

ESCALAMIENTO MULTIDIMENSIONAL: CONCEPTO Y APLICACIONES

Constantino Arce, Cristina de Francisco e Iria Arce

Universidad de Santiago de Compostela

A través del presente artículo se ofrece una visión conceptual, a la vez que operativa, del concepto de escalamiento multidimensional. En la forma de presentación se busca, en primer lugar, que los psicólogos interesados comprendan lo que es el modelo de escalamiento multidimensional a través de varios ejemplos muy sencillos e intuitivos y, en segundo lugar, adquieran competencias que le permitan resolver distintos problemas de escalamiento multidimensional con el uso de software específico. Se pretende igualmente descargar la presentación de fórmulas y métodos matemáticos sin renunciar por ello al rigor metodológico que el tema requiere.

Palabras clave: Escalamiento de objetos, Escalamiento de sujetos, Datos de proximidad, Datos de preferencia, Reducción de la dimensionalidad.

The present article offers a conceptual, and at the same time operative, vision of the concept of multidimensional scaling. In the manner it is presented, the aim is, firstly, to help interested psychologists understand what the multidimensional scaling model is, using a number of simple, intuitive examples; and, secondly, for them to acquire the competence required to resolved different problems in multidimensional scaling through the use of specific software. The aim is also to download the presentation of mathematical formulae and method, without renouncing the methodological rigour that the subject demands.

Key words: Scaling, Proximity data, Preference data, Dimensionality reduction.

El escalamiento multidimensional, en su formulación más básica, pretende representar un conjunto de objetos en un espacio de baja dimensionalidad. La palabra objeto es muy genérica y se refiere, en realidad, a cualquier entidad que deseemos escalar. Otro término equivalente utilizado en Psicología es estímulo. El número de dimensiones, habitualmente reducido (dos, tres, cuatro), las decide el investigador por razones sustantivas, aunque también puede hacerse por criterios estadísticos. Los modelos y métodos de construcción de escalas unidimensionales, que fueron desarrollados en la primera mitad del siglo XX, entre los que cabe citar a Thurstone, Likert, Guttman o Coombs, constituyen los antecedentes de los modelos y métodos más modernos de escalamiento multidimensional y, en muchas ocasiones, pueden éstos últimos considerarse como generalizaciones de aquellos.

El primer autor en desarrollar un modelo y un método de escalamiento multidimensional ha sido Torgerson (1958). A su modelo se le conoce, hoy en día, con el nombre de modelo métrico clásico. La denominación de métrico tiene que ver con la escala de medida que se re-

quiere, o asume, para los datos que es de intervalos, en la jerarquía de Stevens. Pocos años después Shepard (1962) y Kruskal (1964a, 1964b) han propuesto un modelo que permite un descenso en la escala de medida hasta el nivel ordinal. A este modelo se le denomina no-métrico clásico. Carroll y Chang (1970) lograron un avance significativo con la propuesta de un modelo que permite derivar, además del espacio de objetos, un espacio de sujetos sobre el que se representa el peso o ponderación que cada sujeto concede a cada una de las dimensiones del espacio de objetos. El modelo de Carroll y Chang, que se conoce con el nombre de modelo INDS-CAL, tiene gran interés psicológico dado que permite o tiene en cuenta las diferencias individuales en la percepción del espacio de objetos. Existe un espacio de objetos común, compartido por todos los sujetos, pero permite las diferencias entre unos individuos y otros en la percepción de dicha configuración.

Existen programas informáticos específicos para cada uno de los modelos señalados anteriormente pero hoy en día es posible resolver problemas múltiples de escalamiento multidimensional con un único programa de ordenador como, por ejemplo PROXSCAL O ALSICAL, que tienen implementados numerosos modelos y forman ambos parte del paquete estadístico SPSS de uso universal.

Uno de los rasgos que más diferencia al escalamiento

Correspondencia: Constantino Arce, Facultad de Psicología, Universidad de Santiago de Compostela, 15.782 Santiago de Compostela, España. E-mail: constantino.arce@usc.es

multidimensional de otros modelos estadísticos de análisis de datos es la matriz de entrada. En Psicología, estamos habituados a utilizar una matriz de datos rectangular X con n sujetos en las filas y p variables en las columnas, donde un elemento x_{ij} representa la medida obtenida para un sujeto i en una variable j . En su forma más típica, la matriz de entrada para el escalamiento multidimensional es una matriz de datos cuadrada de orden p con una misma entidad representada en las filas y en las columnas: los objetos que intentamos representar en el espacio multidimensional. Un elemento en esta matriz \emptyset representa la distancia o desemejanza entre dos objetos i y j . Lo que tenemos en la matriz es, en realidad, una matriz de distancias o desemejanzas entre todos los pares de objetos.

La diferencia entre distancia (concepto geométrico) y desemejanza (concepto psicológico) está en que el primero, al ser un concepto matemático, no contiene error; mientras el segundo, al ser un concepto psicológico, perceptivo o subjetivo, sí contiene error. Las desemejanzas son, en realidad, distancias que contienen error o distancias distorsionadas por los mecanismos perceptivos de los seres humanos. Los modelos y métodos de escalamiento multidimensional pueden resolver ambos tipos de problemas, con error y sin error en los datos de entrada. En Psicología es más habitual trabajar con datos que contienen error y los modelos de escalamiento multidimensional pueden tratar este problema.

DERIVACIÓN DE UNA CONFIGURACIÓN DE PUNTOS A PARTIR DE UNA MATRIZ DE DISTANCIAS

En la Tabla 1 se ofrece la matriz de distancias quilométricas entre 7 ciudades españolas: A Coruña, Bilbao, Barcelona, Cáceres, Madrid, Sevilla y Valencia.

Nos proponemos elaborar, a partir de dicha matriz, un mapa de España; es decir, obtener una representación espacial de las 7 ciudades sobre un plano, donde uno de los ejes será la dirección norte-sur y otro eje será la dirección este-oeste. Utilizamos, para ello, el procedimiento PROXSCAL, implementado en SPSS.

El resultado que nos ofrece es el que se puede observar en la Figura 1.

Dado que el mapa de España es conocido, podemos valorar subjetivamente el grado en que el mapa derivado por el programa se ajusta al mapa real. Podemos decir que el mapa conseguido es bastante bueno, aunque no perfecto. En la investigación en Psicología, es habitual trabajar con configuraciones que no tienen una con-

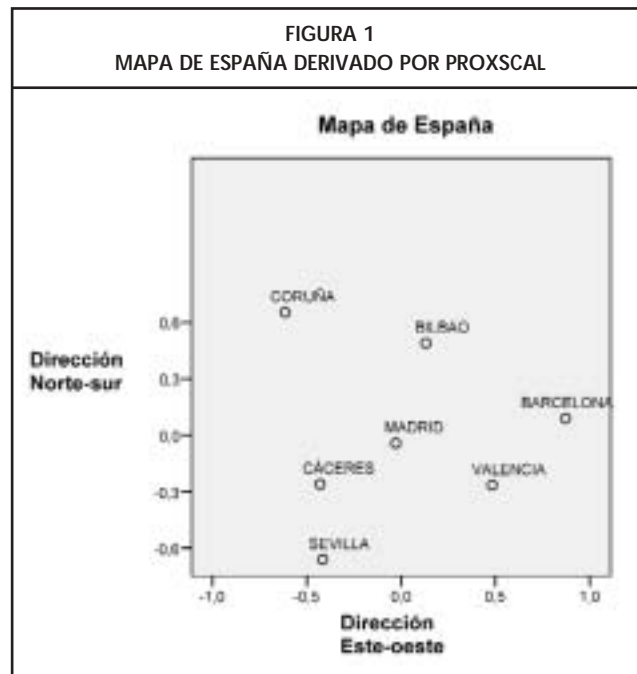
traparte objetiva conocida de antemano. Por eso, cuando el programa nos deriva una solución se vuelve muy importante tener un indicador o, incluso, varios--cuántos más mejor--, del grado en que la configuración derivada por el programa se ajusta a la ideal (desconocida). Todos los programas de escalamiento multidimensional ofrecen al usuario indicadores de ajuste para que pueda valorar lo "buena" que es la solución obtenida por el programa para su problema.

Los indicadores de bondad de ajuste ofrecidos por PROXSCAL para el mapa de España se ofrecen en la Tabla 2.

Hay dos tipos de indicadores. Aquellos para los que el cero representa un ajuste perfecto. De este primer tipo serían los indicadores Stress bruto normalizado, Stress-I, Stress-II y S-Stress. Y aquellos para los que el ajuste perfecto está representado por el 1. De este segundo tipo serían los dos últimos de la Tabla: Dispersión explicada

TABLA 1
DISTANCIAS QUILOMÉTRICAS ENTRE 7 CIUDADES ESPAÑOLAS

	A Coruña	Barcelona	Bilbao	Cáceres	Madrid	Sevilla	Valencia
A Coruña	0						
Barcelona	1050	0					
Bilbao	542	567	0				
Cáceres	617	895	591	0			
Madrid	586	600	379	294	0		
Sevilla	857	971	847	256	507	0	
Valencia	937	341	569	615	352	637	0



(D.A.F.) y Coeficiente de congruencia de Tucker. Observando los valores de unos y otros siempre se llega a la misma conclusión: que el ajuste del modelo es bueno o muy bueno en este caso. Esto es así porque el grado de error en los datos (distancias) era muy pequeño. Las distancias utilizadas como entrada eran las distancias por carretera. Si utilizáramos las distancias lineales el ajuste sería perfecto. Los cuatro primeros índices de ajuste deberían ser iguales a 0 y los dos últimos iguales a 1.

PERCEPCIÓN DE LOS MEDIOS DE TRANSPORTE PÚBLICO

Arce (1993) se propuso obtener un mapa perceptivo de los medios de transporte utilizados por los ciudadanos de Santiago de Compostela. Para ello, elaboró una lista de todos los medios de transporte (públicos y privados) que podrían estar a su disposición en la ciudad, formó con ellos todos los pares posibles y pidió a una muestra

de ciudadanos que juzgaran la desemejanza para cada par de medios de transporte.

Los medios de transporte estudiados fueron nueve: avión, tren, autobús interurbano, autobús urbano, taxi, coche particular, moto, ciclomotor y bicicleta. Con nueve objetos o estímulos (aquí, medios de transporte) se pueden formar 36 pares. Para averiguar el número de pares se utiliza la fórmula $n(n-1)/2$, donde n es el número de objetos o estímulos. Sustituyendo, en este caso, donde $n = 9$, nos queda $9(9-1)/2 = 36$. En la Tabla 3 se ofrecen los 36 pares formados en el estudio, siguiendo un método, denominado rotación estándar, muy útil porque los datos (desemejanzas) ya quedan ordenados en la forma en que luego se van a introducir en la matriz de entrada. El método sigue la secuencia (1,2), (1,3) ... (1,9), (2,3), (2,4) ... (2,9), (3,4) (3,5) ... (3,9) ... (8,9).

Para la formación del número de pares, hemos asumido la simetría de los juicios de desemejanza, queriendo esto decir que para un par dado (p.e. avión/tren), se asume que el juicio emitido por un sujeto sería el mismo si el par se presenta en el orden avión/tren que en el orden tren/avión. Salvo raras excepciones, este supuesto es habitual en la investigación en Psicología.

Una vez que tenemos el listado con todos los pares que queremos que los sujetos nos juzguen, debemos elaborar la escala de respuesta que deben utilizar los sujetos para juzgar la desemejanza de los objetos o estímulos incluidos en cada par. En la mencionada investigación se ha utilizado una escala de nueve puntos, donde 1 indicaba que los medios de transporte incluidos en el par eran muy parecidos y 9 que eran muy distintos. A modo de ejemplo:

Avión/tren								
Muy parecidos		Moderadamente parecidos				Muy distintos		
1	2	3	4	5	6	7	8	9

Los sujetos utilizaron esta escala para juzgar la desemejanza en los 36 pares formados en el estudio.

En el Figura 2 se ofrece el mapa perceptivo de los medios de transporte para un sujeto de la muestra. En la configuración de puntos obtenida ahora tenemos un problema añadido con respecto a la configuración de la Figura 1. En el problema de las distancias entre ciudades conocíamos el significado de los ejes, un eje era la dirección norte-sur y otro eje era la dirección este-oeste. Pero ¿qué significado tienen ahora los ejes de la

Stress bruto normalizado	,00055
Stress-I	,02349
Stress-II	,06824
S-Stress	,00117
Dispersión explicada (D.A.F.)	,99945
Coeficiente de congruencia de Tucker	,99972

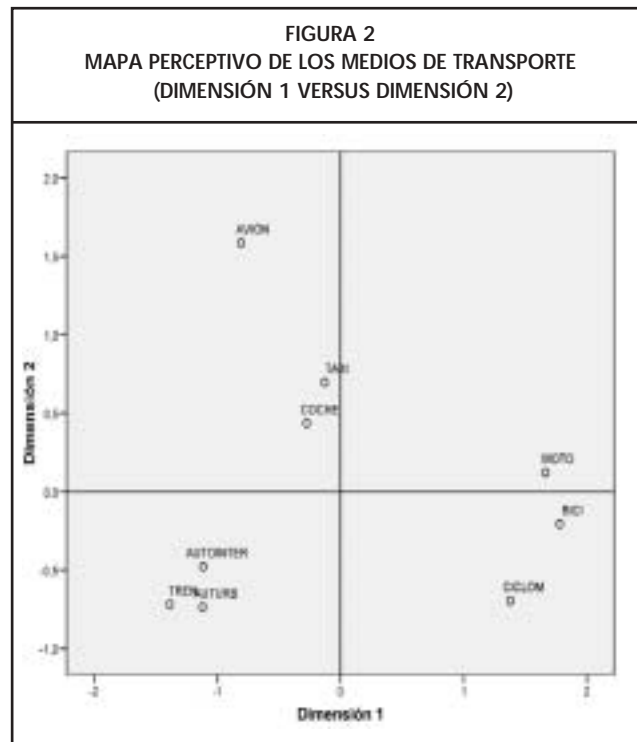
1. Avión/tren 19. Autobús interurbano/moto
2. Avión/autobús interurbano 20. Autobús interurbano/ciclomotor
3. Avión/autobús urbano 21. Autobús interurbano/bicicleta
4. Avión/taxi 22. Autobús urbano/taxi
5. Avión/coche particular 23. Autobús urbano/coche particular
6. Avión/moto 24. Autobús urbano/moto
7. Avión/ciclomotor 25. Autobús urbano/ciclomotor
8. Avión/bicicleta 26. Autobús urbano/bicicleta
9. Tren/autobús interurbano 27. Taxi/coche particular
10. Tren/autobús urbano 28. Taxi/moto
11. Tren/taxi 29. Taxi/ciclomotor
12. Tren/coche particular 30. Taxi/bicicleta
13. Tren/moto 31. Coche particular/moto
14. Tren/ciclomotor 32. Coche particular/ciclomotor
15. Tren/bicicleta 33. Coche particular/bicicleta
16. Autobús interurbano/autobús urbano 34. Moto/ciclomotor
17. Autobús interurbano/taxi 35. Moto/bicicleta
18. Autobús interurbano/coche particular 36. Ciclomotor/bicicleta

configuración que hemos obtenido? El sujeto probablemente haya utilizado en sus juicios distintos ejes o dimensiones para evaluar la desemejanza entre los medios de transporte. Por ejemplo, puede que para juzgar la desemejanza para un par dado se haya fijado en la seguridad de los medios de transporte. Para otro par pudo haberse fijado en el prestigio social, etc. Mediante el escalamiento multidimensional se busca obtener una configuración de puntos, pero también averiguar el significado de cada uno de los ejes o dimensiones de dicha configuración. Existen varios modos para enfrentarse a esta cuestión, pero el más fiable pasa por recoger más datos. En efecto, en la mencionada investigación, además de pedir a los sujetos los juicios de desemejanza, se les ha pedido que evaluaran cada uno de los medios de transporte en un serie de propiedades entre las que figuraban la seguridad, la estabilidad, la resistencia, la fuerza, el peso, la atracción, el prestigio, la puntualidad, el estatus social o la confortabilidad de los medios de transporte. Luego, se ha averiguado si existía algún tipo de relación entre alguna de estas propiedades y el posicionamiento de los medios de transporte en cada una de las dimensiones derivadas. En una primera fase exploratoria se probaron soluciones con 2, 3 y 4 dimensiones. La solución a la que se le encontró mejor significado fue la de 3 dimensiones. Análisis estadísticos de regresión múltiple, donde se tomaba como variable dependiente una propiedad dada de los medios de transporte y como variables independientes las coordenadas de los medios de transporte derivadas por los programas de escalamiento multidimensional, mostraron que la dimensión 1 (eje horizontal) representaba la seguridad percibida de los medios de transporte, la dimensión 2 (eje vertical) se refería al atractivo de los medios de transporte, y la dimensión 3 (eje de profundidad) representaba el prestigio social de los medios de transporte. En la Figura 2, donde se representan las dos primeras dimensiones de la solución tri-dimensional, los medios de transporte situados a la izquierda (tren, autobús interurbano, autobús urbano, etc.) se perciben como más seguros y los situados a la derecha (bicicleta, moto, ciclomotor) como más inseguros. De modo semejante, los medios de transporte situados más arriba (avión, taxi, coche, moto) se perciben como más atractivos y los situados más abajo como menos atractivos (autobús urbano, tren, ciclomotor, autobús interurbano, bicicleta).

**EL CASO DE MÁS DE UNA MATRIZ DE ENTRADA:
EL MODELO INDSCAL**

En los ejemplos utilizados hasta ahora disponíamos de una matriz de entrada. En el primer problema, la matriz de distancias quilométricas entre las siete ciudades españolas; y, en el segundo problema, la matriz de desemejanzas entre los medios de transporte para un sujeto de la muestra. En el primer problema, en realidad, no existía otra posibilidad porque la matriz de distancias es única, pero en el segundo problema disponemos de múltiples sujetos y nos hubiera gustado introducir la matriz de desemejanzas de cada sujeto. De hecho, en la investigación original se ha hecho así. Hoy en día cualquier programa de escalamiento multidimensional permite la obtención de un espacio de objetos común compartido por una muestra de sujetos u otra fuente de datos.

Entre los modelos que tratan el tema de múltiples matrices de entrada, existe uno que merece especial atención porque dispone de propiedades que pueden resultar muy interesantes desde el punto de vista psicológico. Se trata del modelo INDSCAL de Carroll y Chang (1970). Este modelo permite obtener dos espacios: el espacio de objetos, común para todos los sujetos de la muestra, y el espacio de sujetos. El aspecto novedoso del modelo es realmente este último espacio. En el espacio de sujetos se representa el peso, la ponderación o importancia que



cada sujeto de la muestra concede a cada una de las dimensiones de la configuración de objetos. Es decir, los sujetos comparten un mismo espacio de objetos pero el modelo permite que cada uno perciba dicho espacio de manera distinta; permite en definitiva las diferencias individuales entre unos y otros sujetos.

Arce (1994) pidió a dos sujetos que evaluaran la semejanza entre 7 marcas de coches: Ferrari, Porsche, BMW, Mercedes, Renault, Seat y Opel. Obtuvo dos ma-

trices de desemejanzas y utilizó ambas como entrada para un escalamiento multidimensional. Los resultados evidenciaron que la primera dimensión perceptiva eran los rasgos deportivos de la marca de coches y la segunda su confortabilidad. En la Figura 3 se ofrece el espacio de objetos, común, compartido por los dos sujetos. Los coches situados más a la derecha (dimensión 1) se perciben como más deportivos que los situados a la izquierda y los situados más abajo (dimensión 2) como más confortables que los situados más arriba.

En la Figura 4 se ofrece el espacio de sujetos. Mientras el espacio de objetos es común para los dos sujetos, el espacio de sujetos nos indica que el sujeto 1 (SRC_1 en el gráfico) concede más importancia a la dimensión 1, los rasgos deportivos del coche, mientras el sujeto 2 (SRC_2, en el gráfico) concede más importancia a la confortabilidad de los coches. A diferencia del espacio de objetos donde cada objeto se representa por un punto, en el espacio de sujetos, el sujeto se representa por un vector (una línea). Cuanto más cerca esté el vector de una dimensión, más importancia le concede el sujeto a dicha dimensión y cuanto más alejado menos importancia. En efecto, se observa en el gráfico que el sujeto 1 está más cerca de la dimensión 1 (diseño deportivo), concediendo por tanto más importancia a esta dimensión, mientras el sujeto 2 está más cerca de la dimensión 2 indicando que en su caso es ésta la dimensión (confortabilidad de los coches) la que adquiere más peso en sus juicios sobre las marcas de coche.

FIGURA 3
ESPACIO DE OBJETOS

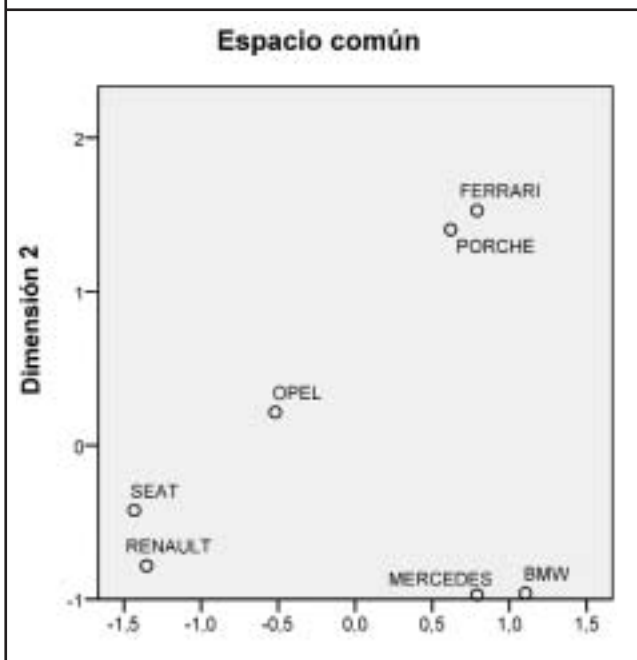
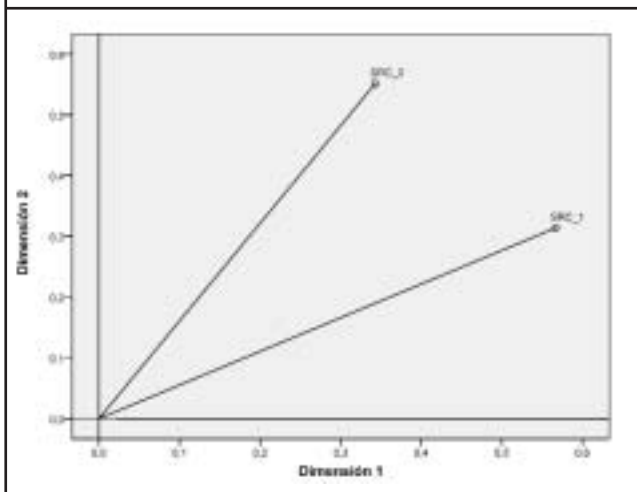


FIGURA 4
ESPACIO DE SUJETOS



ESCALAMIENTO MULTIDIMENSIONAL CON DATOS DE PREFERENCIA

Aunque el escalamiento multidimensional utiliza, en su forma más típica, una matriz de desemejanzas entre objetos como entrada, se han desarrollado modelos y métodos que permiten el escalamiento multidimensional de objetos a partir de datos de preferencia (p.e. Bennett y Hays, 1960; Carroll, 1980; Tucker, 1960). Si tenemos n objetos que queremos escalar, al sujeto simplemente se le pide que los posicione por orden de preferencia, asignándole el número 1 al objeto más preferido, el 2 al segundo más preferido y así sucesivamente hasta el último objeto, al que debe asignar el número n. Estos datos tienen la ventaja de que son más cómodos de obtener que los datos de desemejanza. La tarea suele ser mucho más simple tanto para el sujeto como para el investigador. Los datos de preferencia se ordenan, luego, en una matriz rectangular, con sujetos en las filas y objetos en las

columnas. Cada fila es un sujeto y un elemento de la fila representa el orden (o preferencia) que dicho sujeto ha concedido a un objeto dado.

A modo de ejemplo, supongamos que estuviésemos interesados en obtener un mapa perceptivo de los deportes y actividades físicas que los ciudadanos pueden practicar en su tiempo de ocio. Para hacer el ejemplo manejable, elegimos 8 deportes o actividades físicas y 16 sujetos, a los que les pedimos que nos indiquen sus preferencias marcando con un 1 el deporte o actividad física más preferida para él/ella, con un 2 el deporte o actividad física que prefiere en segundo lugar y así sucesivamente hasta marcar el que prefiere en último lugar, al que le asignaría el número 8.

Los deportes o actividades físicas elegidos en el ejemplo son: fútbol, baloncesto, tenis, atletismo, caminar, nadar, andar en bicicleta y correr.

Las preferencias indicadas por los sujetos se ofrecen en Tabla 4.

En la Figura 5 se ofrece el mapa perceptivo de los deportes y actividades físicas evaluadas por los sujetos de la muestra. Para interpretar el significado de las dimensiones, nos fijamos, en primer lugar, en las propiedades de los deportes o actividades físicas que ocupan los lugares más extremos en cada una de las dimensiones. Es probable que tengan alguna propiedad contrapuesta que nos ayude a interpretar el significado de la dimensión respectiva. Así, se puede observar que en la dimensión 1 (horizontal) a la derecha están situadas las

actividades físicas no competitivas (caminar, nadar, correr y andar en bicicleta) y a la izquierda los deportes competitivos (fútbol, atletismo, tenis y baloncesto). Podría interpretarse, por tanto, la dimensión 1 como la competitividad de los deportes o actividades físicas. De modo semejante, si observamos el posicionamiento de los



TABLA 4
PREFERENCIAS DE LOS SUJETOS

Sujeto	Fútbol	Baloncesto	Tenis	Atletismo	Caminar	Nadar	Andar Bicicleta	Correr
1	8	7	6	5	1	4	3	2
2	7	8	5	6	2	3	4	1
3	8	7	6	5	1	3	2	4
4	7	8	5	6	2	4	3	1
5	6	5	7	8	1	2	3	4
6	5	6	7	8	2	3	4	3
7	6	5	7	8	2	1	3	4
8	5	6	8	7	3	2	4	3
9	1	2	3	4	5	6	7	8
10	2	1	4	3	5	7	6	8
11	1	2	4	3	8	7	6	5
12	2	1	3	4	8	6	7	5
13	3	4	1	2	8	6	7	5
14	4	3	1	2	8	7	6	5
15	4	3	2	1	7	8	5	6
16	3	4	2	1	8	6	7	5

deportes y actividades físicas en la dimensión 2 (eje vertical), podemos apreciar que hacia arriba están los deportes o actividades físicas de naturaleza individual (atletismo, correr, tenis, etc.) y en la parte inferior los deportes colectivos (fútbol, baloncesto). Podría interpretarse, por tanto, esta segunda dimensión como el tipo de deporte o actividad: individual versus colectivo.

RESOLUCIÓN DE PROBLEMAS DE ESCALAMIENTO MULTIDIMENSIONAL CON SPSS

Hasta ahora hemos pasado por encima, sin detenernos, en el proceso de resolución de los problemas de escalamiento por medio de software específico. Vamos a reproducir ahora cómo hemos resuelto el problema de las distancias quilométricas con el procedimiento PROXSCAL implemen-

tado en SPSS. Simultáneamente, en ocasiones, indicaremos cuáles son las diferencias en la toma de decisiones entre este problema y los otros tres que también hemos resuelto.

Paso 1. Creamos el fichero de datos en SPSS con las distancias quilométricas entre las siete ciudades españolas. Debe tener la apariencia que ofrece la Figura 6. Dado que la matriz de entrada es cuadrada, las filas en la matriz tienen el mismo significado que las columnas. La fila 1 es Coruña, la fila 2 Barcelona y así sucesivamente hasta la fila 7 que es Valencia. El hecho de que aparezca el nombre de las ciudades en las columnas y no aparezca en las filas es porque el sistema SPSS permite, utilizando la pestaña Vista de variables, etiquetar o dar nombres a las columnas de la matriz de datos pero no así a las filas.

Paso 2. Elegimos el procedimiento que queremos ejecutar:

Analizar/Escalas/Escalamiento multidimensional (PROXSCAL)

Paso 3. Forma de los datos

El procedimiento PROXSCAL permite dos tipos de datos de entrada:

- (a) datos de proximidad (matriz cuadrada)
- (b) datos de perfil (matriz rectangular)

Las distancias quilométricas, al igual que las desemejanzas entre objetos, son datos de proximidad. Elegimos, por tanto, la opción que indica al programa que los datos son proximidades (ver Figura 7).

Si tuviésemos una matriz rectangular de entrada como, por ejemplo, cuando disponíamos de preferencias en el ejemplo de los deportes o actividades físicas, entonces tendríamos que elegir la opción que indica al programa que debe crear proximidades de los datos.

Paso 4. Número de matrices de entrada

El procedimiento PROXSCAL permite una o más de una matriz de entrada. El número de matrices de las que disponemos en el problema se indica en el recuadro denominado número de fuentes. Como, en este caso, sólo disponemos de una matriz elegimos la opción una fuente matricial (ver Figura 7). En el tercer problema que hemos resuelto aquí, el de las marcas de coche, hemos elegido la opción varias fuentes matriciales, dado que disponíamos de dos matrices, una por sujeto. Las matrices, en el fichero de entrada se sitúan unas debajo de otras, respetando el mismo formato en todas ellas.

FIGURA 6

FICHERO DE ENTRADA PARA SPSS CON LAS DISTANCIAS QUILOMÉTRICAS ENTRE LAS SIETE CIUDADES

	CORUÑA	BARCELONA	BILBAO	CÁCERES	MADRID	SEVILLA	VALENCIA
1	.00						
2	1060.00	.00					
3	542.00	597.00	.00				
4	817.00	895.00	491.00	.00			
5	836.00	600.00	379.00	204.00	.00		
6	867.00	371.00	647.00	256.00	507.00	.00	
7	837.00	341.00	569.00	615.00	362.00	637.00	.00

FIGURA 7

ELECCIÓN DE LA FORMA DE LOS DATOS Y DEL NÚMERO DE MATRICES DE ENTRADA EN EL PROCEDIMIENTO PROXSCAL

Formato de datos:

Los datos son proximidades

Crear proximidades de los datos

Número de fuentes:

Una fuente matricial

Varias fuentes matriciales

Una fuente:

Las proximidades están en una matriz en las columnas.

Las proximidades están en una sola columna.

Varias fuentes:

Las proximidades están en matrices apiladas en las columnas.

Las proximidades están en columnas, una fuente por columna.

Las proximidades están apiladas en una sola columna.

Paso 5. Pulsamos el botón Definir (ver Figura 7)

Paso 6. Seleccionamos los objetos que queremos escalar (aquí ciudades)

Para ello marcamos las 7 ciudades en el recuadro de la izquierda de la Figura 8 y las pasamos al recuadro denominado Proximidades pulsando la flecha que está entre ambos.

Paso 7. Elección del Modelo (pulsamos el botón denominado Modelo)

Dado que tenemos una sola matriz de proximidades (distancias) de entrada, el procedimiento no permite elegir el modelo de escalamiento (ver Figura 9) que, en cualquier caso, será semejante al modelo clásico.

Si tuviésemos más de una matriz de proximidades de entrada, entonces sí que podríamos elegir el modelo de escalamiento. Los más habituales serían el modelo con replicación (denominado modelo de Identidad en el cuadro de diálogo) y el modelo INDSCAL (denominado Euclídeo ponderado en el cuadro de diálogo). En el modelo de replicación los sujetos se consideran replicas unos de otros, lo que significa que las diferencias que puedan existir entre ellos se atribuyen a factores aleatorios. El modelo INDSCAL, por el contrario, permite las diferencias individuales. En el problema de las marcas de coche, hemos elegido este modelo.

Paso 8. Más decisiones sobre la forma y la naturaleza de los datos

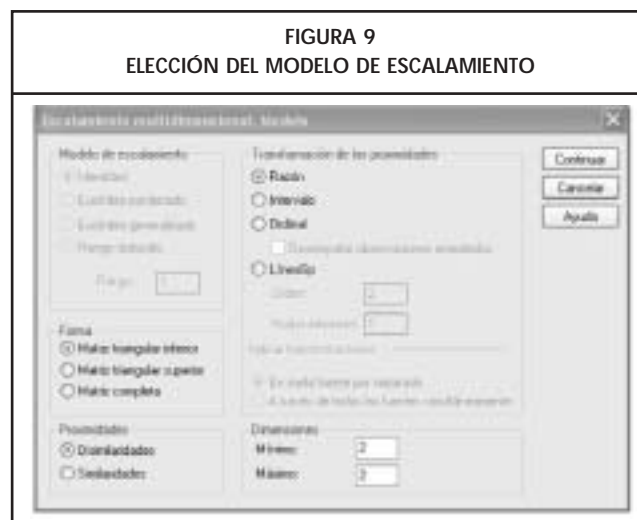
El programa hasta ahora sabe que tenemos una matriz de proximidades de entrada (cuadrada), pero todavía nos pide, bajo la denominación de Forma (ver Figura 9), que le especifiquemos si la información la tenemos en el triángulo inferior, en el superior o en toda la matriz. Como la matriz de distancias es simétrica, hemos optado por disponer la información tan sólo en la mitad inferior (matriz triangular inferior). La opción de matriz completa sólo se usa cuando la matriz de entrada es asimétrica.

Los programas de escalamiento multidimensional proyectan las desemejanzas como distancias en el espacio. Cuanto mayor sea la desemejanza entre objetos mayor será su distancia en el espacio multidimensional. Pero los datos de entrada pueden ser desemejanzas o semejanzas. Si fuesen semejanzas entonces la relación con las distancias sería inversa: cuanto mayor sea la semejanza entre dos objetos en el mundo empírico menor sería la distancia entre ellos en el espacio. Esta es, pues, una es-

pecificación sustancial que el usuario debe hacer al programa. En nuestro caso, en el recuadro Proximidades, debemos elegir Disimilaridades (sinónimo de desemejanzas). Las distancias se conciben como desemejanzas.

Bajo la denominación de Transformación de las proximidades (ver Figura 9), el procedimiento permite al usuario elegir la escala de medida para los datos de entrada. Si elegimos razón o intervalos, el modelo será métrico y si elegimos ordinal el modelo será no-métrico. Rara vez en la investigación en Psicología se elige el nivel de medida de razón; los más frecuentes son intervalos u ordinal. Aquí, en este problema, sin embargo, dado que los datos son distancias y no juicios de los sujetos, elegimos el nivel de medida más alto (razón).

En el problema de los deportes, donde utilizábamos datos de preferencia, hemos especificado que el nivel de



medida era ordinal. En este caso, el programa todavía nos permite hacer una especificación más, bajo la denominación de Desempatar observaciones empatadas. Esta es una decisión muy técnica. Si la elegimos, el programa asumirá que el proceso de medida es continuo y si no la elegimos que es discreto. Esta decisión tan sólo tiene repercusión para aquellos casos en que haya empates. Por defecto, el programa asume que el proceso de medida es discreto y respeta los empates en los datos. Si consideramos que el proceso de medida es continuo debemos especificar al programa que proceda a desempatar las observaciones empatadas. En nuestro problema de los deportes, hemos probado con las dos opciones sin que hayamos notado diferencias en las soluciones derivadas por el programa. De hecho, esto es lo que ocurre en la mayoría de las ocasiones. Es decir, se trata de decisiones que tienen importancia a nivel matemático pero no tanta a nivel sustantivo.

Paso 9. Número de dimensiones

Si tenemos una hipótesis clara de partida podemos elegir un número de dimensiones fijo, y si no la tenemos lo mejor es que probemos a obtener soluciones distintas y a posteriori seleccionemos la solución con el número de dimensiones que sea más interpretable desde un punto de vista sustantivo. En nuestro ejemplo de las distancias, el número de dimensiones que hemos elegido, dado que se trataba de un mapa, fueron dos (ver Figura 9).

En el ejemplo de los medios de transporte, como puede ocurrir en otras muchas investigaciones en Psicología que se realizan con carácter exploratorio, no teníamos una hipótesis tan clara en cuanto al significado de las dimensiones que podríamos obtener. En consecuencia, hemos probado a obtener soluciones en dos, tres y cuatro dimensiones. Luego, a posteriori, hemos intentado buscarles un significado. Hemos comprobado que eran interpretables tres y ésta fue la solución que hemos elegido.

La interpretación de las dimensiones puede hacerla el investigador tratando de analizar, en primer lugar, las propiedades de los objetos que ocupan posiciones más extremas en la dimensión. Cuando el procedimiento separa mucho a los objetos suele ser porque tienen alguna propiedad contrapuesta que, si la identificamos, puede ayudarnos a dar nombre a la dimensión. Este procedimiento lo hemos utilizado en el ejemplo de los deportes. No obstante, esta interpretación basada en la opinión de un experto (el investigador) puede ser discutida por otros investigadores (o expertos). Lo ideal es proceder como

hemos hecho en el problema de los medios de transporte, donde además de los juicios de desemejanza entre objetos se ha pedido a los sujetos que evaluaran cada uno de los medios de transporte sobre una serie de escalas bipolares que representaban hipotéticas propiedades de los medios de transporte. Luego, por métodos estadísticos de correlación y regresión, se ha podido ofrecer evidencia de cuál era el verdadero significado de cada una de las dimensiones retenidas.

Paso 10. Restricciones y Opciones

Tanto el botón de Restricciones como el de Opciones (ver Figura 8) permiten al usuario tomar decisiones de un nivel muy avanzado. En la práctica, se suelen tomar las opciones que el procedimiento tiene implementadas por defecto. En Restricciones, por defecto, el programa asume que debe estimar todas las coordenadas de los objetos (sin restricciones). A veces, de manera excepcional, las coordenadas se conocen y lo único que se busca es proyectar nuevos objetos sobre un espacio ya definido. En este caso, habría que proporcionar al programa las coordenadas que leería en un archivo que nosotros le indiquemos.

De forma semejante, en Opciones, por defecto, el programa toma una determinada configuración inicial (simplex) que nos permite cambiar por otras alternativas (p.e. Torgerson). También nos permite cambiar los criterios para alcanzar la convergencia y el número de iteraciones que realiza el algoritmo. Rara vez se podrán obtener mejores resultados si se cambian las opciones que el programa tiene incorporadas por defecto.

Paso 11. Toma de decisiones sobre la salida

En los botones denominados Gráficos y Resultados (ver Figura 8), el programa permite al usuario elegir lo que quiere que aparezca en la salida. Podemos oscilar desde una salida muy simple, con los elementos sustanciales, hasta una salida muy sobrecargada con todo tipo de detalles técnicos. Es aconsejable, en un primer momento, obtener una salida simple y, luego, si fuese necesario, obtener nuevas salidas con más elementos informativos. Por defecto, en Gráficos, el programa nos ofrece el gráfico más relevante que es el espacio de objetos (espacio común). Adicionalmente podríamos pedirle otros gráficos que nos permitirían visualizar el grado de ajuste del modelo. Por su parte, por defecto, en Resultados, el programa nos ofrece las coordenadas de los objetos (aquí ciudades) y los índices de ajuste del modelo, tal como aparecen en la Tabla 2.

PROGRAMAS DE ORDENADOR PARA EL ESCALAMIENTO MULTIDIMENSIONAL

Existe una amplia lista de programas de ordenador para la solución de problemas de escalamiento multidimensional. Nosotros hemos resuelto todos los problemas anteriores haciendo uso del procedimiento PROXSCAL, implementado en SPSS, pero el mismo paquete estadístico dispone de otro procedimiento, denominado ALSICAL, que también permite la solución de múltiples problemas de escalamiento multidimensional. Para acceder a este procedimiento se debe seguir la secuencia Analizar/Escalas/Escalamiento multidimensional (ALSICAL). Cualquier problema de los que aquí hemos tratado se podría haber resuelto igualmente con ALSICAL.

En la Tabla 5 se ofrece una pequeña relación de programas de ordenador actualmente disponibles en el mercado. Además de los ya señalados, PROXSCAL y ALSICAL, es posible resolver problemas de escalamiento multidimensional con otros programas tales como GGVIS, PERMAP, MULTISCALE o NewMDSX. GGVIS y PERMAP comparten la propiedad de ser interactivos y de poderse adquirir gratuitamente por In-

ternet. MULTISCALE tiene la ventaja de estar también disponible de forma gratuita, pero es de difícil manejo. Su autor, Ramsay, goza de un gran prestigio en la historia del escalamiento multidimensional. Finalmente, NewMDSX es, en realidad, un paquete de programas que permite la solución de problemas de escalamiento multidimensional y de otro tipo de problemas relacionados.

Si desea más información sobre programas de ordenador y, de manera más general, sobre el escalamiento multidimensional en relación con la historia, los modelos, los métodos y las múltiples posibilidades de aplicación en Psicología y temas relacionados pueden servirle de ayuda los libros de Borg y Groenen (2005), el manual más reciente que se haya escrito sobre escalamiento multidimensional a fecha de hoy, Kruskal y Wish (1978), Arabie, Carroll y DeSarbo (1987), Green, Carmone y Smith (1989), o Arce (1993, 1994). Para ejemplos de aplicaciones véase Wish, Deutsch y Kaplan (1976) o Sabucedo y Arce (1990).

AGRADECIMIENTOS

Esta investigación ha sido realizada con la ayuda de la Dirección Xeral de Investigación, Desenvolvemento e Innovación de la Xunta de Galicia (PGIDIT06PXIB211187PR).

REFERENCIAS

- Arabie, P., Carroll, J.D., y DeSarbo, W.S. (1987). *Three-way scaling and clustering*. Newbury Park, CA: Sage.
- Arce, C. (1993). *Escalamiento multidimensional: una técnica multivariante para el análisis de datos de proximidad y preferencia*. Barcelona: Promociones y Publicaciones Universitarias (PPU).
- Arce, C. (1994). *Técnicas de construcción de escalas psicológicas*. Madrid: Editorial Síntesis.
- Bennett, J.F., y Hays, W.L. (1960). Multidimensional unfolding: determining the dimensionality of ranked preference data. *Psychometrika*, 25, 27-43.
- Borg, I., y Groenen, P.J.F. *Modern multidimensional scaling. Theory and applications*. Nueva York: Springer.
- Buja, A., y Swayne, D.F. (2002). Visualization methodology for multidimensional scaling. *Journal of Classification*, 19, 7-44.
- Carroll, J.D. (1980). Models and methods for multidimensional analysis of preferential choice (or other dominance) data. En E.D. Lantermann y H. Feger (Eds.), *Similarity and choice* (pp. 234-289). Viena: Hans Huber.

TABLA 5 PROGRAMAS DE ORDENADOR PARA EL ESCALAMIENTO MULTIDIMENSIONAL		
NOMBRE	DISPONIBILIDAD	DOCUMENTALES
PROXSCAL	En SPSS http://www.spss.com/	Commandeur y Heiser (1993), Meulman, Heiser y SPSS (1999), De Leeuw y Heiser (1980)
ALSICAL	En SPSS http://www.spss.com/	Takane, Young y De Leeuw (1977)
GGVIS	Gratuito, por Internet http://www.ggobi.org E-mail: ggobi-help@ggobi.org	Buja y Swayne (2002)
PERMAP	Gratuito, por Internet http://www.ucs.louisiana.edu/~rbh8900 E-mail: ron@heady.us	Ron B. Heady, University of Louisiana, Lafayette, USA
MULTISCALE	Gratuito, por Internet ftp://ego.psych.mcgill.ca/pub/ramsay/multiscl/ o dirigiéndose al autor, profesor James O. Ramsay, e-mail: ramsay@psych.mcgill.ca	Ramsay (1977)
NewMDSX	http://www.newmdsx.com/	Coxon (2004)

- Carroll, J.D. y Chang, J.J. (1970). Analysis of individual differences in multidimensional scaling via N-way generalization of Eckart-Young decomposition. *Psychometrika*, 35, 283-319.
- Commandeur, J.J.F., y Heiser, W.J. (1993). Mathematical derivations in the proximity scaling (PROXSCAL) of symmetric data matrices. Tech. Rep. No. RR-93-03. Leiden, The Netherlands: Department of Data Theory, Leiden University.
- Coxon, A.P.M. (2004). Multidimensional Scaling. En M.S. Lewis-Beck, A. Bryman, y T. F. Liao (Eds.), *The Sage Encyclopedia of Social Science Research Methods*. Thousand Oaks: Sage.
- De Leeuw, J., y Heiser, W.J. (1980). Multidimensional scaling with restrictions on the configuration. En P.R. Krishnaiah (Ed.), *Multivariate analysis* (Vol. V, pp. 501-522). Amsterdam, Holanda: North-Holland.
- Green, P.E., Carmone, F.J., y Smith, S.M. (1989). *Multidimensional scaling. Concepts and applications*. Boston: Allyn and Bacon.
- Kruskal, J.B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29, 1-27.
- Kruskal, J.B. (1964b). Nonmetric multidimensional scaling: a numerical method. *Psychometrika*, 29, 115-129.
- Kruskal, J.B., y Wish, M. (1978). *Multidimensional scaling*. Newbury Park, CA: Sage.
- Meulman, J.J., Heiser, W.J. y SPSS (1999). *SPSS Categories 10.0*. Chicago: SPSS.
- Ramsay, J.O. (1977). Maximun likelihood estimation in multidimensional scaling. *Psychometrika*, 42, 241-266.
- Sabucedo, J.M. , y Arce, C. (1990). Types of political participation: a multidimensional analysis. *European Journal of Political Research*, 20, 93-102.
- Shepard, R.N. (1962). The analysis of proximities: multidimensional scaling with an unknown distances function (I y II). *Psychometrika*, 27, 125-139, 219-246.
- Takane, Y., Young, F.W., y De Leeuw, J. (1977). Non-metric individual differences multidimensional scaling: an alternating least-squares method with optimal scaling features. *Psychometrika*, 42, 7-67.
- Torgerson, W.S. (1958). *Theory and methods of scaling*. Nueva York: Wiley.
- Tucker, L.R. (1960). Intra-individual and inter-individual multidimensionality, en H. Gulliksen y S. Messick (Eds.), *Psychological scaling: theory and applications* (pp. 155-167). Nueva York: Wiley.
- Wish, M., Deutsch, M. y Kaplan, S.J. (1976). Perceived dimensions of interpersonal relations. *Journal of Personality and social Psychology*, 33, 409-420.

