Introducción.

Tratemos de introducir la problemática y pretensiones de esta técnica, a través de un ejemplo. Si conociéramos el triángulo de distancias entre las capitales de provincias españolas, ¿seríamos capaces de posicionar, en un cierto espacio, el conjunto de puntos que generan esas distancias? ¿Podríamos identificar las coordenadas geográficas precisas de cada una de las capitales, de forma que nos permitan reproducir el triángulo de distancias?

La intuición nos hace pensar primeramente en que el espacio en el que pretendemos encontrar las coordenadas de las capitales de provincias es un espacio de dos dimensiones, si nos conformamos con la representación habitual que de las mismas solemos ver en los diversos tipos de mapas (carreteras, político,...) o quizás de tres si quisiéramos tener información de la altitud de las mismas sobre el nivel del mar. Obsérvese que, en cualquier caso, podríamos dibujar un mapa de las capitales consideradas sobre el plano de dos cualesquiera de las dimensiones resultantes.

Pero, si el caso no fuera tan conocido como el del ejemplo, ¿cuántas dimensiones deberíamos considerar para poder representar adecuadamente los puntos que generan las distancias? Así pues, paralelamente a la investigación de las coordenadas precisas de cada punto, el problema también contempla la determinación del número de dimensiones mínimamente requerido.

La técnica se extiende, además, para el caso en que lo que tengamos de partida sea un triángulo de disimilaridades entre los casos, no necesariamente distancias. En este sentido, la utilidad de esta técnica será máxima cuando consideremos datos sobre los que se han observado características de tipo cualitativo y sobre los que se han construido las distancias o disimilaridades.

Es evidente que algunas técnicas importantes que hemos visto como las del Análisis Discriminante, el Análisis de Componentes Principales o las del Análisis Factorial, no pueden ser aplicadas sobre variables de tipo cualitativo, ya que exigen datos observados sobre escalas de intervalo. En el caso de datos observados en escalas cualitativas, la técnica del Escalado Multidimensional nos permitirá encontrar una cierta "proyección óptima" de los datos sobre un espacio de tipo euclídeo y, en consecuencia, aquéllas técnicas sí podrían ser aplicadas. Esa "proyección óptima" sería aquélla que mejor reproduzca la relación de proximidad existente entre los datos originales y que viene plasmada en la matriz inicial de distancias o de disimilaridades, a través de la distancia euclídea finalmente medida sobre el espacio resultante.

Obsérvese que, en cierto sentido, esto es una generalización de lo que hace el Análisis Factorial de Correspondencia. Allí, a partir de dos características (en el caso del Análisis de Correspondencias Simple) o más (en el caso del Análisis de Correspondencias Múltiple), "proyectábamos" las modalidades analizadas sobre un espacio, donde la distancia euclídea entre puntos reproducía las distancias de Benzecrí entre las modalidades originalmente

observadas a partir de la correspondiente Tabla de Contingencia. Ahora, la Técnica del Escalado Multidimensional trata de hacer algo similar pero con los casos, tratando que la distancia euclídea entre los puntos del nuevo espacio reproduzca la proximidad entre los casos observadas a partir de la correspondiente matriz de distancias o disimilaridades.

Hablaremos de **Escalado Multidimensional Métrico** cuando partamos de matrices de distancias, y de **Escalado Multidimensional No Métrico** cuando partamos de matrices de disimilaridades. En cualquiera de los casos, vamos buscando "proyectar" nuestros casos, aunque estén observados en escalas ordinales o nominales, sobre espacios euclídeos, donde la distancia euclídea reproduzca lo más fielmente posible la similitud o proximidad de los mismos (inicialmente dada en forma de matriz de disimilaridades o distancias no necesariamente euclídeas) y donde poder aplicarles técnicas propias de escalas de intervalo, imposible de aplicar en el espacio inicial.

Escalado Multidimensional Métrico.

La primera solución al problema, cuando la matriz de partida es una matriz de distancias entre los puntos de un espacio, la dio Torgenson en 1952.

Esta primera aproximación parte de considerar n elementos en un cierto espacio en el que se ha definido una cierta distancia y de los que, en consecuencia, se conoce la matriz de distancias entre cada par de ellos. No conocemos nada acerca de ese espacio, sino que se conoce exclusivamente la matriz de distancias. La aportación fundamental de Torgenson es un procedimiento para calcular las coordenadas de n puntos en un espacio real $\Re^n$, cada uno de ellos representando a cada uno de los n elementos iniciales, de tal manera que la distancia euclídea entre cada dos puntos reproducen fielmente las distancias entre los elementos en el espacio inicial.

El problema así planteado se conoce como **Problema Clásico**, el cual conduce a la **Técnica Básica del Escalado Multidimensional Métrico** y que podemos enunciarlo de la siguiente forma:

"Conocida las distancias d_{rs}, r=1,...,n; s=1,...,n, entre cada dos elementos de un conjunto de n en un cierto espacio ¿podríamos obtener unas coordenadas $x_r=(x_{r1},...,x_{rp})'$, representantes de cada elemento, en algún cierto espacio euclídeo $\Re^p$, para alguna cierta dimensión p de tal manera que $d_{rs} = \sqrt{\sum_{j=1}^{p}(x_{rj} - x_{sj})^2}$?"

Supongamos dos individuos genéricos, r y s, para los que conocemos su distancia (y por tanto la distancia euclídea de sus representantes en $\Re^p$), d_{rs}, pero no sus coordenadas, $x_r=(x_{r1},...,x_{rp})'$ y $x_s=(x_{s1},...,x_{sp})'$, que son precisamente nuestras incógnitas finales, además de la dimensión p del espacio final donde se representarán.

Llamando b_{rs} al producto escalar $x_r'x_s$, podemos ver fácilmente, por la simetría de los productos escalares, que: $b_{sr} = x_s'x_r = x_r'x_s = b_{rs}$.

Así, si conociéramos los coeficientes de la matriz $B=((b_{rs}))_{nxn}=XX'$, y fuesen (λ_i,e_i),

i=1,2,...,n sus autovalores y sus respectivos autovectores (columnas) ortonormalizados asociados, entonces podríamos ver que una solución posible para la matriz de coordenadas buscada para los casos sería:

$$X = (\sqrt{\lambda_1}\cdot e_1, \sqrt{\lambda_2}\cdot e_2, ..., \sqrt{\lambda_n}\cdot e_n)$$

Esto es evidente ya que, al estar los autovectores ortonormalizados, se cumple que

$$e_j^{'}\cdot e_j = 1 \quad , \forall j = 1,2,...,n \quad ;; \quad e_i^{'}\cdot e_j = 0 \quad , \forall i \neq j \; ; \; i = 1,2,...,n \; ; \; j = 1,2,...,n$$

y entonces,

$$B\cdot e_j = XX'e_j = (\sqrt{\lambda_1}\cdot e_1, \sqrt{\lambda_2}\cdot e_2, ,..., \sqrt{\lambda_n}\cdot e_n)\cdot(\sqrt{\lambda_1}\cdot e_1, \sqrt{\lambda_2}\cdot e_2, ..., \sqrt{\lambda_n}\cdot e_n)'\cdot e_j =$$
$$= (\sqrt{\lambda_1}\cdot e_1, \sqrt{\lambda_2}\cdot e_2, ..., \sqrt{\lambda_n}\cdot e_n)\cdot(0,...,0,\sqrt{\lambda_j},0,...,0)'\cdot e_j = \lambda_j\cdot e_j$$

lo que asegura que efectivamente, con esta solución para la matriz de coordenadas X, (λ_i, e_i), i=1,2,...,n son los autovalores y autovectores de B=XX'

No olvidemos que nuestro objetivo es encontrar las coordenadas a partir de las distancias, por lo que, si podemos obtener los valores de los b_{rs} a partir de las d_{rs}, aplicando el resultado anterior, tendremos resuelto nuestro problema.

Observemos que con esta notación, la distancia euclídea entre los dos puntos r y s la podemos notar como la raíz cuadrada del producto escalar del vector diferencia de sus coordenadas, por sí mismo:

$$d_{rs} = \sqrt{\sum_{j=1}^{p}(x_{rj} - x_{sj})^2} = \sqrt{(x_r - x_s)^{'}(x_r - x_s)} = \sqrt{x_r^{'}x_r - x_r^{'}x_s - x_s^{'}x_r + x_s^{'}x_s} \quad \forall r,s$$

Con la notación anterior, el cuadrado de la distancia euclídea puede ser puesta como:

$$d_{rs}^2 = b_{rr} + b_{ss} - 2b_{rs} \quad , r = 1,2,...,n \; ; \; s = 1,2,...,n \qquad [1]$$

que constituyen un sistema de ecuaciones lineales con n(n+1)/2 incógnitas independientes, los b_{rs} del triángulo superior derecho y diagonal principal de la matriz B, ya que podemos obtener los b_{rs} del triángulo inferior izquierdo de la matriz B, a partir de su simetría:

$$b_{sr} = b_{rs} \qquad , s = 2,...,n \; ; \; r = 1,...,s-1 \qquad [2]$$

y con n(n-1)/2 ecuaciones linealmente independientes, ya que las ecuaciones correspondientes al triángulo inferior izquierdo de la matriz D son exactamente las mismas que las del triángulo superior derecho,

$$d_{sr}^2 = b_{ss} + b_{rr} - 2b_{rs} = b_{rr} + b_{ss} - 2b_{rs} = d_{rs}^2 \quad , s = 2,...,n \; ; \; r = 1,...,s-1$$

y las que corresponden a la diagonal principal de D no son realmente ecuaciones:

$$0 = d_{rr}^2 = b_{rr} + b_{rr} - 2b_{rr} = 0 \quad , r = 1,2,...,n$$

Así pues, el sistema es indeterminado, por lo que para hacerlo compatible determinado, necesitamos fijar n incógnitas mediante n condiciones iniciales, de alguna forma.

La solución de Torgenson consiste en fijar como origen de coordenadas del nuevo espacio de representación, el centroide de los datos. Esto es,

$$(0,0,...,0)_{1xp} = (\bar{x}_1, \bar{x}_2, ..., \bar{x}_p) = \frac{1}{n}(\sum_{r=1}^{n} x_{r1}, \sum_{r=1}^{n} x_{r2}, ..., \sum_{r=1}^{n} x_{rp}) \Leftrightarrow (0,0,...,0)_{1xp} = (1,1,...,1)_{1xn} X_{nxp}$$

De aquí,

$$(0,0,...,0)_{1xp} X' = (1,1,...,1)_{1xn} XX' = (1,1,...,1)_{1xn} B_{nxn} = (\sum_{r=1}^{n} b_{r1}, \sum_{r=1}^{n} b_{r2}, ..., \sum_{r=1}^{n} b_{rn})$$

de donde finalmente obtenemos las n condiciones iniciales que necesitamos:

$$\sum_{r=1}^{n} b_{rs} = 0 \quad , s = 1,2,...,n \qquad [3]$$

Observemos, que aunque no lo precisamos para encontrar la solución, también se cumple que:

$$\sum_{s=1}^{n} b_{rs} = 0 \quad , r = 1,2,...,n \qquad [4]$$

Así se convierte en sistema compatible determinado y se pueden obtener los coeficientes b_{rs} de la matriz B.

Para ello, a partir del sistema de ecuaciones [1], el procedimiento propuesto por Torgenson es el siguiente:

Si sumamos en s las n ecuaciones expresadas en [1], tenemos en cuenta que las sumas expresadas en [4] valen cero, notamos por T la traza, o suma de los elementos de la diagonal, de la matriz B y adoptamos una notación análoga a la utilizada en las tablas de contingencia, representando la suma de una serie de elementos mediante la sustitución, en el elemento, del índice sumado por un punto (·) , obtenemos:

$$\sum_{s=1}^{n} d_{rs}^2 = \sum_{s=1}^{n} b_{rr} + \sum_{s=1}^{n} b_{ss} - 2\sum_{s=1}^{n} b_{rs} \Leftrightarrow d_{r\cdot}^2 = n \cdot b_{rr} + T \quad , r = 1,2,...,n \qquad [5]$$

Si ahora sumamos en r las n ecuaciones expresadas en [1] y tenemos en cuenta que las sumas expresadas en [3] valen cero, con la misma notación anterior, obtenemos:

$$\sum_{r=1}^{n} d_{rs}^2 = \sum_{r=1}^{n} b_{rr} + \sum_{r=1}^{n} b_{ss} - 2\sum_{r=1}^{n} b_{rs} \Leftrightarrow d_{\cdot s}^2 = T + n \cdot b_{ss} \quad , s = 1,2,...,n \qquad [6]$$

Finalmente, si sumamos simultáneamente en r y s,

$$\sum_{r=1}^{n}\sum_{s=1}^{n} d_{rs}^2 = \sum_{r=1}^{n}\sum_{s=1}^{n} b_{rr} + \sum_{r=1}^{n}\sum_{s=1}^{n} b_{ss} - 2\sum_{r=1}^{n}\sum_{s=1}^{n} b_{rs} \Leftrightarrow d_{\cdot\cdot}^2 = nT + nT = 2nT \qquad [7]$$

Como conocemos las distancias, a partir de [7] calculamos el valor de la traza e B. Así, conocida la traza de B, a partir de [6] (ó [5]) obtenemos uno a uno los elementos de la diagonal principal de B. Y finalmente, conocidos éstos, a partir de las ecuaciones de [1], obtendremos el resto de elementos de B. Osea,

$$T = \frac{d_{\cdot\cdot}^2}{2n} \; ; b_{ss} = \frac{d_{\cdot s}^2 - T}{n} \quad , s = 1,2,...,n \; ; \; b_{rs} = \frac{b_{rr} + b_{ss} - d_{rs}^2}{2} \quad , r = 1,2,...,n \; ; s = 1,2,...,n$$

Con esto, a partir de la matriz de distancias D, habríamos obtenido los elementos de la matriz B; y a partir de ellos, por lo descrito anteriormente, si fuesen (λ_i,e_i), i=1,2,...,n sus autovalores y sus respectivos autovectores (columnas) ortonormalizados asociados, entonces $X = (\sqrt{\lambda_1} \cdot e_1, \sqrt{\lambda_2} \cdot e_2, ..., \sqrt{\lambda_n} \cdot e_n)$ es una posible solución para la matriz de coordenadas buscada para los casos.

Pero la solución no es única. Cualquier rotación ortogonal de la misma, provee una nueva solución:

Si X es solución y R ortogonal $\Rightarrow$ XR' es solución

ya que, al ser R ortogonal, R'=R^{-1}, lo que induce a encontrar nuevas descomposiciones de B como producto de una matriz por su transpuesta:

$$(XR')(XR')' = (XR')(RX') = X(R'R)X' = X(R^{-1}R)X' = XX' = B$$

En consecuencia, lo que nos viene a decir esto es que el procedimiento de Torgenson nos da como solución una nube de puntos, cuyo centro de gravedad es en origen de coordenadas, y que reproduce perfectamente la relación de distancias entre ellos de que se partió. Sin embargo, cualquier rotación de esta nube de punto en el espacio que conserve su centro de gravedad en el origen de coordenadas, igualmente respeta la misma relación de distancias y, por tanto, también es otra posible solución.

Para fijar una única solución, debemos imponer alguna otra condición extra que nos seleccione de entre todas las posibilidades. Por ejemplo, los mapas se suelen presentar de forma única, porque adicionalmente se les añade la información de los puntos cardinales. Si no fuese así, cualquier rotación del mapa obviamente sería una representación válida que respeta las distancias entre las distintas ciudades o sitios.

Escalado Multidimensional No Métrico.

Con la misma finalidad que el Escalado Multidimensional Métrico que acabamos de ver, parte de conocer la relación de proximidad entre casos o elementos para los que se han observado sus características, generalmente en las escalas nominales u ordinales, a partir de matrices de disimilaridades, no necesariamente matrices de distancias como vimos en el caso métrico, para finalmente, como objetivo, tratar de situar aquellos casos sobre un espacio métrico en el que la distancia euclídea entre puntos representantes de los casos originales reproduzcan lo más fielmente posible la correspondiente disimilaridad entre ellos.

El Problema general podría ser introducido de la siguiente forma:

"Conocida las disimilaridades δ_{rs}, r=1,...,n; s=1,...,n, entre cada dos puntos de un conjunto de n, ¿podremos asignar a cada punto unas coordenadas $x_r=(x_{r1},...,x_{rp})'$ en algún cierto espacio, donde sus respectivas distancias euclídeas reproduzcan las disimilaridades originales entre cada dos individuo?"

Para asimilar mejor la diferencia con la técnica anterior, planteemos el siguiente ejemplo:

Si conocemos un triángulo de distancias aproximadas (disimilaridades) entre capitales de provincias españolas confeccionadas como promedio de las opiniones al respecto de un conjunto de personas, ¿podremos identificar unas coordenadas de las mismas de forma que nos permitan dibujar un mapa aproximado de sus posiciones?

Observemos que el proceso de promediado de las opiniones que tienen al aspecto las personas preguntadas, influidas claramente por el mayor o menor conocimiento de la realidad que cada uno tenga y por el kilometraje de las distancias por carretera (no en línea recta) entre cada dos ciudades, que sin duda será la base más general de las respuestas, confieren al resultado subjetivo obtenido un carácter de medida de disimilaridad. Lógicamente, bajo la suposición de un cierto conocimiento general sobre la realidad, las personas tenderán a dar distancias mayores a las capitales más alejadas y viceversa; pero en ningún caso sus respuestas son resultado de una medida real de distancia existente entre capitales (y menos de la distancia euclídea) sino que reflejarán una medida de la proximidad o lejanía (disimilaridad) construida a partir de sus opiniones. El resultado dibujado sobre un plano no corresponderá tanto a la localización de las ciudades sobre el mapa físico, político o de carreteras de España, sino a la idea de cercanía o lejanía que subjetivamente tenemos entre todos de estas ciudades.

Nuestro objetivo será en 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 que reproduzcan esas creencias.

Algoritmo Básico del Escalado Multidimensional No Métrico

Es algoritmo es un proceso iterativo que comienza por asignar, mediante algún método, unas primeras coordenadas a cada punto, y a continuación revisa la adecuación de las mismas comprobando si las distancias euclídeas entre las mismas son compatibles con las disimilaridades iniciales. Si lo fueran, habríamos encontrado una solución. Si no lo fueran,

tendríamos que cambiar las coordenadas al menos de algún punto para hacer que la compatibilidad entre distancias y similaridades sea mejor y volveríamos a reiterar el proceso.

Ha llegado el momento de ver con más detalle qué significan las expresiones "las distancias euclídeas en el nuevo espacio reproduzcan las disimilaridades" o que "las distancias euclídeas en el nuevo espacio sean compatibles con las disimilaridades originales" que intuitivamente estamos empleando en este apartado.

Obviamente, si unas son distancias y otras son disimilaridades sin más, pueden presentar propiedades diferentes que no necesariamente tienen que cumplir las dos. Pensemos por ejemplo en la propiedad triangular o en el hecho de que sólo es nula una distancia cuando se compara un elemento consigo mismo. Estas propiedades serán siempre verificadas por la distancia euclídea, ya que, considerada como medida de proximidad, es una distancia y verifica todas las propiedades que vimos en el capítulo dedicado a las proximidades. Sin embargo, las disimilaridades de que partimos no tienen porqué verificar estas propiedades.

En conclusión, esa "reproducción" o "compatibilidad" de las disimilaridades con las distancias, no significa que las distancias calculadas en el nuevo espacio den como resultado los mismos valores que las disimilaridades originales. Lo que buscamos es que, en el nuevo espacio, se presenten adecuadamente las mismas relaciones de proximidad que nos muestra la matriz de disimilaridades. Es decir, que si las disimilaridades nos dicen que dos individuos son más disimilares (diferentes) que otros dos, las distancias euclídeas entre las coordenadas de los dos primeros debe ser mayor que la que se calcule entre los segundos; y esto, análogamente, debe cumplirse para cualesquiera puntos comparados.

Matemáticamente podemos expresarlo de la siguiente forma:

$$\forall i,j,k,l \qquad \delta_{ij} \leq \delta_{kl} \Leftrightarrow d_{ij} \leq d_{kl}$$

Así pues, el Algoritmo Básico de Escalado Multidimensional No Métrico opera de la siguiente forma:

1ª etapa: En un primer paso estimamos unas coordenadas iniciales $x_r=(x_{r1},...,x_{rp})'$ para cada elemento r, a partir de las disimilaridades $((\delta_{rs}))$.

Diversos procedimientos pueden ser empleados para ello: desde la distribución uniforme entorno al origen de coordenadas de los puntos, hasta la asignación aproximada previa en base a nuestro conocimiento. Sin embargo, la aproximación mecánica más utilizada suele ser la de aplicar la técnica del escalado multidimensional métrico, aún a sabiendas de que nuestra matriz inicial no es de distancias euclídeas, sino de disimilaridades. La solución que aporta lógicamente puede considerarse como una primera aproximación de la localización de los puntos en el espacio.

2ª etapa: Para comprobar la adecuación de la solución anterior, debemos cotejar que las distancias euclídeas entre cada dos puntos, conservan las relaciones de proximidad que expresaban las disimilaridades iniciales, como ya hemos expuesto anteriormente. Para poder realizar esto en la siguiente etapa, calcularemos en ésta esas distancias euclídeas entre cada dos puntos, con las que construiremos la matriz de distancias $((d_{rs}))$.

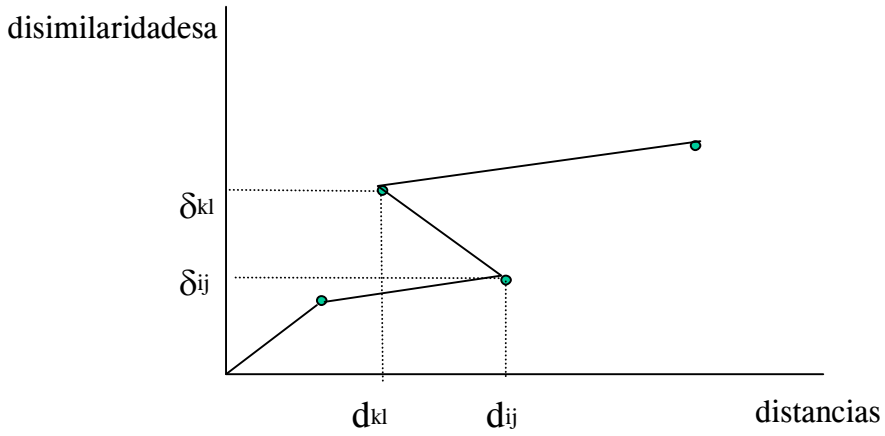
3ª etapa: Ahora ya podemos comprobar la adecuación de la nueva estructura. Y para ello, comprobamos si se cumple la condición pertinente de que sean :

$$\delta_{ij} \leq \delta_{kl} \Leftrightarrow d_{ij} \leq d_{kl} \quad \forall i,j,k,l$$

Si se cumpliera esto para todas las posibles comparaciones, habríamos encontrado una estructura de puntos adecuada a la información inicial, y terminaríamos admitiéndola como solución. Si no fuera así, debemos seguir con el proceso y tratar de mejora la estructura actual. Iríamos por tanto a la siguiente etapa.

Gráficamente existe un método de representar la adecuación de la nueva estructura, conocido como Diagrama de Shepard. Consiste en la representación cartesiana de todos los posibles pares (d_{ij},δ_{ij}) , considerando pues como eje de abcisas, el de las distancias en el nuevo espacio y como eje de ordenadas, el de las disimilaridades originales, en la que se unen los puntos representados, mediante segmentos, ordenadamente de menor a mayor componente de disimilaridad. A continuación presentamos una representación esquemática del Diagrama de Shepard.

Diagrama de Shepard



A medida que nos desplazamos a la derecha en el eje de abcisas (distancias), nos iremos encontrando con distancias que miden la proximidad de puntos cada vez más lejanos entre sí en el nuevo espacio. Análogamente, si nos movemos de abajo a arriba sobre el eje de ordenadas, iremos encontrándonos con disimilaridades que corresponden a puntos cada vez más dispares, según la estructura inicial

Si la estructura de coordenadas encontrada fuera adecuada, cabría esperar que los puntos de esta representación, pares (d_{ij},δ_{ij}), fueran describiendo una trayectoria monótona no decreciente, desplazándose siempre hacia la derecha, sin movimientos de zig-zag en la misma, ya que al crecer los δ_{ij}, también lo deberían hacer los d_{ij}.

Observemos que cuando existe un movimiento de zig-zag, como el representado en la figura al movernos del punto (d_{ij},δ_{ij}) al (d_{kl},δ_{kl}), se contradice la adecuación de la estructura, cumpliéndose que $\delta_{ij} \leq \delta_{kl}$ y sin embargo $d_{ij} > d_{kl}$

A este zigzageo o "nerviosismo" de la poligonal de Shepard, se le conoce con el nombre de

Stress de la Solución. Y son estos movimientos de zig-zag los que nos informan de los pares de puntos que no respetan, en el nuevo espacio, las relaciones de proximidad que nos daba la matriz de disimilaridades inicial, y por tanto los que deben ser corregidos de una forma especial, lo que haremos en la etapa siguiente.

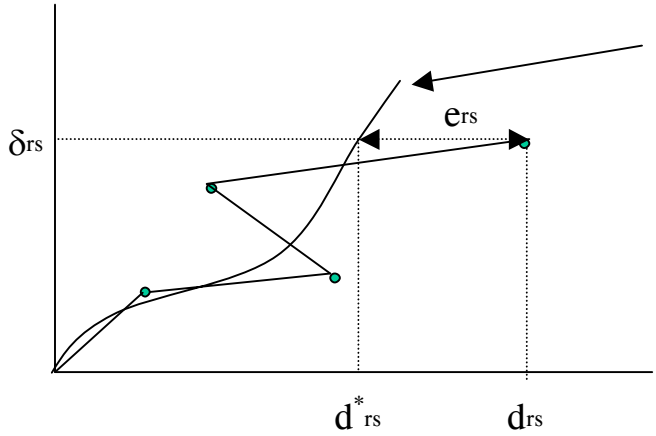
4ª etapa: Como hemos dicho, una estructura de puntos que representase adecuadamente las relaciones de proximidad que se desprende de su matriz de disimilaridades, debería conducir a una representación de la poligonal del diagrama de Shepard en forma de función monótona no decreciente.

Así pues, una estructura de puntos que condujese a una función de tal tipo, sería mejor que la estructura actual, corrigiendo las claras muestras de inadecuación que manifiestan los zig-zag de la curva de Shepard.

La idea es por tanto estimar, como nuevas distancias de una estructura potencialmente mejor, unos nuevos valores, $((d_{rs}^*))$, que puedan deducirse como función, $d_{rs}^*=f(\delta_{rs})$, monótona y no decreciente de las disimilaridades. Estas nuevas cantidades d_{rs}^* que satisfacerían las exigencias de una solución compatible con la estructura de disimilaridades inicialmente observada reciben el nombre de *disparidades.*

Existen muchas funciones monótonas no decrecientes que podrían servir de modelo en esta búsqueda de estas *disparidades*: rectas, polinomios, ... Quizás la más utilizasa de ellas sea la curva logística, por su forma de S muy adaptable a través de sus tres parámetros.Y para estimarla se emplean las técnicas usuales de regresión de las distancias reproducidas por la solución actual como función de las disimilaridades iniciales.

En el siguiente gráfico encontramos descrita la situación del problema de ajuste que tenemos planteado:



Caso de estimación de una curva logística:

$$d_{rs}^* = f(\delta_{rs}) = \frac{1}{\beta_0 + \beta_1 \cdot \beta_2^{\delta_{rs}}}$$

Para nuestro problema, las disimilaridades δ_{rs} hacen el papel de variables independientes o explicativas, mientras que las distancias reproducidas d_{rs} juegan el papel de variables dependientes o explicadas en el modelo $d_{rs}^*=f(\delta_{rs})$, siendo por tanto las disparidades d_{rs}^* los valores teóricos estimados por el modelo.

Los residuos $e_{rs}= d_{rs} - d_{rs}^*$, muestran las desviaciones entre la distancia reproducida por la solución actual y la disparidad ajustada para la consecución de una mejor estructura espacial. Por lo que, a partir de ellos podemos construir una serie de medidas que nos

sinteticen la bondad o adecuación de la estructura actual de puntos, en comparación con la ideal que obtendríamos mediante el ajuste de la curva. Estas medidas reciben el nombre de **Medidas de Stress**, y entre las principales, destacamos las siguientes:

Medidas de Stress

Coef. de Determinación: $$R^2 = 1 - \frac{\sum e_{rs}^2}{\sum (d_{rs} - \bar{d}_{rs})^2}; \quad 0 \le R^2 \le 1$$

Stress 2: $$S_2 = \left(\frac{\sum e_{rs}^2}{\sum (d_{rs} - \bar{d}_{rs})^2} \right)^{1/2} = \sqrt{1 - R^2}; \quad 0 \le S_2 \le 1$$

Stress 1 (de Kruskal): $$S_1 = \left(\frac{\sum e_{rs}^2}{\sum d_{rs}^2} \right)^{1/2} ; \quad 0 \le S_1$$

Stress de Young-Takane-de Leeuw: $$S = \left(\frac{\sum \left(d_{rs}^2 - \bar{d}_{rs}^2 \right)^2}{\sum \bar{d}_{rs}^4} \right)^{1/2} ; \quad 0 \le S \le 1$$

La interpretación de las dos primeras medidas se realiza de la forma habitual ya vista en regresión, sabiendo que R^2 es el coeficiente de determinación. La estructura será tanto más adecuada cuanto más se aproxime el valor de R^2 a 1, o el valor de S_2 a 0.

La medida S_1 fue introducida por Kruskal (1964), quien caracteriza la adecuación de la estructura de puntos, en base a su experiencia, de la siguiente forma según sus valores:

0-perfecta; 0,025-excelente;0,05-buena; 0,1-aceptable; 0,2-pobre

Sin embargo, los valores teóricos que la regresión proporciona para las disparidades d_{rs}^*, ya no tiene que verificar las propiedades de distancia que los d_{rs} sí verificaban; por lo que serán en general simplemente medidas de disimilaridad.

Así pues es esta 4ª etapa, estimamos la función que deberían seguir puntos de la curva de Shepard para que la estructura de puntos en el espacio pudiera ser considerada como adecuada y, a partir de ella, calculamos las medidas de Stress que nos informen del grado de adecuación. Si el grado indica que hemos llegado a una estructura tolerablemente adecuada, terminaríamos el proceso y daríamos como solución la última estructura de puntos obtenida.

Si la medida de Stress indica no haberse alcanzado una estructura espacial mínimamente adecuada, debemos estimar otras coordenadas $x_r=(x_{r1},...,x_{rp})'$ para cada elemento r, a partir de las nuevas disparidades $((d_{rs}^*))$.

Para ello evaluaríamos cuanto cambiaría el Stress de la solución si variáramos un poco cada punto de la solución actual en cada una de las posibles direcciones

$$\frac{dS}{dx_{ij}}$$

Obsérvese que cuando esta cantidad es negativa, una pequeña variación positiva de la coordenada j del punto i disminuye la medida de stress S y mejoraría la solución actual. Si fuera positiva, deberíamos disminuir un poco dicha coordenada para mejorar la solución actual.

Así pues, si la medida de Stress indica no haberse alcanzado una estructura espacial mínimamente adecuada, debemos volver al paso 1 tratando de partir de unas coordenadas mejores obtenidas de acuerdo con la información anterior; para lo que se estiman, en consecuencia, otras coordenadas $x_i=(x_{i1},...,x_{ip})'$ para cada elemento i-ésimo modificando cada coordenada actual de acuerdo a los siguientes incrementos

$$\Delta x_{ij} = -\alpha \frac{dS}{dx_{ij}}$$

siendo α una cantidad positiva, llamada paso de la iteración, que va variando de magnitud (generalmente de mayor a menor) a medida que avanza el número de iteraciones, para proveer una adecuada convergencia del algoritmo. El proceso vuelve a reiterarse desde la etapa 1ª, para luego volver a calcular nuevas distancias reproducidas por la nueva solución en la etapa 2ª, evaluar su adecuación en la etapa 3ª, etc., como ya hemos expuesto anteriormente.

Obviamente, el proceso iterativo deberá seguir mientras se vaya mejorando sustancialmente el Stress de la solución obtenida en cada iteración (mejor adecuación de la estructura de puntos a la matriz inicial de disimilaridades). El proceso deberá parar cuando se haya encontrado una verdadera solución final que es totalmente compatible con las proximidades originalmente observadas (cuando $S_1=0$; $S_2=0$; $R^2=1$), o cuando no es posible mejorar más la solución encontrada aunque no sea totalmente compatible con las proximidades originalmente observadas (cuando $S_1>0$; $S_2>0$; $R^2<1$) siendo la mejoría del Stress que se obtienen en cada nueva iteración prácticamente despreciable.

Dimensionalidad:

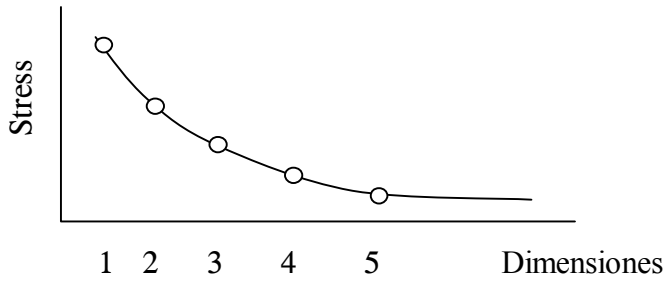
Recordemos que, al emplear la técnica del Escalado multidimensional Métrico de Torgenson para n datos, inicialmente se podían obtener coordenadas sobre un espacio de dimensión n, siendo el número de autovalores de la matriz B estrictamente positivos los que marcaban finalmente el número mínimo de dimensiones necesarias para ser representados adecuadamente.

Ahora bien, en el caso del Escalado Multidimensional No Métrico, la técnica de Torgenson se utiliza para aproximar una solución, por lo que, en este caso, no tenemos definitivamente resuelto el problema de la dimensionalidad, necesitando estudiar cuál es la dimensión del espacio de representación mas adecuada.

Adicionalmente, incluso en el caso del escalado multidimensional métrico, el número de dimensiones teórica del espacio de representación suele ser alto, en cuanto se parte de un alto número de casos, por lo que, también en esta situación, se hace necesario estudiar a partir de qué número de dimensiones podemos prescindir de las demás, por ofrecer aquéllas una representación aceptablemente buena de las disimilaridades iniciales y no aportar las demás prácticamente nada nuevo. Dos son las principales opciones que suelen emplearse:

La primera opción consiste en realizar un Análisis de Componentes Principales sobre la matriz de disimilaridades ((δ_{rs})) o sobre los datos proyectados en un espacio de un número suficientemente alto de dimensiones. El número de componentes que debemos considerar y retener indicará el número de dimensiones que debe emplearse para el Escalado Multidimensional.

La segunda opción ha sido propuesta por Kruskal, quien sugiere realizar el análisis con varias dimensiones y considerar el Stress conseguido en cada caso, para quedarnos con aquella dimensión a partir de la cual no se reduce significativamente éste. Para facilitar el proceso de decisión, sugiere representar gráficamente los resultados, a modo de gráfico de sedimentación ya visto en el Análisis de Componentes Principales, y aprender a apreciar en él dónde comienza la zona de sedimentación (número de dimensiones) a partir de la cual el Stress no desciende prácticamente. Para finalizar, presentamos un gráfico ilustrativo de este tipo, en el que vemos que, a partir de la 5ª dimensión, prácticamente no disminuye nada el Stress.



Todo lo que acabamos de ver en este tema, suele recogerse en los programas informáticos bajo el bajo el nombre genérico de Técnica ALSCAL (composición derivada del término anglosajón con el que se conoce esta técnica, multidimensioAL SCALing).

El modelo de diferencias individuales y otros modelos de Escalado Multidimensional

Lo visto hasta el momento, siempre se ha basado en matrices de distancias o de disimilaridades en las que se comparaban dos a dos todos los elementos del conjunto. Por tanto, esas matrices de partida siempre han sido cuadradas y generalmente simétricas.

Sin embargo las técnicas de escalado multidimensional pueden considerar, con análoga filosofía, matrices cuadradas no simétricas e incluso matrices rectangulares a las que nos conducen ciertas situaciones reales; por ejemplo, la proximidad entre dos ciudades medida por el tiempo de desplazamiento en coche a una determinada hora, conduce a una matriz cuadrada generalmente no simétricas, ya que, dependiendo de las horas, las densidades de tráfico son mayores en un sentido que en otro; o las apreciaciones de distintos jueces acerca de la actividad de distintos individuos como medidas de afinidad o proximidad entre los jueces y los individuos (caso en el que no todos los individuos se comparan entre sí, sino que se comparan los elementos de un grupo con los de otro grupo diferente), nos conduce a matrices generalmente rectangulares que ni siquiera tienen que ser cuadradas.

La técnica del ALSCAL puede abordar todos estos casos anteriores. Pero existen además ciertas generalizaciones de este procedimiento básico, accesibles en los programas estadísticos más comunes para ordenadores. La Técnica que recibe el nombre de INDSCAL (composición derivada del término anglosajón con el que se conoce esta técnica, INDividual differences multidimensioal SCALing), consiste en considerar en los razonamientos expuestos, la distancia euclídea ponderada, en lugar de la distancia euclídea, permitiendo establecer ponderaciones o pesos capaces de establecer diferencias individuales en la consideración de las variables o de los casos, según el enfoque que se adopte. Por su parte la técnica que recibe el nombre de PROXSCAL, desarrollada por la Universidad holandesa de Leiden, contiene como casos particulares al ALSCAL y al INDSCAL y puede considerar simultáneamente el mismo problema para distintos conjuntos de datos con un patrón común de comportamiento, afectado de un factor que determina la segmentación de dichos datos. El procedimiento realiza el escalado dimensional en cada espacio particular de cada conjunto de dato y en el espacio común de todos juntos, para permitir estudiar las diferencias de comportamientos entre ellos.

En cualquier caso, el modelo de diferencias individuales fue desarrollado por Carroll, J.D y Chang, J. en 1970, y se emplea cuando las disimilaridades entre unos mismos individuos se presentan en diversas tablas de estructura similar (p.e. distintos años). En este caso, la distancia ajustada es:

$$d_{ij}^{t} = \sqrt{\sum_{k=1}^{p} w_{kt}(x_{ik} - x_{jk})^2}$$

Y como resultado, se llegan a conocer los siguientes resultados:

– las coordenadas de cada objeto en un espacio común
– los pesos de paso del espacio común a los individuales
– las coordenadas de cada objeto en cada espacio individual