Ejemplo

Si conocemos el triángulo de distancias por carretera entre ciertas capitales de provincias españolas, ¿podremos identificar unas coordenadas de las mismas de forma que nos permitan dibujar un mapa aproximado de sus posiciones?

Observemos que la distancia por carreteras entre cada dos ciudades, no es precisamente la distancia euclídea (no se mide en línea recta). Pero puede comprobarse que sí son realmente distancias en el sentido estudiado en el capítulo de las proximidades al cumplir todas las propiedades exigidas a éstas.

Nuestro objetivo será pues el de encontrar unas coordenadas para cada ciudad de forma que la distancia euclídea entre ellas reproduzca lo mejor posible las disimilaridades anteriores entre cada dos de ellas. Partamos para ello del siguiente triángulo de distancias entre las siguientes capitales de provincias españolas, análogo al que pudieran obtener consultando un mapa de carreteras, por ejemplo.

Distancias d_{rs}	La Coruña	San Sebastián	Barcelona	Almería	Cádiz	Salamanc	Madrid
La Coruña	0						
San Sebastián	763	0					
Barcelona	1118	529	0				
Almería	1172	1032	809	0			
Cádiz	1072	1132	1284	484	0		
Salamanca	473	469	778	763	599	0	
Madrid	609	469	621	563	663	212	0

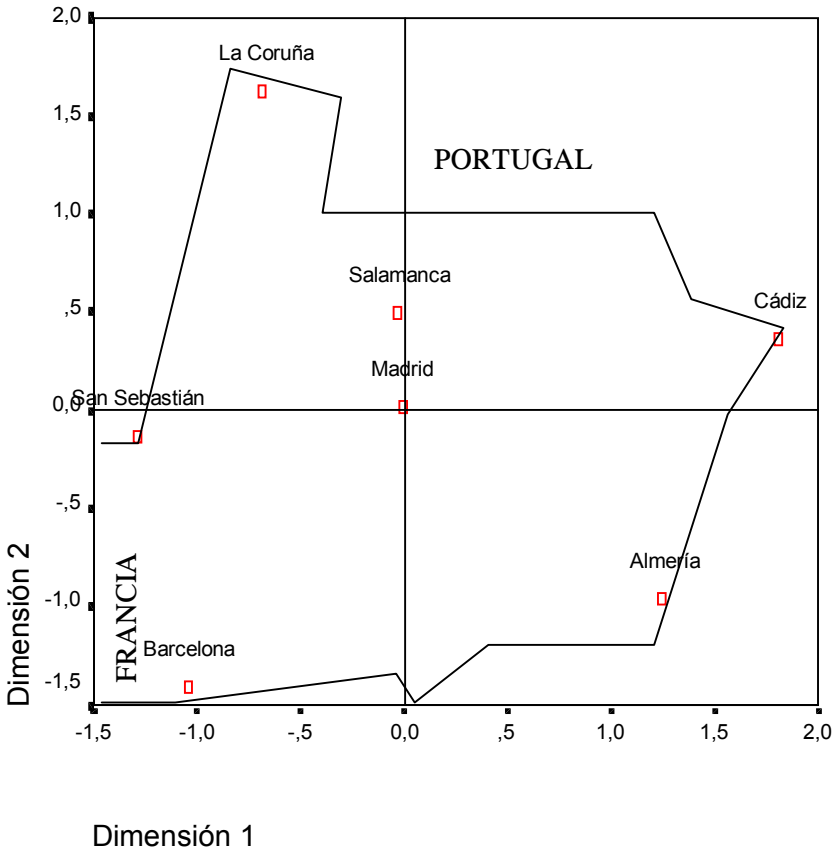
Obsérvese que la diagonal principal es nula, ya que la distancia de un punto a sí mismo lo es; y el resto de elementos nos marca la distancia en kilómetros entre las ciudades de la fila y columnas en las que están situados. Obsérvese también que no es necesario más que un triángulo de distancias de la matriz de distancias, ya que el otro es su simétrico.

¿Seríamos capaces de obtener a partir de aquí las coordenadas de cada ciudad, como para poder representarlas en un mapa? Este es el problema que nos resuelve la técnica del escalado multidimensional métrico. Así, tras aplicar el algoritmo de Torgenson, obtendríamos las siguientes coordenadas:

Coordenadas de las Capitales

	Dimensión 1	Dimensión 2
La Coruña	-0,6884	1,6225
San Sebastián	-1,2945	-0,1263
Barcelona	-1,0401	-1,4116
Almería	1,2471	-0,9566
Cádiz	1,8101	0,3605
Salamanca	-0,0286	0,4996
Madrid	-0,0058	0,0118

lo que nos permitiría representar las ciudades como sigue:



Observamos que efectivamente las posiciones relativas de las ciudades son correctas. Las distancias entre los puntos que las representan son compatibles con las que conocemos, con la salvedad de que las distancias de que partíamos no eran realmente distancias euclídeas (en línea recta), sino distancias por carretera (con sus curvas y desniveles sobre el terreno), por lo que el mapa sale algo distorsionado con respecto a los que podemos consultar. Pero en cualquier caso, vemos que sus situaciones sobre el papel difieren por otro motivo de las que estamos habituados a ver en los mapas de España. El mapa está desorientado, la Costa Norte con la frontera de Francia están orientadas a lo que tradicionalmente es el Oeste, y lo que es más, el mapa está como si le hubiésemos dado la vuelta o mirado a través de un espejo.

Todo ello se debe precisamente, a que la solución inicialmente encontrada no es la que estamos acostumbrados a contemplar, sino una rotación de ella. Para obtener la solución convencional, debemos pues rotarla en el sentido de llevar la Costa Norte a la parte alta del papel y darle la vuelta al mapa para que la Costa Atlántica con Portugal se encuentren al lado izquierdo del papel.

Una pregunta nos debe asaltar. ¿Por qué hemos representado estos datos en un plano de solamente dos dimensiones si como hemos visto, inicialmente, la solución de Torgenson, $X = (\sqrt{\lambda_1} \cdot e_1, \sqrt{\lambda_2} \cdot e_2, ..., \sqrt{\lambda_n} \cdot e_n)$, proporciona coordenadas en un espacio de dimensión n?

Esto tiene que ver con el problema de la dimensionalidad de la solución. En principio, la matriz B es una matriz simétrica y definida positiva de dimensión n, de la que efectivamente podemos extraer n posibles autovalores y autovectores ortonormalizados asociados. Al ser la matriz B simétrica y definida positiva, los autovalores serán no negativos. Además, análogamente a como ocurría en el análisis de Componentes Principales, su magnitud está relacionada con la variabilidad de los datos en la dimensión que su autovector asociado marca. A mayor magnitud, mayor variación de los datos en esa dimensión. Así pues los autovalores podrían ordenarse de mayor a menor magnitud, induciendo una ordenación de las dimensiones en función de su importancia en cuanto a la absorción de variabilidad de los datos.

Así pues, aunque teóricamente hay n dimensiones posibles (no más), es habitual que exista un número determinado de autovalores de B nulos, por lo que las dimensiones correspondientes serán inoperantes, presentando el valor constante (0) para todos los datos y podemos prescindir de ellas. Así pues, la dimensión exacta del espacio donde se representa la nube de puntos, viene proporcionada indirectamente por el método y no es otra que el número de autovalores no nulos de la matriz B,

En la práctica, autovalores muy pequeños proporcionarían valores de coordenadas muy pequeños para todos los datos, por lo que generalmente los despreciamos, fijándonos solamente en aquellas dimensiones que efectivamente expresan bien la mayor parte de la variabilidad de los datos.

Es por ello por lo que el mapa lo hemos representado sólo en dos dimensiones. Hemos seleccionado aquellas dos dimensiones que explican la mayor variabilidad (localización en superficie) ignorando y despreciando otras dimensiones menos importantes que, por ejemplo, nos hablaría de la altitud o de la esfericidad de la tierra.