

La práctica del análisis de correspondencias

MICHAEL GREENACRE

Catedrático de Estadística en la Universidad Pompeu Fabra

Separata del capítulo 19

Análisis de correspondencias conjunto

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

Análisis de correspondencias conjunto

Ampliar el AC simple de una tabla de contingencia de dos entradas a muchas variables no es tan sencillo. Como hemos visto en el capítulo anterior, la estrategia habitual es aplicar el AC a la matriz binaria o la matriz de Burt. Sin embargo, la geometría de estas aproximaciones ya no es tan clara; además, aparte de que la inercia total del AC tiene poco sentido, los porcentajes de inercia explicados por el mapa son bajos. En el mapa de ACM basado en la matriz de Burt, intentamos visualizar la matriz completa y, sin embargo, en realidad estamos interesados sólo en las tablas de contingencia que se hallan fuera de la diagonal. Es decir, en las tablas que cruzan pares de variables distintas. En este capítulo veremos que el análisis de correspondencias conjunto (ACCo) intenta superar este inconveniente, centrándose en las tablas que se hallan fuera de la diagonal, lo que lleva a una mejor determinación de la inercia total y a una mejor representación de los datos en los mapas.

Contenido

El ACM se ajusta mal debido a que se ha exagerado la inercia total	195
Omisión de las tablas de la diagonal: AC conjunto	196
Resultados del ACCo	197
Los resultados del ACCo no se hallan anidados	197
Ajuste de los resultados del ACM para ajustar las tablas de fuera de la diagonal	198
Ajuste simple del ACM	199
Inercia ajustada = inercia media de las tablas situadas fuera de la diagonal	199
Ajuste de cada inercia principal	200
Porcentajes de inercia del ajuste simple del ACM	200
Conjunto de datos 10: interés por las noticias en Europa	201
Puntos adicionales en ACCo y en ACM ajustado	202
RESUMEN: Análisis de correspondencias conjunto	203

La tabla de la imagen 18.6 proporciona las inercias de cada una de las tablas que componen la matriz de Burt, así como la de su media, que es la inercia total de la matriz de Burt del ejemplo sobre el trabajo de las mujeres. El valor de esta media viene dado en gran medida por las inercias de las tablas de la diagonal, cuyos valores son iguales

El ACM se ajusta mal debido a que se ha exagerado la inercia total

Imagen 19.1:
 Porcentaje de inercia de cada una de las 16 tablas de la matriz de Burt explicado por el mapa bidimensional del ACM

PREGUNTAS	Pregunta 1	Pregunta 2	Pregunta 3	Pregunta 4	Porcentaje pregunta
Pregunta 1	51,9	78,4	82,5	80,4	61,2
Pregunta 2	78,4	55,5	88,2	76,6	65,3
Pregunta 3	82,5	88,2	59,6	86,6	69,7
Pregunta 4	80,4	76,6	86,6	54,6	63,5

al número de categorías de la variable correspondiente menos 1 ($4 - 1 = 3$, para las variables del ejemplo que consideramos). Como el análisis intenta explicar la inercia de toda la tabla, las elevadas inercias de la diagonal afectarán gravemente el ajuste de la tabla. Por ejemplo, en la tabla de la imagen 19.1 mostramos el porcentaje de inercia de cada tabla explicado por el mapa bidimensional del ACM. Dicho mapa llega a explicar el 65% de la inercia total, observamos que las tablas situadas fuera de la diagonal se explican mucho mejor que las situadas en la diagonal. Sumando las inercias explicadas de las tablas situadas fuera de la diagonal y expresando esta suma con relación a la suma de total, vemos que el mapa bidimensional del ACM explica el 83,2% de la inercia de estas tablas (podemos obtener los valores de inercia de cada tabla a partir de los porcentajes de la tabla de la imagen 19.1 y de los valores de las inercias totales de la tabla de la imagen 18.6). De forma similar, podemos ver que el mapa bidimensional del ACM explica el 55,4% de la inercia de las tablas situadas en la diagonal. Si sólo estamos interesados en las tablas situadas fuera de la diagonal, deberíamos considerar que la inercia explicada por el mapa bidimensional del ACM, es del 83,2%, y no del 65,0%. Como veremos más adelante existe la posibilidad de mejorar este valor del 83,2%.

Omisión de las tablas de la diagonal: AC conjunto

Queda claro que, en la matriz de Burt, la inclusión de las tablas de la diagonal degrada los resultados globales del ACM, ya que intentamos visualizar de forma innecesaria las tablas de la diagonal que, además, presentan inercias muy elevadas. De hecho, las tablas de la diagonal tienen los máximos valores de inercia que se pueden alcanzar. Ignorando las tablas de la diagonal podríamos mejorar el cálculo de la inercia explicada por el mapa. El *análisis de correspondencias conjunto* (ACCo) es un algoritmo iterativo que lleva a cabo el AC de la matriz de Burt, buscando el ajuste sólo para las tablas situadas fuera de la diagonal. En este procedimiento, partimos de los resultados del ACM para, a continuación, sustituir las tablas de la diagonal por valores estimados a partir de estos resultados mediante la fórmula de reconstitución (13.4). Dado que la matriz de Burt es simétrica, las coordenadas y las masas de filas y de columnas son iguales. La fórmula toma la siguiente expresión para la solución bidimensional:

$$\hat{p}_{jj'} = c_j c_{j'} \left(1 + \sqrt{\lambda_1} \gamma_{j1} \gamma_{j'1} + \sqrt{\lambda_2} \gamma_{j2} \gamma_{j'2} \right) \tag{19.1}$$

donde $\hat{p}_{jj'}$ es el valor estimado de la frecuencia relativa de la (j, j') -ésima celda de la matriz de Burt. Utilizando esta fórmula, podemos sustituir las tablas de la dia-

PREGUNTAS	Pregunta 1	Pregunta 2	Pregunta 3	Pregunta 4	Porcentaje pregunta
Pregunta 1	—	97,8	95,8	77,8	88,2
Pregunta 2	97,8	—	87,4	97,0	91,8
Pregunta 3	95,8	87,4	—	96,7	91,9
Pregunta 4	77,8	97,0	96,7	—	88,5

Imagen 19.2:

Porcentaje de inercia de cada una de las 12 tablas fuera de la diagonal de la matriz de Burt explicado por el mapa bidimensional del ACCo

gonal de la matriz de Burt por sus valores estimados, obteniendo así una *matriz de Burt modificada*. A continuación, llevamos a cabo un nuevo AC de la matriz de Burt modificada para obtener una nueva solución. Una vez obtenidos los resultados, volvemos a reemplazar las tablas de la diagonal por sus estimaciones de (19.1) y obtenemos una nueva matriz de Burt modificada. Y así sucesivamente hasta llegar a la convergencia; en cada iteración mejoramos el ajuste de las tablas situadas fuera de la diagonal.

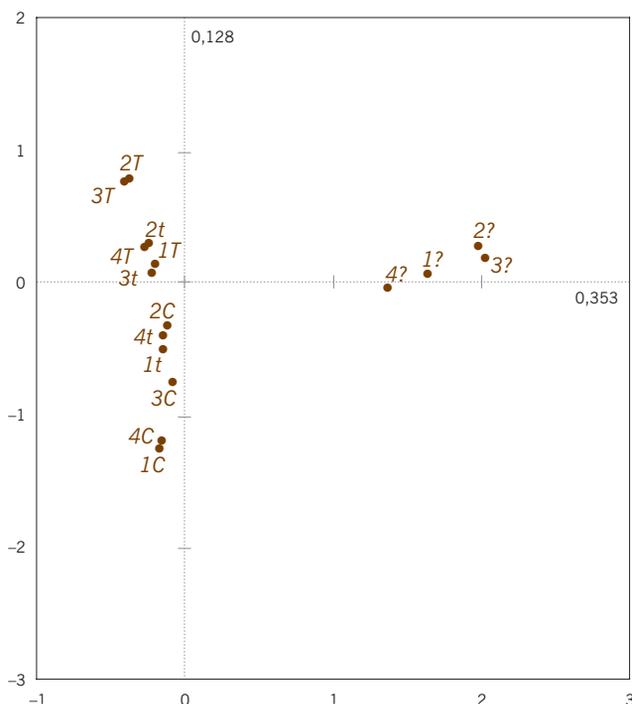
La aplicación del ACCo a los datos correspondientes a las cuatro variables sobre el trabajo de las mujeres nos lleva a los siguientes resultados: 90,2% de inercia explicada, y los porcentajes de inercia explicada de cada tabla que mostramos en la tabla de la imagen 19.2. Estos resultados son claramente mejores que los anteriores. Todas las tablas quedan muy bien representadas, la peor es la correspondiente al cruce de las preguntas 1 y 4, para la que la inercia explicada es del 77,8%. La imagen 19.3 muestra el mapa del ACCo, en el que la escala se ha mantenido de forma intencionada idéntica a la de la tabla 18.5 para poderla comparar. El resultado es prácticamente idéntico, la única diferencia observada es la disminución de la escala. En la imagen 18.5, las inercias de los dos ejes principales eran de 0,481 y 0,263, respectivamente, mientras que ahora son de 0,353 y de 0,128. Por tanto, una vez más se ha producido una disminución de las inercias principales, especialmente del segundo eje. Ello también ocurrió cuando en el ACM pasamos de la matriz binaria (imagen 18.2) a la matriz de Burt (imagen 18.5). Sin embargo, en aquella ocasión las coordenadas estándares de los dos análisis eran idénticas; aquí el resultado del ACCo es distinto del resultado del ACM, como podemos ver examinando en detalle el mapa de la imagen 19.3 y comparándolo con los mapas del ACM del capítulo 18.

En el ACCo, como ocurre en el ACM, los ejes principales no se hallan anidados (es decir, el resultado de dos dimensiones no contiene exactamente la mejor solución unidimensional como eje principal). No obstante, en la práctica se produce un anidamiento aproximado. Este es el motivo por el cual en el mapa de la imagen 19.3 no damos los porcentajes de inercia de los ejes —sólo podemos dar el porcentaje de inercia del resultado global, en este caso el 90,2%—. Este hecho también impide dar las contribuciones de los ejes a la inercia de un punto. A pesar de que cada punto tiene una cierta calidad de re-

Resultados del ACCo**Los resultados del ACCo no se hallan anidados**

Imagen 19.3:

Mapa del ACCo de la matriz de Burt correspondiente a las cuatro preguntas sobre el trabajo de las mujeres; porcentajes de inercia del mapa: 90,2%. El porcentaje de inercia es la suma de las inercias explicadas de cada tabla (obtenidos de las imágenes 19.2 y 18.6) y expresados como un porcentaje de la suma de los valores de las inercias de las tablas situadas fuera de la diagonal (consúltese el apéndice teórico, A)



presentación en el mapa, no la podemos dividir en las partes correspondientes a cada eje.

Ajuste de los resultados del ACM para ajustar las tablas de fuera de la diagonal

Nuestra experiencia nos ha enseñado que casi siempre existe una gran semejanza entre los resultados del ACCo y del ACM. Este hecho sugiere que, en realidad, lo que distingue los resultados del ACCo de los del ACM es sólo un cambio de escala. Por tanto, podríamos investigar este cambio de escala como alternativa para mejorar el ACM. Dadas las coordenadas estándares del resultado del ACM, ¿cómo podríamos cambiar de escala (es decir, definir las coordenadas principales) para reconstruir de forma óptima los datos de las tablas de la matriz de Burt que se hallan fuera de la diagonal? Se trata de un problema de regresión, utilizando, una vez más, la fórmula de reconstitución (19.1), pero considerando como factores de escala (es decir, las raíces cuadradas de las inercias principales) los coeficientes desconocidos del modelo de regresión β_1 y β_2 (para una solución bidimensional) obtenemos:

$$\frac{p_{jj'}}{c_j c_{j'}} - 1 = \beta_1 \gamma_{j1} \gamma_{j'1} + \beta_2 \gamma_{j2} \gamma_{j'2} + e_{jj'} \tag{19.2}$$

Llevamos a cabo la regresión de manera que obtengamos los valores de la «variable respuesta», disponiendo en forma de vector todos los valores situados a la

izquierda de (19.2), sólo para las celdas de las tablas situadas fuera de la diagonal —en nuestro ejemplo de cuatro variables, con seis tablas de 4×4 situadas fuera de la diagonal, tendremos, en el vector, $6 \times 16 = 96$ valores—. Como «variable explicativa» tenemos los correspondientes productos de $\gamma_{j_1}\gamma_{j_1}$ y $\gamma_{j_2}\gamma_{j_2}$. Llevamos a cabo una regresión de mínimos cuadrados ponderada, sin ordenada en el origen, con pesos iguales de valores respectivos $c_j c_j$. —en nuestro ejemplo, obtenemos unos coeficientes estimados $\hat{\beta}_1 = 0,5922$ y $\hat{\beta}_2 = 0,3532$ —. Los cuadrados de estos valores proporcionan, como «inercias principales», los valores óptimos 0,351 y 0,125, respectivamente, para los que la inercia explicada es del 89,9% (el coeficiente de determinación R^2 de la regresión). Es lo mejor que podemos hacer con los resultados del ACM; fijémonos en la similitud entre estos valores y los de las inercias principales del mapa de ACCo de la imagen 19.3, cuyos valores eran 0,353 y 0,128. Para representar las categorías en el mapa, calculamos las coordenadas principales multiplicando las coordenadas estándares del ACM de los dos primeros ejes por los factores de escala $\hat{\beta}_1$ y $\hat{\beta}_2$. De nuevo, la solución obtenida no es anidada, los resultados dependen de la dimensionalidad de la solución; si llevamos a cabo el mismo cálculo para una solución tridimensional, los dos primeros coeficientes de regresión no serán exactamente los que acabamos de obtener. El anidamiento se mantendrá sólo cuando las «variables explicativas» de (19.2) no estén correlacionadas. Ignorando las correlaciones, podríamos obtener de manera más simple un ajuste anidado (pero subóptimo), como describimos a continuación.

Vamos a describir un ajuste más simple de las inercias principales que cumple la condición de anidamiento, fácil de calcular y que implica los siguientes pasos: 1) recalcular la inercia total sólo para las tablas que se hallan fuera de la diagonal, y 2) un ajuste simple de las inercias principales derivado del ACM. Según nuestra experiencia, en general con este ajuste simplificado obtenemos resultados muy similares a los del ajuste óptimo que nos permiten expresar las inercias principales de la forma habitual, como porcentajes de inercia.

En el ACM de la matriz de Burt \mathbf{B} , la inercia total es la media de las inercias de todas las tablas que la componen, incluyendo las de la diagonal. Sin embargo con el ACCo, la inercia total es la media de las inercias de las tablas situadas fuera de la diagonal. Lo podemos calcular fácilmente a partir de la inercia total de \mathbf{B} , ya que conocemos de forma exacta cuáles son los valores de las inercias de las tablas de la diagonal: $J_q - 1$, donde J_q es el número de categorías de la variable q -ésima. Por tanto,

$$\text{suma de las inercias de las } Q \text{ tablas de la diagonal} = J - Q \quad (19.3)$$

mientras que

$$\text{suma de las inercias de todas las tablas de doble entrada} = Q^2 \times \text{inercia}(\mathbf{B}) \quad (19.4)$$

Ajuste simple del ACM

Inercia ajustada =
inercia media de las
tablas situadas fuera de
la diagonal

Restando (19.3) de (19.4) obtenemos la suma de las inercias de las tablas situadas fuera de la diagonal, y luego dividiendo por $Q(Q-1)$ obtenemos la media, lo que lleva a:

$$\text{media de la inercia fuera de la diagonal} = \frac{Q}{Q-1} \times \left(\text{inercia}(\mathbf{B}) - \frac{J-Q}{Q^2} \right) \quad (19.5)$$

Utilizando como ejemplo nuestros datos sobre el trabajo de las mujeres:

$$\text{inercia media fuera de la diagonal} = \frac{4}{3} \times \left(1,1452 - \frac{12}{16} \right) = 0,5269$$

Otra manera de calcular este valor es promediar directamente la media de las inercias de todas las tablas proporcionadas por la tabla de la imagen 18.6. Lo calculamos sólo en uno de los triángulos de la tabla, ya que solamente hay $\frac{1}{2}Q(Q-1) = 6$ pares de tablas distintas:

$$\frac{1}{6}(0,3657 + 0,4262 + 0,6457 + 0,8942 + 0,3477 + 0,4823) = 0,5269$$

Ajuste de cada inercia principal

Supongamos que en el ACM de la matriz de Burt \mathbf{B} indicamos como λ_k , para $k = 1, 2$, etc., las inercias principales (valores propios). Calcularemos las inercias principales ajustadas λ_k^{adj} de la siguiente manera:

$$\lambda_k^{\text{adj}} = \left(\frac{Q}{Q-1} \right)^2 \times \left(\sqrt{\lambda_k} - \frac{1}{Q} \right)^2, \quad k = 1, 2, \dots \quad (19.6)$$

En nuestro ejemplo, las dos primeras inercias principales ajustadas son:

$$\frac{16}{9} \times \left(0,6934 - \frac{1}{4} \right)^2 = 0,3495 \quad \text{y} \quad \frac{16}{9} \times \left(0,5132 - \frac{1}{4} \right)^2 = 0,1232$$

(fijémonos en la similitud de los valores óptimos 0,351 y 0,125, que vimos anteriormente).

Porcentajes de inercia del ajuste simple del ACM

Para obtener los porcentajes de inercia de cada eje principal, expresamos las inercias ajustadas con relación a la inercia total ajustada:

$$100 \times \frac{0,3495}{0,5269} = 66,3\% \quad \text{y} \quad 100 \times \frac{0,1232}{0,5269} = 23,4\%$$

En consecuencia, el porcentaje de inercia de la solución bidimensional del ajuste simple del ACM es del 89,7%, sólo el 0,2% menor que la del ajuste óptimo (no anidado) y el 0,5% menor que la del ACCo. Con los ejemplos anteriores, hemos visto,

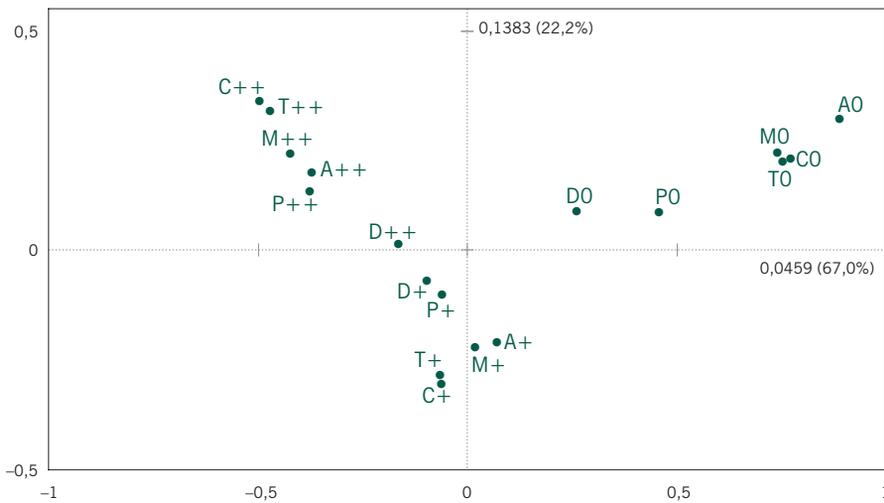


Imagen 19.4:
 Mapa del ACM ajustado correspondiente a los datos sobre los intereses de los europeos. Porcentaje de inercia del mapa: 89,2% (si se hubiera llevado a cabo el ACM con la matriz binaria, la inercia explicada sería sólo el 41,1%)

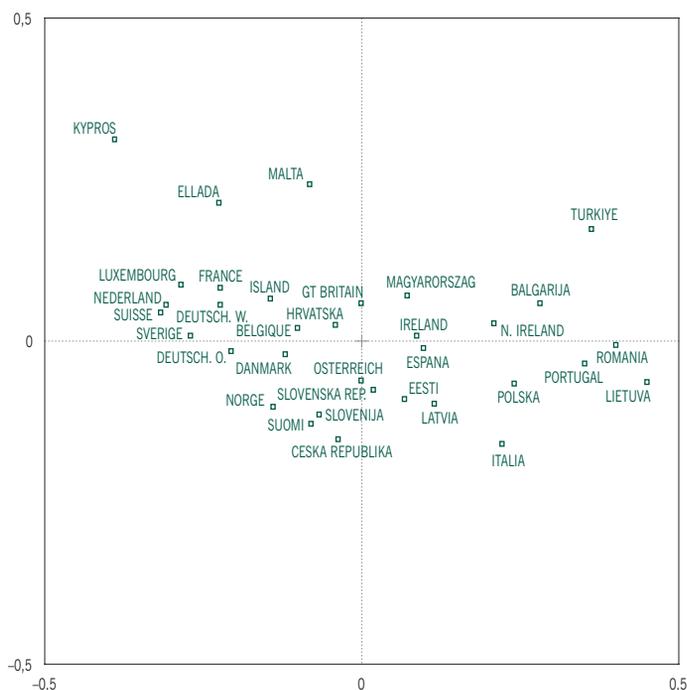
pues, que los porcentajes calculados a partir de este ajuste simple proporcionan un porcentaje de inercia global que es un límite inferior del porcentaje óptimo obtenido con el ACCo. Por tanto, cuando demos el resultado de un ACM, lo mejor es expresar el ajuste de la manera que acabamos de ver. Y en consecuencia, para obtener las coordenadas principales a partir de las coordenadas estándares utilizaremos como factores de escala las raíces cuadradas de las inercias principales ajustadas de forma simple. No vamos a representar otra vez el mapa, ya que las posiciones relativas de los puntos son las mismas que vimos en los mapas de las imágenes 18.2 y 18.5; solamente la escala es distinta, más parecida a la del mapa de la imagen 18.2.

Como ejemplo adicional y también para ilustrar otros aspectos del ACCo, consideremos un conjunto de datos derivado de la encuesta del Eurobarómetro de 2005. Entre otros aspectos, se preguntó a los encuestados sobre su interés por las noticias sobre: deportes (D), política (P), descubrimientos médicos (M), contaminación ambiental (A), innovación tecnológica (T) y descubrimientos científicos (C). Las posibilidades de respuesta eran: «muy interesados» (++) , «moderadamente interesados» (+) y «nada interesados» (0). Siguiendo esta notación, vamos a simbolizar las respuestas por: A+, por ejemplo para «moderadamente interesados en la contaminación atmosférica» y P0 para «nada interesados en política». Con el objetivo de evitar el efecto de las no respuestas, que, como vimos en el ejemplo anterior, tiene un fuerte efecto sobre los resultados, hemos omitido los encuestados con respuestas tipo «no sabe» o con respuestas «no contesta». Ello ha significado una reducción de la muestra de 33.190 a 29.652, es decir, una reducción del 10,7% (en el capítulo 21 trataremos específicamente sobre las no respuestas). En la imagen 19.4, mostramos el mapa de ACM ajustado correspondiente a estos datos. El mapa muestra que los «nada interesados» forman su propia diagonal de dispersión a la derecha

Conjunto de datos 10:
 interés por las noticias
 en Europa

Imagen 19.5:

Países europeos representados como puntos adicionales en el mapa del ACM ajustado correspondiente a datos sobre los intereses de los europeos. (Se muestran los nombres de los países como aparecen en el Eurobarómetro.)



de los «interesados», que asimismo van de «moderadamente interesados» abajo, a «muy interesados» arriba a la izquierda. Estamos ante un ejemplo de mapa que podría mejorar con una rotación de ejes, si queremos que estas dos líneas de dispersión coincidan más con los ejes principales. (En el epílogo hacemos algunos comentarios sobre las rotaciones.) El primer eje explica el 67,0% de la inercia y define una escala de interés general sobre las noticias. El segundo eje (22,0%) muestra el interés por los descubrimientos científicos y la innovación tecnológica en los extremos, lo que indica una elevada correlación entre ambos. Sin embargo, los dos puntos correspondientes al interés por los deportes se hallan cerca del centro de dispersión, lo que indica que, por ejemplo, un elevado interés por los deportes (D++) va asociado con un interés moderado en los otros temas, y viceversa. Recordemos que lo que estamos visualizando son las asociaciones de las categorías de una determinada variable con las categorías de las restantes variables.

Puntos adicionales en ACCo y en ACM ajustado

A pesar de que no mostramos las posiciones de los 29.652 casos de los datos, los podemos imaginar como puntos adicionales. Es decir, si añadiéramos como filas adicionales la gran matriz binaria de 29.652×18 a la matriz de Burt, cada encuestado tendría una posición en el mapa del ACM (sin embargo, fijémonos en que sólo existen $3^6 = 729$ respuestas distintas, por tanto, las respuestas de los encuestados situarán en los puntos que representan cada tipo de respuesta). Dado que en las tres ver-

siones distintas del ACM (binario, Burt y ajustado), las coordenadas estándares son iguales, las posiciones de las coordenadas principales de los encuestados serán las mismas en los tres. Como vimos en el capítulo 18, podemos mostrar las categorías adicionales añadiendo las tablas correspondientes a sus cruzamientos por las variables activas como filas adicionales de la matriz de Burt. Por ejemplo, con los datos que nos ocupan, tenemos muestras de 34 países europeos. Los encuestados de los distintos países tienen una posición en el mapa, la de cada país se halla en la posición media de los puntos correspondientes a los encuestados de ese país. El mapa de la imagen 19.5 muestra las posiciones de los países etiquetadas con los nombres locales (podemos imaginar este mapa superpuesto al mapa de la imagen 19.4). TURKIYE (Turquía) es, de todos los países, el que se halla más en la posición de «nada de interés»: un 40% de los encuestados turcos no mostró interés alguno en ningún tema, excepto la contaminación ambiental (22%); sin embargo KYPROS (Chipre), ELLADA (Grecia) y MALTA parecen ser los países más interesados; así, por ejemplo, Chipre tenía los mayores porcentajes en «muy interesado» en temas como contaminación ambiental (75%), descubrimientos médicos (62%), innovación tecnológica (53%) y descubrimientos científicos (55%).

1. Podemos definir el ACM como el AC de la matriz de Burt, formada por todas las tablas de contingencia de dos entradas derivadas del cruzamiento de un conjunto de variables, incluyendo las de una variable con ella misma, lo que conlleva una sobreestimación de la inercia total.
2. El *análisis de correspondencias conjunto* (ACCo) busca el mapa que mejor explica las tablas de contingencia correspondientes a los cruces de todos los pares de variables, ignorando las tablas que se hallan en la diagonal de la matriz de Burt. Esta aproximación implica un nuevo algoritmo iterativo que conduce a una solución óptima no anidada.
3. En el ACCo, la inercia total corresponde a la media de las inercias de todas las tablas que se hallan fuera de la diagonal de la matriz de Burt.
4. Una solución intermedia consiste en hallar las coordenadas estándares del ACM, mediante regresión de mínimos cuadrados ponderada de las tablas de contingencia de interés. Sin embargo, de nuevo, esta solución no es anidada.
5. El ACM *ajustado* es una solución simple, anidada, y que, por tanto, mantiene todas las buenas propiedades del AC, al mismo tiempo que resuelve el problema de la baja inercia. Consiste en aplicar algunos ajustes a las inercias principales y a la inercia total del ACM.
6. En todos los tipos de ACM, representamos las variables adicionales de la misma manera. Concretamente, cruzamos las variables adicionales con las variables activas, y añadimos las tablas de contingencia resultantes como filas adicionales a la matriz de Burt (o en la matriz de Burt modificada en ACCo).