trabajo, en el año 2005 es mayor en Andalucía que en las demás ciudades y menor en Galicia.

Con respecto a la interacción entre los grupos y las causas de mortalidad, se observa que la mayor asociación se presenta entre mortalidad prematura en mujeres y el cáncer de colon, y la menor asociación entre mortalidad no prematura y esta misma causa. Hay alta asociación entre este último grupo y la mortalidad por cáncer de pulmón.

TABLA 1: Causas de mortalidad en España

		Género masculino									
		MP					MNP				
Per.	Comunidad	15	34	44	35	36	34	36	15	35	40
94-96	Andalucía	98.0	83.1	43.3	21.8	32.8	776.2	722.8	492.4	505.8	539.3
	Cataluña	85.1	62.1	30.7	26.4	24.6	639.3	567.0	415.4	446.3	440.3
	Madrid	78.6	59.2	29.1	36.0	19.5	581.4	431.2	440.2	418.8	337.4
	Valencia	86.8	83.3	38.4	20.1	32.7	780.5	786.4	437.0	448.7	469.8
	Galicia	87.4	71.8	34.1	24.9	29.7	589.6	620.1	369.9	541.6	439.5
04-06	Andalucía	79.6	62.2	28.5	21.1	25.3	653.5	545.4	432.8	372.2	393.9
	Cataluña	76.7	42.7	21.2	18.6	16.8	469.7	348.2	384.1	327.0	360.3
	Madrid	64.7	40.9	15.0	13.1	13.6	466.5	262.9	375.2	275.5	260.8
	Valencia	82.1	64.8	24.3	19.7	21.2	659.4	417.5	387.5	305.7	381.2
	Galicia	87.8	56.7	22.3	22.3	20.8	484.8	377.0	328.3	344.7	325.0

		Género femenino									
		FP					FNP				
Per.	Comunidad	17	36	34	11	15	36	35	34	28	27
94-96	Andalucía	36.1	15.3	18.0	10.4	5.8	675.1	554.7	475.0	106.7	176.6
	Cataluña	37.3	11.6	10.8	10.3	7.2	464.0	414.5	334.7	174.9	104.2
	Madrid	37.2	9.2	8.0	10.1	7.8	363.8	381.9	291.8	125.3	97.3
	Valencia	36.5	14.3	14.3	10.5	6.0	712.1	474.0	423.1	149.1	174.2
	Galicia	32.3	15.4	10.0	8.8	6.7	589.1	524.6	314.4	124.9	97.0
04-06	Andalucía	27.8	11.9	12.3	9.1	9.2	488.6	389.4	378.4	103.3	128.4
	Cataluña	26.2	7.9	6.8	8.8	12.9	274.0	298.8	238.4	199.5	96.2
	Madrid	24.0	6.8	6.8	8.0	13.8	222.0	270.9	240.2	95.3	55.9
	Valencia	26.4	10.5	12.7	8.4	12.9	374.3	323.7	375.7	143.7	123.3
	Galicia	25.8	8.8	8.1	8.8	11.8	343.5	301.2	242.5	107.0	74.5

Causas de mortalidad: 15-cáncer de pulmón, 34-isquemia cardíaca, 44-enfermedad crónica del hígado, 35-otras enfermedades cardíacas, 36- cerebro vascular, 40- enfermedad pulmonar obstructiva, 17-cáncer de seno, 11-cáncer de colon, 28-desórdenes mentales o de comportamiento, 27- diabetes

A continuación se realiza el ACS que permite descomponer los residuales de este modelo. En el histograma de valores propios (figura 3) se observa que los datos tienen una primera dirección de inercia muy importante y los tres primeros valores propios retienen el 90.6 % de la inercia. Por esta razón se realiza el análisis con base en los tres primeros ejes factoriales.

Según lo visto anteriormente, este análisis permite estudiar las interacciones $año \times grupo$, $año \times causa$, $comunidad \times grupo$, $comunidad \times causa$ y las de orden superior. En la figura 4 se observa que Andalucía y Valencia están muy asociadas a isquemia cardíaca, especialmente para el año 2005, en 1995 también se observa alta incidencia de mortalidad por enfermedades cerebro vasculares. Galicia presenta alta mortalidad debido a otras causas cardíacas en 1995 y a cáncer de seno y enfermedad pulmonar obstructiva en 2005. Madrid es la más asociada a mortalidad por cáncer de pulmón, y la asociación es mayor para el año 2005. En Cataluña se presentan más casos de mortalidad por desórdenes mentales o



FIGURA 3: Valores propios del ACS

de comportamiento que en el promedio para el año 2005, y en este año está muy asociada a mortalidad por cáncer de pulmón y colon.

La interacción año, comunidad y grupo no es clara en este plano. Este análisis está dominado por la diferencia que se presenta en las causas de mortalidad entre los dos años, como se observa en la figura 5.

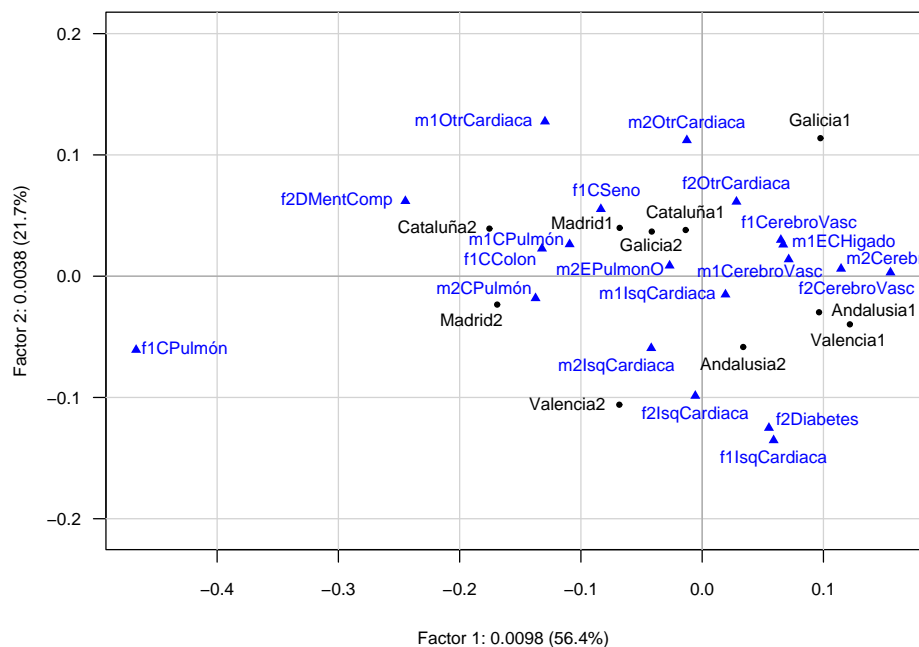


FIGURA 4: Primer plano factorial del ACS

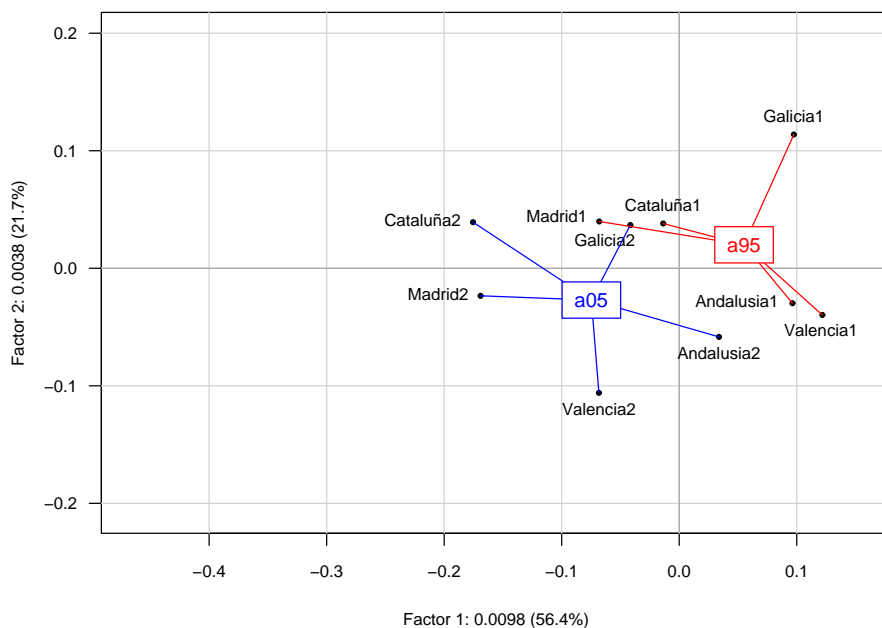


FIGURA 5: Primer plano factorial del ACS mostrando años

3.2. Análisis de correspondencias para el modelo de independencia intra bloques

El modelo log-lineal asociado al AC intra bloques de la tabla 1, que presenta anidamiento entre las causas y los grupos poblacionales, está dado por la fórmula (24). En la tabla 4 del apéndice se presenta la estadística de razón de verosimilitud, que indica que el modelo es significativo. Las interacciones entre $\text{año} \times \text{comunidad}$, $\text{año} \times \text{grupo}$, $\text{año} \times \text{comunidad} \times \text{grupo}$ y $\text{año} \times \text{grupo} \times \text{causa}$ no son significativas.

En la tabla 6 se presentan las estimaciones de los parámetros y su significancia. Se observa que las causas de mortalidad más asociadas al año 2005 son el cáncer de colon en mujeres de edad avanzada y cáncer de pulmón en mujeres jóvenes. Además, es baja la asociación entre mortalidad no prematura por cáncer de pulmón en mujeres con este año.

La mayor asociación entre grupo y causa de mortalidad se presenta entre mujeres jóvenes y cáncer de colon, y también es alta la mortalidad por enfermedades cerebro vasculares para este grupo. Por el contrario, la mortalidad por esta causa en hombres jóvenes es baja.

En cuanto a la interacción entre año y causa de mortalidad, se observa que la mayor asociación se presenta entre el año 2005 y enfermedades cerebro vasculares, y las demás asociaciones con este año no son significativas al 5%.

A continuación se realiza el AC intra bloques que permite descomponer los residuales de este modelo. En el histograma de valores propios (figura 6) se observa que los datos tienen una primera dirección de inercia muy importante y los dos primeros valores propios retienen el 73.3% de la inercia. Por esta razón se realiza el análisis con base en los primeros dos ejes factoriales.

En la figura 7 se presenta el primer plano factorial del AC intra bloques, donde se centra por años y grupos. Esto permite ver asociaciones entre las causas y las comunidades con más claridad, centrando cada bloque en el promedio de la representación.

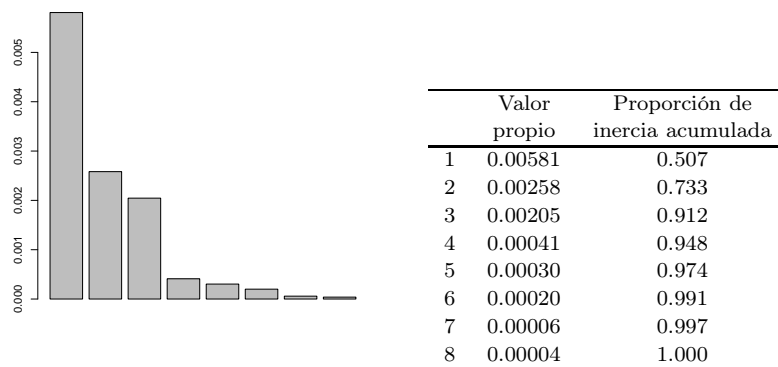


FIGURA 6: Valores propios del AC intra bloques

En este plano se observa que Galicia está muy asociada a mortalidad no prematura por otras enfermedades cardíacas, mientras que Madrid presenta mayor mortalidad por cancer de pulmón que el promedio en hombres jóvenes. Por otra parte, en 1995 Cataluña está muy asociada a mortalidad por cáncer de seno y colon en mujeres jóvenes y cáncer de pulmón en hombres jóvenes, pero en 2005 esta comunidad se asocia más a otras causas cardíacas en hombres jóvenes y cáncer de pulmón en mujeres jóvenes.

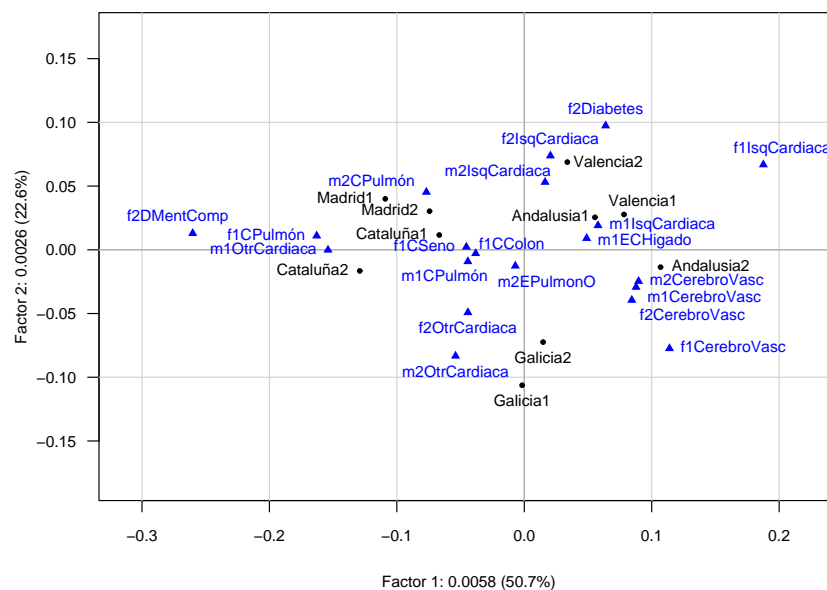


FIGURA 7: Primer plano factorial del AC intra bloques

Valencia y Andalucía tienen perfiles de mortalidad muy similares en 1995, y están muy asociadas a mortalidad por isquemia cardíaca y enfermedad crónica del hígado en hombres jóvenes. Pero en 2005 se presentan mayores diferencias entre las causas de mortalidad de las dos ciudades. Valencia se encuentra más asociada a mortalidad no prematura por isquemia cardíaca y diabetes en mujeres, mientras que Andalucía presenta alta mortalidad por causas cerebro vasculares en todos los grupos en estudio.

4. Conclusiones

- La exploración de las relaciones existentes entre los modelos log-lineales y el análisis de correspondencias permite encontrar una metodología para el estudio de tablas múltiples que facilita la interpretación de los resultados obtenidos mediante cada una de las técnicas. Al ver el AC como una herramienta para el análisis de los residuales de un modelo log-lineal se facilita la interpretación de las interacciones presentes en los planos factoriales.
- Es posible aplicar la metodología de análisis combinado utilizando extensiones del AC que se encuentran en la literatura, y que pueden extraer interacciones que ocultan otras en las cuales también pueden haber interés.
- El ACS para la tabla de tasas de mortalidad de algunas comunidades autónomas de España mostró las causas más asociadas a cada año. En 1995 se presenta alta mortalidad por enfermedades cerebro vasculares y otras causas cardíacas, mientras que el año 2005 está más asociado a mortalidad por cáncer de pulmón, cáncer de colon e isquemia cardíaca. Por otra parte, el modelo log-lineal asociado al AC intra mostró que en el grupo de mujeres jóvenes se presenta alta mortalidad por cáncer de colon y enfermedades cerebro vasculares, mientras que esta causa es poco frecuente en hombres jóvenes. También es alta la mortalidad por enfermedades cerebro vasculares en 2005. Con el AC intra se vio alta mortalidad en Galicia por otras causas cardíacas, mientras que Madrid y Cataluña están más asociadas a mortalidad por cáncer de pulmón. Valencia y Andalucía son las comunidades más asociadas a isquemia cardíaca y enfermedades cerebro vasculares.

Software

Para el cálculo de las estimaciones de los modelos log-lineales se utilizó el procedimiento CATMOD del programa SAS (SAS Institute 2002).

Los resultados del análisis de correspondencias simples se obtuvieron a través del paquete *ade4* (Chessel et al. 2004) del programa R (R Development Core Team 2008) y los del análisis de correspondencias intra a través del paquete *pamctdp* (Pardo 2008). Las gráficas de estos dos análisis se realizaron con el paquete *FactoClass* (Pardo & DelCampo 2007).

Referencias

- Agresti, A. (2002), *Categorical Data Analysis*, John Wiley & Sons, New York.
- Chessel, D., Dufour, A.-B. & Thioulouse, J. (2004), 'The ade4 package-I- One-table methods', *R News* **4**, 5–10.
- Escofier, B. (1984), 'Analyse factorielle en reference a un modele. Application a l'analyse de tableaux d'echange', *Revue de Statistique Appliquée* **32**(4), 25–36.
- Escofier, B. & Pagès, J. (1992), *Análisis factoriales simples y múltiples. Objetivos, métodos e interpretación*, Universidad del País Vasco, Bilbao.
- Instituto Nacional de Estadística (2008), *Series históricas de población*, España. Consultado en septiembre de 2008.
*<http://www.ine.es/jaxi/menu.do?type=pcaxis&path=%2Ft20%2Fe245%2Fp05&file=inebase&L=&divi=&his=>
- Jouglà, E., Salem, G., Gancel, S. & Michel, V. (2003), Health statistics - Atlas on Mortality in the European Union, Technical report, Eurostat.
- Lebart, L., Morineau, A. & Piron, M. (1995), *Statistique exploratoire multidimensionnelle*, Dunod, Paris.

- Niederlaender, E. (2006), Causes of Death in the EU, Technical report, Eurostat. Consultado en septiembre de 2008.
*http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-NK-06-010/EN/KS-NK-06-010-EN.PDF
- Pardo, C. & DelCampo, P. (2007), ‘Combinación de métodos factoriales y de análisis de conglomerados en R: el paquete FactoClass’, *Revista Colombiana de Estadística* **30**(2), 231–245.
*www.matematicas.unal.edu.co/revcoles
- Pardo, C. E. (2008), *pamctdp: Principal Axes Methods for Contingency Tables With Partition Structures on Rows and Columns*. R package version 0.0.3.
*cepardot@unal.edu.co
- R Development Core Team (2008), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
*<http://www.R-project.org>
- SAS Institute (2002), *SAS/STAT User’s Guide: Statistics.*, Cary, NC, USA.
- Van der Heijden, P., de Falguerolles, A. & de Leeuw, J. (1989), ‘A Combined Approach to Contingency Table Analysis using Correspondence Analysis and Log-linear Analysis’, *Applied Statistics* **38**, 249–292.
- Van der Heijden, P. & de Leeuw, J. (1985), ‘Correspondence Analysis used Complementary to Log-linear Analysis’, *Psychometrika* **50**, 429–447.

Apéndice A. Descripción de los cálculos para la tabla 1.

Las tasas brutas de mortalidad para las comunidades autónomas de España, promediadas cada tres años consecutivos, son suministradas por la Eurostat. Estas tasas de mortalidad están disponibles por grupos de edad, que tienen un rango de cinco años.

Estos datos presentan alta influencia de la estructura de edades en la población (Jougla et al. 2003), por lo tanto es útil estandarizar las tasas de mortalidad a una población de referencia, en este caso la de la Comunidad Europea (tabla 2).

Las tasas de mortalidad estandarizadas por el método directo se calculan a partir de la fórmula

$$SDR_j = \sum_{i=1}^n P_i T_{ji}$$

donde n es el número de grupos de edad, SDR_j es la tasa de mortalidad estandarizada para la subpoblación j , P_i es el porcentaje del grupo de edad i en la población estándar europea y T_{ji} es la tasa de mortalidad bruta en la subpoblación j para el grupo i de edad.

Así, para los cuatro grupos poblacionales definidos en la sección 3, las tasas de mortalidad estandarizadas se calculan de la siguiente forma, utilizando los porcentajes de cada grupo de edad en las poblaciones prematura y no prematura, presentados en la tabla 2:

$$SDR_{MP} = 0.179T_{M,35-39} + 0.179T_{M,40-44} + 0.179T_{M,45-49} + 0.179T_{M,50-54} + 0.154T_{M,55-59} + 0.128T_{M,60-64}$$

$$SDR_{MNP} = 0.364T_{M,65-69} + 0.273T_{M,70-74} + 0.182T_{M,75-79} + 0.091T_{M,80-84} + 0.091T_{M,85+}$$

$$SDR_{FP} = 0.179T_{F,35-39} + 0.179T_{F,40-44} + 0.179T_{F,45-49} + 0.179T_{F,50-54} + 0.154T_{F,55-59} + 0.128T_{F,60-64}$$

$$SDR_{FNP} = 0.364T_{F,65-69} + 0.273T_{F,70-74} + 0.182T_{F,75-79} + 0.091T_{F,80-84} + 0.091T_{F,85+}$$

TABLA 2: Población estándar europea.

Grupo de edad (años)	Población estándar europea	P_P (%)	P_{NP} (%)
0	1600		
1-4	6400		
5-9	7000		
10-14	7000		
15-19	7000		
20-24	7000		
25-29	7000		
30-34	7000		
35-39	7000	17.9	
40-44	7000	17.9	
45-49	7000	17.9	
50-54	7000	17.9	
55-59	6000	15.4	
60-64	5000	12.8	
65-69	4000		36.4
70-74	3000		27.3
75-79	2000		18.2
80-84	1000		9.1
85+	1000		9.1
Total	100000		

Fuente: Jouglá et al. (2003)

Apéndice B. Código en R para el ACS y el AC intra.

```
library(ade4)
AC2=dudi.coa(datos$tab,scannf = F, nf = 3)
library(FactoClass)
planfac(AC2,cframe=1.0, all.point=T,cex.row=0.8,cex.col=0.8)

#Proyección de los centros de gravedad de los años
planfac(AC2,cframe=1.0, all.point=T,cex.row=0.8,cex.col=0.8,Tcol=F)
s.class(AC2$li,datos$rbfac,wt=AC2$lw,cpoint = 0,cellipse = 0,
        col=c("blue","red","green","black"),sub="Plano1-2 Grupos especies",
        add.p = TRUE)

library(pamctdp) ww <- witwit.model(AC2, datos$rb1, datos$cb1,scannf=FALSE,nf=4,model="B")
planfac(ww)
```

Apéndice C. Código en SAS para la estimación de los modelos log-lineales.

```
/*Modelo log-lineal asociado al ACS*/
proc catmod data=Datos; weight tasa;
  model a_no*region*grupo*causa=_response_ /
    missing=structural zero=sampling
    freq pred=freq /*noparm noresponse oneway*/;
  loglin a_no region grupo causa a_no*region grupo*causa;
  response /out=salidas_simple_no_anida; quit;

/*Modelo intra*/
proc catmod data=Datos; weight tasa;
  model a_no*region*grupo*causa=_response_ /
    missing=structural zero=sampling
    freq pred=prob /*noparm noresponse oneway*/;
  loglin a_no region grupo causa region*grupo a_no*region a_no*grupo
  a_no*causa causa*grupo a_no*grupo a_no*region*grupo a_no*causa*grupo;
  response /out=salidas_intra_no_anida; quit;
```


Apéndice D. Principales resultados de los modelos log-lineales.

TABLA 3: Bondad de ajuste del modelo log-lineal asociado al ACS

Fuente	gl	Chi-cuadrado	Pr>ChiSq
año	1	903.86	<.0001
comunidad	4	725.25	<.0001
grupo	3	4749.74	<.0001
causa	9	2807.51	<.0001
año×comunidad	4	10.57	0.0319
grupo×causa	7*	1014.45	<.0001
Razón de verosimilitud	171	683.12	<.0001

Nota: los efectos señalados con * contienen uno o más parámetros redundantes o restringidos

TABLA 4: Bondad de ajuste del modelo log-lineal intra bloques

Fuente	gl	Chi-cuadrado	Pr>ChiSq
año	1	1.78	0.1816
comunidad	4	106.74	<.0001
grupo	3	4671.25	<.0001
causa	9	2766.00	<.0001
región×grupo	12	33.57	0.0008
año×comunidad	4	1.12	0.8908
año×grupo	3	1.27	0.7360
año×causa	9	60.86	<.0001
grupo×causa	7*	985.49	<.0001
año×comunidad×grupo	12	7.41	0.8290
año×grupo×causa	7*	9.77	0.2022
Razón de verosimilitud	128	448.60	<.0001

Nota: los efectos señalados con * contienen uno o más parámetros redundantes o restringidos

TABLA 5: Estimación de los parámetros del modelo log-lineal asociado al ACS

Parámetro		Estimación	Error estándar	Chi-cuadrado	Pr >ChiSq	
año	a05	-0.1530	0.00509	903.86	<.0001	
comunidad	Andalucía	0.1836	0.00947	375.95	<.0001	
	Cataluña	-0.0558	0.0103	29.10	<.0001	
	Galicia	-0.0355	0.0103	11.91	0.0006	
	Madrid	-0.2077	0.0110	357.55	<.0001	
grupo	f1	-2.3055	0.0731	994.18	<.0001	
	f2	1.1202	0.0293	1460.12	<.0001	
	m1	-0.5451	0.0388	197.30	<.0001	
causa	CColon	-3.6003	0.2854	159.08	<.0001	
	CPulmón	1.1078	0.0573	374.19	<.0001	
	CSeno	1.8859	0.0949	395.06	<.0001	
	CerebroVasc	0.6276	0.0414	229.95	<.0001	
	DMentComp	-0.0824	0.0601	1.88	0.1705	
	Diabetes	-0.2472	0.0610	16.43	<.0001	
	ECHigado	0.0494	0.0697	0.50	0.4785	
	EPulmonO	0.3956	0.0557	50.35	<.0001	
	IsqCardiaca	0.8309	0.0520	254.89	<.0001	
	año×comunidad	a05 Andalucía	0.0251	0.00947	7.01	0.0081
		a05 Cataluña	0.00987	0.0103	0.91	0.3401
		a05 Galicia	-0.0208	0.0103	4.07	0.0437
		a05 Madrid	-0.00725	0.0110	0.44	0.5092
grupo×causa	f1 CColon	4.2856	0.2702	251.63	<.0001	
	f1 CPulmón	-0.4129	0.1507	7.51	0.0061	
	f1 CSeno	
	f1 CerebroVasc	0.2388	0.1033	5.34	0.0208	
	f1 DMentComp	
	f1 Diabetes	
	f1 ECHigado	
	f1 EPulmonO	
	f1 IsqCardiaca	
	f2 CColon	-3.5766	0.1885	360.18	<.0001	
	f2 CPulmón	1.0964	0.1776	38.12	<.0001	
	f2 CSeno	
	f2 CerebroVasc	0.5106	0.0427	143.26	<.0001	
	f2 DMentComp	
	f2 Diabetes	
	f2 ECHigado	
	f2 EPulmonO	
	f2 IsqCardiaca	
	m1 CColon	
	m1 CPulmón	
	m1 CSeno	
	m1 CerebroVasc	-0.7693	0.0668	132.53	<.0001	
	m1 DMentComp	
	m1 Diabetes	
	m1 ECHigado	
	m1 EPulmonO	
	m1 IsqCardiaca	

TABLA 6: Estimación de los parámetros del modelo log-lineal intra bloques

Parámetro	Estimación	Error estándar	Chi-cuadrado	Pr >ChiSq
año05	-0.0521	0.039	1.78	0.1816
Andalucía	0.1505	0.0215	48.95	<.0001
Cataluña	-0.0542	0.0227	5.71	0.0169
Galicia	-0.0187	0.0225	0.69	0.4072
Madrid	-0.1775	0.0234	57.66	<.0001
f1	-2.3009	0.0738	971.73	<.0001
f2	1.1148	0.0296	1417.82	<.0001
m1	-0.5473	0.0393	194.27	<.0001
CColon	-3.5777	0.2894	152.85	<.0001
CPulmón	1.1167	0.0578	373.69	<.0001
CSeno	1.8816	0.0959	384.59	<.0001
CerebroVasc	0.6124	0.042	212.36	<.0001
DMentComp	-0.0772	0.0607	1.62	0.2036
Diabetes	-0.2534	0.0617	16.86	<.0001
ECHigado	0.034	0.0711	0.23	0.6325
EPulmonO	0.3922	0.0564	48.35	<.0001
IsqCardiaca	0.8303	0.0527	248.33	<.0001
Andalucía f1	-0.0648	0.0558	1.35	0.2456
Andalucía f2	0.0704	0.0243	8.43	0.0037
Andalucía m1	-0.0249	0.0362	0.47	0.4912
Cataluña f1	0.0303	0.0582	0.27	0.6033
Cataluña f2	-0.0163	0.0259	0.4	0.5288
Cataluña m1	-0.0232	0.0388	0.36	0.5489
Galicia f1	-0.0263	0.0585	0.2	0.6531
Galicia f2	-0.0191	0.0257	0.55	0.4564
Galicia m1	0.0692	0.0374	3.42	0.0645
Madrid f1	0.0946	0.0596	2.52	0.1127
Madrid f2	-0.0898	0.027	11.08	0.0009
Madrid m1 -	0.00332	0.0403	0.01	0.9344
05 Andalucía	0.0125	0.0215	0.34	0.5614
año05 Cataluña	0.0026	0.0227	0.01	0.9086
año05 Galicia	0.00119	0.0225	0	0.9577
año05 Madrid	-0.0231	0.0234	0.97	0.3235
05 f1 -0	0.00883	0.0738	0.01	0.9048
año05 f2	0.0142	0.0296	0.23	0.6303
año05 m1	-0.026	0.0393	0.44	0.5082
05 CColon	0.6916	0.2894	5.71	0.0168
año05 CPulmón	0.02	0.0578	0.12	0.7285
año05 CSeno	-0.0995	0.0959	1.08	0.2996
año05 CerebroVasc	-0.1601	0.042	14.52	0.0001
año05 DMentComp	0.0124	0.0607	0.04	0.8378
año05 Diabetes	-0.1163	0.0617	3.55	0.0596
año05 ECHigado	-0.1535	0.0711	4.65	0.031
año05 EPulmonO	-0.0981	0.0564	3.03	0.0819
año05 IsqCardiaca	-0.0736	0.0527	1.95	0.1622
f1 CColon	4.2687	0.2747	241.44	<.0001
f1 CPulmón	-0.4566	0.1545	8.73	0.0031
f1 CSeno
f1 CerebroVasc	0.2464	0.1046	5.55	0.0185
f1 DMentComp
f2 Diabetes
f2 ECHigado
f2 EPulmonO
f2 IsqCardiaca

TABLA 6. (Continuación) Estimación de los parámetros del modelo log-lineal intra bloques

Parámetro	Estimación	Error estándar	Chi-cuadrado	Pr > ChiSq
m1 CColon
m1 CPulmón
m1 CSeno
m1 CerebroVasc	-0.7593	0.0678	125.57	<.0001
m1 DMentComp
m1 Diabetes
m1 ECHigado
m1 EPulmonO
m1 IsqCardiaca
año05 Andalucía f1	-0.0221	0.0558	0.16	0.6926
año05 Andalucía f2	0.0169	0.0243	0.49	0.4854
año05 Andalucía m1	-0.00859	0.0362	0.06	0.8125
año05 Cataluña f1	-0.0185	0.0582	0.1	0.7503
año05 Cataluña f2	0.0223	0.0259	0.74	0.3893
año05 Cataluña m1	-0.00375	0.0388	0.01	0.9229
año05 Galicia f1	0.015	0.0585	0.07	0.797
año05 Galicia f2	-0.0441	0.0257	2.95	0.086
año05 Galicia m1	0.0459	0.0374	1.5	0.2205
año05 Madrid f1	0.0137	0.0596	0.05	0.8182
año05 Madrid f2	0.0203	0.027	0.56	0.4529
año05 Madrid m1	-0.0527	0.0403	1.71	0.1916
año05 f1 CColon	-0.7061	0.2747	6.61	0.0102
año05 f1 CPulmón	0.3371	0.1545	4.76	0.0291
año05 f1 CSeno
año05 f1 CerebroVasc	0.0408	0.1046	0.15	0.6965
año05 f1 DMentComp
año05 f1 Diabetes
año05 f1 ECHigado
año05 f1 EPulmonO
año05 f1 IsqCardiaca
año05 f2 CColon	0.4771	0.1922	6.16	0.0131
año05 f2 CPulmón	-0.2866	0.1811	2.5	0.1136
año05 f2 CSeno
año05 f2 CerebroVasc	-0.0528	0.0432	1.49	0.222
año05 f2 DMentComp
año05 f2 Diabetes
año05 f2 ECHigado
año05 f2 EPulmonO
año05 f2 IsqCardiaca
año05 m1 CColon
año05 m1 CPulmón
año05 m1 CSeno
año05 m1 CerebroVasc	0.0573	0.0678	0.72	0.3977
año05 m1 DMentComp
año05 m1 Diabetes
año05 m1 ECHigado
año05 m1 EPulmonO
año05 m1 IsqCardiaca