

ESCALADO MULTIDIMENSIONAL

El escalamiento multidimensional es un método que trata de encontrar la estructura de un conjunto de medidas de distancia entre objetos o casos. Esto se logra asignando observaciones a posiciones específicas de un espacio conceptual (normalmente de dos o tres dimensiones) de modo que las distancias entre los puntos en el espacio concuerden al máximo con las disimilitudes dadas.

Para realizar esta práctica, vamos a analizar los datos que se encuentran en el fichero **Distancias_coche.sav**. Este fichero contiene las distancias kilométricas por carretera entre doce capitales de provincia españolas. Al abrir el fichero de datos, nos encontraremos con la pantalla de la Figura 1, en la que aparece una matriz de distancias que es triangular inferior (cabe recordar que las matrices de disimilitud son simétricas):

ciudad	barcelon	bilbao	la coruña	madrid	oviedo	salamanc	san seba	santande	sevilla	valencia	valladol	zaragoza	var
1 Barcelona	.00												
2 Bilbao	620.00	.00											
3 La Coruña	1118.00	644.00	.00										
4 Madrid	621.00	395.00	609.00	.00									
5 Oviedo	902.00	304.00	340.00	451.00	.00								
6 Salamanca	778.00	395.00	473.00	212.00	315.00	.00							
7 San Sebastián	529.00	119.00	763.00	469.00	423.00	469.00	.00						
8 Santander	694.00	107.00	548.00	394.00	206.00	364.00	228.00	.00					
9 Sevilla	1046.00	933.00	947.00	538.00	789.00	474.00	1007.00	837.00	.00				
10 Valencia	349.00	633.00	961.00	352.00	803.00	564.00	594.00	673.00	697.00	.00			
11 Valladolid	664.00	281.00	456.00	194.00	253.00	116.00	353.00	249.00	588.00	546.00	.00		
12 Zaragoza	296.00	324.00	833.00	325.00	604.00	482.00	268.00	397.00	863.00	326.00	367.00	.00	
13													
14													
15													
16													
17													
18													
19													
20													
21													
22													
23													
24													
25													
26													
27													
28													
29													
30													
31													
32													
33													

Figura 1: Matriz de distancias entre doce capitales de provincia seleccionadas.

Dentro de SPSS, el procedimiento que realiza el escalado multidimensional se encuentra en el submenú **Escalas** del menú **Analizar**, como se muestra en la Figura 2.

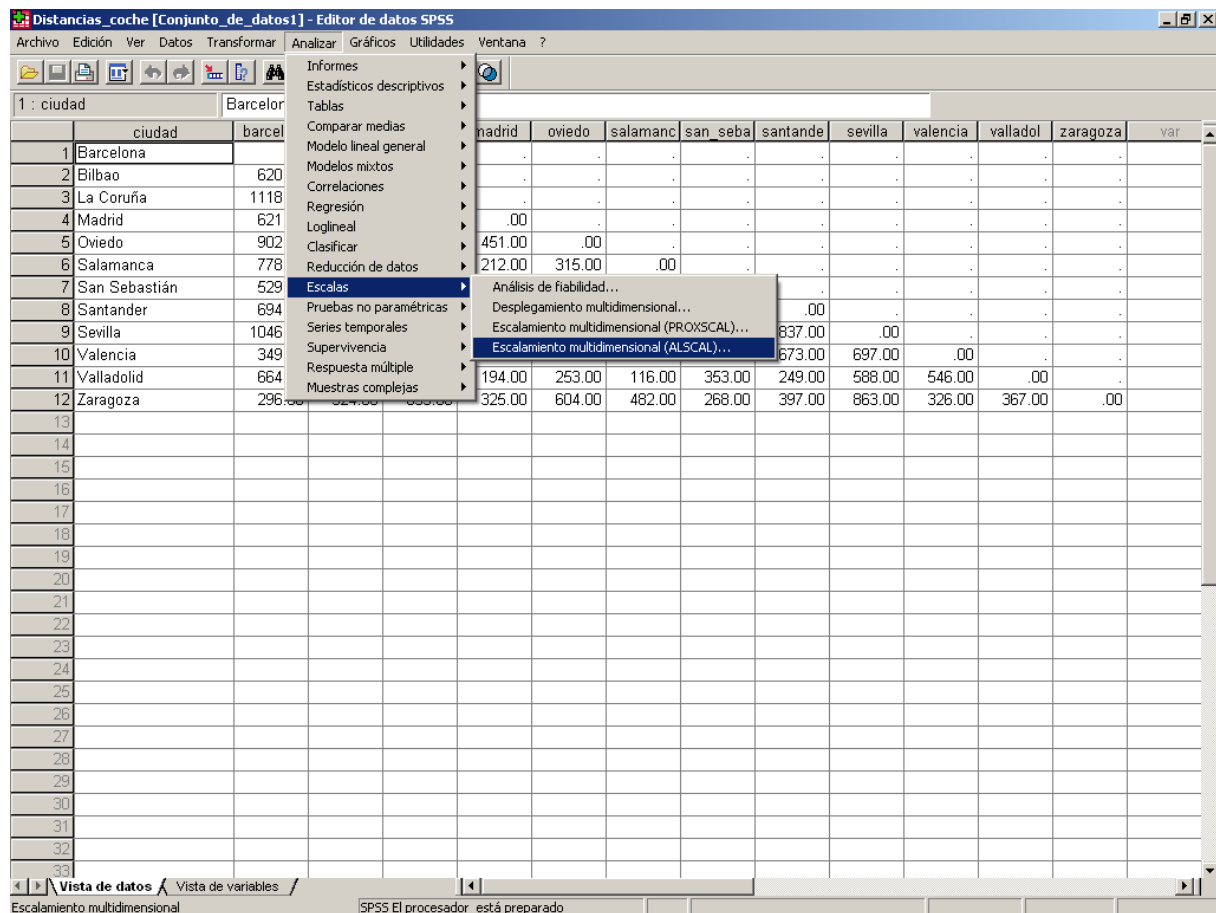


Figura 2: Selección del procedimiento **Escalamiento multidimensional (ALSCAL)**.

Al pulsar en dicha opción, el cuadro de diálogo que aparece tiene el aspecto de la Figura 3:

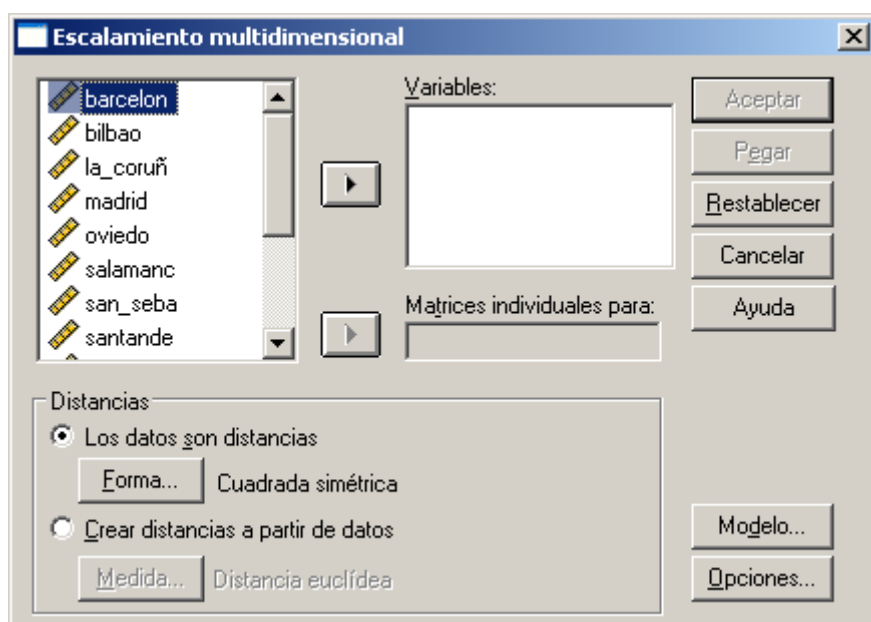


Figura 3: Cuadro de diálogo del procedimiento **Escalamiento multidimensional**.

donde figuran todas las opciones que permite este procedimiento. Para empezar, hay que seleccionar las variables que vayan a ser incluidas en el análisis (en nuestro ejemplo, los doce casos). Si el archivo de datos representa distancias entre un conjunto de casos (como es nuestro ejemplo), se debe especificar la forma de la matriz de distancias pulsando el botón **Forma** (para nuestro ejemplo, seleccionaremos la forma cuadrada simétrica).

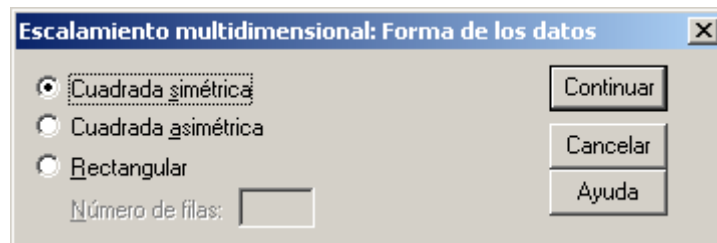


Figura 4: Opciones de **Forma** de la matriz de distancias del editor de datos.

Si los datos no fueran distancias, SPSS puede crear la matriz de distancias a partir de las variables observadas seleccionando dicha opción en el cuadro de diálogo del procedimiento. Si en este caso pulsamos el botón de **Medida**, podemos seleccionar el tipo de medida más conveniente para nuestro análisis:

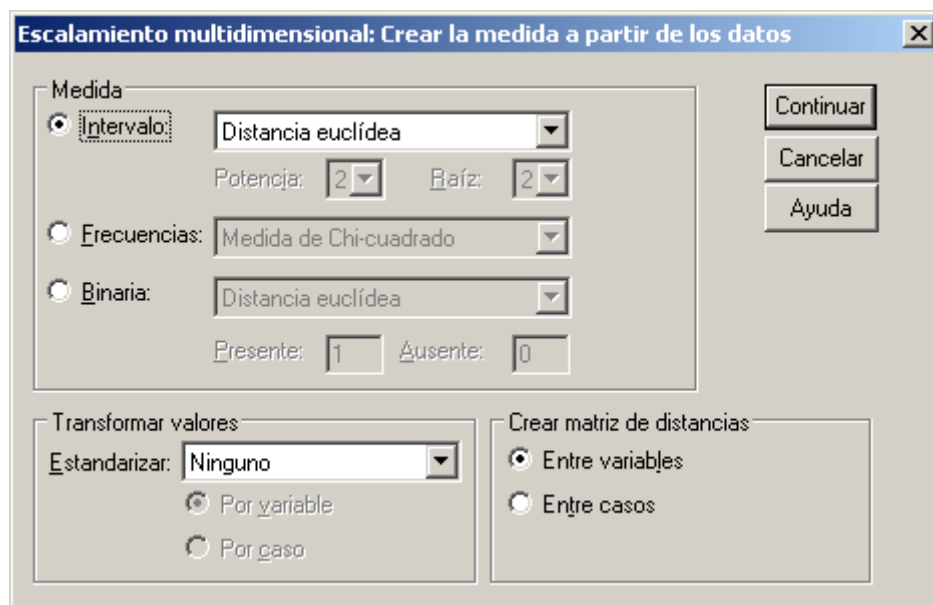


Figura 5: Especificación del tipo de medida para crear la matriz de distancias a partir de los datos.

La estimación correcta de un modelo de escalamiento multidimensional depende de los aspectos de los datos y del modelo en sí. En nuestro procedimiento, pulsando el botón de **Modelo**, aparece la siguiente ventana, en la que podemos seleccionar todas las opciones necesarias:

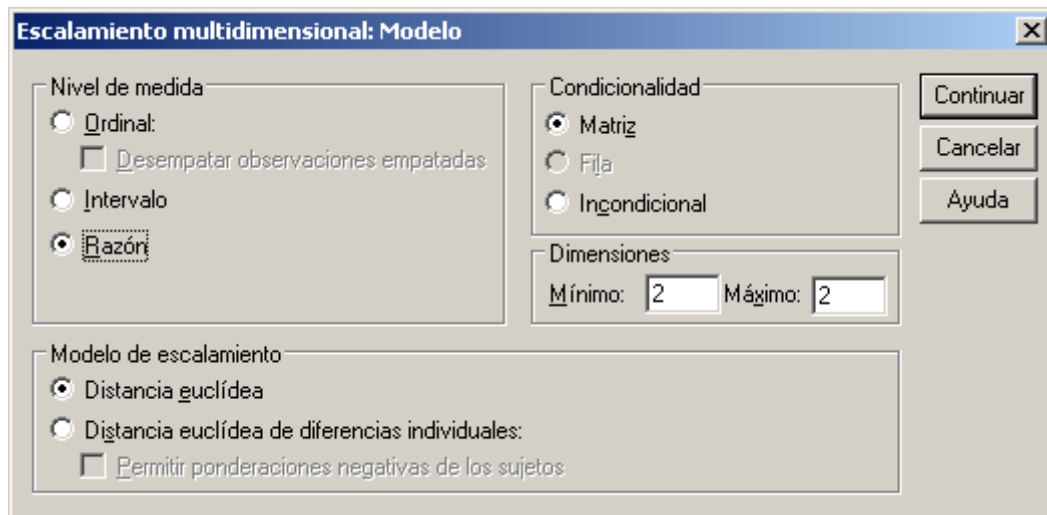


Figura 6: Determinación del modelo en el procedimiento de **Escalamiento multidimensional**.

El modelo euclídeo puede utilizarse con cualquier tipo de matriz. Si los datos constituyen una única matriz, como en nuestro ejemplo, SPSS ejecuta un análisis de escalamiento multidimensional clásico (CMDS). Si los datos están constituidos por dos o más matrices se genera un análisis de escalamiento multidimensional replicado (RMDS). Elegimos como nivel de medida el de razón, pues tenemos un conjunto de distancias kilométricas entre los casos que estamos analizando.

También se pueden especificar diversas opciones para el análisis de escalamiento multidimensional pulsando el botón de **Opciones**. Entre estas opciones se encuentran los diferentes resultados que se pueden visualizar, así como fijar cuál debe ser el comportamiento del algoritmo del escalado multidimensional.

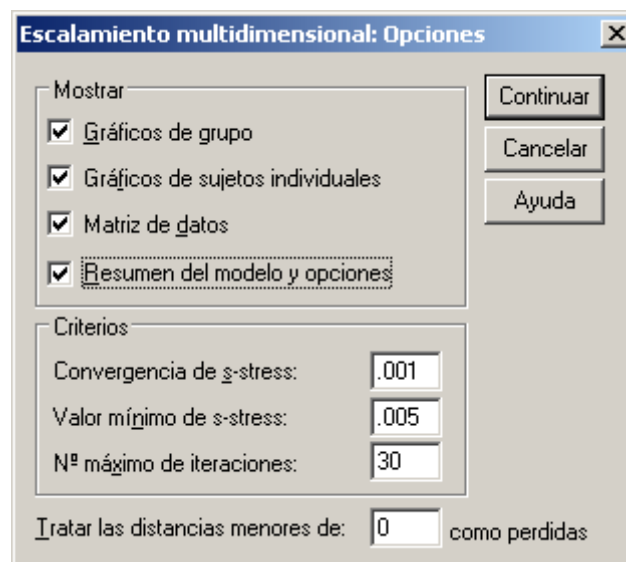


Figura 7: Opciones del procedimiento de **Escalamiento multidimensional**.

Al realizar el procedimiento, los resultados que se obtienen son los siguientes:

Escalamiento multidimensional

Para empezar, se presentan todas las opciones que han sido seleccionadas para realizar el procedimiento:

Alscal Procedure Options

Data Options-

Number of Rows (Observations/Matrix).	12
Number of Columns (Variables) . . .	12
Number of Matrices	1
Measurement Level	Ratio
Data Matrix Shape	Symmetric
Type	Dissimilarity
Approach to Ties	Leave Tied
Conditionality	Matrix
Data Cutoff at000000

Model Options-

Model	Euclid
Maximum Dimensionality	2
Minimum Dimensionality	2
Negative Weights	Not Permitted

Output Options-

Job Option Header	Printed
Data Matrices	Printed
Configurations and Transformations .	Plotted
Output Dataset	Not Created
Initial Stimulus Coordinates . . .	Computed

Algorithmic Options-

Maximum Iterations	30
Convergence Criterion00100
Minimum S-stress00500
Missing Data Estimated by	Ulbounds

A continuación, figura la matriz inicial de distancias entre los casos de estudio, según el orden de entrada de la matriz en el editor de datos.

Raw (unscaled) Data for Subject 1					
	1	2	3	4	5
1	.000				
2	620.000	.000			
3	1118.000	644.000	.000		
4	621.000	395.000	609.000	.000	
5	902.000	304.000	340.000	451.000	.000
6	778.000	395.000	473.000	212.000	315.000
7	529.000	119.000	763.000	469.000	423.000
8	694.000	107.000	548.000	394.000	206.000
9	1046.000	933.000	947.000	538.000	789.000
10	349.000	633.000	961.000	352.000	803.000
11	664.000	281.000	456.000	194.000	253.000
12	296.000	324.000	833.000	325.000	604.000
	6	7	8	9	10
6	.000				
7	469.000	.000			
8	364.000	228.000	.000		
9	474.000	1007.000	837.000	.000	
10	564.000	594.000	673.000	697.000	.000
11	116.000	353.000	249.000	588.000	546.000
12	482.000	268.000	397.000	863.000	326.000
	11	12			
11	.000				
12	367.000	.000			

Más tarde, aparece el proceso iterativo seguido por el procedimiento del Escalamiento multidimensional:

Iteration history for the 2 dimensional solution (in squared distances)

Young's S-stress formula 1 is used.

Iteration	S-stress	Improvement
1	.08614	
2	.07507	.01107
3	.07486	.00020

Iterations stopped because
S-stress improvement is less than .001000

Stress and squared correlation (RSQ) in distances

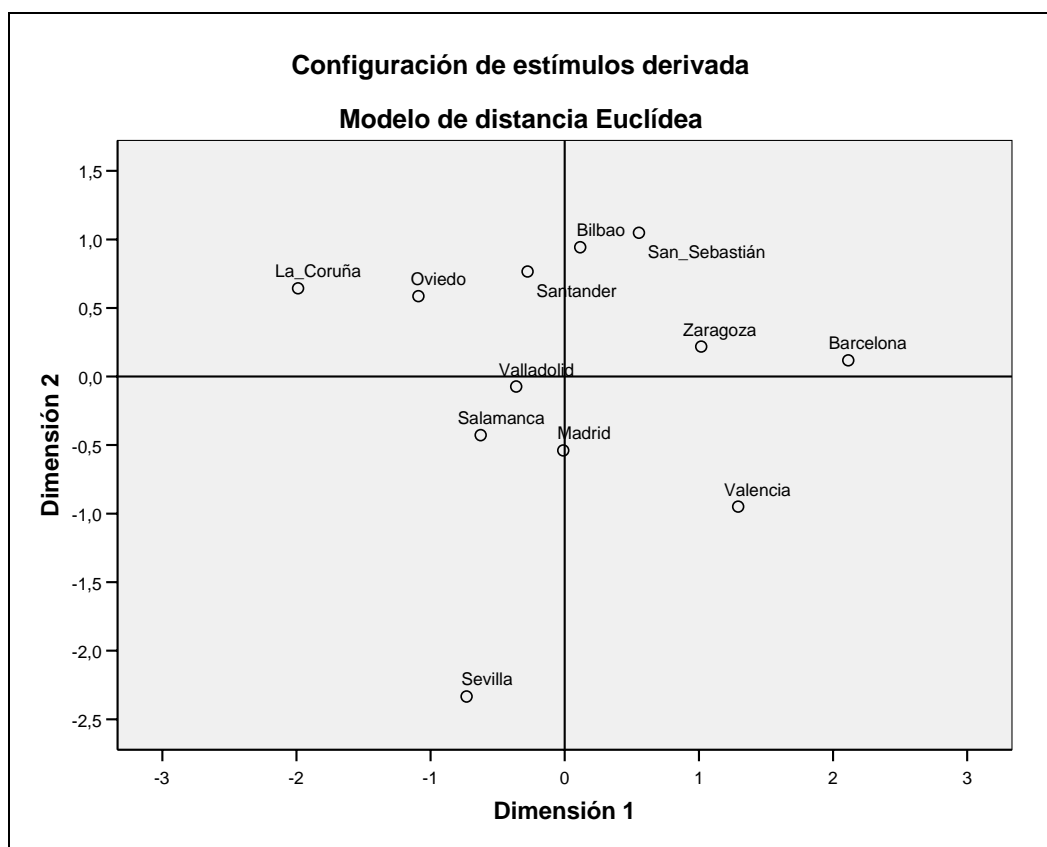
RSQ values are the proportion of variance of the scaled data (disparities) in the partition (row, matrix, or entire data) which is accounted for by their corresponding distances.
Stress values are Kruskal's stress formula 1.

For matrix
Stress = .05193 RSQ = .98515

Con estos resultados, la configuración óptima derivada en dos dimensiones, que representa la matriz de distancias inicial, es:

Configuration derived in 2 dimensions			
Stimulus Coordinates			
Dimension			
Stimulus Number	Stimulus Name	1	2
1	Barcelon	2.1142	.1178
2	Bilbao	.1147	.9431
3	La_Coruñ	-1.9887	.6432
4	Madrid	-.0114	-.5393
5	Oviedo	-1.0915	.5860
6	Salamanc	-.6272	-.4276
7	San_Seba	.5522	1.0488
8	Santande	-.2771	.7659
9	Sevilla	-.7326	-2.3339
10	Valencia	1.2928	-.9491
11	Valladol	-.3626	-.0734
12	Zaragoza	1.0171	.2184

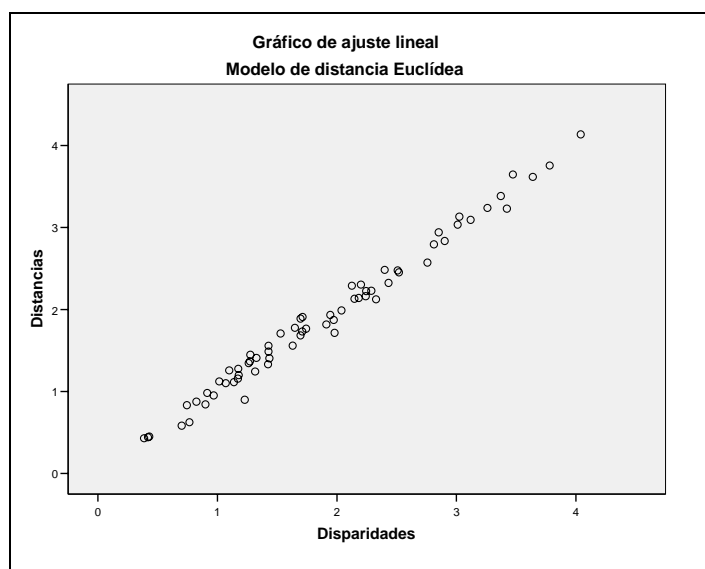
La representación de dicho sistema de coordenadas es la que se presenta en el gráfico siguiente, sobre el que se podría representar el mapa de España de forma aproximada, lo cual corrobora que la matriz de distancias ha sido bien transformada por el procedimiento:



La matriz de distancias calculada para las coordenadas en el nuevo sistema de coordenadas generado queda:

Optimally scaled data (disparities) for subject 1					
	1	2	3	4	5
1	.000				
2	2.241	.000			
3	4.040	2.327	.000		
4	2.244	1.428	2.201	.000	
5	3.260	1.099	1.229	1.630	.000
6	2.812	1.428	1.709	.766	1.138
7	1.912	.430	2.757	1.695	1.529
8	2.508	.387	1.980	1.424	.744
9	3.780	3.372	3.422	1.944	2.851
10	1.261	2.288	3.473	1.272	2.902
11	2.400	1.016	1.648	.701	.914
12	1.070	1.171	3.010	1.175	2.183
	6	7	8	9	10
6	.000				
7	1.695	.000			
8	1.315	.824	.000		
9	1.713	3.639	3.025	.000	
10	2.038	2.147	2.432	2.519	.000
11	.419	1.276	.900	2.125	1.973
12	1.742	.969	1.435	3.119	1.178
	11	12			
11	.000				
12	1.326	.000			

En el siguiente gráfico se comprueba que la aproximación a la matriz de distancias a la que hemos llegado con el escalamiento multidimensional es bastante buena, ya que la relación entre unas distancias y otras es aproximadamente lineal:



Como se aprecia en el gráfico derivado, la configuración de los casos que se ha representado en dos dimensiones se aproxima bastante al plano de las ciudades españolas consideradas. Sin embargo, este método trata de encontrar una configuración en una dimensión determinada que proporciona una matriz de distancias derivada, de tal forma que se repliquen las relaciones de distancia de partida entre los casos de análisis.

Para ver un ejemplo de una situación en la que la configuración derivada de los estímulos no coincide con la situación real, sino que se obtiene una rotación de la misma, se va a aplicar esta misma metodología al fichero de datos **Distancias_avion.sav**, en el que se han incorporado las distancias entre algunas capitales europeas, medidas por el tiempo (en horas) que se tarda en llegar de una a otra por avión.

Los resultados son los siguientes:

Escalamiento multidimensional

Alscal Procedure Options

Data Options-

Number of Rows (Observations/Matrix).	12
Number of Columns (Variables) . . .	12
Number of Matrices	1
Measurement Level	Ratio
Data Matrix Shape	Symmetric
Type	Dissimilarity
Approach to Ties	Leave Tied
Conditionality	Matrix
Data Cutoff at000000

Model Options-

Model	Euclid
Maximum Dimensionality	2
Minimum Dimensionality	2
Negative Weights	Not Permitted

Output Options-

Job Option Header	Printed
Data Matrices	Printed
Configurations and Transformations .	Plotted
Output Dataset	Not Created
Initial Stimulus Coordinates . . .	Computed

Algorithmic Options-

Maximum Iterations	30
Convergence Criterion00100
Minimum S-stress00500
Missing Data Estimated by	Ulbounds

Raw (unscaled) Data for Subject 1

	1	2	3	4	5
1	.000				
2	2.083	.000			
3	4.833	3.667	.000		
4	2.333	1.333	4.167	.000	
5	2.333	2.000	3.833	2.500	.000
6	2.833	1.667	2.833	1.667	2.083
7	2.150	1.917	5.000	1.333	3.250
8	3.500	2.333	2.000	2.417	2.333
9	3.167	1.917	2.583	1.917	2.583
10	1.167	2.500	5.500	2.500	3.000
11	2.167	1.500	3.833	1.917	1.167
12	2.417	1.000	3.750	1.167	2.250
	6	7	8	9	10
6	.000				
7	2.333	.000			
8	1.333	2.750	.000		
9	1.083	2.250	1.500	.000	
10	3.500	2.667	4.333	3.583	.000
11	1.667	2.750	2.167	2.083	2.500
12	1.417	1.667	2.083	1.583	2.750
	11	12			
11	.000				
12	1.583	.000			

Iteration history for the 2 dimensional solution (in squared distances)

Young's S-stress formula 1 is used.

Iteration	S-stress	Improvement
1	.13832	
2	.12357	.01475
3	.12237	.00120
4	.12232	.00005

Iterations stopped because
S-stress improvement is less than .001000

Stress and squared correlation (RSQ) in distances

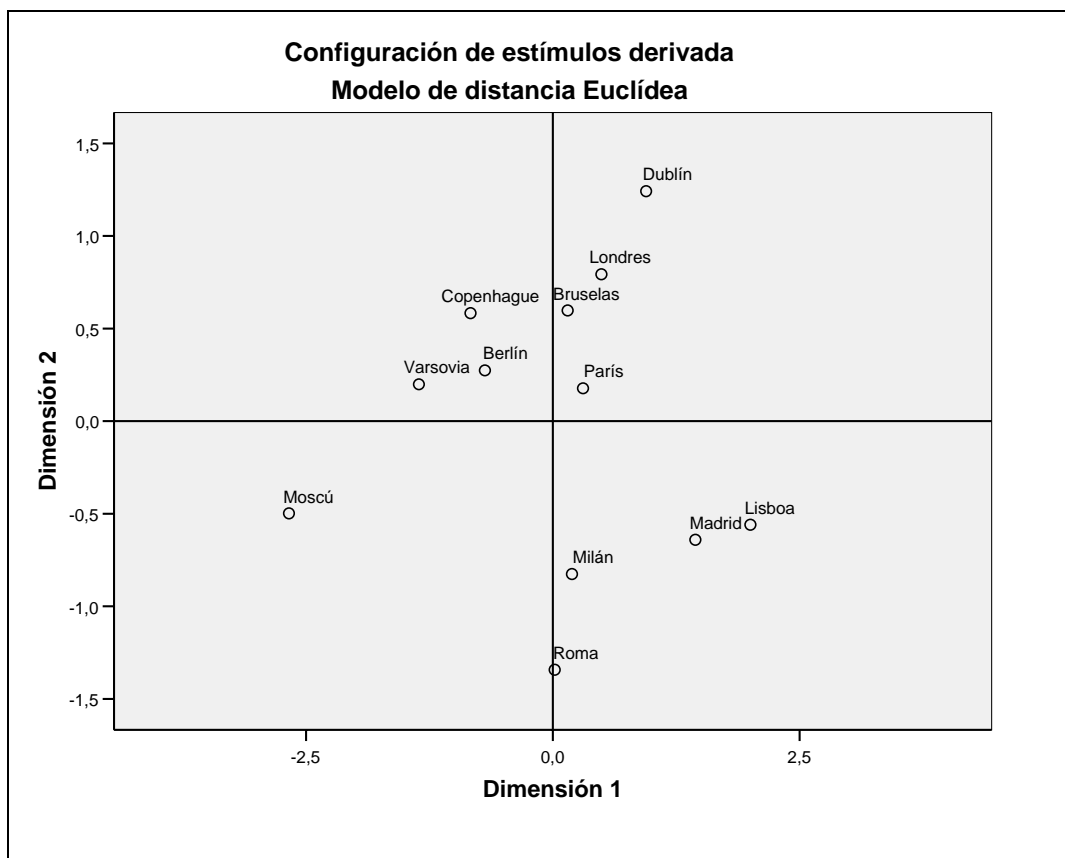
RSQ values are the proportion of variance of the scaled data (disparities) in the partition (row, matrix, or entire data) which is accounted for by their corresponding distances.
Stress values are Kruskal's stress formula 1.

For matrix
Stress = .11528 RSQ = .96111

Configuration derived in 2 dimensions			
Stimulus Coordinates			
Dimension			
Stimulus Number	Stimulus Name	1	2
1	Madrid	1.4443	-.6406
2	París	.3059	.1774
3	Moscú	-2.6725	-.4979
4	Londres	.4920	.7931
5	Roma	.0185	-1.3428
6	Berlín	-.6885	.2742
7	Dublín	.9450	1.2418
8	Varsovia	-1.3565	.1990
9	Copenhag	-.8333	.5834
10	Lisboa	2.0006	-.5596
11	Milán	.1938	-.8259
12	Bruselas	.1507	.5978

Optimally scaled data (disparities) for subject 1					
	1	2	3	4	5
1	.000				
2	1.691	.000			
3	3.924	2.977	.000		
4	1.894	1.082	3.383	.000	
5	1.894	1.624	3.112	2.030	.000
6	2.300	1.353	2.300	1.353	1.691
7	1.745	1.556	4.059	1.082	2.639
8	2.842	1.894	1.624	1.962	1.894
9	2.571	1.556	2.097	1.556	2.097
10	.947	2.030	4.465	2.030	2.436
11	1.759	1.218	3.112	1.556	.947
12	1.962	.812	3.044	.947	1.827
	6	7	8	9	10
6	.000				
7	1.894	.000			
8	1.082	2.233	.000		
9	.879	1.827	1.218	.000	
10	2.842	2.165	3.518	2.909	.000
11	1.353	2.233	1.759	1.691	2.030
12	1.150	1.353	1.691	1.285	2.233
	11	12			
11	.000				
12	1.285	.000			

Y la representación gráfica, en dos dimensiones, de los casos analizados, resulta ser:



A pesar de que se aprecia que la dimensión 1 está invertida, en relación con la situación real de los casos en las representaciones cartográficas habituales, la disposición relativa de las ciudades sí coincide con la situación en términos de las distancias entre las capitales, medidas por tiempo de vuelo entre ellas:

