

# La práctica del análisis de correspondencias

**MICHAEL GREENACRE**

Catedrático de Estadística en la Universidad Pompeu Fabra

---

Separata del capítulo 18

## **Análisis de correspondencias múltiples**

Primera edición: julio 2008

ISBN: 978-84-96515-71-0

Traducción: Jordi Comas Angelet  
Revisión: Carles M. Cuadras Avellana

© **Michael Greenacre, 2008**  
© de la edición en español, **Fundación BBVA, 2008**

[www.fbbva.es](http://www.fbbva.es)



## Análisis de correspondencias múltiples

Hasta ahora hemos analizado la relación entre dos variables categóricas, o entre dos conjuntos de variables categóricas, en las que las variables fila eran diferentes de las variables columna. En cambio, en este capítulo y en los dos siguientes, analizaremos la relación existente entre variables similares, mediante el *análisis de correspondencias múltiples*, de forma abreviada ACM. Investigaremos el tipo de asociación existente entre variables y su intensidad. Podemos llevar a cabo el ACM sobre una matriz que contenga los datos codificados de forma binaria, la *matriz binaria*, o bien sobre una matriz formada por todos los cruzamientos posibles entre las variables, la *matriz de Burt*. Ambas posibilidades, muy relacionadas entre sí, presentan algunos inconvenientes que intentaremos solucionar en el capítulo 19, en el que presentamos versiones mejoradas del ACM.

### Contenido

Un conjunto de variables categóricas «homogéneas» .....	185
Matriz binaria .....	186
Definición 1 del ACM: AC de la matriz binaria .....	186
Inercia de la matriz binaria .....	187
Matriz de Burt .....	189
Definición 2 del ACM: AC de la matriz de Burt .....	189
Comparación del ACM de las matrices binaria y de Burt .....	190
Inercia de la matriz de Burt .....	191
Situación de variables adicionales en el mapa .....	191
Interpretación de los puntos adicionales .....	192
RESUMEN: Análisis de correspondencias múltiples .....	193

En este capítulo analizaremos la relación existente entre más de dos variables, generalmente en el contexto de un solo fenómeno de interés. Por ejemplo, las cuatro variables que vimos en el capítulo 17, sobre el trabajo de las mujeres, podrían ser nuestros datos de interés. También podrían ser las respuestas a preguntas relacionadas con la opinión de la gente sobre la ciencia, o datos que describan

Un conjunto de variables  
categóricas  
«homogéneas»

las condiciones ambientales de una serie de localidades. Lo importante es que las variables sean «homogéneas», es decir, que sean sustantivamente similares. Por ejemplo, no debemos mezclar variables de opinión con variables demográficas.

**Matriz binaria**

Consideremos las mismas cuatro variables que analizamos en el capítulo 17. Para evitar grandes diferencias culturales entre países, utilizaremos sólo datos de Alemania, incluyendo las muestras de Alemania del Este y Alemania Occidental, y así llegamos a un total de 3418 encuestados ( $N = 3418$ ). (Hemos omitido, de las muestras originales, tres casos para los que faltaban algunos datos demográficos: véase el apéndice de cálculo, B.) Nos centraremos en las cuatro preguntas etiquetadas de 1 a 4 sobre el trabajo de las mujeres. Cada pregunta puede tomar cuatro valores categóricos, que al igual que antes etiquetamos de la manera siguiente:  $T$  (trabajo a tiempo completo),  $t$  (trabajo a tiempo parcial),  $C$  (permanecer en casa) y  $?$  (no sabe/no contesta). La *matriz binaria* resultante será una matriz de  $3418 \times 16$  en la que hemos codificado todas respuestas de forma binaria. En la tabla de la imagen 18.1 mostramos la codificación para las seis primeras filas. Las 16 columnas de la derecha corresponden a la codificación binaria de las 16 posibles respuestas. Para el primer individuo, por ejemplo, las respuestas a las cuatro primeras preguntas son: 1, 3, 2 y 2, es decir  $T, t, C$  y  $?$ , que hemos codificado como 1 0 0 0, 0 0 1 0, 0 1 0 0 y 0 1 0 0, respectivamente.

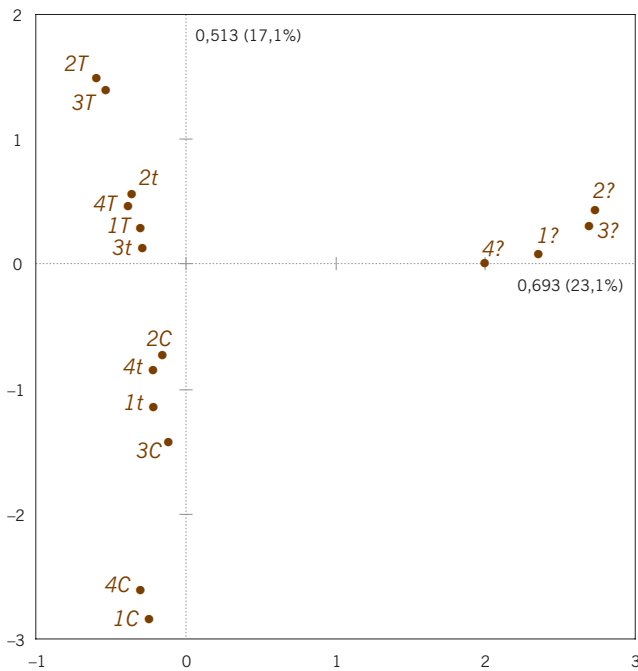
**Definición 1 del ACM: AC de la matriz binaria**

Podemos definir el ACM como el AC de la matriz binaria. Este análisis proporciona coordenadas para las 3418 filas y las 16 columnas. En el mapa de la imagen 18.2 mostramos las posiciones de las 16 categorías. El primer eje principal muestra que las cuatro categorías de «no respuesta» se hallan juntas, oponiéndose a todas las variables sustantivas. En el análisis anterior de estas preguntas (mapa de la imagen 17.4), en el que relacionamos las respuestas con variables demográficas, las no respuestas no ejercían un papel importante en ninguno de los dos primeros ejes. Sin embargo ahora, debido a que estamos interesados en la relación

**Imagen 18.1:**  
 Datos originales y codificación binaria correspondiente a los seis primeros encuestados de  $N = 3418$

Preguntas				Pregunta 1				Pregunta 2				Pregunta 3				Pregunta 4			
1	2	3	4	T	t	C	?	T	t	C	?	T	t	C	?	T	t	C	?
1	3	2	2	1	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0
2	3	3	2	0	1	0	0	0	0	1	0	0	0	1	0	0	1	0	0
4	3	3	2	0	0	0	1	0	0	1	0	0	0	1	0	0	1	0	0
4	4	4	4	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
4	4	4	4	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
1	3	2	1	1	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

... y así sucesivamente para 3418 filas



**Imagen 18.2:**  
 Mapa del ACM correspondiente a las cuatro preguntas sobre el trabajo de las mujeres; inercia total = 3, porcentaje de inercia del mapa: 40,2%

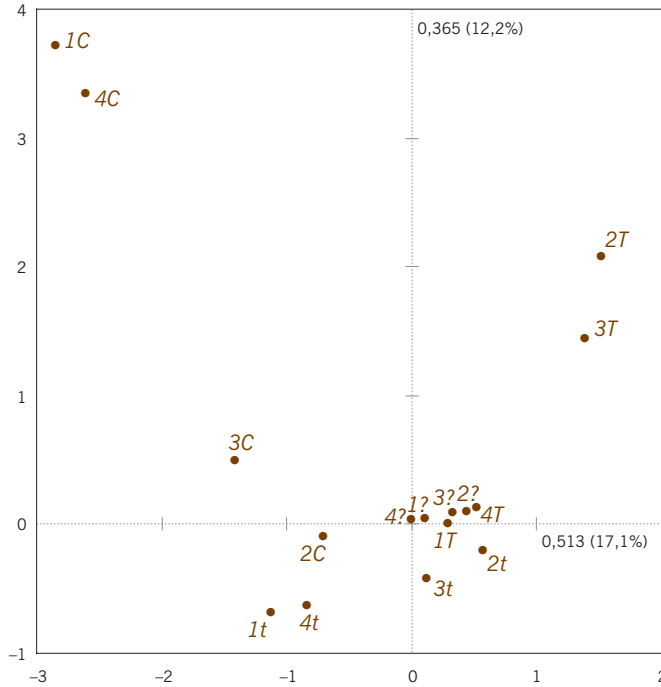
entre las variables respuesta entre sí, las no respuestas son el hecho más destacable: las personas que no responden una pregunta tienden a hacer lo mismo con las otras (por ejemplo, entre los seis primeros individuos de la tabla de la imagen 18.1 ya hay dos personas con no respuesta para las cuatro preguntas). En el segundo eje del mapa de la imagen 18.2 aparecen alineadas las categorías sustantivas; de las opiniones tradicionales, abajo, a las más liberales, arriba. En el mapa de la imagen 18.3 mostramos la segunda y la tercera dimensión del mapa, que dejan fuera la mayor parte del efecto de las no respuesta. Podemos ver que aquí las posiciones de los puntos son muy similares a las del mapa de la imagen 17.4. El hecho de que el lado liberal de la dimensión horizontal se halle ahora a la derecha no tiene consecuencia alguna sobre la interpretación; siempre es posible invertir un eje (multiplicando todas las coordenadas por  $-1$ ).

El cálculo de la inercia total de la matriz binaria es muy simple. Depende sólo del número de preguntas y del número de respuestas categóricas. No depende de sus valores concretos. Supongamos que tenemos  $Q$  variables y que cada variable  $q$ , tiene  $J_q$  categorías,  $J$  indica el número total de categorías:  $J = \sum_q J_q$  (en nuestro ejemplo,  $Q = 4$ ,  $J_q = 4$ ,  $q = 1, \dots, Q$  y  $J = 16$ ). La matriz binaria, simbolizada por  $\mathbf{Z}$ , con  $J$  columnas, es una matriz compuesta formada por tablas  $\mathbf{Z}_q$  agrupadas lateralmente, una para cada variable. En cada tabla, los valores marginales de las filas

Inercia de la matriz binaria

**Imagen 18.3:**

Mapa del ACM correspondiente a las cuatro preguntas sobre el trabajo de las mujeres, mostrando las dimensiones tercera y cuarta; inercia total = 3, porcentaje de inercia del mapa: 29,3%



son iguales a una columna de unos. Por tanto, podemos aplicar el resultado de (17.1) del capítulo 17: la inercia total de la matriz binaria es igual a la media de la inercia de las tablas que la componen. Cada tabla  $Z_q$  tiene un solo uno en cada fila, los restantes valores son ceros. Por tanto, estamos ante un ejemplo de matriz para la que todos los perfiles fila se hallan en los vértices, las asociaciones más extremas posibles entre filas y columnas. En consecuencia, en todas las tablas, las inercias de todos los ejes principales serán iguales a 1. Y, por tanto, la inercia total de la tabla  $Z_q$  será igual a su dimensionalidad, es decir, igual a  $J_q - 1$ . La inercia de  $Z$  será la media de las inercias de las tablas que la componen:

$$\text{inercia}(\mathbf{Z}) = \frac{1}{Q} \sum_q \text{inercia}(\mathbf{Z}_q) = \frac{1}{Q} \sum_q (J_q - 1) = \frac{J - Q}{Q} \quad (18.1)$$

Dado que  $J - Q$  es la dimensionalidad de  $Z$ , la inercia media por dimensión será  $1/Q$ . Fijémonos en que, en los mapas de las imágenes 18.2 y 18.3, las primeras tres dimensiones que hemos interpretado tienen inercias principales iguales a 0,693, 0,513 y 0,365, respectivamente, todas por encima de la media de  $1/4 = 0,25$ . Utilizamos el valor  $1/Q$  como umbral para decidir para qué ejes es interesante interpretar el ACM (similar al valor umbral de 1 de los valores propios en el análisis de componentes principales).

<i>1T</i>	<i>1t</i>	<i>1C</i>	<i>1?</i>	<i>2T</i>	<i>2t</i>	<i>2C</i>	<i>2?</i>	<i>3T</i>	<i>3t</i>	<i>3C</i>	<i>3?</i>	<i>4T</i>	<i>4t</i>	<i>4C</i>	<i>4?</i>
2501	0	0	0	172	1107	1131	91	355	1710	345	91	1766	538	40	157
0	476	0	0	7	129	335	5	16	261	181	18	128	293	17	38
0	0	79	0	1	6	72	0	1	17	61	0	14	21	38	6
0	0	0	362	1	57	108	196	7	96	55	204	51	45	2	264
172	7	1	1	181	0	0	0	127	48	4	2	165	15	0	1
1107	129	6	57	0	1299	0	0	219	997	61	22	972	239	13	75
1131	335	72	108	0	0	1646	0	24	989	573	60	760	616	84	186
91	5	0	196	0	0	0	292	9	50	4	229	62	27	0	203
355	16	1	7	127	219	24	9	379	0	0	0	360	14	1	4
1710	261	17	96	48	997	989	50	0	2084	0	0	1348	567	23	146
345	181	61	55	4	61	573	4	0	0	642	0	202	286	73	81
91	18	0	204	2	22	60	229	0	0	0	313	49	30	0	234
1766	128	14	51	165	972	760	62	360	1348	202	49	1959	0	0	0
538	293	21	45	15	239	616	27	14	567	286	30	0	897	0	0
40	17	38	2	0	13	84	0	1	23	73	0	0	0	97	0
157	38	6	264	1	75	186	203	4	146	81	234	0	0	0	465

**Imagen 18.4:**  
*Matriz de Burt que contiene todos los cruces posibles de las cuatro variables del ejemplo sobre la opinión de la gente sobre el trabajo de las mujeres. En la diagonal se hallan los cruces de las variables por ellas mismas*

Una estructura alternativa de datos para el ACM es la matriz compuesta por todas las tablas resultantes de cruzar todas las variables de interés dos a dos, la *matriz de Burt*, que mostramos en la tabla de la imagen 18.4 para los datos del ejemplo que estamos considerando. En este caso, la matriz de Burt es una matriz compuesta de  $4 \times 4$ , formada por 16 tablas. Con excepción de las tablas de la diagonal, las restantes 12 se obtienen cruzando los valores de dos variables de los 3418 encuestados. La matriz de Burt es simétrica, por tanto, fuera de la diagonal, sólo hay seis cruzamientos distintos que se transponen a ambos lados de la diagonal de la matriz compuesta. Las tablas de la diagonal corresponden a los cruces de las variables por ellas mismas, son matrices diagonales con las frecuencias marginales de la variable en su diagonal. Por ejemplo, las frecuencias marginales de la pregunta 1 son: 2501 para *T*, 476 para *t*, 79 para *C* y 362 para *?*. La matriz de Burt, **B**, se relaciona, de forma sencilla, con la matriz binaria **Z** de la manera siguiente:

$$\mathbf{B} = \mathbf{Z}^T \mathbf{Z} \tag{18.2}$$

La otra forma «clásica» de definir el ACM es el AC de la matriz de Burt **B**. Dado que **B** es una matriz simétrica, las soluciones de filas y de columnas son idénticas, por tanto sólo mostramos una de ellas (mapa de la imagen 18.5). Debido a la relación directa (18.2), no es ninguna sorpresa que los resultados de los dos análisis sean similares. De hecho, a primera vista, el mapa de la imagen 18.5 tiene el

**Matriz de Burt**

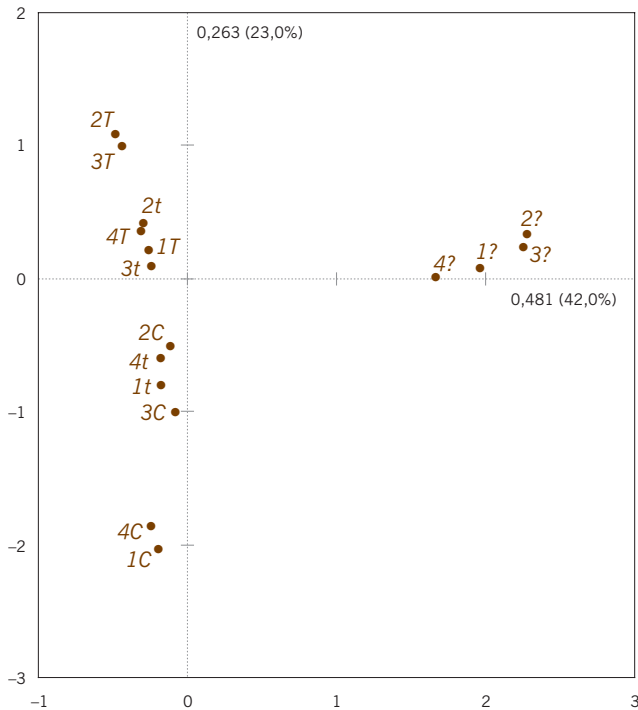
---

**Definición 2 del ACM:** AC de la matriz de Burt

---

**Imagen 18.5:**

Mapa del ACM de la matriz de Burt correspondiente a las cuatro preguntas sobre el trabajo de las mujeres, que muestra la primera y la segunda dimensión; inercia total = 1,145, porcentaje de inercia del mapa: 65,0%



mismo aspecto que el mapa de la imagen 18.2, sólo observamos un ligero cambio de escala en los dos ejes. Es la única diferencia entre ambos análisis: la versión Burt del ACM genera coordenadas principales en una escala reducida en comparación con las de la versión binaria. La reducción es relativamente mayor en el segundo eje en comparación con el primero.

#### Comparación del ACM de las matrices binaria y de Burt

Las dos formas de definir el ACM se relacionan de la siguiente manera:

- En los dos análisis, las coordenadas estándares de las respuestas categóricas son idénticas: es un resultado directo de la relación (18.2).
- También, como resultado de (18.2), las inercias principales del análisis de Burt son los cuadrados de los de la matriz binaria.
- Dado que las inercias principales son menores de 1, sus cuadrados darán valores más pequeños (y en consecuencia las inercias principales más pequeñas tendrán cuadrados relativamente más pequeños). Las coordenadas principales son las coordenadas estándares multiplicadas por la raíz cuadrada de las inercias principales, lo que explica la reducción de escala del mapa de la imagen 18.5 con relación al de la imagen 18.2.



PREGUNTAS	<i>Pregunta 1</i>	<i>Pregunta 2</i>	<i>Pregunta 3</i>	<i>Pregunta 4</i>	<i>Media</i>
<i>Pregunta 1</i>	3,0000	0,3657	0,4262	0,6457	1,1094
<i>Pregunta 2</i>	0,3657	3,0000	0,8942	0,3477	1,1519
<i>Pregunta 3</i>	0,4262	0,8942	3,0000	0,4823	1,2007
<i>Pregunta 4</i>	0,6457	0,3477	0,4823	3,0000	1,1189
<i>Media</i>	1,1094	1,1519	1,2007	1,1189	1,1452

**Imagen 18.6:**  
Inercias, obtenidas aplicando el AC de cada una de las 16 subtablas, de la matriz de Burt

- En consecuencia, los porcentajes de inercia serán siempre mayores en el análisis de Burt.

Todas las subtablas que componen la matriz de Burt tienen los mismos valores marginales totales de filas y columnas. Así, se cumple de forma exacta el resultado (17.1): la inercia de  $\mathbf{B}$  será la media de las inercias de las subtablas  $\mathbf{B}_{qs}$  que lo componen. En la tabla de la imagen 18.6 mostramos las 16 inercias individuales de la matriz de Burt, así como las medias de sus filas y de sus columnas. La media global es igual a la inercia total de  $\mathbf{B}$ , es decir 1,145. En esta tabla, las inercias de las matrices de la diagonal son exactamente igual a 3. Las inercias cumplen lo que vimos en (18.1) para las inercias de las tablas de la matriz binaria —hay  $J_q \times J_q$  tablas de dimensionalidad  $J_q - 1$  con una asociación perfecta fila-columna—; por tanto, la inercia máxima es igual al número de dimensiones. Los altos valores de las inercias de las matrices de la diagonal de la tabla de la imagen 18.6 explican porqué la inercia total de la matriz de Burt es tan alta, ello también explica los bajos porcentajes de inercia de los ejes. En el próximo capítulo retomaremos este tema.

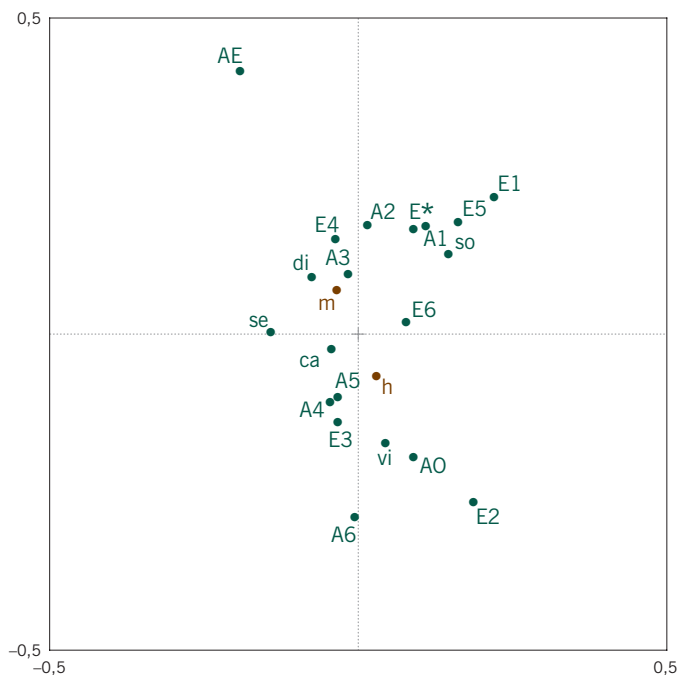
Inercia de la matriz de Burt

Supongamos que queremos relacionar las variables demográficas (género, edad, etc.) con las asociaciones observadas en los mapas de ACM. Existen dos maneras similares de hacerlo. La primera posibilidad consiste en codificar estas variables adicionales como variables binarias y añadirlas como columnas adicionales a la matriz binaria. La segunda posibilidad es cruzar las variables demográficas con las cuatro preguntas, como hicimos en el análisis de matrices compuestas del capítulo 17, y añadir estas tablas de contingencia como filas adicionales de la matriz binaria o como filas (o columnas) adicionales a la matriz de Burt. En el próximo capítulo veremos que, en la práctica, la segunda opción es mejor, ya que nos permite utilizar las versiones mejoradas del ACM. Ambas opciones proporcionan las mismas posiciones de los puntos adicionales, además de tener la misma interpretación como posiciones medias de los casos pertenecientes a una determinada categoría demográfica. En el mapa de la imagen 18.7 mostramos las posiciones de las cinco variables demográficas que vimos anteriormente, las podríamos sobreponer a los mapas de las imágenes 18.2 y 18.5.

Situación de variables adicionales en el mapa

**Imagen 18.7:**

*VARIABLES ADICIONALES CON RELACIÓN A LOS DOS PRIMEROS EJES PRINCIPALES, QUE PODRÍAMOS SUPERPONER A LOS MAPAS DE LA IMAGENES 18.2 O 18.5. ESTOS PUNTOS OCUPAN UNA PEQUEÑA ÁREA DEL MAPA (FIJÉMONOS EN LA ESCALA); DE TODAS FORMAS, PRESENTARÁN UNA MAYOR DISPERSIÓN EN EL MAPA DE LA MATRIZ DE BURT QUE EN EL DE LA MATRIZ BINARIA*



### Interpretación de los puntos adicionales

Las posiciones de las respuestas en las dos primeras dimensiones del mapa de la imagen 18.2 (igual que en el de la imagen 18.5) indican que cuanto más a la derecha se halle una categoría demográfica, mayor será la frecuencia de no respuestas. Cuando más arriba se halle una categoría, más liberales serán las opiniones, y cuanto más abajo, más tradicionales serán las opiniones. En consecuencia, Alemania Occidental es más tradicional y tiene una mayor proporción de no respuestas que Alemania del Este. Situación prácticamente idéntica al contraste hombre-mujer (h-m), pero no tan pronunciado como la diferencia entre las dos regiones alemanas. Los grupos de edad muestran la misma tendencia que vimos anteriormente con los jóvenes (A1) en la parte de arriba (liberal) y los de mayor edad (A6) abajo (tradicionales). Los niveles de educación más bajos tienen frecuencias de no respuesta más elevadas, mientras que los niveles educativos más altos tienden a tener opiniones más liberales, lo que no ocurre con los niveles educativos más bajos E1 y E2. Con relación a los estados civiles, los solteros (so) están por encima de la media con relación a la no respuesta y a las actitudes liberales, oponiéndose a los separados (se) que presentan una ocurrencia de no respuestas. Sin embargo, se hallan en la media respecto a la dimensión liberal-tradicional.

1. El ACM se ocupa de las relaciones entre un conjunto de variables; en general, variables homogéneas en cuanto hacen referencia a un mismo tema, siendo además las escalas de respuesta iguales.
2. Podemos recodificar las variables en la matriz binaria, que tiene tantas filas como casos y tantas columnas como categorías de respuesta. En las filas (es decir, los casos), los valores son todos 0 con excepción de un 1 que indica una categoría en particular de las variables.
3. La *matriz de Burt*, una matriz cuadrada simétrica, está formada por tablas de contingencia de dos entradas resultantes del cruce de todos los pares de variables. En la diagonal se hallan los cruces de las variables por ellas mismas.
4. Las dos definiciones alternativas de ACM, resultantes de la aplicación del AC a la matriz binaria o a la matriz de Burt, son prácticamente equivalentes. Las dos dan coordenadas estándares idénticas para los puntos correspondientes a las categorías.
5. La diferencia entre las dos definiciones se halla en las inercias principales: las de la matriz de Burt son los cuadrados de las de la matriz binaria. En consecuencia, los porcentajes de inercia del análisis de Burt son siempre más elevados que los del análisis binario.
6. Sin embargo, la codificación hace que los porcentajes de inercia de las dimensiones de los mapas sean artificialmente bajos, lo que conlleva una subestimación de la calidad de los mapas.