

Análisis de datos. Técnicas aplicadas a datos de proximidad

Tema 5: Análisis de datos de proximidad con MDS.

Introducción.

Denotaremos por Δ el conjunto de datos observados y por Δ^2 el conjunto de datos observados al cuadrado, por \hat{D} las disparidades y por $\widehat{D^2}$ las disparidades u observaciones optimamente representadas al cuadrado y finalmente, por D^2 las distancias al cuadrado. Supondremos que hay n estímulos que queremos representar en dimensión K y R individuos de forma que sus opiniones se denotarán por δ_{ijr} , las disparidades por \widehat{d}_{ijr} y las distancias por d_{ijr} , $i, j = 1, \dots, n$, $r = 1, \dots, R$. Asociado por tanto a cada individuo habrá una matriz de datos y por tanto Δ^2 será en general una colección de matrices de datos al cuadrado Δ_r^2 para cada r ; $\widehat{D^2}$ será la colección de matrices de disparidades al cuadrado $\widehat{D_r^2}$ para cada individuo r y D^2 una colección de matrices de distancias D_r^2 al cuadrado, cuyos elementos serán

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

donde supondremos $w_r \geq 0$, aunque esta restricción es opcional. Además fijando $w_{rk} = 1, \forall r, k$ se reduce al modelo sin ponderaciones.

Los modelos estudiados hasta ahora han sido clasificados atendiendo a la naturaleza métrica o no métrica de los datos, al número de modas y de vías. No obstante existen otros aspectos referentes a los datos de especial consideración y que además de los anteriores condicionan el modelo de MDS empleado. Estos aspectos reflejan fielmente la idea de Fisher de *representación óptima* de que para la representación de los datos, (a) el modelo debería ajustarse a los datos tanto como fuese posible en sentido mínimo cuadrático y (b) las características de las medidas que originan los datos deberían ser estrictamente preservadas. Es por ello que para la descripción del tipo de datos, además del *nivel de medida* (métrico o no métrico si nos referimos a modelos o nominal, ordinal, intervalo y razón si es referente a los datos) expuesto en las secciones anteriores, resulta necesario introducir los siguientes aspectos:

- Réplicas en los datos.
- Presencia de datos faltantes.
- Simetría.
- Proceso de medida.
- Condicionalidad.

Los cuatro primeros apartados se corresponden con la forma que presentan los datos. Hasta ahora, los modelos expuestos presuponen datos de disimilaridad ordinales o continuos, simétricos, sin observaciones faltantes y con o sin réplicas, aunque alguno de ellos puede generalizarse sin preservar las propiedades de medida de los datos. Por tanto para el tratamiento de datos que incumplan alguna de esas hipótesis surge la necesidad de un modelo eficiente.

Para la descripción del último aspecto y del concepto de representación óptima de Fisher, podemos definir el problema de la *representación óptima* como el problema de encontrar una transformación t ,

$$t: \Delta \mapsto \hat{D},$$

de forma que t sea una función que mantenga las restricciones de medida y que exista una relación mínimo cuadrática entre las disparidades y las distancias. Desde este punto de vista las restricciones de medida de nivel, proceso y condicionalidad se corresponden con tres aspectos de las observaciones categóricas.

La *condicionalidad* es un aspecto que está ligado con la idea de Coombs (1964), de que las características de las medidas de las observaciones estén condicionadas a algunos aspectos de la situación experimental de forma que algunas observaciones no puedan ser comparadas con otras. En este caso se distinguirán tres niveles de condicionalidad asociados con tres niveles de comparación de observaciones y por tanto con tres tipos de restricciones en la transformación.

- *Matriz-condicional*. Es el caso más general y se corresponde a la situación por defecto en que cada matriz de observaciones representa la disimilaridad entre los pares de estímulos para un individuo. En este caso, los datos para cada matriz son comparables dentro de la matriz ya que se supone que el individuo usó el mismo criterio para la emisión de su juicio. Sin embargo no serían en principio comparables los datos de una matriz con los de otra ya que un valor de disimilaridad, por ejemplo de 6, en una escala de 1 a 9 para el individuo r indica menor disimilaridad que uno de 7 para ese individuo pero podría indicar mayor disimilaridad si el valor de 7 correspondiese a otro individuo s cuyo punto de vista esté desplazado respecto del primero. En este caso esta restricción hace que sean consideradas R transformaciones t_r , una para cada matriz individual de datos, en lugar de una sola t para todo el conjunto de datos. Así, el dominio de la transformación se subdivide para cada individuo.
- *Fila-condicional*. Este tipo de condicionalidad se manifiesta cuando solo son comparables las observaciones de cada fila pero no las observaciones entre distintas filas. Por ejemplo, puede suceder si la matriz de disimilaridad es asimétrica y para cada fila se ordenan sus elementos (salvo el valor diagonal) en orden creciente de disimilaridad. En este caso solo podrán compararse los elementos dentro de cada fila y no entre filas. Otro caso típico ocurre cuando la matriz de datos no es cuadrada sino rectangular, representando las filas un conjunto como pueden ser los estímulos y las columnas variables o individuos. En este caso solo serían comparables las filas o las columnas, pero no entre sí. Sería el caso de un modelo *Unfolding*. En este caso la restricción se traduce en

considerar una transformación para cada fila de cada matriz, es decir el dominio sería cada fila de cada matriz y la denotamos por t_{ir} , para cada $i=1,\dots,n$, $r=1,\dots,R$.

- *Incondicional.* En este caso, todos los datos son comparables entre si sin distinción alguna y la transformación t tendrá dominio en todas las observaciones.

Teniendo en cuenta las propiedades anteriormente descritas, veamos a continuación el análisis de datos de proximidad mediante diferentes modelos de MDS en los entornos de R y SPSS.

El paquete smacof de R

Actualmente y salvo **smacof**, no existe en R un paquete específico para MDS. Existen funciones en **mass** para hacer MDS clásico y mds no métrico, como hemos visto anteriormente, así como algunas otras específicas en diferentes paquetes. En el paquete **smacof**, las funciones principales son `smacofSym()` para MDS básico sobre datos simétricos, `smacofIndDiff()` para MDS a tres vías, así como otras funciones específicas para soluciones con restricciones y en espacios no Euclídeos. Además de los resultados numéricos, pueden representarse en dos dimensiones la configuración, el diagrama de Shepard, los residuos y el diagrama de descomposición del STRESS. La configuración también puede representarse en tres dimensiones.

La solución básica de MDS con **smacof** para datos simétricos se realiza mediante la función (ver la ayuda actualizada en R):

```
smacofSym(delta, ndim = 2, weightmat = NULL, init = NULL, type = "ratio", ties = "primary", verbose = FALSE, relax = 1, modulus = 1, itmax = 1000, eps = 1e-06)
```

Argumentos (puede variar con cada versión)

<code>delta</code>	Una matriz de disimilaridades simétrica o un objeto de la clase "dist"
<code>ndim</code>	Numero de dimensiones
<code>weightmat</code>	Opcional matriz con los pesos de las disimilaridades
<code>init</code>	Matrix con la configuración inicial (opcional)
<code>type</code>	Ratio, interval mspline, modelos métricos. Ordinal no-métrico
<code>ties</code>	Especificación de los datos empatados en MDS no métrico: "primary", "secondary", or "tertiary"
<code>verbose</code>	Si es TRUE, se muestran los valores intermedios del STRESS
<code>relax</code>	Actualización usando el método smacof con relajaciones.
<code>modulus</code>	Numero de iteraciones para la regression monótona.
<code>itmax</code>	Maximo numero de iteraciones
<code>eps</code>	Criterio de convergencia

Resultados (pueden cambiar con las versiones)

<code>obsdiss</code>	Disimilaridades observadas, normalizadas
<code>confdiss</code>	Disimilaridades asociadas a la configuración
<code>conf</code>	Matriz de la configuración final
<code>stress.m</code>	Valor del stress para MDS métrico
<code>stress.nm</code>	Valor del stress para MDS no métrico (si es calculado)
<code>ndim</code>	Numero de dimensiones
<code>model</code>	Tipo de modelo smacof
<code>niter</code>	Numero de iteraciones
<code>nobj</code>	Numero de objetos

Ejemplo 5.1:

```
## Solución simple de SMACOF para los datos de kinship
data(kinshipdelta)
res = smacofSym(kinshipdelta)
res
summary(res)

## Solución no métrica 3D de SMACOF para los datos de trading
data(trading)
res = smacofSym(trading, ndim = 3, type="ordinal", ties = "secondary")
res
```

Ejercicio 5.1: Realiza el análisis anterior y comenta los resultados obtenidos.

Solución de **smacof** con diferencias individuales.

```
smacofIndDiff(delta, ndim = 2, weightmat = NULL, init = NULL, type = "ratio",
ties = "primary", constraint = NULL, verbose = FALSE, modulus = 1, itmax =
1000, eps = 1e-6)
```

Argumentos (Ver la ayuda actualizada en R)

<code>delta</code>	Una lista de matrices de disimilaridad o de objetos de la clase <code>dist</code>
<code>ndim</code>	Numero de dimensiones
<code>weightmat</code>	Optional matrix with dissimilarity weights
<code>init</code>	Matrix con la configuración inicial (opcional)
<code>type</code>	Ratio, interval mspline, modelos métricos. Ordinal no-métrico
<code>ties</code>	Especificación de los datos empatados en MDS no métrico
<code>constraint</code>	Puede ser "indscale", "idioscale", or "identity" (ver detalles)
<code>verbose</code>	Si es <code>TRUE</code> , se muestran los valores intermedios del STRESS

modulus	Actualización usando el método smacof con relajaciones.
itmax	Numero de iteraciones para la regression monótona.
eps	Maximo numero de iteraciones

Detalles

INDSCAL es utilizado con matrices de pesos identidad. Pueden imponerse una restricción adicional usando “identity”, que hace que las configuraciones para los individuos/replicas/vías sean iguales. Los modelos “diagonal” que hace que las matrices de pesos de la configuración sean diagonales, mientras que “idioscal” no impone ninguna restricción.

Resultados (Ver la ayuda actualizada en R)

obsdiss	Lista de las disimilaridades normalizadas
confdiss	Lista de las disimilaridades asociadas a las configuraciones
conf	Lista de matrices de configuración final
gspace	Configuración conjunta o espacio general de estímulos
cweights	Pesos de la configuración
stress	Valor del stress-1 para MDS métrico
spp	Stress por punto
sps	Stress por sujeto
ndim	Numero de dimensiones
model	Tipo de modelo smacof
niter	Numero de iteraciones
nobj	Numero de objetos

Ejemplo 5.2:

```
data(perception)
res <- smacofIndDiff(perception)
res
summary(res)

res.id <- smacofIndDiff(perception, constraint = "identity")
res.diag <- smacofIndDiff(perception, constraint = "diagonal")
res.idio <- smacofIndDiff(perception, constraint = "idioscal")
```

Ejercicio 5.2: Realiza el análisis anterior y comenta los resultados obtenidos.

ALSCAL con SPSS

ALSCAL es un modelo extremadamente versátil que puede realizar MDS métrico y no métrico con o sin diferencias individuales, a la vez que puede realizar MDU

(*Multidimensional Unfolding*) para datos de preferencia. En esta sección se describe el análisis mediante *Multidimensional Scaling* de datos de disimilaridad con ALSCAL. Actualmente, ALSCAL es el modelo de MDS más utilizado dada su flexibilidad y sobre todo por ser el único que puede realizar simultáneamente análisis tanto métricos como no métricos, por lo que ha sido el primer algoritmo de MDS implementado en los paquetes estadísticos SAS y SPSS, lo que permite complementar el análisis de los datos con otras técnicas como regresión o cluster.

Aunque la versión para Windows de SPSS permite la selección del modelo deseado para el análisis de los datos mediante menús, no todas las opciones están disponibles como es usual en los paquetes estadísticos de alto nivel, por lo que a continuación se describe su estructura de sintaxis. Mediante ALSCAL pueden analizarse con MDS datos del tipo *matriz-condicional* así como *fila-condicional* e incluso *incondicional*. Considera datos provenientes de las cuatro escalas de medida *nominal*, *ordinal*, *intervalo* y *razón*, pudiendo además en el caso *ordinal* analizar datos tanto de *similitud* como de *disimilitud*. También hay que destacar que permite considerar datos de tipo *discreto* (*tied*) o de tipo *continuo* (*untied*), pudiendo realizar el análisis en 6 dimensiones como máximo.

Los párrafos y comandos más importantes son:

ALSCAL. Este párrafo es el único obligatorio y en él hay que especificar necesariamente las columnas de la matriz o matrices de disimilaridad que serán leídas para su análisis. Como mínimo cada matriz de datos deberá tener cuatro filas y cuatro columnas. Por defecto, ALSCAL realiza un análisis de MDS Euclídeo no métrico en dimensión dos. Se supone una o varias matrices de disimilaridad simétricas para analizar, cuyos datos son disimilaridades obtenidas en una escala ordinal, asumiendo condicionalidad tipo matriz y los datos ligados no son desligados. Además, los valores negativos serán tratados como datos faltantes. La única orden asociada al párrafo es:

- **VARIABLES=***lista de variables* o bien **NVAR1 TO NVARq**: En lista de variables se escriben los nombres de las columnas separados por espacios, o bien pueden indicarse indicando el nombre de la primera variable y el de la última usando como nexo la orden *TO*.

/FILE=fichero. Este párrafo indica el nombre del fichero en el que se encuentra la información adicional que queremos incluir en el análisis: una configuración inicial o fija para las coordenadas de los estímulos y/o para los pesos de las matrices de disimilaridad. Aquellos modelos en los que se utiliza se denominan *externos* ya que la información extra que se le suministra al modelo condiciona el análisis efectuado. El fichero puede ser creado con el párrafo *OUTFILE* o con cualquier editor y debe incluir algunas órdenes específicas que lo determinan. Las variables que se corresponden con las dimensiones deben denominarse *DIM1*,..., *DIMK*. Además debe contener la orden *TYPE_* cuyas opciones, en el orden en que deben ser incluidas son:

TYPE_ (CON o ROW)	Configuración
SUB	Pesos para los sujetos
STI	Pesos para los estímulos.

Las siguientes órdenes son opcionales en el párrafo *FILE* y cada una puede ser identificada con la *opción INITIAL* (la opción por defecto) que indica que sea utilizada como configuración o pesos iniciales, o *FIXED* que indica que sea utilizada como valores fijos, es decir, que solo serán estimados el resto de parámetros del modelo.

- **CONFIG** (opción). Leerá la configuración inicial del fichero especificado. Es válida para matrices cuadradas de datos simétricos y asimétricos. La opción asociada para *TYPE_* es *CON* o *ROW*.
- **SUBJWGHT** (opción). Lee los pesos de los individuos o matrices de datos. La opción de *TYPE_* asociada es *SUB*. Esta opción solo es válida para los modelos *INDSCAL* y *AINDS*.
- **STIMWGHT** (opción). Lee los pesos de los estímulos. La opción asociada para *TYPE_* es *STI*. Solo puede usarse si el modelo empleado es *AINDS* o *ASCAL*.

/INPUT=ROWS (opciones). Este párrafo especifica cuantas filas de datos queremos leer por matriz de disimilaridad. *ALSCAL* lee las matrices de disimilaridad fila a fila (estando especificadas las columnas mediante la opción *VARIABLES* de *FILE*). Las opciones asociadas son

- *n*. Indicando el número de filas consecutivas a ser leídas. El mínimo valor para *n* debe ser 4.
- *ALL*. Indicará que se leerán todas las filas (es la opción por defecto).

/SHAPE=opciones. Este párrafo permite indicar la estructura de los datos dentro de las matrices de disimilaridad. Las opciones son:

- **SYMMETRIC**. Es la opción por defecto incluso si se omite el párrafo e indicará que los datos dentro de cada matriz son simétricos. *ALSCAL* solo mirará los valores de cada triangular inferior sin la diagonal, por lo que en ese caso basta con escribir las matrices triangulares de datos de disimilaridad.
- **ASYMMETRIC**. En este caso las matrices son disimilaridades asimétricas y se leerán las matrices completas salvo los valores diagonales.

/LEVEL= opciones. Este párrafo especifica el nivel de medida de los datos de disimilaridad. Las opciones condicionan la fase de representación óptima del algoritmo *ALSCAL* y son:

- **NOMINAL**. Esta opción permite indicar que los datos de disimilaridad han sido obtenidos en escalas nominales, lo que conlleva a producir un análisis no métrico que es útil cuando existen pocas categorías, cuando hay muchas observaciones en las categorías y/o cuando el orden de las categorías es desconocido.
- **ORDINAL [(UNTIE) (SIMILAR)]**. Esta opción indica que los datos de disimilaridad han sido obtenidos en una escala de tipo ordinal. Optativamente pueden añadirse a ésta dos nuevas opciones. Por defecto los datos se suponen

ordinales y discretos aunque si se desea que sean continuos, ésta es la única opción que lo permite, pudiendo asumirse la continuidad mediante la opción *UNTIE*. Además es la única opción que permite el análisis de datos de similaridad, lo cual es indicado empleando la opción *SIMILAR*. Es la opción por defecto si se omite el párrafo completo.

- **INTERVAL [(d)].** Esta opción indica que los datos de disimilaridad provienen de una escala de tipo intervalo. Optativamente permite indicar el grado d del polinomio que en este caso está asociado a la fase de representación óptima del algoritmo. Por defecto, $d=1$, es decir, la transformación será lineal, admitiendo como valor máximo 4.
- **RATIO [(d)].** Esta opción indica que los datos de disimilaridad provienen de una escala de tipo razón. Al igual que la anterior, optativamente puede decidirse el grado de un polinomio de transformación (aunque con constante de traslación cero en este caso), a lo sumo 4 y por defecto 1.

/CONDITION= opciones. Este párrafo permite decidir la condicionalidad de los datos de disimilaridad y por tanto los conjuntos de datos entre los que se pueden establecer comparaciones, lo que condiciona el dominio de las transformaciones del nivel de medida. Las opciones son:

- **MATRIX.** Es la opción por defecto si no se indica ninguna en el párrafo e indica que son comparables los datos entre matrices. Si cada matriz representa un sujeto indicará por tanto que se establece comparación entre sujetos.
- **ROW.** Esta opción indica el grado fila-condicional y solo puede estar asociado con datos asimétricos y por tanto con los modelos *ASCAL* o *AINDS*.
- **UNCONDITIONAL.** Indica el grado {em indondicional}, es decir, todos los datos son comparables entre si independientemente de la fila o matriz en que se encuentren.

/MODEL= modelo. Este párrafo define el modelo empleado para el análisis de los datos. Para el análisis de datos de disimilaridad puede realizarse mediante *ALSCAL* uno de los siguientes cuatro modelos.

- **EUCLID.** Este modelo es el empleado para el análisis de datos de disimilaridad simétricos de una matriz o de varias matrices de disimilaridad si se consideran éstas réplicas. Es el modelo por defecto.
- **INDSCAL.** Este modelo es el empleado para el análisis de datos de disimilaridad simétricos referente a varios sujetos con ponderaciones. Es el modelo *ALSCAL* a tres vías con ponderaciones.
- **ASCAL.** Este modelo es el empleado para el análisis de datos de disimilaridad asimétricos de una matriz o de varias matrices si se consideran éstas réplicas.

- AINDS. Este modelo es el empleado para el análisis de datos de disimilaridad asimétricos para varios individuos con ponderaciones. Es el equivalente al caso anterior pero a tres vías y ponderado.

/CRITERIA=opciones. Mediante este párrafo pueden especificarse varias opciones para el control del algoritmo ALSCAL. Las opciones son:

- CONVERGE(c). Permite indicar el valor c deseado de forma que el algoritmo se detenga cuando el incremento entre dos valores consecutivos del SSTRESS sea menor o igual al valor indicado. Por defecto, $c=0.001$ y constituye uno de los tres criterios de parada del algoritmo.
- ITER(ni). Permite indicar el número máximo de iteraciones que realizará el algoritmo en el caso de no obtener convergencia. Por defecto, $ni=30$ y es otro criterio de parada.
- STRESSMIN(s). Permite indicar un valor s , de forma que el algoritmo se detendrá si el valor del SSTRESS es menor o igual que s . Por defecto, $s=0.05$ y éste constituye otro criterio de parada del algoritmo.
- NEGATIVE. Si se indica esta orden, el modelo permitirá la presencia de pesos negativos en los modelos con diferencias individuales, es decir, INDSCAL y AINDS. También permitirá pesos negativos asociados a los estímulos, como ocurre en el modelo ASCAL.
- CUTOFF(c). Permite determinar el origen para los datos de disimilaridad. Por defecto, $c=0$, lo que indica que los valores de disimilaridad o similitud negativas serán tratados como datos faltantes. Mediante esta orden podemos indicar el origen de la escala de forma que los valores mayores o iguales a éste serán considerados para el análisis.
- NOUBL. Cuando está presente, indica que no sean estimadas las cotas del intervalo asociado a cada dato faltante para el cálculo de la configuración inicial. Esto no afecta al algoritmo ya que los datos faltantes no son considerados por el algoritmo durante el proceso de estimación.
- DIMENS(min,[max]). Permite indicar la dimensión mínima y si se desea también la máxima del modelo. Un único valor entre paréntesis indicará la única dimensión sobre la que trabajar. El máximo número de dimensiones es 6 y por defecto se analizan los datos en dimensión 2.
- TIESTORE(n). Permite indicar la cantidad de memoria reservada para el tratamiento de datos ligados. Por defecto, $n=1000$ pero si durante el análisis aparece un mensaje referente a la falta de memoria por este motivo deberá incrementarse esa cantidad hasta que sea suficiente.

/PRINT=opciones. Este párrafo permite indicar la información que se desea aparezca como resultado del análisis. Las opciones son:

- DATA. Si se indica, aparecerán los datos de disimilaridad originales y las disparidades estimadas durante el proceso.
- HEADER. Esta opción muestra un resumen del modelo, del algoritmo y de las opciones seleccionadas durante el análisis.

/PLOT=opciones. Este párrafo permite controlar los gráficos que serán generados durante el análisis. Las opciones asociadas son:

- DEFAULT. Esta opción producirá los gráficos por defecto asociados a cada modelo.
- ALL. Esta opción producirá todos los gráficos asociados a cada modelo.

/OUTFILE=fichero. Mediante este párrafo se le indica al programa que guarde las coordenadas (y pesos, según el modelo), en un fichero de SPSS, con formato específico de forma que solo es posible leer los datos de este fichero empleando el párrafo *FILE*.

/MATRIX=IN (opción). Cuando se usa SPSS, por defecto, se supone que los datos que se pretenden analizar son los datos correspondientes a la ventana de datos activa en ese momento, la cual será una matriz de disimilaridades. No obstante, este párrafo permite el análisis de datos que no necesariamente tienen que estar en el editor de datos en forma de matriz, es decir, permite el análisis de datos de disimilaridad obtenidos indirectamente mediante el comando PROXIMITIES. Así, si el fichero de datos activo en el editor, es por ejemplo, rectangular formado por diversas variables y queremos realizar MDS de las correlaciones entre las variables, basta con usar el comando PROXIMITIES y guardar la matriz de correlaciones que será de similitud en un fichero. Posteriormente, bastará indicar con el comando MATRIX a ALSCAL cual es el nombre del fichero donde se encuentran los datos en forma de matriz para ser analizados. Las opciones asociadas a este párrafo son:

- fichero. Se indica el nombre del fichero donde están los datos creados con PROXIMITIES. Esta opción debe ser utilizada cuando el fichero que queremos analizar no es el fichero activo.
- Se utiliza esta opción cuando el fichero que se va a analizar coincide con el fichero activo. En este caso puede eliminarse el párrafo MATRIX ya que es superfluo.

Ejemplo 5.3. Consideremos los datos del Ejemplo 4.1 recogidos en la siguiente Tabla.

Tabla 5.1. 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.

Tipo pan	Brea d1.1	Brea d1.2	Brea d2.1	Brea d2.2	Brea d3.1	Brea d3.2	Brea d4.1	Brea d4.2	Brea d5.1	Brea d5.2
Bread 1.1	0.000 0	3.316 6	6.855 7	6.782 3	5.291 5	7.615 8	6.708 2	6.324 6	6.403 1	6.164 4
Bread 1.2	3.316 6	0.000 0	8.246 2	8.775 0	7.416 2	8.306 6	8.000 0	7.810 2	8.000 0	7.416 2
Bread 2.1	6.855 7	8.246 2	0.000 0	3.873 0	4.795 8	3.000 0	5.099 0	4.795 8	5.099 0	5.567 8
Bread 2.2	6.782 3	8.775 0	3.873 0	0.000 0	4.899 0	4.000 0	6.557 4	6.480 7	5.385 2	4.690 4
Bread 3.1	5.291 5	7.416 2	4.795 8	4.899 0	0.000 0	6.164 4	3.605 6	2.828 4	3.316 6	3.464 1
Bread 3.2	7.615 8	8.306 6	3.000 0	4.000 0	6.164 4	0.000 0	6.082 8	6.324 6	5.567 8	5.291 5
Bread 4.1	6.708 2	8.000 0	5.099 0	6.557 4	3.605 6	6.082 8	0.000 0	1.732 1	2.000 0	5.000 0
Bread 4.2	6.324 6	7.810 2	4.795 8	6.480 7	2.828 4	6.324 6	1.732 1	0.000 0	3.000 0	4.899 0
Bread 5.1	6.403 1	8.000 0	5.099 0	5.385 2	3.316 6	5.567 8	2.000 0	3.000 0	0.000 0	3.873 0
Bread 5.2	6.164 4	7.416 2	5.567 8	4.690 4	3.464 1	5.291 5	5.000 0	4.899 0	3.873 0	0.000 0
Bread 1.1	0.000 0	3.162 3	9.798 0	10.44 03	9.486 8	6.245 0	9.539 4	8.000 0	8.888 2	10.34 41
Bread 1.2	3.162 3	0.000 0	8.831 8	9.643 7	8.602 3	5.744 6	9.000 0	7.746 0	8.426 1	9.949 9
Bread 2.1	9.798 0	8.831 8	0.000 0	3.605 6	5.831 0	6.403 1	8.062 3	7.483 3	6.855 7	7.000 0
Bread 2.2	10.44 03	9.643 7	3.605 6	0.000 0	6.557 4	7.483 3	9.165 2	8.888 2	6.782 3	6.164 4
Bread 3.1	9.486 8	8.602 3	5.831 0	6.557 4	0.000 0	5.385 2	6.855 7	7.071 1	3.873 0	4.795 8
Bread 3.2	6.245 0	5.744 6	6.403 1	7.483 3	5.385 2	0.000 0	8.246 2	7.141 4	6.633 2	7.071 1
Bread 4.1	9.539 4	9.000 0	8.062 3	9.165 2	6.855 7	8.246 2	0.000 0	3.000 0	4.690 4	6.164 4
Bread 4.2	8.000 0	7.746 0	7.483 3	8.888 2	7.071 1	7.141 4	3.000 0	0.000 0	5.385 2	7.000 0
Bread 5.1	8.888 2	8.426 1	6.855 7	6.782 3	3.873 0	6.633 2	4.690 4	5.385 2	0.000 0	2.828 4
Bread 5.2	10.34 41	9.949 9	7.000 0	6.164 4	4.795 8	7.071 1	6.164 4	7.000 0	2.828 4	0.000 0
Bread 1.1	0.000 0	4.899 0	6.000 0	6.403 1	8.000 0	6.480 7	8.062 3	9.591 7	6.403 1	8.000 0
Bread 1.2	4.899 0	0.000 0	7.348 5	6.855 7	8.246 2	7.348 5	7.937 3	10.09 95	6.708 2	8.124 0
Bread 2.1	6.000 0	7.348 5	0.000 0	2.236 1	4.242 6	4.000 0	5.744 6	7.071 1	4.582 6	5.656 9
Bread 2.2	6.403 1	6.855 7	2.236 1	0.000 0	3.873 0	3.000 0	4.690 4	6.245 0	4.242 6	5.000 0
Bread 3.1	8.000 0	8.246 2	4.242 6	3.873 0	0.000 0	4.472 1	5.000 0	7.211 1	5.196 2	6.480 7
Bread 3.2	6.480 7	7.348 5	4.000 0	3.000 0	4.472 1	0.000 0	2.645 8	4.000 0	3.316 6	3.464 1
Bread 4.1	8.062 3	7.937 3	5.744 6	4.690 4	5.000 0	2.645 8	0.000 0	3.605 6	4.000 0	3.316 6
Bread 4.2	9.591 7	10.09 95	7.071 1	6.245 0	7.211 1	4.000 0	3.605 6	0.000 0	6.082 8	3.741 7

Bread 5.1	6.403 1	6.708 2	4.582 6	4.242 6	5.196 2	3.316 6	4.000 0	6.082 8	0.000 0	3.316 6
Bread 5.2	8.000 0	8.124 0	5.656 9	5.000 0	6.480 7	3.464 1	3.316 6	3.741 7	3.316 6	0.000 0
Bread 1.1	0.000 0	3.162 3	9.434 0	7.615 8	7.937 3	8.124 0	9.486 8	8.246 2	9.643 7	8.306 6
Bread 1.2	3.162 3	0.000 0	8.426 1	6.000 0	7.000 0	6.324 6	8.944 3	7.746 0	7.937 3	7.000 0
Bread 2.1	9.434 0	8.426 1	0.000 0	4.795 8	3.162 3	4.582 6	5.196 2	4.582 6	4.899 0	3.464 1
Bread 2.2	7.615 8	6.000 0	4.795 8	0.000 0	3.316 6	3.162 3	7.348 5	5.831 0	6.082 8	4.358 9
Bread 3.1	7.937 3	7.000 0	3.162 3	3.316 6	0.000 0	4.123 1	5.744 6	4.123 1	6.000 0	3.162 3
Bread 3.2	8.124 0	6.324 6	4.582 6	3.162 3	4.123 1	0.000 0	6.164 4	5.477 2	5.196 2	3.000 0
Bread 4.1	9.486 8	8.944 3	5.196 2	7.348 5	5.744 6	6.164 4	0.000 0	2.828 4	5.196 2	5.385 2
Bread 4.2	8.246 2	7.746 0	4.582 6	5.831 0	4.123 1	5.477 2	2.828 4	0.000 0	5.385 2	4.795 8
Bread 5.1	9.643 7	7.937 3	4.899 0	6.082 8	6.000 0	5.196 2	5.196 2	5.385 2	0.000 0	4.899 0
Bread 5.2	8.306 6	7.000 0	3.464 1	4.358 9	3.162 3	3.000 0	5.385 2	4.795 8	4.899 0	0.000 0
Bread 1.1	0.000 0	3.000 0	8.062 3	6.633 2	7.348 5	5.916 1	9.746 8	8.000 0	8.000 0	7.483 3
Bread 1.2	3.000 0	0.000 0	8.124 0	6.082 8	7.416 2	5.831 0	9.486 8	7.681 1	7.280 1	6.855 7
Bread 2.1	8.062 3	8.124 0	0.000 0	3.316 6	2.236 1	3.741 7	5.099 0	6.403 1	4.582 6	4.795 8
Bread 2.2	6.633 2	6.082 8	3.316 6	0.000 0	3.741 7	2.645 8	5.916 1	6.324 6	3.741 7	3.464 1
Bread 3.1	7.348 5	7.416 2	2.236 1	3.741 7	0.000 0	3.000 0	5.196 2	5.477 2	4.472 1	4.472 1
Bread 3.2	5.916 1	5.831 0	3.741 7	2.645 8	3.000 0	0.000 0	6.324 6	6.403 1	4.582 6	3.873 0
Bread 4.1	9.746 8	9.486 8	5.099 0	5.916 1	5.196 2	6.324 6	0.000 0	5.385 2	6.855 7	5.744 6
Bread 4.2	8.000 0	7.681 1	6.403 1	6.324 6	5.477 2	6.403 1	5.385 2	0.000 0	6.324 6	5.656 9
Bread 5.1	8.000 0	7.280 1	4.582 6	3.741 7	4.472 1	4.582 6	6.855 7	6.324 6	0.000 0	2.449 5
Bread 5.2	7.483 3	6.855 7	4.795 8	3.464 1	4.472 1	3.873 0	5.744 6	5.656 9	2.449 5	0.000 0
Bread 1.1	0.000 0	3.605 6	5.656 9	8.185 4	7.000 0	7.000 0	8.717 8	8.888 2	7.348 5	9.165 2
Bread 1.2	3.605 6	0.000 0	6.245 0	8.717 8	8.717 8	8.246 2	9.000 0	9.055 4	8.185 4	9.539 4
Bread 2.1	5.656 9	6.245 0	0.000 0	6.245 0	6.708 2	7.280 1	6.782 3	6.557 4	4.690 4	5.656 9
Bread 2.2	8.185 4	8.717 8	6.245 0	0.000 0	6.164 4	5.291 5	6.855 7	5.291 5	5.916 1	6.855 7
Bread 3.1	7.000 0	8.717 8	6.708 2	6.164 4	0.000 0	2.828 4	5.196 2	4.899 0	5.744 6	6.082 8
Bread 3.2	7.000 0	8.246 2	7.280 1	5.291 5	2.828 4	0.000 0	6.245 0	5.099 0	6.708 2	7.549 8
Bread 4.1	8.717 8	9.000 0	6.782 3	6.855 7	5.196 2	6.245 0	0.000 0	3.873 0	5.656 9	4.000 0
Bread	8.888	9.055	6.557	5.291	4.899	5.099	3.873	0.000	5.385	4.358

4.2	2	4	4	5	0	0	0	0	2	9
Bread 5.1	7.3485	8.1854	4.6904	5.9161	5.7446	6.7082	5.6569	5.3852	0.0000	3.4641
Bread 5.2	9.1652	9.5394	5.6569	6.8557	6.0828	7.5498	4.0000	4.3589	3.4641	0.0000
Bread 1.1	0.0000	6.0000	9.3808	8.3066	6.6332	10.0995	9.3274	8.8882	7.6158	9.0000
Bread 1.2	6.0000	0.0000	9.1652	8.5440	7.6158	9.6954	10.9087	11.0000	7.4833	8.3066
Bread 2.1	9.3808	9.1652	0.0000	4.5826	5.2915	3.4641	9.3274	9.9499	5.8310	6.5574
Bread 2.2	8.3066	8.5440	4.5826	0.0000	2.6458	4.1231	6.1644	7.3485	4.1231	5.2915
Bread 3.1	6.6332	7.6158	5.2915	2.6458	0.0000	5.6569	5.3852	7.0000	3.1623	4.5826
Bread 3.2	10.0995	9.6954	3.4641	4.1231	5.6569	0.0000	9.2195	9.3274	5.6569	6.4031
Bread 4.1	9.3274	10.9087	9.3274	6.1644	5.3852	9.2195	0.0000	4.6904	6.8557	6.4807
Bread 4.2	8.8882	11.0000	9.9499	7.3485	7.0000	9.3274	4.6904	0.0000	8.5440	7.0711
Bread 5.1	7.6158	7.4833	5.8310	4.1231	3.1623	5.6569	6.8557	8.5440	0.0000	5.1962
Bread 5.2	9.0000	8.3066	6.5574	5.2915	4.5826	6.4031	6.4807	7.0711	5.1962	0.0000
Bread 1.1	0.0000	3.4641	10.7703	11.9583	9.5394	11.1803	11.4891	12.3693	11.0454	10.7703
Bread 1.2	3.4641	0.0000	10.6771	11.0000	9.5394	10.1489	10.9545	11.4455	10.9545	10.6771
Bread 2.1	10.7703	10.6771	0.0000	4.1231	7.1414	6.4031	9.1652	9.6437	6.9282	8.2462
Bread 2.2	11.9583	11.0000	4.1231	0.0000	7.7460	5.0990	7.9373	8.2462	6.8557	7.8102
Bread 3.1	9.5394	9.5394	7.1414	7.7460	0.0000	5.4772	6.7082	6.7823	5.9161	4.7958
Bread 3.2	11.1803	10.1489	6.4031	5.0990	5.4772	0.0000	4.3589	4.2426	3.8730	4.3589
Bread 4.1	11.4891	10.9545	9.1652	7.9373	6.7082	4.3589	0.0000	3.0000	4.0000	3.1623
Bread 4.2	12.3693	11.4455	9.6437	8.2462	6.7823	4.2426	3.0000	0.0000	4.3589	4.1231
Bread 5.1	11.0454	10.9545	6.9282	6.8557	5.9161	3.8730	4.0000	4.3589	0.0000	2.8284
Bread 5.2	10.7703	10.6771	8.2462	7.8102	4.7958	4.3589	3.1623	4.1231	2.8284	0.0000

En primer lugar importamos los datos con SPSS usando en el menú *Archivo* la opción *Leer datos de texto*. A continuación efectuamos el análisis de los datos para un modelo de diferencias individuales con matriz de pesos diagonal.

ALSCAL

VARIABLES= Bread1.1 Bread1.2 Bread2.1 Bread2.2 Bread3.1 Bread3.2
Bread4.1 Bread4.2 Bread5.1 Bread5.2
/SHAPE=SYMMETRIC

```
/LEVEL=RATIO
/CONDITION=MATRIX
/MODEL=INDSCAL
/CRITERIA=CONVERGE(.00001) STRESSMIN(.0005) ITER(100) CUTOFF(0)
DIMENS(2,2)
/PLOT=DEFAULT ALL
/PRINT=DATA HEADER .
```

Los resultados obtenidos se aprecian a continuación, donde se puede ver que solo fueron necesarias 5 iteraciones para obtener la convergencia.

Alscal Procedure Options

Data Options-

```
Number of Rows (Observations/Matrix).  10
Number of Columns (Variables) . . .  10
Number of Matrices . . . . .  8
Measurement Level . . . . .  Ratio
Data Matrix Shape . . . . .  Symmetric
Type . . . . .  Dissimilarity
Approach to Ties . . . . .  Leave Tied
Conditionality . . . . .  Matrix
Data Cutoff at . . . . .  .000000
```

Model Options-

```
Model . . . . .  Indscal
Maximum Dimensionality . . . . .  2
Minimum Dimensionality . . . . .  2
Negative Weights . . . . .  Not Permitted
```

Output Options-

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

Algorithmic Options-

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

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

Young's S-stress formula 1 is used.

Iteration	S-stress	Improvement
0	.37473	
1	.37473	
2	.35227	.02247
3	.35156	.00071
4	.35153	.00003
5	.35153	.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.

Matrix	Stress	RSQ	Matrix	Stress	RSQ
1	.305	.623	2	.302	.672
3	.330	.618	4	.319	.639
5	.319	.641	6	.364	.541
7	.318	.582	8	.265	.738

Averaged (rms) over matrices
 Stress = .31632 RSQ = .63175



Configuration derived in 2 dimensions

Stimulus Coordinates

Dimension

Stimulus Stimulus 1 2

Number	Name		
1	Bread1.1	1.8302	.8668
2	Bread1.2	1.8546	.8697
3	Bread2.1	.0171	-1.4471
4	Bread2.2	.0446	-1.3751
5	Bread3.1	-.3026	-.6244
6	Bread3.2	.0457	-1.1688
7	Bread4.1	-1.0340	1.0595
8	Bread4.2	-.9371	1.1980
9	Bread5.1	-.6737	.3217
10	Bread5.2	-.8447	.2997



Subject weights measure the importance of each dimension to each subject. Squared weights sum to RSQ.

A subject with weights proportional to the average weights has a weirdness of zero, the minimum value.

A subject with one large weight and many low weights has a weirdness near one.

A subject with exactly one positive weight has a weirdness of one, the maximum value for nonnegative weights.

Subject Weights

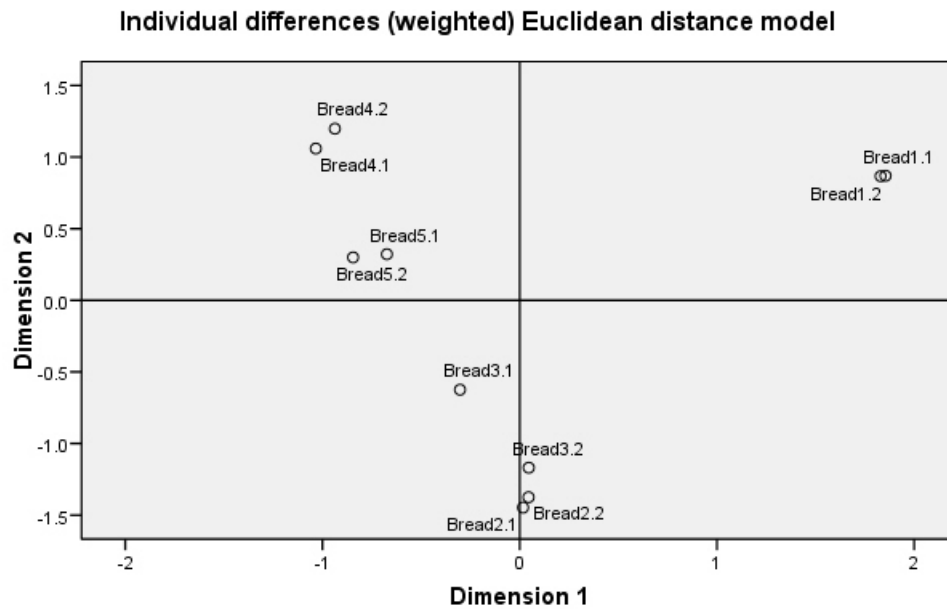
Subject Number	Weirdness	Dimension	
		1	2
1	.1596	.5995	.5137
2	.2314	.5900	.5691
3	.2309	.7146	.3280
4	.1426	.7060	.3747
5	.1219	.7019	.3851
6	.1129	.6429	.3578
7	.2230	.5526	.5256
8	.0764	.7398	.4363

Overall importance of each dimension: .4344 .1973

La configuración obtenida para el modelo INDSCAL queda de la siguiente forma, donde puede apreciarse las agrupaciones para los tipos de pan, excepto para el tipo 3 en el que se marca una clara diferencia respecto a la opinión

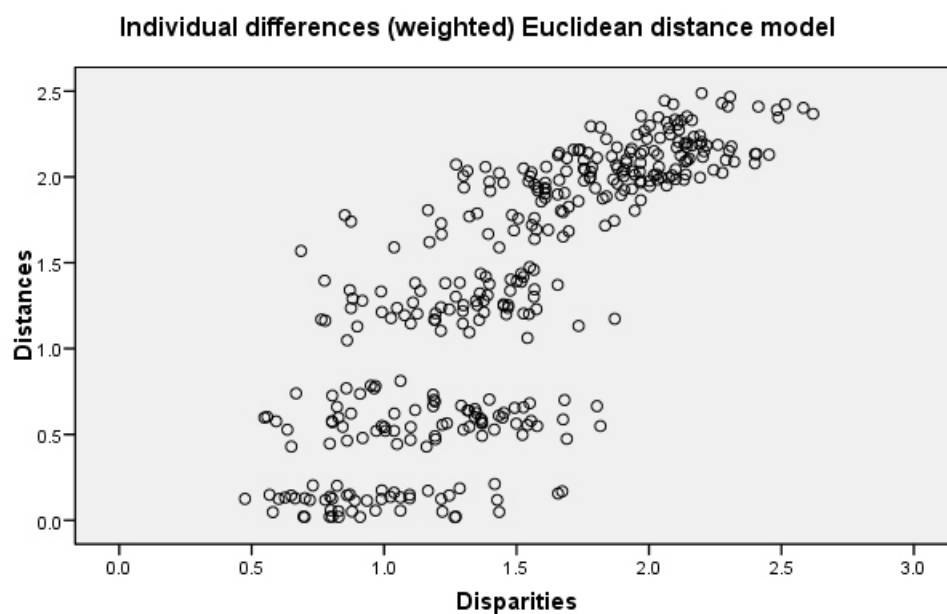
emitida en ambas réplicas.

Derived Stimulus Configuration

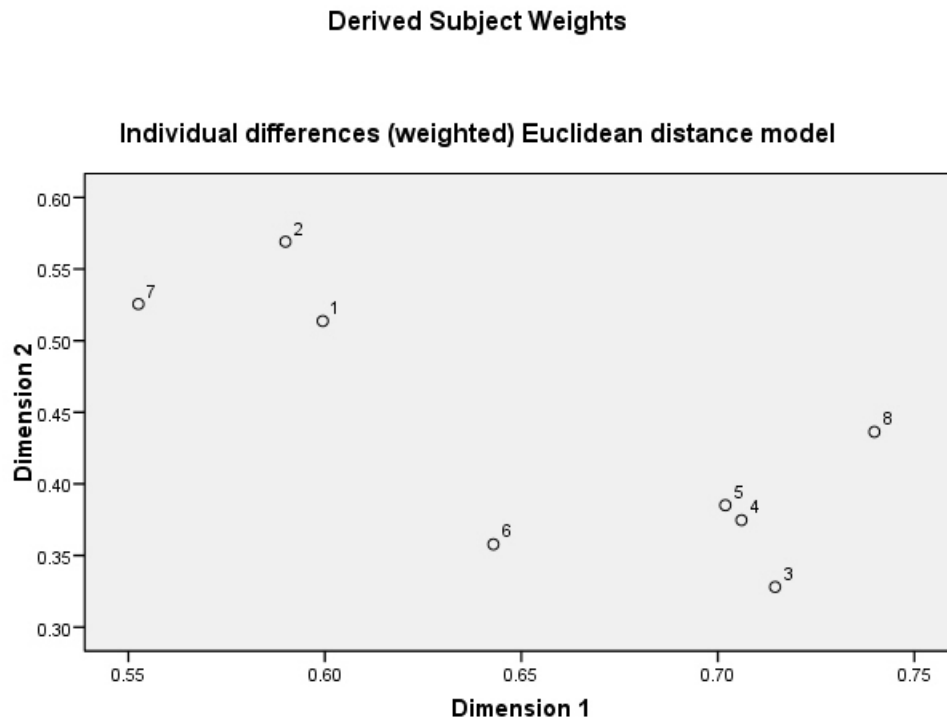


El diagrama de Shepard muestra igualmente un ajuste desigual respecto al tamaño de las disimilaridades, como en los modelos para pesos iguales anteriormente analizados.

Scatterplot of Linear Fit



La gráfica de los pesos indica que los individuos 7, 1 y 2 poseen opiniones parecidas, aunque diferentes de las del resto. Además los individuos 2 y 8 fueron los más influyentes en el análisis.



Ejercicio 5.3. Usando los datos de la Tabla 4.1 de colors de Helm, (1959):

- Leer los datos con SPSS.
- Realizar el análisis de los datos usando ALSCAL SPSS para el modelo identidad.
- Realizar el análisis de los datos con ALSCAL de SPSS para el modelo de diferencias individuales.
- Compara los resultados con los obtenidos mediante PROXSCAL.

Bibliografía

- Coombs CH (1950). Psychological Scaling Without a Unit of Measurement." Psychological Review, 57, 145-158.
- Helm (1959). A multidimensional ratio scaling analysis of color relations. (Tech. Rep.). Princeton. NJ: Princeton University and Educational Testing Service.

