

# La práctica del análisis de correspondencias

**MICHAEL GREENACRE**

Catedrático de Estadística en la Universidad Pompeu Fabra

---

Separata del capítulo 22

## **Análisis de tablas cuadradas**

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 tablas cuadradas

En este capítulo vamos a considerar el caso particular de las tablas de frecuencias cuadradas. Es decir, cuando filas y columnas hacen referencia a los mismos objetos en dos circunstancias distintas. Encontramos este tipo de datos en muchas situaciones. Por ejemplo, en las tablas sobre la movilidad social, en las matrices de confusión utilizadas en psicología, en tablas de cambio de marcas de los consumidores, en investigación de mercados, en el estudio de referencias cruzadas de artículos científicos entre revistas, en matrices de transición entre comportamientos y en tablas de migración. A menudo este tipo de tablas se caracterizan por tener valores relativamente elevados en la diagonal. Estos valores indican una asociación muy fuerte que fácilmente enmascara las asociaciones más sutiles situadas fuera de la diagonal, que no quedan reflejadas en los ejes principales. Una aproximación para poder aplicar el AC a tablas cuadradas es dividir el análisis en dos partes: 1) un análisis de la parte *simétrica* de la tabla, que absorbe la mayor componente de la inercia, incluyendo la diagonal y 2) un análisis de la parte *antisimétrica* de la tabla, que contiene la información situada fuera de la diagonal. Es la visualización de esta última componente la que muestra la magnitud y el sentido del «flujo» de filas a columnas y viceversa.

### Contenido

Conjunto de datos 12: movilidad social y profesiones de padres e hijos .....	226
AC de tablas cuadradas .....	226
La diagonal de la tabla domina el AC .....	227
Simetría y antisimetría en la tabla cuadrada .....	228
AC de la parte simétrica .....	229
AC de la parte antisimétrica .....	230
AC simultáneo de las partes simétrica y antisimétrica .....	230
Visualización de las partes simétrica y antisimétrica .....	231
RESUMEN: Análisis de tablas cuadradas .....	233

Conjunto de datos 12:  
movilidad social y  
profesiones de padres e  
hijos

Para entrar rápidamente en materia, consideremos unos datos clásicos sobre movilidad social. Son unos datos sobre profesiones de padres e hijos que fueron publicados por Karl Pearson hace más de 100 años. Los mostramos en la tabla de la imagen 22.1. Para crear esta tabla se hizo un recuento de las profesiones de padres e hijos. Dado que muchos hijos tenían la misma profesión que sus padres, los valores de la diagonal son elevados, algo que suele ser habitual en este tipo de tablas. Sin embargo, también suelen aparecer asimetrías destacables. Así, en el ejemplo que nos ocupa, podemos ver que la suma de los valores de la primera fila (padres militares) es 50, mientras que la suma de la primera columna (hijos militares) es 84. El flujo de hijos hacia el ejército se produce de padres propietarios rurales (fila 7) y comerciantes (fila 10). Por otra parte, los hijos de padres comerciantes experimentan un gran flujo de salida hacia otras profesiones. Así, observamos que hay 106 padres comerciantes, en cambio, los hijos que han elegido el comercio son 24. El flujo de salida se produce hacia artistas, teólogos, escritores y militares.

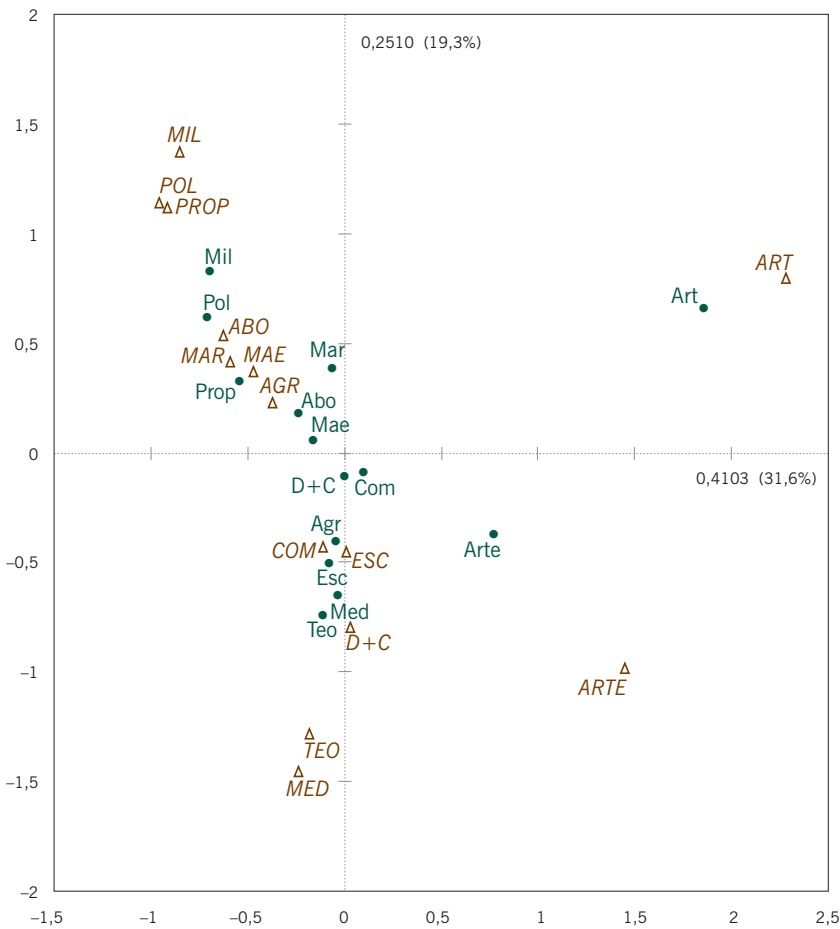
AC de tablas cuadradas

Podemos utilizar el AC para visualizar esta tabla de contingencia (imagen 22.2). La tabla tiene una inercia elevada (1,297), sin duda debido a las fuertes asociaciones existentes entre filas y columnas. Por tanto es adecuado utilizar el mapa asimétrico. Es decir, expresaremos los perfiles de los padres en coordenadas principales, y los de los hijos en coordenadas estándares. Si el perfil de una profesión de los padres tiene todo ceros excepto el valor de la diagonal, ello indica que esa profesión se halla en el vértice de esa profesión. Así, la segunda fila, correspondiente a artistas, es casi de este tipo, tienen el valor relativo más elevado (51 de

Imagen 22.1:

Tabla de contingencia correspondiente a las profesiones de padre e hijos. Vemos, por ejemplo, que las profesiones de los hijos de los 50 padres militares son: 28 militares, 4 maestros, 1 propietario rural, 3 abogados, etc.

PROFESIÓN PADRES	PROFESIÓN HIJOS														Sumas
	MIL	ART	MAE	ARTE	TEO	AGR	PROP	ABO	ESC	COM	MED	MAR	PO	D+C	
Militares	28	0	4	0	0	0	1	3	3	0	3	1	5	2	50
Artistas	2	51	1	1	2	0	0	1	2	0	0	0	1	1	62
Maestros	6	5	7	0	9	1	3	6	4	2	1	1	2	7	54
Artisanos	0	12	0	6	5	0	0	1	7	1	2	0	0	10	44
Teólogos	5	5	2	1	54	0	0	6	9	4	12	3	1	13	115
Agricultores	0	2	3	0	3	0	0	1	4	1	4	2	1	5	26
Propietarios rurales	17	1	4	0	14	0	6	11	4	1	3	3	17	7	88
Abogados	3	5	6	0	6	0	2	18	13	1	1	1	8	5	69
Escritores	0	1	1	0	4	0	0	1	4	0	2	1	1	4	19
Comerciantes	12	16	4	1	15	0	0	5	13	11	6	1	7	15	106
Médicos	0	4	2	0	1	0	0	0	3	0	20	0	5	6	41
Marinos	1	3	1	0	0	0	1	0	1	1	1	6	2	1	18
Políticos	5	0	2	0	3	0	1	8	1	2	2	3	23	1	51
Docentes y científicos	5	3	0	2	6	0	1	3	1	0	0	1	1	9	32
Sumas	84	108	37	11	122	1	15	64	69	24	57	23	74	86	775



**Imagen 22.2:**  
 Mapa asimétrico del AC sobre los datos de movilidad de la imagen 22.1, las filas en coordenadas principales. Porcentaje de inercia explicado: 51,0%.

62, el 82%) de padres e hijos con la misma profesión. En el mapa de la imagen 22.2 vemos que esta situación se refleja por el hecho de que artistas (Art) quedan muy alejados. Los padres artistas (Art) casi alcanzan el vértice correspondiente a los hijos artistas (ART). La fila de artesanos (Arte) se halla entre los vértices ART y D+C (docentes y científicos) ya que, en términos relativos, bastantes hijos de padres artesanos acaban en estas dos profesiones (fila 4 de la imagen 22.1)

El problema que encontramos al tratar de visualizar una matriz cuadrada de este tipo es la presencia de una poderosa diagonal que tiende a dominar el análisis. Dado que el AC trata de explicar tanta inercia como sea posible, no es de extrañar que el resultado del análisis se vea muy influido por la gran fuente de inercia que representa la diagonal, en detrimento del resto de la tabla que, sin embargo, contiene interesantes flujos entre profesiones de padres e hijos. Para apoyar esta

La diagonal de la tabla domina el AC

afirmación con algunos números, basta con ver que los 14 valores de la diagonal explican el 70,9% de la inercia total, mientras que los 182 valores que se hallan fuera de la diagonal contribuyen con sólo el 29,1% de la inercia; es decir, podemos descomponer la inercia total de la manera siguiente:

$$\begin{aligned} \text{inercia total} &= \text{inercia de la diagonal} + \text{inercia fuera de la diagonal} \\ 1,2974 &= 0,9100 + 0,3774 \\ 100\% &= 70,9\% + 29,1\% \end{aligned} \tag{21.1}$$

El mapa de la imagen 22.2 explica una inercia de 0,6613, que es el 51,0% de la inercia total. Podemos repartir este valor entre los elementos de la diagonal y los que quedan fuera de ésta, de la siguiente manera:

$$\begin{aligned} \text{inercia explicada} &= \text{inercia explicada de la diagonal} + \text{inercia explicada fuera de la diagonal} \\ 0,6613 \text{ (51,0\%)} &= 0,5483 \text{ (59,6\%)} + 0,1130 \text{ (29,9\%)} \end{aligned}$$

Hemos expresado los porcentajes entre paréntesis con relación a las inercias de la expresión (22.1). Vemos que los elementos de fuera de la diagonal se explican pobremente en comparación con los de la diagonal.

Simetría y antisimetría en la tabla cuadrada

Podemos descomponer la tabla en dos partes: una que contenga la parte *simétrica* de la tabla, es decir, el flujo medio entre filas y columnas, y otra parte que incluya la componente *antisimétrica* que cuantifique el flujo diferencial. Podemos expresar la descomposición de la tabla original, simbolizada por **N**, de la siguiente manera:

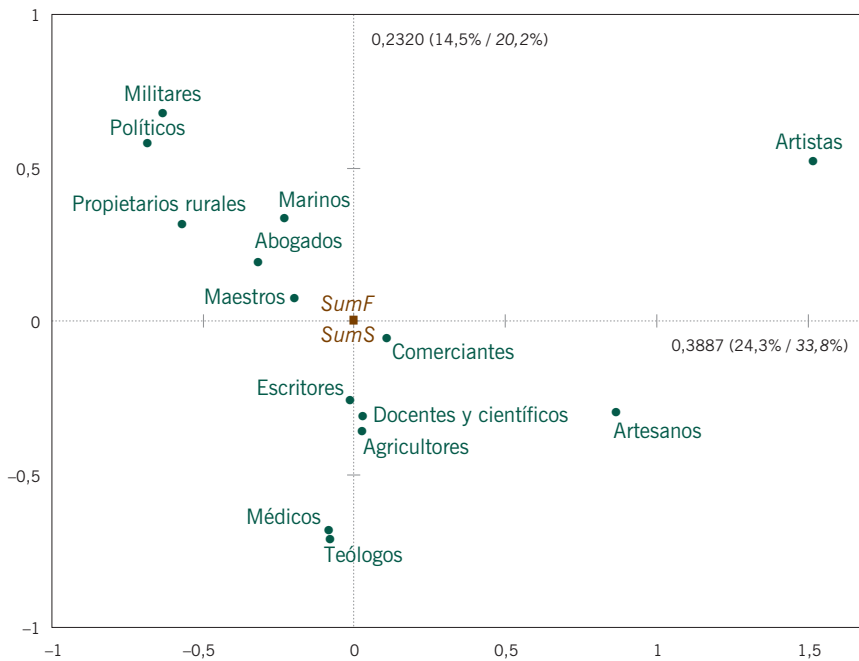
$$\begin{aligned} \mathbf{N} &= \frac{1}{2}(\mathbf{N} + \mathbf{N}^T) + \frac{1}{2}(\mathbf{N} - \mathbf{N}^T) \\ &= \mathbf{S} + \mathbf{T} \end{aligned} \tag{22.2}$$

donde **S** es la parte simétrica, que contiene las medias de los valores situados en los lados opuestos de la diagonal, y **T** la componente antisimétrica, con la mitad de las diferencias:

$$s_{ij} = \frac{1}{2}(n_{ij} + n_{ji}) \quad t_{ij} = \frac{1}{2}(n_{ij} - n_{ji}) \tag{22.3}$$

La siguiente expresión ilustra esta descomposición para los datos situados arriba a la izquierda en la tabla de la imagen 22.1:

$$\begin{bmatrix} 28 & 0 & 4 & 0 & \dots \\ 2 & 51 & 1 & 1 & \dots \\ 6 & 5 & 7 & 0 & \dots \\ 0 & 12 & 0 & 6 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} = \begin{bmatrix} 28 & 1 & 5 & 0 & \dots \\ 1 & 51 & 3 & 6,5 & \dots \\ 5 & 3 & 7 & 0 & \dots \\ 0 & 6,5 & 0 & 6 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} + \begin{bmatrix} 0 & -1 & -1 & 0 & \dots \\ 1 & 0 & -2 & -5,5 & \dots \\ 1 & 2 & 0 & 0 & \dots \\ 0 & 5,5 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$



**Imagen 22.3:**  
 AC de la parte simétrica de la imagen 22.1. Los primeros porcentajes se han calculado con relación a la inercia total de 1,5991, mientras que los porcentajes en cursiva se han calculado con relación a la inercia de la parte simétrica de 1,1485

Por ejemplo, transformamos el valor 1 correspondiente a la segunda fila 1 (padre artista) y cuarta columna (hijo artesano) y el valor 12 correspondiente a la cuarta fila (padre artesano) y segunda columna (hijo artista), en su valor medio 6,5 que situados en las celdas correspondientes de **S**, y las desviaciones de la media  $\pm 5,5$  que situamos en las correspondientes celdas de **T**. La matriz simétrica tiene la misma diagonal que la tabla original y como hemos visto  $s_{ij} = s_{ji}$ . En cambio, la matriz antisimétrica tiene ceros en la diagonal. La antisimetría consiste en que los elementos situados en los lados opuestos de la diagonal tienen los mismos valores absolutos pero signos opuestos, es decir,  $t_{ij} = -t_{ji}$ .

Aplicaremos el AC a la matriz simétrica y a la matriz antisimétrica, de forma separada. En el mapa de la imagen 22.3 mostramos el mapa de la matriz simétrica, con sólo unas coordenadas, ya que las coordenadas de filas y columnas son idénticas. Salvo por el hecho de que cada profesión tiene un solo punto, el mapa se parece mucho al de la imagen 22.2. El mapa muestra la asociación global entre profesiones de padres e hijos. En los ejes, el primer porcentaje, entre paréntesis, hace referencia a la inercia explicada con relación a la tabla asimétrica original, mientras que el porcentaje en cursiva, también entre paréntesis, hace referencia a la inercia explicada con relación a la inercia total de la parte simétrica **S** visualizada. Fijémonos en que los valores marginales de filas y de columnas de **S** son medias de los valores marginales de la matriz asimétrica **N**. De esta manera, si las masas

[AC de la parte simétrica](#)

de filas y de columnas de esta última matriz son  $\mathbf{r}$  y  $\mathbf{c}$ , respectivamente, entonces, las masas de las filas y de las columnas de  $\mathbf{S}$  son  $\mathbf{w} = \frac{1}{2}(\mathbf{r} + \mathbf{c})$ .

AC de la parte  
antisimétrica

Antes de que podamos aplicar el AC a la matriz antisimétrica  $\mathbf{T}$ , tenemos que superar dos problemas. El primero es que  $\mathbf{T}$  tiene valores positivos y valores negativos. En realidad, la suma de los elementos de la matriz es cero. Por tanto, no tiene sentido centrarla con relación a sus valores marginales, el primer paso del algoritmo del AC. Por tanto, tenemos que cambiar el algoritmo de manera que el AC analice los datos sin centrarlos, y así realizar directamente la normalización que conduce a las distancias  $\chi^2$ . Sin embargo, ello nos conduce al segundo problema: la suma de filas y de columnas, como masas, no tiene sentido. La solución obvia es adoptar las mismas masas que en  $\mathbf{S}$ , es decir, las masas  $\mathbf{w}$  definidas anteriormente. Parece, pues, que necesitamos un algoritmo especial para analizar  $\mathbf{T}$ . No obstante, existe un procedimiento que nos permite obtener la solución de forma simple recodificando los datos.

AC simultáneo de las  
partes simétrica y  
antisimétrica

El mencionado procedimiento de recodificación evita tener que implementar un algoritmo especial para hallar la solución del AC de la matriz antisimétrica. La idea es formar una matriz que sea cuatro veces la tabla original  $\mathbf{N}$  con el siguiente formato (es fácil de obtener con R: véase la página 317, o con una hoja de cálculo):

$$\begin{bmatrix} \mathbf{N} & \mathbf{N}^T \\ \mathbf{N}^T & \mathbf{N} \end{bmatrix} \quad (22.4)$$

Es decir, formamos una nueva matriz compuesta en la que situamos la tabla original  $\mathbf{N}$ , arriba a la izquierda y abajo a la derecha, y su transpuesta arriba a la derecha y abajo a la izquierda. A continuación llevamos a cabo el AC en esta nueva matriz compuesta. Si  $\mathbf{N}$  es una matriz  $I \times I$ , entonces la nueva matriz compuesta es una matriz  $2I \times 2I$ , con  $2I - 1$  dimensiones,  $I - 1$  de las cuales corresponden exactamente a las dimensiones de la matriz simétrica  $\mathbf{S}$  y el resto a las de la matriz antisimétrica  $\mathbf{T}$ . Es fácil saber qué dimensiones corresponden a cada parte, ya que las dimensiones de la matriz antisimétrica siempre ocurren en pares de inercias principales iguales. Por ejemplo, en la tabla de la imagen 22.4, mostramos las 27 inercias principales (valores propios) correspondientes a los datos sobre la movilidad social, donde  $I = 14$ . Los siete pares de dimensiones con inercias principales iguales (que mostramos en negrita) corresponden al análisis antisimétrico y son: 3 y 4, 10 y 11, 14 y 15, 17 y 18, 19 y 20, 22 y 23, y finalmente 25 y 26. Las restantes 13 dimensiones corresponden al análisis simétrico. La inercia total de la matriz simétrica es la suma de sus respectivas 13 inercias principales:  $0,3887 + 0,2320 + 0,1439 + \dots = 1,1485$ , lo que corresponde al 71,8% de la inercia total de 1,5991. La inercia total de la matriz antisimétrica es la suma de 7 pares:  $2 \times 0,1584 + 2 \times 0,0418 + \dots = 0,4508$ , lo que corresponde al 28,2% del total (fijémonos en que la inercia total de 1,5991, es



<i>Dim.</i>	<i>Inercia principal</i>	<i>Dim.</i>	<i>Inercia principal</i>	<i>Dim.</i>	<i>Inercia principal</i>
1	0,38868	10	<b>0,04184</b>	19	<b>0,00309</b>
2	0,23204	11	<b>0,04184</b>	20	<b>0,00309</b>
3	<b>0,15836</b>	12	0,02287	21	0,00166
4	<b>0,15836</b>	13	0,02205	22	<b>0,00115</b>
5	0,14391	14	<b>0,01287</b>	23	<b>0,00115</b>
6	0,12376	15	<b>0,01287</b>	24	0,00062
7	0,08184	16	0,01036	25	<b>0,00038</b>
8	0,07074	17	<b>0,00759</b>	26	<b>0,00038</b>
9	0,04984	18	<b>0,00759</b>	27	0,00015

**Imagen 22.4:**

*Inercias principales de las 27 dimensiones del análisis de la matriz compuesta de  $28 \times 28$  (22.4) formada a partir de los datos sobre movilidad social. Las inercias principales que ocurren en pares corresponden a la parte antisimétrica*

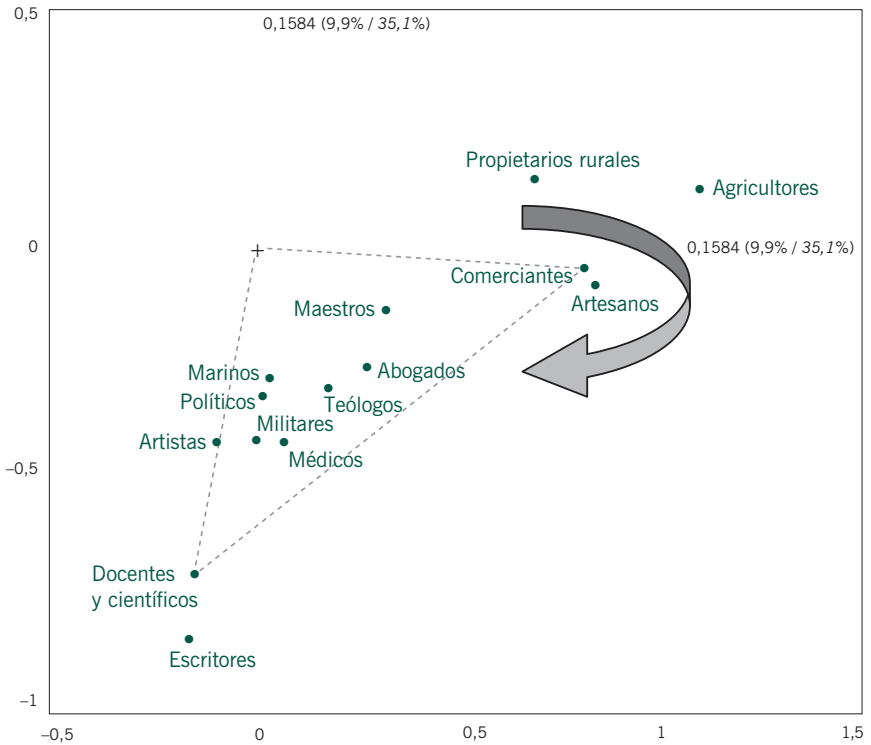
mayor que la inercia de la matriz original, de 1,2974 (22.1), ya que hemos centrado la tabla con relación a los valores marginales medios  $w$  de filas y columnas).

Como podemos ver en la tabla de la imagen 22.4, las dimensiones 1 y 2 son las que mejor visualizan la matriz simétrica. Explican una inercia de 0,6217 de un total de 1,1485, el 54,0%. El AC de la matriz compuesta proporciona los resultados de filas y de columnas repetidos; por tanto, utilizaremos sólo un conjunto de coordenadas principales para dibujar el mapa (para las dimensiones 1 y 2, hemos utilizado el primer conjunto de coordenadas de la imagen 22.6; véase también el apéndice de cálculo). Las dimensiones 3 y 4 son las mejores para visualizar la matriz antisimétrica. Explican 0,3167 de 0,4506, es decir, el 70,3% de la inercia de la parte antisimétrica. El mapa de la parte antisimétrica que mostramos en la imagen 22.5 tiene algunas propiedades particulares. La primera es que, debido a la igualdad de las inercias principales, las coordenadas pueden girar libremente en el mapa bidimensional. No se identifican con relación a los ejes principales, por tanto, no dibujamos estos ejes en el mapa. La segunda es que, debido a la antisimetría de la matriz, los resultados del AC de la matriz compuesta aparecen repetidos, pero con el signo cambiado. En el mapa dibujaremos un sólo conjunto de puntos —para dibujar el mapa de la imagen 22.5 hemos utilizado el segundo conjunto de coordenadas para las dimensiones 3 y 4 de la tabla de la imagen 22.6—. Para interpretar este tipo de mapas no utilizamos las distancias entre puntos, sino que interpretamos las áreas triangulares formadas por pares de puntos y el origen. Por ejemplo, comerciantes y docentes-científicos definen un gran triángulo con relación al origen, que interpretamos como un fuerte flujo diferencial entre estas dos profesiones. La flecha en el sentido de las agujas del reloj, que hemos dibujado en el mapa, indica el sentido del flujo de padres a hijos: los hijos de padres comerciantes son con relativa frecuencia docentes-científicos (en la tabla de la imagen 22.1 podemos ver que la frecuencia de padres comerciantes – hijos docente-científicos es 15, mientras que en sentido contrario la frecuencia es 0). Por tanto, generalizando, podemos decir que propietarios rurales, agricultores, comerciantes y artesanos experimentan flujos de salida hacia es-

#### Visualización de las partes simétrica y antisimétrica

**Imagen 22.5:**

AC de la parte antisimétrica de la tabla de la imagen 22.1. Hemos calculado los primeros porcentajes con relación a la inercia total de 1,5991, mientras que los porcentajes en cursiva se han calculado con relación a la inercia de la parte antisimétrica de 0,4506



critores y docentes-científicos. Algunos pares de profesiones definen, con relación al origen, áreas triangulares muy pequeñas como, por ejemplo, militares, políticos y marinos, lo que indica que no se produce un flujo entre estas profesiones. Sin embargo, podrían experimentar flujos de entrada de agricultores, artesanos, etc.

**Imagen 22.6:**

Coordenadas principales de algunas filas de la matriz compuesta de  $28 \times 28$  (22.4) de los datos sobre movilidad social. Las coordenadas de las dimensiones simétricas (en este caso las dos primeras) son simples repeticiones, mientras que las de la parte asimétrica (las dimensiones 3 y 4) son iguales, pero de signo opuesto

PROFESIÓN	<i>Dim. 1</i>	<i>Dim. 2</i>	<i>Dim. 3</i>	<i>Dim. 4</i>	...
Militares	-0,632	0,671	-0,011	0,416	...
Artistas	1,521	0,520	0,089	0,423	...
Maestros	-0,195	0,073	-0,331	0,141	...
Artesanos	0,867	-0,298	-0,847	0,092	...
Teólogos	-0,077	-0,709	-0,189	0,305	...
⋮	⋮	⋮	⋮	⋮	⋮
Militares	-0,632	0,671	0,011	-0,416	...
Artistas	1,521	0,520	-0,089	-0,423	...
Maestros	-0,195	0,073	0,331	-0,141	...
Artesanos	0,867	-0,298	0,847	-0,092	...
Teólogos	-0,077	-0,709	0,189	-0,305	...
⋮	⋮	⋮	⋮	⋮	⋮

1. Las tablas cuadradas con el mismo tipo de elementos en filas y en columnas son especiales debido a que los valores de su diagonal desempeñan un papel predominante en el análisis que, a menudo, enmascara las estructuras de la tabla que quedan fuera de la diagonal.
2. Una alternativa al AC habitual es dividir la tabla en dos partes: una parte *simétrica* y otra *antisimétrica*, de manera que esta última —en general, de menor inercia que la parte simétrica— capture las asimetrías de la tabla.
3. Analizamos la parte simétrica de la forma habitual, sin embargo, para analizar la parte antisimétrica utilizamos un algoritmo de cálculo del AC que elimina el centrado y la normalización de la tabla con respecto a sus valores marginales; para estos datos no tendría sentido utilizar los valores marginales como masas.
4. En ambos análisis, las masas que utilizamos para ponderar y para las distancias  $\chi^2$  son las medias de las masas de filas y de columnas de la tabla original.
5. Otra posibilidad de análisis es que apliquemos el AC a una matriz compuesta formada por la tabla original arriba a la izquierda y abajo a la derecha, y su transpuesta arriba a la derecha y abajo a la izquierda. Las dimensiones correspondientes a la parte simétrica tienen inercias principales únicas, mientras que las de la parte antisimétrica ocurren en pares iguales.
6. Interpretamos el mapa de la parte simétrica de la forma habitual, es decir, este mapa muestra la asociación global entre los elementos de la tabla.
7. Sin embargo, interpretamos el mapa de la parte antisimétrica, que posee una geometría especial, con relación a las áreas de triángulos formados por pares de puntos y el origen. Dichas áreas indican la intensidad de las asimetrías entre estos pares de elementos de la tabla. La dirección de la asimetría es la misma para todos los pares de puntos del mapa.