

Análisis de datos. Técnicas aplicadas a datos de proximidad

Tema 4: Diferencias individuales en MDS.

Introducción.

El método de Torgerson fue el primer procedimiento métrico para modelos de MDS unimodales a dos vías y aun hoy día es utilizado para construir la matriz de configuración inicial de los procedimientos iterativos asociados a la mayoría de los demás modelos de MDS.

Dentro de los modelos métricos, bimodales y a tres vías más importantes está sin duda INDSCAL, (Carrol y Chang, 1970). Se trata de conseguir la representación de los objetos o estímulos en un espacio de dimensión baja, pero teniendo en cuenta las diferencias o influencia de las opiniones introducidas por los individuos o sujetos encuestados. Para ello, el modelo de distancia introducido es el Euclídeo ponderado, el cual además, hará que la solución no sea invariante mediante rotaciones aunque si lo será frente a homotecias. Los datos a analizar se supondrán simétricos y de tipo intervalo o razón.

Supondremos ahora que hay n estímulos u objetos que queremos representar en dimensión K , R individuos a los cuales les pediremos su opinión y por tanto, R matrices de disimilaridad entre los n estímulos, cuyos valores denotaremos por δ_{ijr} , siendo (i,j) el par de estímulos juzgado y r el individuo que juzga. El modelo de distancia considerado será,

$$d_{ijr} = \sqrt{\sum_{k=1}^K w_{rk} (x_{ik} - x_{jk})^2},$$

donde w_{rk} será la componente k -ésima del vector w_r que contiene los pesos asociados a las opiniones del individuo r . Esos vectores se organizan en una matriz de ponderación W de dimensión $R \times K$ que habrá que estimar en el modelo. Así, siguiendo el modelo usual mínimo cuadrático de MDS, las disimilaridades son convertidas en distancias estimadas y posteriormente se estiman los pesos y las coordenadas de los puntos mediante un proceso iterativo que minimice el STRESS.

El método INDSCAL

Para formalizar matemáticamente el desarrollo de INDSCAL hay que hacer una distinción en primer lugar, en cuanto al espacio de representación de los estímulos analizados. En todo modelo ponderado pueden considerarse dos espacios de representación; uno el denominado *espacio de representación general* o global, en el

que se representan los estímulos considerando simultáneamente toda la información proveniente de la totalidad de los individuos estudiados y otro el denominado *espacio de percepción privado* que está asociado con cada individuo de forma separada, en el que las coordenadas de los puntos se han obtenido teniendo en cuenta exclusivamente la información proveniente del individuo en cuestión.

Teniendo presente la separación entre esos dos conceptos, consideremos x_{ik} la coordenada en la dimensión k del estímulo i en el espacio general de los estímulos X . Por otra parte, podemos considerar igualmente la coordenada en la dimensión k del estímulo i pero expresada no en el espacio general sino en el espacio particular asociado al sujeto r y que notaremos por x_{ikr} . Debido a que tenemos la relación entre distancias

$$d_{ijr}^{*2} = \sum_{k=1}^K w_{rk} (x_{ik} - x_{jk})^2 = \sum_{k=1}^K (w_{rk}^{1/2} x_{ik} - w_{rk}^{1/2} x_{jk})^2,$$

y denotando por $x_{ikr} = w_{rk}^{1/2} x_{ik}$, se tiene que,

$$d_{ijr}^{*2} = \sum_{k=1}^K (x_{ikr} - x_{jkr})^2.$$

Así, la relación entre x_{ik} y x_{ikr} viene dada por la expresión $x_{ikr} = w_{rk}^{1/2} x_{ik}$, que ofrece una interpretación alternativa de la métrica euclídea ponderada, de forma que el modelo INDSCAL puede expresarse en términos de una clase particular de transformaciones del espacio conjunto general, además del cálculo de la métrica euclídea clásica. La clase de las transformaciones puede describirse algebraicamente como lineales con matriz de transformación diagonal.

La relación que anteriormente expresábamos entre las coordenadas consideradas en el espacio general o en el particular de un estímulo, para el sujeto r , puede expresarse matricialmente de la forma $X_r = W_r^{1/2} X$, donde X_r es la configuración asociada al individuo r y X es la configuración global. Además, W_r es la matriz diagonal de los pesos asignados por el individuo r a las dimensiones de representación que serán obligatoriamente las mismas para cada individuo.

Dadas las distancias entre los estímulos para cada individuo r , las cuales pueden obtenerse por ejemplo por el procedimiento descrito en el apartado anterior, podemos obtener los productos escalares asociados mediante la relación establecida por el método clásico. Dados pues los productos escalares b_{ijr} podemos expresar éstos en relación con las coordenadas de los estímulos de la forma,

$$B_r = X_r X_r^t = X W_r^{1/2} (W_r^{1/2})^t X^t = X W_r X^t$$

y esa relación puede expresarse en las coordenadas de la forma

$$b_{ijr} = \sum_{k=1}^M w_{rk} x_{ik} x_{jk}.$$

Se trata por tanto de resolver la ecuación

$$B_r = XW_rX^t$$

El método INDSCAL trabaja bajo una hipótesis simple: Sea y_{ijr} un dato de una matriz tridimensional. Cada uno de esos valores es una función de un conjunto de parámetros de la forma

$$y_{ijr} = \sum_{k=1}^M a_{rk} b_{ik} c_{jk}.$$

En primer lugar, si comparamos esa expresión con la del producto escalar, se trata de resolver el sistema trilineal, donde $a_{rk} = w_{rk}$, $b_{ik} = x_{ik}$ y $c_{jk} = x_{jk}$. Para resolver el problema se puede proceder de la siguiente forma: En primer lugar, se sitúa cada matriz B_r en un vector fila con n^2 elementos, y situamos todos esos vectores por filas en una matriz B , $R \times n^2$. De igual forma, almacenamos las diagonales de las R matrices de pesos W_r en una matriz W , $R \times K$. Por último, se construye una matriz V de la siguiente forma; la primera columna se construye para $i=1$ y para $j=1$, calculando los productos $x_{1k}x_{1k}$, $\forall k=1, \dots, K$; la segunda columna de V se calcula de forma análoga pero $i=1$, $j=2$. Así se procede hasta completar los $K \times n^2$ valores de V .

Una vez obtenidas las matrices anteriores, podemos escribir la relación conjunta, para todos los individuos simultáneamente, de los productos escalares de la forma

$$B = WV.$$

Así, dada una configuración X , se tendrán B y V , con lo que la matriz W puede estimarse mediante regresión mínimo cuadrática de la forma:

$$W = BV^t(VV^t)^{-1}$$

Con W se obtienen las W_r y de ahí se obtienen los productos W_rX^t para cada r y se sitúan las $(RK \times n)$ matrices en una matriz H , $(K \times nR)$. Se crea igualmente una matriz \dot{B} con las R matrices B_r de forma idéntica a la anterior. Con $B_r = XW_rX^t$ se encuentra una nueva matriz $X^{(1)}$, $(n \times K)$, de la forma

$$X^{(1)} = \dot{B}H^t(HH^t)^{-1}.$$

Con esta nueva matriz $X^{(1)}$ se determina la matriz $W^{(1)}$ y después otra matriz $X^{(2)}$ con la que comenzaremos el proceso nuevamente hasta obtener convergencia y una solución final de X y W .

Diferencias individuales con SMACOF

Dado un conjunto de disimilaridades, δ_{ijr} entre n objetos ($i,j=1,2,\dots,n$) para R individuos ($k=1,2,\dots,R$). SMACOF también permite determinar R configuraciones X_r

de orden $n \times K$, de tal manera que las distancias Euclídeas $d_{ijr}(X_r)$ medidas entre los puntos de las representaciones obtenidas sean lo más parecidas posible a las disimilaridades δ_{ijr} de inicio $\forall i,j = 1 \dots n; r=1 \dots R$. Formalmente, el problema que resuelve SMACOF en el caso más general es la minimización de la función de pérdidas de mínimos cuadrados STRESS, de Kruskal (1964),

$$f(X_1, X_2, \dots, X_R, \Phi) = \frac{1}{R} \sum_{r=1}^R \sum_{i < j}^n w_{ijr} \left[\Phi(\delta_{ijk} - d_{ij}(X_k)) \right]^2,$$

siendo

- $d_{ij}(X_r)^2 = (x_{ir} - x_{jr})^t (x_{ir} - x_{jr})$, con x_{ir} y x_{jr} , vectores columna de orden $K \times 1$, definidos a partir de las filas i y j de la matriz de X_r .
- w_{ijr} , una ponderación no negativa dada, en la mayoría de los casos diferente para cada terna (i,j,r) . Estas ponderaciones desempeñan dos objetivos. En primer lugar facilitan la manipulación de datos faltantes (en el caso de faltar alguna medida de disimilaridad, la ponderación en tal caso vale 0) y, en segundo lugar, permiten dar más énfasis en el análisis a unas disimilaridades que a otras.

Las características más interesantes del algoritmo SMACOF pueden resumirse en los siguientes aspectos.

Las matrices desconocidas X_r , pueden estar sujetas o no a restricciones. Tales restricciones en SMACOF son del tipo $X_r = ZA_r$, con Z una matriz desconocida de orden $Rn \times K$ representando el espacio común y A_r matrices también desconocidas, de orden $R \times R$, conteniendo los pesos. Tanto la matriz Z como las matrices A_r , nuevamente pueden tener restricciones adicionales. Las distintas restricciones sobre las matrices A_r , dan lugar a los diferentes modelos de MDS.

- Modelo Euclídeo generalizado: IDIOSCAL. A_r no singular.
- Modelo de rango reducido: A_r es una matriz de rango $p < K$.
- Modelo Euclídeo ponderado: INDSCAL. A_r , es una matriz diagonal.
- Modelo identidad: A_r , es la matriz identidad.

Las restricciones sobre la matriz Z pueden ser de dos tipos:

- Algunas o todas las coordenadas del espacio común están fijas por el usuario.
- El usuario restringe el espacio común a una combinación lineal de variables externas. En este caso se considera, $ZA_r = QBA_r$, con Q , de orden $n \times s$, conteniendo las variables externas y B , de orden $s \times K$, con los pesos de regresión. En este proceso de ajuste, SMACOF también determina las

transformaciones óptimas para las variables externas, considerando diferentes transformaciones en función del tipo de variable con el que se esté trabajando..

El algoritmo general de SMACOF.

La versión implementada en SPSS de SMACOF que se denomina PROXSCAL, permite elegir entre cinco transformaciones implementadas: transformaciones monótonas, lineales, polinómicas, splines o cambios de escala. Se pueden considerar también estas transformaciones para cada fuente por separado (condicionamiento matricial) o a todos los datos globalmente (nivel de medida incondicional). Todas las transformaciones se determinan mediante procedimientos de mínimos cuadrados alternantes.

Comparado con otros programas de MDS, PROXSCAL presenta muchas ventajas. El método del subgradiente usado en PROXSCAL garantiza la convergencia del proceso, frente al método del gradiente empleado por otros métodos como MDSCAL, KYST, SSA y MINISSA.

Otros algoritmos, como ALSCAL, que también garantizan convergencia monótona en la mayoría de las ocasiones, presentan el inconveniente de que en la primera etapa del análisis transforman las disimilaridades en distancias al cuadrado o productos escalares. Con PROXSCAL no se produce ninguna transformación a priori de los datos y las disimilaridades se aproximan directamente a las distancias.

Supongamos que X_r^l denota una matriz arbitraria satisfaciendo las restricciones del modelo. Entonces, $X_r^{l*} = V_r^{-1} B(X_r^l) X_r^l$, denota la actualización sin restricciones dada la matriz X_r^l , mientras que X_r^{l+1} , denota una nueva solución para X_r , mejor que la anterior, satisfaciendo además las mismas restricciones que X_r^l , siendo V_r , una matriz, función de las ponderaciones de los datos y $B(X_r^l)$, una función matricial dependiente de las distancias entre los puntos en la configuración anterior, $d_{ij}(X_r^l)$.

Una actualización restringida se puede obtener minimizando la siguiente función de pérdidas,

$$h(X_1, X_2, \dots, X_R) = \frac{1}{R} \sum_{r=1}^R \text{tr} \left[(X_r^{l*} - X_r)^t V_r (X_r^{l*} - X_r) \right].$$

La minimización de esta función se conoce como un problema de proyección métrica, dado que engloba la proyección de las matrices X_r^l , en el espacio de las soluciones restringidas mediante la métrica V_r . La minimización del STRESS se lleva a cabo a través del siguiente algoritmo:

- 1 Para $l=0$ se consideran configuraciones iniciales X_r^o , $\forall k=1,2,\dots,R$ y se evalúa la ecuación,

$$f(X_1^l, X_2^l, \dots, X_R^l, \Phi) = \frac{1}{R} \sum_{r=1}^R \sum_{i < j}^n w_{ijr} \left[\Phi(\delta_{ijr}) - d_{ij}(X_r^l) \right]^2.$$

2 Para cada r se calcula la transformada de Guttman,

$$X_r = V_r^- B(X_r^l) X_r^l.$$

3 Se resuelve la ecuación del paso 1 para el problema de proyección métrica.

4 Se evalúa la función de pérdida,

$$f(X_1^{l+1}, X_2^{l+1}, \dots, X_R^{l+1}, \Phi) = \frac{1}{R} \sum_{r=1}^R \sum_{i < j}^n w_{ijr} \left[\Phi(\delta_{ijr}) - d_{ij}(X_r^{l+1}) \right]^2$$

5 Se considera $l=l+1$. Si la diferencia entre los valores de la función de pérdidas en las dos estimaciones consecutivas es mayor que algún criterio determinado de antemano, se vuelve al paso 2, en caso contrario, el algoritmo se detiene.

Aunque no se conoce en profundidad como afecta el problema de los mínimos locales a este procedimiento, se sabe que se producen y que en muchas ocasiones estos vienen motivados por la configuración inicial que se considere. Como consecuencia, PROXSCAL dispone de un procedimiento especial para determinar la configuración inicial y comenzar el algoritmo.

Aplicando el método SIMPLEX, se determina una solución en dimensión alta, sin considerar transformaciones en los datos, excepto para el caso de datos ordinales, en los que se consideraría el rango de los datos, y se proyecta dicha configuración sobre un subespacio de dimensión predeterminada para la configuración final, considerando todas las restricciones exigidas sobre las coordenadas.

Para obtener dicha proyección, se minimiza

$$g(X_1, \dots, X_R) = \frac{1}{R} \sum_{r=1}^R \text{tr}((X_r^+ - X_r H_r^t)^t (X_r^+ - X_r H_r^t)),$$

con X_1, \dots, X_R , satisfaciendo las restricciones, y siendo H_r , la matriz de proyección ortonormal y X_1^+, \dots, X_R^+ , las transformadas de Guttman de las estimaciones anteriores.

Ejemplo 4.1: Consideremos el conjunto de datos descrito en Bro (1998), que consisten en la opinión de 8 catadores respecto a 11 atributos, emitidas sobre 10 muestras de pan formadas con 5 tipos de pan (horneados con diferente contenido de sal), donde cada pan es presentado dos veces para ser clasificado. Así el pan3.2 es el pan del tipo 3 en la réplica 2. Los datos analizados son las 8 matrices de disimilaridad simétricas obtenidas entre las 10 muestras.

1. Veamos el análisis usando matrices de pesos diagonales.

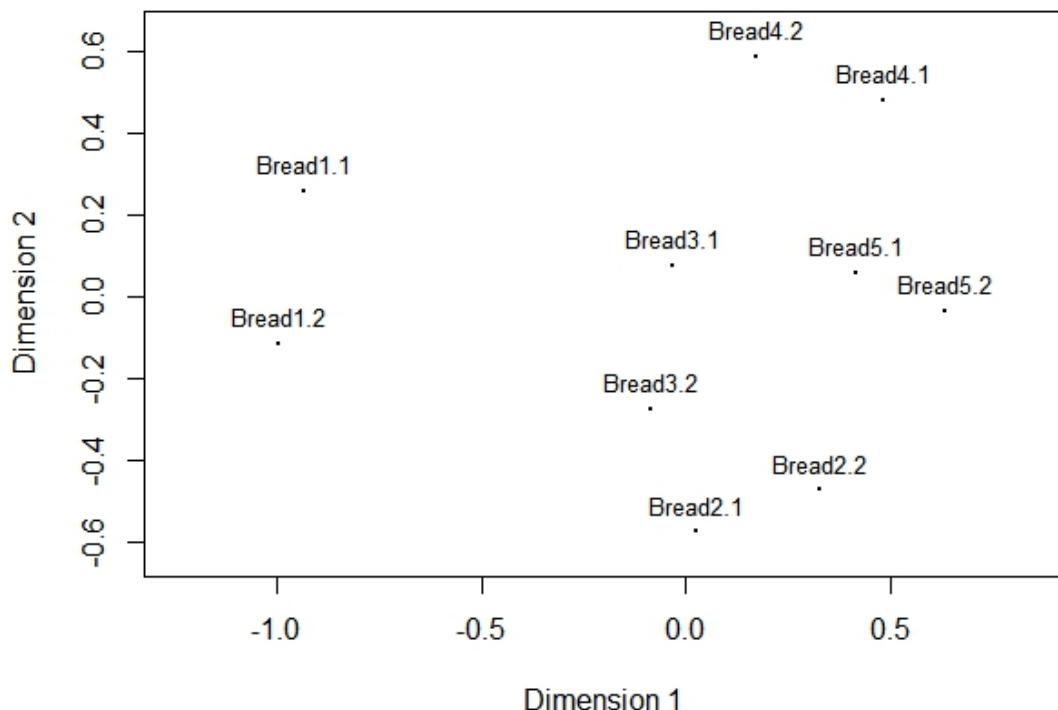
#1) Modelo indscal. Matrices de pesos diagonales.

```

res=smacofIndDiff(bread,constraint="indscal",ndim=2)
res
summary(res)
plot(res, main="Modelo Indscal",asp=1)
res$cweight
res$conf
res$gspace
dev.new()
plot(res,plot.type="resplot",main="Residuos sin restricciones",asp=1)

```

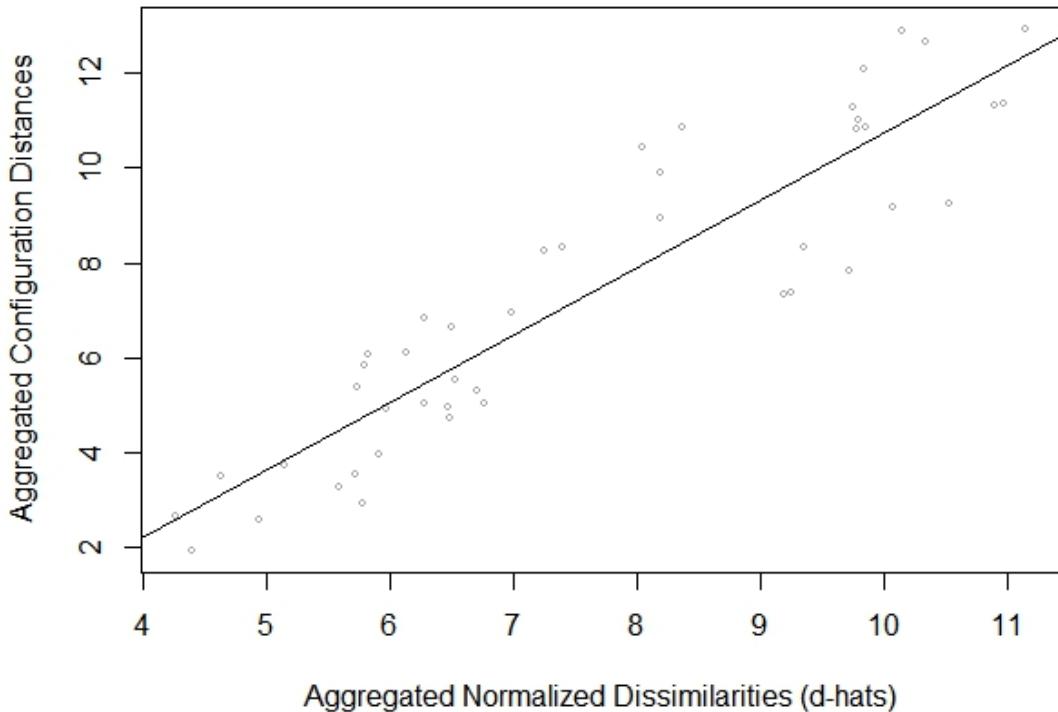
Modelo Indscal



La solución sin restricciones refleja claramente los pares de réplicas en los tipos de pan, donde pueden apreciarse cómo los puntos para cada tipo de pan están próximos entre sí.

Veamos a continuación la gráfica de residuos para esta solución.

Residuos Indscal



La gráfica de residuos muestra que pequeñas distancias son sistemáticamente sobreestimadas, mientras que las grandes distancias son infraestimadas (creciente en dispersión).

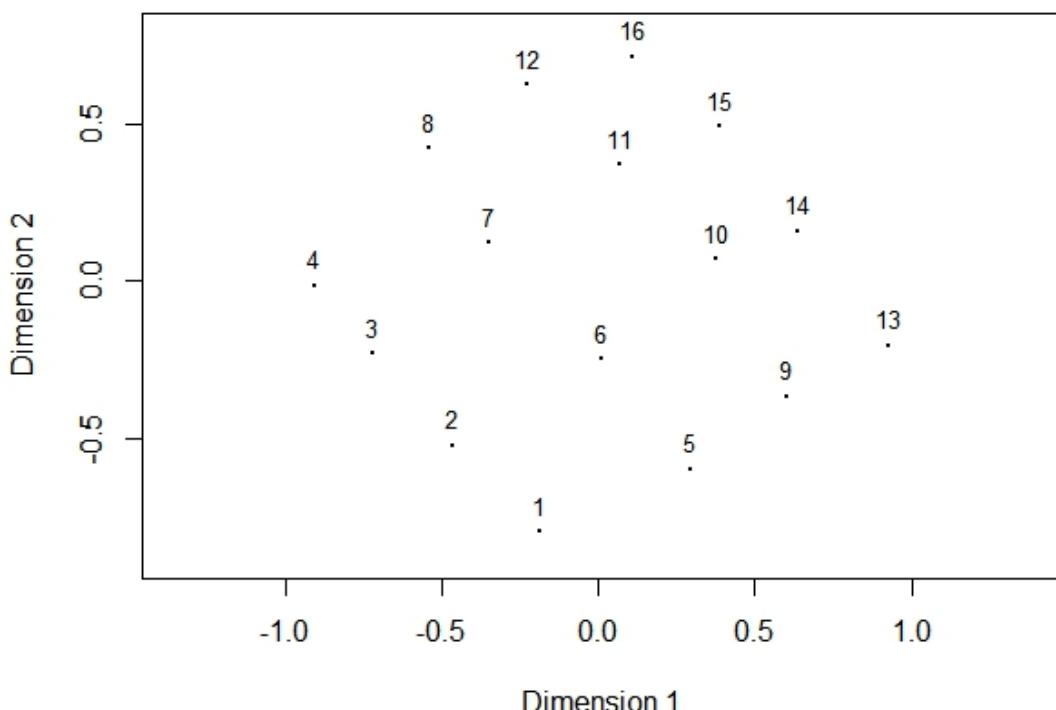
Ejemplo 4.2: Consideremos el conjunto de datos “perception” de Smacof. Se asignaron 42 sujetos a dos grupos de 21 personas. Se presentaron 120 pares de rectángulos como estímulo. Para el primer grupo (ancho-alto; WH), los rectángulos se construyeron de acuerdo con un diseño mientras que para el segundo grupo (tamaño-forma; SS), los rectángulos se construyeron de acuerdo con otro diseño. Todos los sujetos tuvieron que juzgar la similitud de los rectángulos en una escala de 0 a 9. Veamos el análisis usando matrices diagonales.

1. Veamos el análisis de los datos sin restricciones.
2. Veamos el análisis de los datos para el modelo INDSCAL

#1) Restricción identidad.

```
res.identidad=smacofIndDiff(perception,constraint="identity",ndim=2)
res.identidad
summary(res.identidad)
plot(res.identidad, main="Modelo identidad",asp=1)
res.identidad$cweight
res.identidad$conf
res.identidad$gspace
plot(res.identidad,plot.type="resplot",main="Residuos. Restricción
+ identidad",asp=1)
```

Modelo Identidad



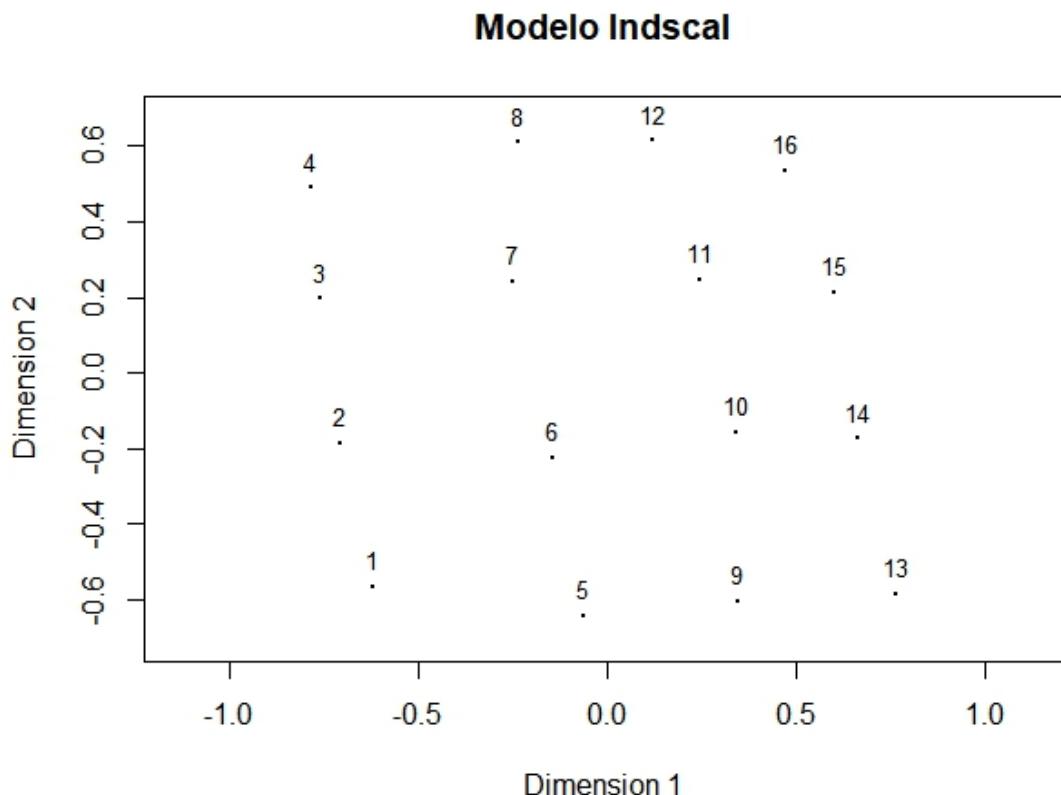
De los resultados se aprecia que ambas configuraciones son idénticas a la configuración global como cabía esperar y los pesos estimados son la matriz de identidad. El gráfico presenta la disposición en un espacio no ponderado de las opiniones de los individuos.

Veamos ahora la solución del modelo INDSCAL.

```
#2) Modelo indscal. Matrices de pesos diagonales.  
res=smacofIndDiff(perception,constraint="indscal",ndim=2)  
res  
summary(res)  
plot(res, main="Modelo Indscal",asp=1)  
res$cweight  
res$conf  
res$gspace  
dev.new()  
plot(res,plot.type="resplot",main="Residuos sin restricciones",asp=1)
```

De los resultados se aprecia que ambas configuraciones son idénticas a la configuración global como cabía esperar y los pesos estimados son la matriz de identidad. El gráfico presenta la disposición en un espacio no ponderado de las opiniones de los individuos.

Veamos ahora la solución del modelo INDSCAL.



La configuración claramente difiere de la anterior por la influencia de los pesos estimados en los ejes. Puede apreciarse ahora que las configuraciones privadas son distintas y las matrices de pesos no son la identidad.

Ejercicio 4.1. La siguiente Tabla contiene las distancias entre los 45 pares de colores formados con 10 colores, emitidas por 11 individuos normales (Helm, 1959).

Leer los datos con SPSS.

1. Realizar el análisis de los datos usando PROXSCAL SPSS para el modelo identidad.
 2. Realizar el análisis de los datos con PROXSCAL de SPSS para el modelo de diferencias individuales.
 3. Compara los resultados anteriores.

Tabla 4. 1. Distancias estimadas para 11 individuos normales sobre los 45 pares de colores normales de Helm (1959).

11.4	11.2	5.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0
13.3	13.5	8.2	4.0	0.0	0.0	0.0	0.0	0.0	0.0
12.0	12.9	9.6	5.8	3.8	0.0	0.0	0.0	0.0	0.0
12.3	12.0	12.7	6.8	5.4	3.6	0.0	0.0	0.0	0.0
10.6	11.5	13.7	9.3	7.9	5.6	4.2	0.0	0.0	0.0
4.9	8.2	13.4	10.5	9.9	9.0	8.2	5.1	0.0	0.0
3.5	6.3	11.7	12.2	13.2	10.4	9.8	6.8	3.8	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10.5	4.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
13.4	12.2	4.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0
14.0	14.8	8.3	3.5	0.0	0.0	0.0	0.0	0.0	0.0
13.2	14.6	10.7	4.7	3.6	0.0	0.0	0.0	0.0	0.0
11.7	14.1	12.8	8.8	6.9	4.1	0.0	0.0	0.0	0.0
10.2	13.4	14.1	11.0	9.4	6.9	4.1	0.0	0.0	0.0
6.4	9.7	12.9	11.8	12.4	10.6	10.0	4.1	0.0	0.0
3.5	7.9	10.9	11.7	13.7	12.2	11.1	6.9	3.4	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10.8	7.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9.9	8.9	6.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
11.1	10.7	8.7	3.9	0.0	0.0	0.0	0.0	0.0	0.0
10.3	10.8	9.6	6.8	5.0	0.0	0.0	0.0	0.0	0.0
8.8	10.6	10.1	9.4	8.3	4.3	0.0	0.0	0.0	0.0
7.6	10.4	10.8	9.7	9.0	7.3	4.9	0.0	0.0	0.0
5.8	9.0	11.7	10.4	10.9	9.0	7.2	4.7	0.0	0.0
3.0	7.5	9.4	9.7	9.6	8.8	7.6	5.6	3.5	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9.1	4.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10.2	7.9	5.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0
12.1	10.4	8.3	3.9	0.0	0.0	0.0	0.0	0.0	0.0
12.5	11.2	10.2	6.5	4.6	0.0	0.0	0.0	0.0	0.0
9.7	12.6	11.3	8.7	7.8	6.3	0.0	0.0	0.0	0.0
9.8	11.4	12.2	10.3	9.9	9.6	4.8	0.0	0.0	0.0
8.3	11.3	11.9	10.7	11.2	10.6	6.8	4.6	0.0	0.0
6.7	10.4	10.7	12.6	11.6	11.6	9.1	7.4	5.2	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9.4	7.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9.5	9.5	7.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9.5	9.5	8.9	3.8	0.0	0.0	0.0	0.0	0.0	0.0
9.8	9.9	9.8	5.3	4.8	0.0	0.0	0.0	0.0	0.0
8.7	10.6	10.5	7.3	6.2	4.7	0.0	0.0	0.0	0.0
6.7	10.6	10.7	7.6	8.2	6.7	4.5	0.0	0.0	0.0
4.9	8.5	9.7	9.2	9.1	8.8	7.2	4.0	0.0	0.0
4.1	7.9	10.2	10.1	9.7	9.9	6.8	5.3	3.4	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10.8	5.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9.7	8.2	4.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10.1	9.4	6.7	3.7	0.0	0.0	0.0	0.0	0.0	0.0

10.3	10.1	9.8	6.6	4.0	0.0	0.0	0.0	0.0	0.0
9.7	10.5	11.3	8.7	7.5	5.4	0.0	0.0	0.0	0.0
9.0	10.8	11.9	10.6	9.9	9.3	5.6	0.0	0.0	0.0
6.6	11.2	11.5	10.0	10.9	9.9	8.2	5.3	0.0	0.0
4.6	10.5	10.2	7.7	10.6	9.7	9.7	6.3	3.4	0.0

ALSCAL

ALSCAL (Alternating Least Squared sCALing) es un algoritmo desarrollado por Takane, Young & de Leew (1977) que mediante mínimos cuadrados alternantes permite desarrollar *Multidimensional Scaling* y *Multidimensional Unfolding*.

ALSCAL es un modelo que permite analizar datos mediante un modelo Euclídeo simple o ponderado que pueden ser: (a) nominales, ordinales de intervalo o razón; (b) poseer observaciones faltantes; (c) simétricos o asimétricos; (d) condicionales o incondicionales; (e) con o sin replicaciones y (f) discretos o continuos. Así pues puede considerarse como uno de los modelos más completos. ALSCAL ha sido de los primeros modelos de MDS implementados en SAS y en SPSS, de ahí que su utilización se haya difundido pasando a ser el modelo más empleado.

Al igual que la mayoría de los modelos de MDS, ALSCAL es un método iterativo. Cada iteración en ALSCAL consiste en dos fases principales, una que transforma los datos, *fase de representación óptima* y otra que estima las coordenadas de los estímulos y los pesos de los sujetos (cuando sea el caso), *fase de estimación del modelo*. Ambas fases van alternándose hasta que se obtiene la convergencia global del proceso. Puesto que en ALSCAL, cada fase es un procedimiento mínimo cuadrático, es por lo que recibe ese nombre.

La fase de *representación óptima* consiste en la transformación de las observaciones según antes exponíamos para que sean preservadas las propiedades de los datos originales. Por otra parte, la fase de *estimación del modelo* es más complicada. La dificultad se debe al hecho de que esta fase conlleva dos bloques; la *estimación de las coordenadas de los estímulos* y la *estimación de los pesos de los individuos*, cuando sea el caso. En el bloque de *estimación de la configuración*, cada coordenada es estimada sola por mínimos cuadrados mientras se mantienen fijos los valores del resto. Por ello, puesto que la estimación de cada coordenada depende de los valores de todas las demás, resulta necesario iterar para su estimación hasta que se estabiliza la localización de todas las coordenadas del punto en cuestión en \mathbb{R}^k . Cuando esto sucede lo que en realidad ha ocurrido es que se ha movido el punto en cuestión a la mejor localización posible, manteniéndose fijos todos los demás puntos. El proceso continua sucesivamente para todos los puntos.

Una vez estimados los puntos, se estiman los pesos. El criterio de optimización en que se basa ALSCAL es el denominado SSTRESS, que es un criterio mínimo cuadrático dado principalmente en función de las distancias al cuadrado de la forma,

$$SS(X, W, \widehat{D}^2) = \sum_{r=1}^R \sum_{i=1}^n \sum_{j=1}^{i-1} (\widehat{d}_{ijr}^2 - d_{ijr}^2)^2,$$

salvo un factor de normalización. Este criterio es semejante en su forma al criterio STRESS de Kruskal, salvo que utiliza las disparidades y las distancias al cuadrado. De ahí su nombre *Squared STRESS*. No obstante el STRESS y el SSTRESS no son equivalentes, por lo que es necesario tener presente que \widehat{d}_{ijr}^2 es el estimador mínimo cuadrático de d_{ijr}^2 y no el cuadrado del estimador mínimo cuadrático de d_{ijr} .

Ejercicio 4.2: Si consideramos $e = \delta_{ij} - d_{ij}$, estudiar la relación entre $\delta_{ij} - d_{ij}$ y $\delta_{ij}^2 - d_{ij}^2$.

El algoritmo ALSCAL

El algoritmo ALSCAL, al igual que la mayoría de los métodos de mínimos cuadrados alternantes, subdivide los parámetros en diversos bloques para su estimación mediante un proceso iterativo. El algoritmo ALSCAL está formado por dos fases principales y dos fases secundarias. La primera fase principal es denominada *fase de representación óptima* y conlleva la obtención mediante mínimos cuadrados de las disparidades, \widehat{D} bajo la hipótesis de que la configuración, X y los pesos, W , son constantes, minimizando el SSTRESS supuesto que X y W son constantes. Matemáticamente formularemos este problema como el cálculo del

$$\min_{\{\widehat{D}\}} [SS(\widehat{D} \mid X, W)].$$

La segunda fase principal se denomina *fase de estimación del modelo* y consiste en dos subfases, la primera resuelve el problema,

$$\min_{\{W\}} [SS(W \mid X, \widehat{D})]$$

y la segunda el problema

$$\min_{\{X\}} [SS(X \mid W, \widehat{D})].$$

Las dos fases secundarias se denominan *fase de inicialización* y *fase de terminación*. Esquemáticamente el proceso sigue los siguientes pasos:

1. *Fase de inicialización.* Calcula los valores iniciales de X y de W partiendo directamente de las disimilaridades observadas, Δ mediante un procedimiento descrito en Young et al. (1978).
2. *Fase de representación óptima.* Esta fase está compuesta a su vez por tres etapas:

- a. Calcula las distancias euclídeas ponderadas al cuadrado D^2 usando X y W .
 - b. Obtiene el estimador mínimo cuadráticos de las disparidades \hat{D} a partir de las distancias D , de las disimilaridades Δ , y de las restricciones de medida asociadas.
 - c. Normaliza las disparidades adecuadamente.
3. *Fase de terminación.* Se determina si el proceso debe terminar, controlando los valores del SSTRESS. Si finaliza el proceso se exponen los resultados y si no, se pasa al siguiente punto.
 4. *Fase de estimación del modelo.* En esta fase se estiman los parámetros del modelo de la siguiente forma:
 - a. Se calcula un nuevo estimador mínimo cuadrático de los pesos W a partir de la X anterior y de las nuevas \hat{D} obtenidas y normalizadas.
 - b. Se imponen restricciones de no negatividad sobre W si fuese necesario.
 - c. Se calcula el estimador mínimo cuadrático de X a partir de los pesos calculados y usando las disparidades obtenidas en la etapa anterior, mediante un procedimiento modificado de Newton-Raphson.
 - d. Se vuelve a la fase de representación óptima.

Finalmente, conviene tener presente que el algoritmo ALSCAL expuesto anteriormente engloba diferentes situaciones sin más que considerar el modelo euclídeo ponderado para datos de disimilaridad simétricos y sin datos faltantes, al ser fácilmente descritas el resto de situaciones teniendo presente que:

- *El modelo euclídeo sin ponderar* puede ajustarse sin más que obviar la fase de estimación de W , fijando $W = I$.
- *Los datos asimétricos* pueden tratarse sin más que considerar en los sumandos todos los índices y no solo la matriz triangular inferior, cambiando además los rangos de las matrices.
- *La presencia de datos faltantes* puede manejarse excluyendo todos los datos faltantes del criterio de optimización y estimando los datos faltantes al final del proceso a partir de los valores obtenidos de los parámetros.

Ejemplo 4.2. La siguiente tabla recoge las distancias aéreas entre las siguientes 12 ciudades de USA: ATLANTA BOSTON CINCINNATI COLUMBUS DALLAS INDNPLIS LITTRICK LOSANGEL MEMPHIS STLOUIS SPOKANE y TAMPA.

Tabla 4.2. Distancias aéreas entre 12 ciudades americanas.

A	B	C	D	E	F	G	H	I	J	K	L
0											
1 068	0										
461	867	0									
549	769	107	0								
805	1 819	943	1 050	0							
508	941	108	172	882	0						
505	1 494	618	725	325	562	0					
2 197	3 052	2 186	2 245	1 403	2 080	1 701	0				
366	1 355	502	586	464	436	137	1 831	0			
558	1 178	338	409	645	234	353	1 848	294	0		
2 467	2 747	2 067	2 131	1 891	1 959	1 988	1 227	2 042	1 820	0	
467	1 379	928	985	1 077	975	912	2 480	779	1 016	2 821	0

Veamos un análisis MDS mediante el módulo ALSCAL de SPSS en dos y tres dimensiones para un modelo métrico.

ALSCAL

```
VARIABLES= ATLANTA BOSTON CINCINNATI COLUMBUS DALLAS INDNPLIS LITTRICK
LOSANGEL MEMPHIS STLOUIS SPOKANE TAMPA
/SHAPE=SYMMETRIC
/LEVEL=RATIO
/CONDITION=MATRIX
/MODEL=EUCLID
/CRITERIA=CONVERGE(.00001) STRESSMIN(.0005) ITER(100) CUTOFF(0) DIMENS(2,3)
/PLOT=DEFAULT
/PRINT=HEADER .
```

Los resultados obtenidos son:

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 3
Minimum Dimensionality 2
Negative Weights Not Permitted

Output Options-

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

Algorithmic Options-

Maximum Iterations 100
Convergence Criterion00001
Minimum S-stress00050
Missing Data Estimated by . . . Ulbounds

▼

Iteration history for the 3 dimensional solution (in squared distances)

Young's S-stress formula 1 is used.

Iteration	S-stress	Improvement
1	.03693	
2	.01918	.01776
3	.01863	.00055
4	.01863	.00000

Iterations stopped because
S-stress improvement is less than .000010

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 = .01894 RSQ = .99924

▼

Configuration derived in 3 dimensions

Stimulus Coordinates

Dimension

Stimulus Number	Stimulus Name	1	2	3
--------------------	------------------	---	---	---

1	ATLANTA	.9705	-.5054	.1103
2	BOSTON	2.0510	1.2685	-.0555
3	CINCNATI	.6667	.4112	-.0599
4	COLUMBUS	.7822	.5262	-.0260
5	DALLAS	-.5142	-.8813	.1272
6	INDNPLIS	.4678	.4128	-.0614
7	LITTROCK	-.0116	-.5122	.2486
8	LOSANGEL	-3.1367	-.8390	.0607
9	MEMPHIS	.2154	-.4061	.0466
10	STLOUIS	.1046	.1948	-.0082
11	SPOKANE	-3.0241	1.5006	-.2188
12	TAMPA	1.4285	-1.1701	-.1636

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

Young's S-stress formula 1 is used.

Iteration	S-stress	Improvement
1	.02057	
2	.01899	.00158
3	.01893	.00006
4	.01893	.00000

Iterations stopped because
S-stress improvement is less than .000010

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 = .01893 RSQ = .99909

▼

Configuration derived in 2 dimensions

Stimulus Coordinates

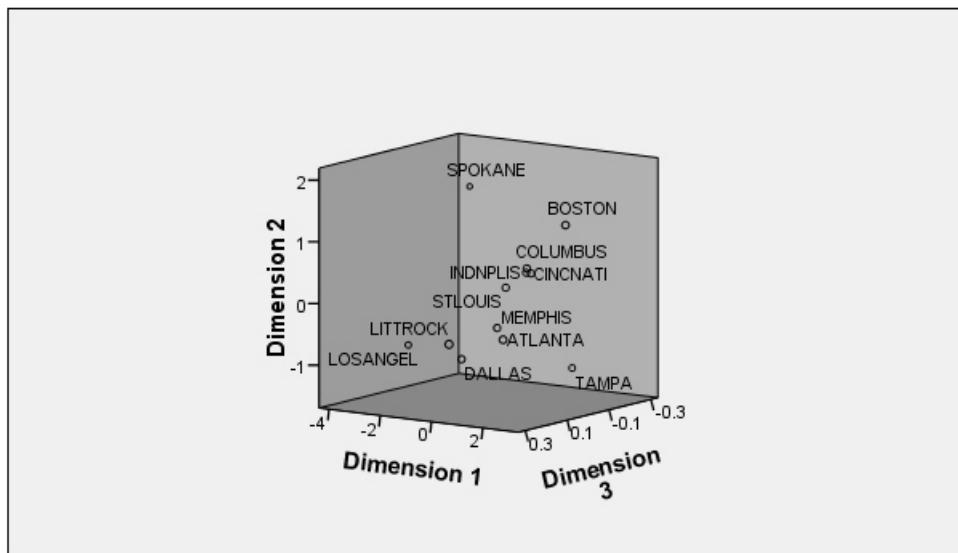
Dimension

Stimulus Number	Stimulus Name	1	2
1	ATLANTA	.7960	-.4129
2	BOSTON	1.6697	1.0482
3	CINCNATI	.5425	.3446
4	COLUMBUS	.6364	.4364
5	DALLAS	-.4171	-.7279
6	INDNPLIS	.3799	.3449
7	LITTROCK	-.0071	-.4414
8	LOSANGEL	-2.5597	-.7003
9	MEMPHIS	.1774	-.3307
10	STLOUIS	.0842	.1621
11	SPOKANE	-2.4789	1.2233
12	TAMPA	1.1768	-.9463

Como puede apreciarse, dos dimensiones explican perfectamente la configuración obtenida.

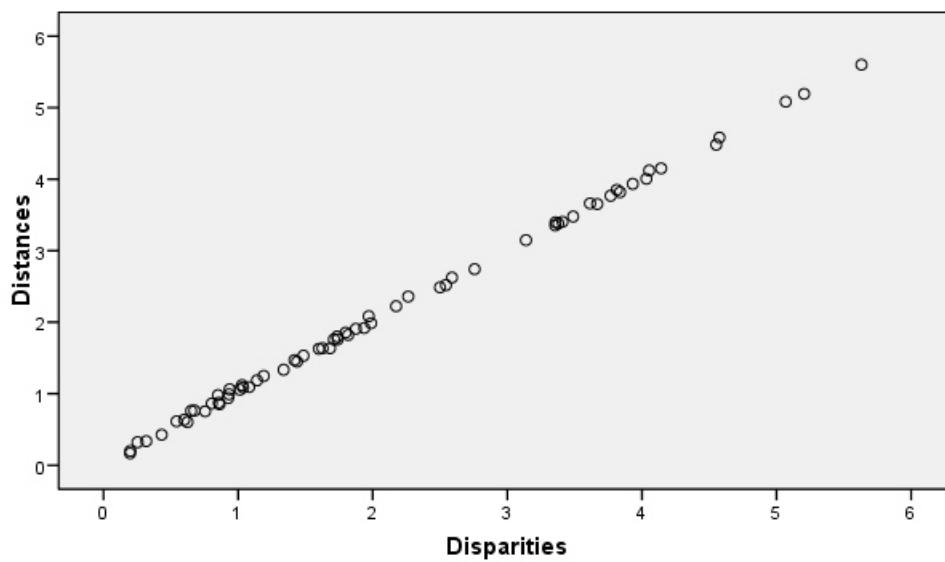
Derived Stimulus Configuration

Euclidean distance model



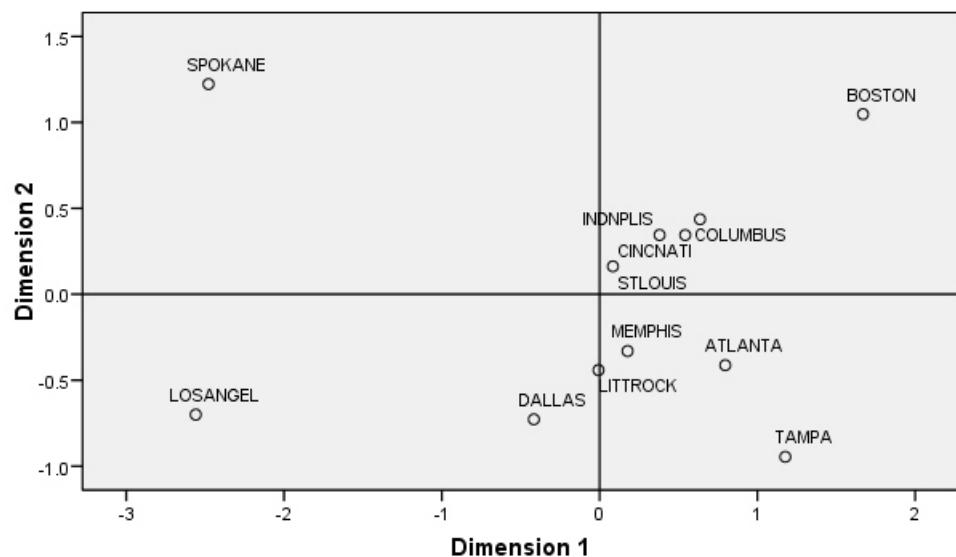
Scatterplot of Linear Fit

Euclidean distance model



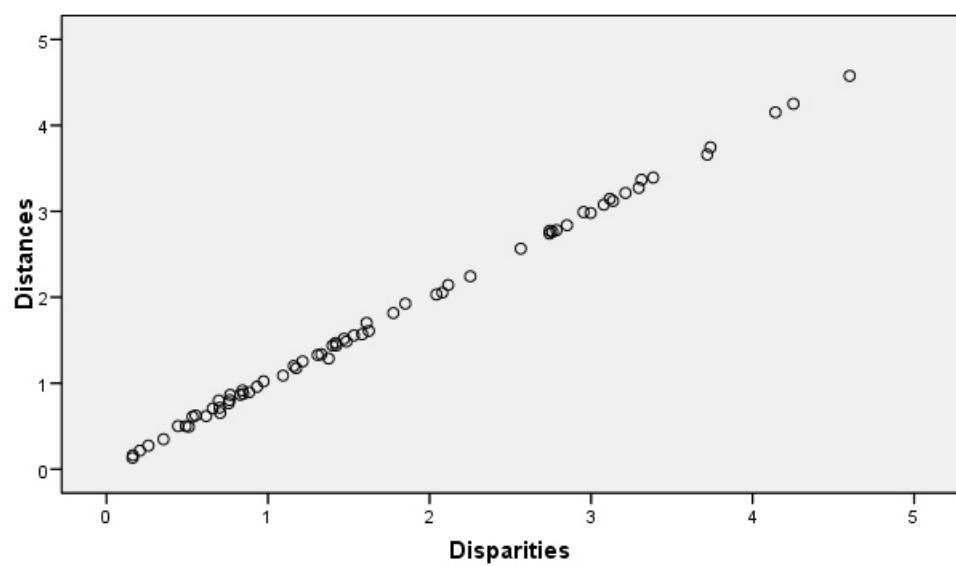
Derived Stimulus Configuration

Euclidean distance model



Scatterplot of Linear Fit

Euclidean distance model



Bibliografía

- Bro R (1998). Multi-Way Analysis in the Food Industry: Models, Algorithms, and Applications. Ph.D. thesis, University of Amsterdam (NL) & Royal Veterinary and Agricultural University (DK).
- Carroll JD, Chang JJ (1970). "Analysis of Individual Differences in Multidimensional Scaling Via an N-Way Generalization of Eckart-Young Decomposition." *Psychometrika*, 35, 283-320.
- de Leeuw J (1977a). Applications of Convex Analysis to Multidimensional Scaling." In JR Barra, F Brodeau, G Romier, B van Cutsem (eds.), *Recent Developments in Statistics*, pp. 133{145. North Holland Publishing Company, Amsterdam.
- de Leeuw J, Heiser WJ (1980). Multidimensional Scaling with Restrictions on the con_guration. In P Krishnaiah (ed.), *Multivariate Analysis, Volume V*, pp. 501{522. North Holland Publishing Company, Amsterdam.
- Helm (1959). A multidimensional ratio scaling analysis of color relations. (Tech. Rep.). Princeton. NJ: Princeton University and Educational Testing Service.
- Kruskal, J. B. (1964).- Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29, 1-28, 115-129.
- Takane, Y., Young, F. & de Leeuw, J. (1977).- Nonmetric Individual Differences MDS: An Alternating Least Squares Method with Optimal Scaling Features. *Psychometrika*, 42, 7-67.