

La práctica del análisis de correspondencias

MICHAEL GREENACRE

Catedrático de Estadística en la Universidad Pompeu Fabra

Separata del capítulo 9

Representaciones bidimensionales

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

Representaciones bidimensionales

Hasta ahora hemos estudiado con bastante extensión, las proyecciones de una nube de perfiles sobre un solo eje principal, la recta que mejor se ajusta. Sin embargo, en la práctica encontraremos que la mayor parte de representaciones del AC son bidimensionales. Es habitual representar el primer eje principal horizontalmente (eje x) y el segundo eje principal verticalmente (eje y). Aunque podemos proyectar la nube de puntos sobre cualquier subespacio de pocas dimensiones, las proyecciones bidimensionales son especiales debido a que son nuestra forma habitual de representar gráficos sobre papel o en las pantallas de ordenador. De todas formas, en el apéndice de cálculo mostramos algunos ejemplos de cómo utilizar el lenguaje de programación R para hacer representaciones en tres dimensiones (imagen B.5, en pág. 306).

Contenido

Conjunto de datos 4: hábitos fumadores de grupos de empleados	95
Análisis de filas	96
Interpretación de los perfiles fila y de los vértices columna	97
Anidado de los ejes principales	98
Interpretación de la segunda dimensión	98
Verificación de la interpretación perfiles-vértices	98
Mapas asimétricos	100
Mapa simétrico	101
Verificación de la distancia ji -cuadrado entre los puntos en un mapa simétrico	102
El peligro de interpretar las distancias entre las filas y las columnas en un mapa simétrico	102
RESUMEN: Representaciones bidimensionales	103

El ejemplo que veremos a continuación, que apareció originalmente en mi libro de 1984, *Theory and Applications of Correspondence Analysis*, ha sido utilizado como ejemplo para ilustrar el AC en los principales programas estadísticos comerciales. Este ejemplo, a pesar de que corresponde a un conjunto de datos ficticios, se ha citado en bastantes artículos científicos y todavía lo podemos utilizar como intro-

Conjunto de datos 4:
hábitos fumadores de
grupos de empleados

Imagen 9.1:

Clasificación de los empleados de una empresa según su nivel profesional y sus hábitos fumadores, que muestra los perfiles de las filas, el perfil fila medio, entre paréntesis, así como las masas de las filas

GRUPO DE EMPLEADOS	TIPOS DE FUMADORES				Total de las filas	Masas
	No fumadores	Fumadores ocasionales	Fumadores medios	Fumadores compulsivos		
Directivos séniors	4	2	3	2	11	0,057
DS	(0,364)	(0,182)	(0,273)	(0,182)		
Directivos jóvenes	4	3	7	4	18	0,093
DJ	(0,222)	(0,167)	(0,389)	(0,222)		
Empleados séniors	25	10	12	4	51	0,279
ES	(0,490)	(0,196)	(0,235)	(0,078)		
Empleados jóvenes	18	24	33	13	88	0,456
EJ	(0,205)	(0,273)	(0,375)	(0,148)		
Secretarías	10	6	7	2	25	0,130
SC	(0,400)	(0,240)	(0,280)	(0,080)		
Total	61	45	62	25	193	
Perfil medio	(0,316)	(0,233)	(0,321)	(0,130)		

ducción a las representaciones bidimensionales. Los datos tratan sobre una encuesta a 193 empleados de una empresa que tiene como objetivo conocer los hábitos de los fumadores de la empresa. Clasificamos a los empleados de la empresa de acuerdo con su nivel profesional (cinco grupos) y sus hábitos fumadores (cuatro grupos) (en la imagen 9.1 reproducimos la correspondiente tabla de contingencia). Dado que se trata de una tabla de 5×4 , sus perfiles fila y sus perfiles columna se hallan exactamente en un espacio tridimensional.

Análisis de filas

Como hemos visto anteriormente, podemos ver esta tabla como un conjunto de filas o como un conjunto de columnas. Supongamos que el análisis de filas es más relevante, es decir, estamos interesados en representar, para cada grupo de empleados, los porcentajes de no fumadores, de fumadores ocasionales, etc. El espacio de perfiles constituye un símplex de cuatro puntos, es decir, un tetraedro de tres dimensiones que es el equivalente tridimensional al espacio triangular que vimos anteriormente (lo podemos visualizar utilizando las representaciones tridimensionales que describimos en el apéndice de cálculo, B). Para reducir la dimensionalidad de los perfiles, los podemos proyectar sobre el plano que mejor se ajuste (imagen 6.6). En el mapa de la imagen 9.2 también representamos gráficamente las proyecciones de los cuatro vértices que representan los hábitos fumadores de los empleados. Fijémonos en que, como es habitual, hemos situado el primer eje principal horizontalmente y el segundo eje principal verticalmente. Junto a los ejes indicamos las inercias principales (de 0,07476 y de 0,01002, respectivamente), así como los correspondientes porcentajes de inercia. Podemos sumar estos valores para conocer el porcentaje de inercia explicado por esta representación. Así vemos que la inercia explicada por el plano es de

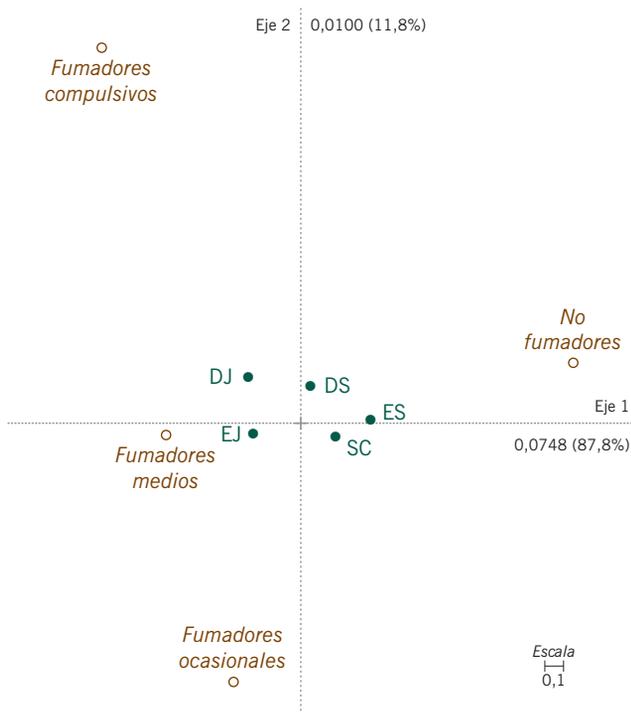


Imagen 9.2:

Mapa óptico del AC bidimensional de los datos sobre los hábitos de los fumadores de la imagen 9.1, con las filas en coordenadas principales (proyecciones de los perfiles) y las columnas en coordenadas estándares (proyecciones de los vértices)

0,08478, lo que representa el 99,5% de la inercia total de 0,08519. Es decir, sacrificando una dimensión, hemos perdido sólo el 0,5% de la inercia de los perfiles. Está claro, pues, que los cinco perfiles fila se hallan muy cerca del plano representado, tan cerca, que cuando exploremos sus posiciones relativas, podremos ignorar las distancias de éstos al plano.

Si nos fijamos sólo en las posiciones de los perfiles, podemos comprobar que los grupos que se hallan más separados son, por un lado, los empleados jóvenes (EJ) y los directivos jóvenes (DJ) situados a la izquierda, y, por otro, los empleados sénior (ES) situados a la derecha; por tanto, las mayores diferencias en los hábitos de los fumadores se hallan entre estos dos extremos. Los directivos sénior (DS) se hallan entre los directivos jóvenes y los empleados sénior, mientras que las secretarías (SC) se hallan muy cerca de los empleados sénior. No obstante, para poder explicar las similitudes y las diferencias entre los grupos de empleados, es necesario que nos fijemos en las posiciones de los perfiles con relación a las de los vértices. Dado que las tres categorías de fumadores se hallan a la izquierda y la de no fumadores se halla a la derecha, la distinción entre derecha e izquierda es equivalente a la distinción entre fumadores y no fumadores. Los grupos EJ y DJ son diferentes del grupo de ES, ya que los primeros son relativamente fumadores,

[Interpretación de los perfiles fila y de los vértices columna](#)

mientras que el grupo ES es relativamente más «no fumador». El centro de este tipo de representaciones es siempre el perfil medio, de manera que podemos considerar las desviaciones de los grupos de empleados en distintas direcciones a partir del perfil medio, las mayores desviaciones se producen de izquierda a derecha.

Anidado de los ejes principales

La representación bidimensional también contiene la mejor representación unidimensional. Si proyectáramos todos los puntos de la imagen 9.2 verticalmente sobre el eje horizontal, esta representación unidimensional sería la misma que habríamos obtenido si, de entrada, hubiésemos estado interesados sólo en la mejor representación unidimensional. Decimos que los ejes están *anidados*. Es decir, la representación óptima de una determinada dimensionalidad contiene todas las representaciones óptimas de menor dimensionalidad. Fijémonos en que las proyecciones, sobre el primer eje, de los tres grupos de fumadores situados a la izquierda, quedan muy cerca entre sí y bastante separadas del punto situado a la derecha correspondiente a los no fumadores. Ésta es la característica más importante de los datos. Utilizando la terminología que vimos en el capítulo 7, podemos decir que la «escala de fumadores» que mejor diferencia los cinco grupos de empleados no es la que asigna intervalos iguales a las cuatro categorías de fumadores, sino la que sitúa a los tres grupos de fumadores muy cerca y muy separados del grupo de no fumadores. Efectivamente, existe una dicotomía entre fumadores y no fumadores.

Interpretación de la segunda dimensión

Continuando con la interpretación bidimensional, vemos que el segundo eje principal (eje vertical) separa los tres grupos de fumadores. Como indica el porcentaje de inercia explicada por el eje vertical, muy inferior, los perfiles no difieren tanto vertical como horizontalmente. No obstante, a pesar de que los EJ y los DJ tienen porcentajes similares de fumadores, como se puede ver por su posición similar en el eje horizontal, llegamos a la conclusión de que el perfil de los EJ tiene relativamente más fumadores ocasionales que fumadores compulsivos en comparación con el perfil de los DJ. Podemos verificar fácilmente estas conclusiones a partir de los datos originales de la imagen 9.1.

Verificación de la interpretación perfiles-vértices

Midiendo las distancias entre los perfiles y los vértices de la imagen 9.2, y comparando posteriormente dichas mediciones con los valores de los perfiles, podemos verificar la interpretación sobre las posiciones de los perfiles con relación a la de los vértices. Tenemos que hacer esta verificación vértice a vértice, por ejemplo, midiendo las cinco distancias de los grupos de empleados al vértice *fumadores ocasionales*. Como regla general, suponiendo que la representación sea de buena calidad, lo que es cierto en esta ocasión, cuanto más cerca se halle un perfil de un vértice, más se identifica este perfil con el grupo representado por el vértice. Así, por ejemplo, en el párrafo anterior dijimos que debido a que EJ se

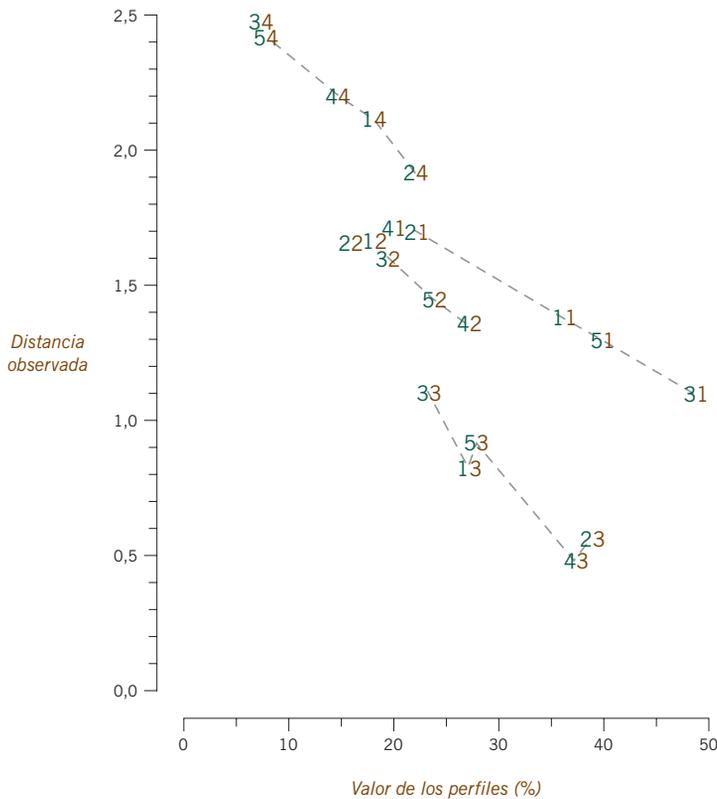
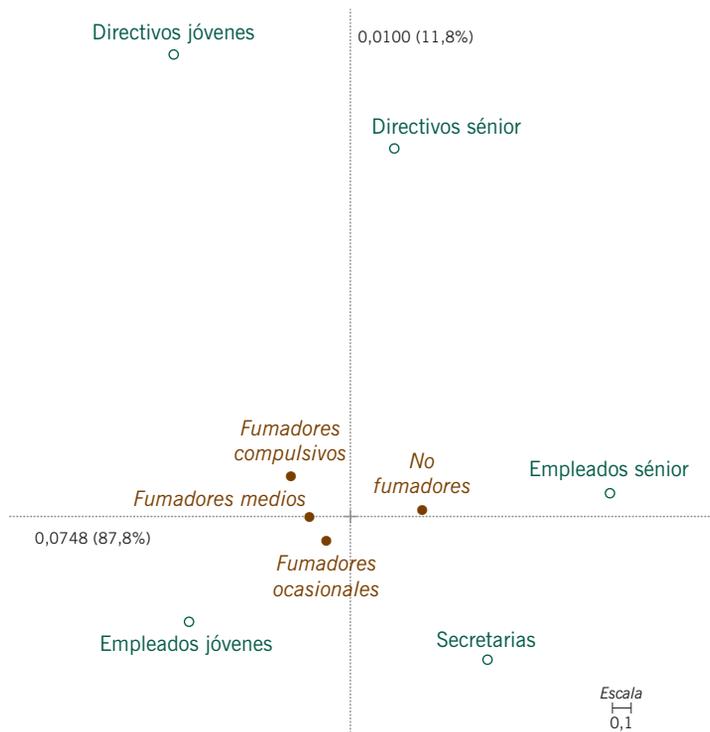


Imagen 9.3: Distancias observadas de los perfiles a los vértices de la imagen 9.2, representadas con relación a los correspondientes valores de los perfiles fila de la imagen 9.1. Hemos etiquetado cada par fila-columna con sus números de categoría correspondiente; por ejemplo, el perfil fila 3 (empleados séniores) y el vértice columna 4 (fumadores compulsivos) se denota como 34. Fijémonos en que, en cada vértice, salvo alguna excepción, a medida que aumentan los valores de los perfiles disminuyen las distancias

halla más cerca del vértice *fumadores ocasionales* que DJ, EJ debe contener relativamente más fumadores ocasionales que DJ. Los datos muestran que 24/88, el 27%, de los individuos de EJ son fumadores ocasionales, mientras que sólo 3/18, el 17%, de los DJ lo son, lo que concuerda con nuestra interpretación. En la imagen 9.3 comparamos, de forma gráfica, las distancias de los perfiles a los vértices con los correspondientes valores de los elementos de los perfiles expresados como porcentaje. Utilizamos la abreviación 42, para indicar la distancia observada de EJ-a-fumadores ocasionales (fila 4, columna 2) y 22 para indicar la de DJ-a-fumadores ocasionales (fila 2, columna 2). Así vemos que los EJ quedan más cerca del vértice *fumadores ocasionales* que los DJ, para esta categoría el valor del elemento correspondiente del perfil es de 0,27 para los EJ y sólo de 0,17 para los DJ. En cada vértice, los elementos de los perfiles se relacionan de forma *monotónicamente inversa* con las distancias de los perfiles a los vértices. Gráficamente (imagen 9.3) ello significa que, en cada vértice, los cinco perfiles se disponen de forma descendente, de izquierda a derecha. Por ejemplo, en el cuarto vértice (*fumadores habituales*), los perfiles con etiquetas 34, 54, 44, 14 y 24, se disponen en este tipo de secuencia descendente.

Imagen 9.4:

Mapa asimétrico del AC de los datos sobre los hábitos de los fumadores de la tabla 9.1, con las columnas en coordenadas principales y las filas en coordenadas estándares



Mapas asimétricos

Decimos que el mapa de la imagen 9.2 es un *mapa asimétrico*, o un mapa con *escalas asimétricas*, ya que es una representación conjunta de perfiles y vértices. En un mapa asimétrico, representamos las filas en coordenadas principales, y las columnas en coordenadas estándares o viceversa. Es decir, si estuviéramos más interesados en el análisis de las columnas que en el de las filas, representaríamos las columnas en coordenadas principales, y las filas en coordenadas estándares. Lo que dijimos en el capítulo 8 sobre el factor de escala entre las filas y las columnas se cumple para todos los ejes principales. En consecuencia, la representación bidimensional de los perfiles columna sería una versión encogida de las posiciones de los vértices mostrados en el mapa de la imagen 9.2. Sin embargo los «factores de contracción» (es decir, las correlaciones canónicas, que son iguales a las raíces cuadradas de las inercias principales) de los dos ejes no son los mismos: $\sqrt{0,07476} = 0,273$ y $\sqrt{0,01002} = 0,1000$, respectivamente. Por tanto, el factor de contracción del primer eje es de 0,273 (una contracción de poco menos de cuatro veces), y el factor de contracción del segundo eje es de 0,1 (diez veces). Siguiendo el mismo razonamiento, para pasar de los perfiles fila de la imagen 9.2 a las posiciones de sus vértices, simplemente tenemos que expandirlos aproximadamente cuatro veces en el primer eje y diez veces en el segundo eje. Aparte de estos factores de escala, las posiciones relativas de los perfiles y los vértices son las mismas. En la ima-

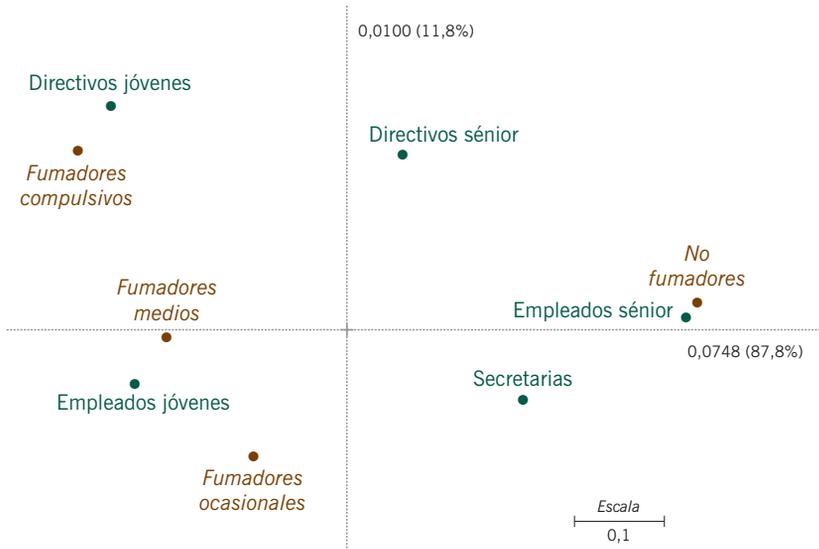


Imagen 9.5: Mapa simétrico de los datos sobre los hábitos de los fumadores. Hemos representado tanto las filas como las columnas en coordenadas principales

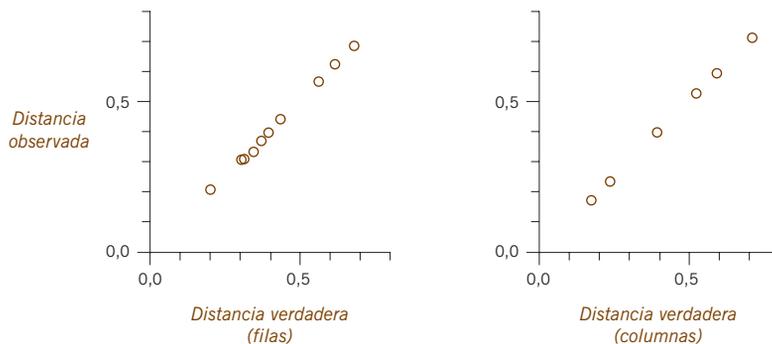
gen 9.4 podemos ver otro posible mapa asimétrico, en el que hemos representado las columnas como perfiles en coordenadas principales, y las filas como vértices en coordenadas estándares. En este último mapa, las posiciones de los perfiles columna se hallan a medias ponderadas de los vértices de las filas, los pesos son los elementos de los perfiles de las columnas. Al mapa asimétrico de la imagen 9.2 lo llamamos mapa en *filas principales* (ya que expresamos las filas en coordenadas principales), en cambio al mapa de la imagen 9.4 lo denominamos mapa en *columnas principales*.

Una vez examinada con bastante profundidad la explicación geométrica de las representaciones asimétricas, vamos ahora a introducir otra posibilidad de representación de los resultados, el *mapa simétrico*. Esta opción es, de lejos, la más popular en la literatura sobre el AC, especialmente entre los investigadores franceses. En los mapas simétricos solapamos en la misma representación, los perfiles fila y los perfiles columna, a pesar de que, en sentido estricto, las configuraciones de filas y columnas emanan de espacios distintos. Por tanto, en los mapas simétricos representamos tanto las filas como las columnas en coordenadas principales. Por ejemplo, el mapa de la imagen 9.5 es un mapa simétrico concerniente a los datos sobre los hábitos de los fumadores, en el que hemos solapado los dos conjuntos de puntos «interiores», que representamos mediante los círculos llenos en los mapas de las imágenes 9.2 y 9.4. La justificación de la representación conjunta de filas y de columnas hay que buscarla en la estrecha relación que existe entre el análisis de filas y el de columnas. Es decir, en la existencia de un solo factor de escala entre las filas y sus correspondientes vértices. La conveniencia de

Mapa simétrico

Imagen 9.6:

Distancias observadas entre las filas y las columnas en la imagen 9.5, representadas con relación a las correspondientes verdaderas distancias χ^2 entre los perfiles fila y los perfiles columna de la tabla 9.1



este tipo de representación radica en que cualquiera que sea el nivel de asociación, en los mapas simétricos la dispersión de los perfiles fila y perfiles columna es similar. Por tanto, es más difícil que en la representación gráfica se produzca un solapamiento de etiquetas. En cambio, en los mapas asimétricos, los perfiles (que en general son los puntos de principal interés) se hallan apretados en el centro de la representación, lejos de los vértices exteriores, lo que hace que la visualización sea menos estética.

Verificación de la distancia ji-cuadrado entre los puntos en un mapa simétrico

En el mapa de la imagen 9.5, en la que hemos representado conjuntamente los perfiles de las filas y los de las columnas, las distancias entre las filas que aparecen en el mapa son aproximadamente distancias χ^2 , de la misma manera que las distancias entre las columnas, son también aproximadamente distancias χ^2 . Al ser la representación de las filas idéntica a la de la imagen 9.2, podemos aplicar la misma interpretación sobre las distancias entre las filas (fijémonos, sin embargo, en la diferencia de escala de estos dos mapas), o sea, lo mismo es aplicable a las columnas de la imagen 9.4. Esta similitud de distancias entre puntos se puede verificar representando gráficamente las distancias observadas *versus* las verdaderas (imagen 9.6). Vemos que existe una excelente concordancia, esperable en tanto que, en ambos casos, la calidad de la representación de los perfiles es del 99,5%.

El peligro de interpretar las distancias entre las filas y las columnas en un mapa simétrico

La comodidad de los mapas simétricos, como el de la imagen 9.5, tiene un precio que deriva del riesgo de querer interpretar, de forma directa, las distancias entre filas y columnas. En estos mapas, no hemos definido ni tenemos previsto definir este tipo de distancias. Se trata de una peculiaridad del AC que, a menudo, es mal comprendida y que, frecuentemente, causa confusión entre los usuarios de los mapas simétricos a los que les gustaría realizar grupos formados por filas y columnas (en este sentido, véase el epílogo de la obra). De forma rigurosa, no es posible deducir a partir de la proximidad entre un punto fila y un punto columna, que la fila y la columna correspondientes presenten una asociación elevada. Este tipo de interpretación es, hasta cierto punto, posible sólo en el caso de mapas asi-

métricos como el de la imagen 9.3. Una regla de oro para la interpretación de este tipo de mapas es que podemos interpretar la distancia entre puntos siempre que éstos se hallen en el mismo espacio, como es el caso de los perfiles fila y de los vértices columna en el espacio de perfiles fila. Cuando interpretemos mapas simétricos, tenemos que tener siempre bien presente que un mapa simétrico no es más que el «solapamiento de dos mapas distintos». En el capítulo 13, describiremos «el biplot»; un mapa que nos permite interpretar de forma más precisa la visualización conjunta de las filas y las columnas.

1. Cuando en una representación gráfica aumenta la dimensionalidad de un subespacio, se incrementa la precisión de la representación de los perfiles. Sin embargo, al aumentar la dimensionalidad, la visualización de los puntos es más y más compleja. En general, preferimos las representaciones bidimensionales.
2. Los ejes principales están *anidados*; es decir, el eje principal de una representación unidimensional es idéntico al primer eje principal de una solución bidimensional, y así sucesivamente. Aumentar la dimensionalidad de una representación implica, simplemente, que añadimos nuevos ejes principales a los que ya hemos hallado.
3. Un *mapa asimétrico* es aquel en que representamos las filas y las columnas en escalas distintas, por ejemplo, las filas en coordenadas principales y las columnas en coordenadas estándares (son los vértices columna). Existen dos posibilidades, según sean de interés principal las filas o las columnas.
4. En un mapa asimétrico en el que, por ejemplo, representamos las filas en coordenadas principales (es decir, un análisis de filas), las distancias entre las filas son, aproximadamente, distancias χ^2 . Las distancias entre las filas y un vértice columna son, en general, inversamente proporcionales a los valores de los elementos del perfil de esa columna.
5. Sin embargo, en los *mapas simétricos*, la representación gráfica más frecuente, expresamos tanto las filas como las columnas en coordenadas principales.
6. En los mapas simétricos, las distancias entre las filas y las distancias entre las columnas son aproximadamente distancias χ^2 de sus respectivos perfiles. No obstante, en un mapa simétrico no existe una interpretación específica para las distancias entre las filas y las columnas.