

**El escalamiento multidimensional (MDS) puede considerarse un término genérico para nombrar a una familia muy amplia de procedimientos de análisis multivariado**

## Objetivos generales

- Resumir los datos mediante un pequeño conjunto de nuevas variables, construidas como transformaciones de las originales, con la mínima pérdida de información.
- Encontrar grupos en los datos, si existen.
- Clasificar nuevas observaciones en grupos definidos y relacionar dos conjuntos de variables.
- Encontrar posibles datos atípicos (outliers)

# Escalamiento Multidimensional

## Objetivo específico

En un término específico y que es el adoptado comunmente, el escalamiento multidimensional se define como una técnica de análisis multivariante cuyo **objetivo** es representar las *proximidades* entre un conjunto de objetos o estímulos como distancias en un espacio de baja dimensión (generalmente de dos o tres dimensiones).

## Proximidades como materia prima del MDS

Las proximidades pueden ser datos de distinto tipo, tales como medidas de similaridad o disimilaridad, distancias, correlaciones, entre otras.

## Ejemplo ilustrativo: Ubicar $n$ ciudades en su espacio geográfico (mapa de las ciudades)

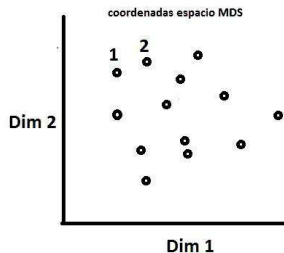
- Si se dispone de las coordenadas geográficas de las ciudades, la representación es trivial, pues simplemente localizamos las coordenadas en el espacio y de esta manera construimos el mapa.
- Supongamos que se dispone únicamente de una matriz cuadrada ( $n \times n$ ) con las distancias entre cada par de ciudades y con esta información queremos reconstruir el mapa de las ciudades
- La primer solución a este problema se debe a los trabajos de Schoenberg (1935) y Young & Houselholder (1938), que se formalizaron en el teorema del MDS clásico.

# Objetivo del MDS

**Matriz de distancias**

$$D = \begin{bmatrix} d_{11} & d_{12} & . & . & . & . & d_{1n} \\ d_{21} & d_{22} & . & . & . & . & d_{2n} \\ d_{31} & d_{32} & . & . & . & . & d_{3n} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ d_{n1} & d_{n2} & . & . & . & . & d_{nn} \end{bmatrix}$$

MDS



## Teorema: MDS clásico

Sea  $\Delta_{(n \times n)} = \{\delta_{ij}\}$  una matriz de distancias entre  $n$  puntos en un espacio de configuración de dimensión  $K$  y sea  $B_{(n \times n)}$  la matriz dada por  $B = HAH$ , siendo  $H_{(n \times n)}$  dada por  $H = I - n^{-1}\mathbf{1}\mathbf{1}^t$  y  $A_{(n \times n)}$  la matriz cuyos elementos vienen dados a través de  $a_{rs} = -\frac{1}{2}\delta_{rs}^2$ . Entonces,  $\Delta$  es una matriz de distancias Euclidianas sii  $B$  es semidefinida positiva. Además se tiene:

- Si  $\Delta$  es la matriz de distancias Euclidianas para una configuración dada por  $Z_{(n \times K)} = (\mathbf{z}_1, \dots, \mathbf{z}_n)^t$ , entonces  $B$  se puede representar como  $B = (HZ)(HZ)^t$ , es decir,

$$b_{ij} = (\mathbf{z}_i - \bar{\mathbf{z}})^t(\mathbf{z}_j - \bar{\mathbf{z}}) \quad \forall i, j = 1, \dots, n,$$

de donde  $B \geq 0$ .  $B$  será la matriz centrada de productos escalares de  $Z$ .

# Modelo clasico de MDS (continuación)

- 1 Inversamente, si  $\mathbf{B}$  es positiva semidefinida de rango  $K$ , entonces se puede construir una configuración  $\mathbf{X}$  asociada a  $\mathbf{B}$  de la siguiente forma:

Sean  $\lambda_1 > \dots > \lambda_K$  los  $K$  valores propios positivos de  $\mathbf{B}$  y sus vectores propios asociados  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_K$  se organizan como columnas formando  $\mathbf{X}_{(n \times K)} = (\mathbf{x}_{(1)}, \dots, \mathbf{x}_{(K)})$ , normalizados según la condición

$$\mathbf{x}_{(i)}' \mathbf{x}_{(i)} = \lambda_i, \quad \forall i = 1, \dots, K.$$

Entonces  $d(\mathbf{x}_i, \mathbf{x}_j) = \delta_{ij}$ ,  $\forall \mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^K$ . Además esa configuración está centrada en  $\bar{\mathbf{x}} = \mathbf{0}$  y  $\mathbf{B}$  es la matriz de productos escalares de esa configuración, es decir  $\mathbf{B} = \mathbf{X}\mathbf{X}'$

# Objetivo real de MDS

- El teorema del MDS clásico establece que si  $\Delta$  es Euclidiana o  $B$  es p.s, podemos encontrar una  $X$  en un espacio de dimensión  $K$ , donde  $K$  es el numero de valores propios positivos de  $B$ , tal que  $d(x_i, x_j) = \delta_{ij}$ ,  $\forall i, j = 1, \dots, n$ .
- El problema que resuelve MDS surge cuando se pretende una representación en un espacio de dimensión  $< K$  y/o la información de proximidad entre los elementos a representar no viene dada en términos de distancias euclídeas.
- En el ejemplo de las ciudades, distancias no Euclidianas podría corresponder al caso en que quisieramos reconstruir el mapa (dos dimensiones) utilizando distancias por carretera o algun coeficiente de proximidad entre las ciudades basado en criterios de tipo económico, social, etc.

# Objetivo real de MDS

**Objetivo real de MDS:** Dado  $\mathbf{O} = \{o_1, \dots, o_n\}$  y  $\Delta = \{\delta_{ij}\}$  entre ellos, construir una configuración  $\mathbf{X} = (\mathbf{x}'_1, \dots, \mathbf{x}'_n)'$ ,  $\mathbf{x}_i \in \mathbb{R}^M$ ,  $M \ll n$ , tal que  $d(\mathbf{x}_i, \mathbf{x}_j) \approx \delta_{ij}$ .

- Al igual que en el caso de componentes principales, si asumimos que los primeros  $m$  valores propios de  $\mathbf{B}$  son positivos y muy grandes y los restantes  $K - m$  valores propios son mucho menores, podemos obtener una representación aproximada de las disimilaridades utilizando solo los  $m$  valores propios positivos mas grandes y sus vectores propios asociados.
- El procedimiento para obtener la solución clásica de MDS mediante los vectores y valores propios de  $\mathbf{B}$  se conoce como *coordenadas principales*



# Construcción de la configuración MDS mediante coordenadas principales (Modelo clásico)

Supongamos que tenemos una matriz de distancias o disimilaridades al cuadrado  $\Delta$ . El procedimiento para obtener las coordenadas principales es el siguiente:

- Se construye la matriz  $\mathbf{B} = -\frac{1}{2}\mathbf{H}\Delta\mathbf{H}$ , de productos cruzados, donde  $\mathbf{H} = \mathbf{I} - \mathbf{n}^{-1}\mathbf{1}\mathbf{1}^t$
- Se obtienen los valores propios de  $\mathbf{B}$ , tomándose los  $m$  valores propios positivos mas grandes, de tal forma que los restantes  $K - m$  valores propios sean cercanos a cero.
- Si  $\mathbf{B}$  no es positiva semidefinida, se podrían ignorar los valores propios negativos y tomar solo los  $m$  valores propios positivos mas grandes y seguir el procedimiento.

# Construcción de la configuración MDS mediante coordenadas principales (Modelo clásico)

- Considerando los  $m$  valores propios positivos y sus vectores propios asociados, por la descomposición espectral podemos aproximar la matriz  $\mathbf{B}$  por

$$\mathbf{B} \approx \mathbf{U}_m \mathbf{\Lambda}_m \mathbf{U}_m' = (\mathbf{U}_m \mathbf{\Lambda}_m^{1/2})(\mathbf{\Lambda}_m^{1/2} \mathbf{U}_m') = (\mathbf{U}_m \mathbf{\Lambda}_m^{1/2})(\mathbf{U}_m \mathbf{\Lambda}_m^{1/2})' = \mathbf{X}_m \mathbf{X}_m',$$

entonces las coordenadas principales de los puntos están dadas por

$$\mathbf{X}_m = (\mathbf{U}_m \mathbf{\Lambda}_m^{1/2}) = [\sqrt{\lambda_1} \mathbf{e}_1 | \cdots | \sqrt{\lambda_m} \mathbf{e}_m]$$

- Al igual que en CP se puede elegir un número apropiado de dimensiones  $m$ , usando el criterio de la proporción de varianza explicada por las primeras  $m$  dimensiones, dado por

$$\frac{\sum_{i=1}^m \lambda_i}{\sum_{i=1}^n |\lambda_i|}$$

# Relación entre coordenadas principales y componentes principales

- El análisis de coordenadas principales o MDS clasico, está muy relacionado con componentes principales. En ambos casos tratamos de reducir la dimensionalidad de los datos.
- En componentes principales partimos de la matriz de covarianza muestral  $\mathbf{S} = \frac{1}{n}\mathbf{X}'\mathbf{X}$
- Obtenemos sus valores y vectores propios y luego proyectamos las variables sobre las direcciones dadas por los vectores propios para obtener los valores de los componentes principales.
- Esas componentes son equivalentes a las coordenadas principales obtenidas directamente como los vectores propios (normalizados por  $\sqrt{\lambda_i}$ ) de  $\mathbf{B} = \mathbf{X}\mathbf{X}'$ .

# Relación entre coordenadas principales y componentes principales

- Cuando los datos originales están dados por la matriz  $\mathbf{X}$  y construimos la matriz  $\mathbf{\Delta}$  de distancias utilizando las distancias euclidianas entre los puntos de  $\mathbf{X}$ , las coordenadas principales obtenidas de la matriz  $\mathbf{B}$  son equivalentes a los componentes principales obtenidos de la matriz  $\mathbf{X}$ .
- Sin embargo, el concepto de coordenadas principales se puede aplicar a una gama mas amplia de problemas que componentes principales, ya que las coordenadas principales pueden obtenerse siempre, aunque las distancias de partida no hayan sido generadas exactamente a partir de las variables  $\mathbf{X}$ .

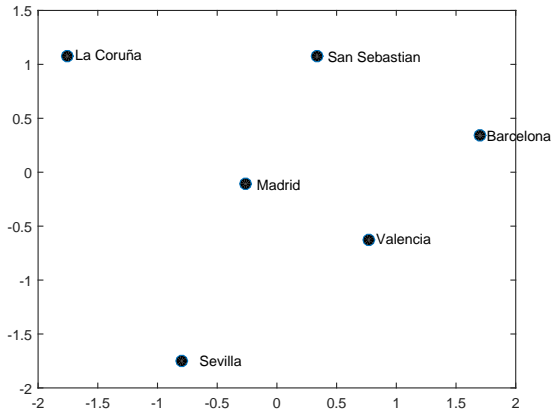
## Ejemplo: Obtención de las coordenadas principales

- Las distancias en kilómetros por carretera entre 6 ciudades españolas: Madrid(M), Barcelona(B), Valencia(V), Sevilla(S), San Sebastián(SS) y La Coruña(LC), están dadas en el siguiente cuadro, que llamaremos matriz de distancias  $\Delta$ .

$$\Delta = \begin{array}{c|cccccc} & M & B & V & S & SS & LC \\ \hline M & 0 & 627 & 351 & 550 & 488 & 603 \\ B & 627 & 0 & 361 & 1043 & 565 & 1113 \\ V & 351 & 361 & 0 & 567 & 564 & 954 \\ S & 550 & 1043 & 567 & 0 & 971 & 950 \\ SS & 488 & 565 & 564 & 971 & 0 & 713 \\ LC & 603 & 1113 & 954 & 950 & 713 & 0 \end{array}$$

- Construye un mapa de estas ciudades mediante coordenadas principales.

# Ejemplo: Obtención de las coordenadas principales



# Las proximidades como materia prima del MDS

## Antecedentes del MDS en el área de psicología

- Al igual que otras técnicas de reducción de dimensionalidad, MDS tuvo su origen en el área de psicología.
- La terminología empleada en MDS ha sido elaborada fundamentalmente en este ámbito y frecuentemente puede no resultar usual en estadística.
- El MDS trata indistintamente los términos: objetos, individuos, sujetos o estímulos.

## Concepto de proximidad

- Desde un punto de vista general el término proximidad indica el concepto de cercanía en espacio, tiempo o cualquier otro contexto.
- Desde un punto de vista matemático, ese término hace referencia al concepto de disimilaridad o similaridad entre dos elementos.

# Las proximidades como materia prima del MDS

## Definición y Propiedades de las disimilaridades

Sea  $\mathbf{O}$  un conjunto de elementos (individuos, estímulos sujetos u objetos) sobre los que queremos definir una disimilaridad.

Dados dos elementos  $o_i, o_j \in \mathbf{O}$  y  $\delta$  una función. Se dice que  $\delta$  es una *disimilaridad* si verifica:

- 1  $\delta_{ij} = \delta_{ji}, \forall i, j$
- 2  $\delta_{ij} \geq 0, \forall i, j$
- 3  $\delta_{ii} = 0, \forall i.$

## Similaridad como concepto opuesto a disimilaridad

Una similaridad  $s$  es un término opuesto al de disimilaridad por lo que deberá existir alguna transformación monótona  $t$  tal que  $t(s) = \delta$ . Una transformación de ese tipo podría ser  $\delta = 1 - s$  si  $0 \leq s \leq 1$ .



# Las proximidades como materia prima del MDS

## Medidas de proximidad para datos binarios

- Coeficiente de Jaccard (1908)
- Roger and Tanimoto (1960)
- Gower and Legendre (1986)

## Medidas de proximidad para datos continuos

- Distancia Euclidea (1908)
- Distancia City block
- Distancia de Minkowski

Proximidades obtenidas como juicios emitidos por sujetos sobre pares de objetos

Las proximidades son producto del valor subjetivo que le asigna una persona a pares de objetos.

# Clasificación de las proximidades de acuerdo al número de modos y vías

Las proximidades se clasifican de acuerdo al número de **modos** y al número de **vías**

**Ejemplo: 10 botellas de vino de diferentes bodegas y catadores**

- Un catador proporciona una puntuación entera entre 0 y 10 para cada par de botellas de acuerdo a la proximidad que a su juicio presentan entre ellas, generándose disimilaridades de la forma  $\delta_{ij}$
- Un conjunto de  $R$  catadores da una puntuación entera entre 0 y 10 para cada par de botellas de acuerdo a la proximidad que a su juicio presentan entre ellas, generándose  $\delta_{ijr}$  donde  $i, j$  se refieren a los vinos y  $r$  al  $r$ -ésimo catador.

# Clasificación de las proximidades de acuerdo al número de modos y vías

## Concepto de modo y vía en MDS

Considerando los dos conjuntos distintos implicados en su obtención -objetos e individuos-, las proximidades suelen clasificarse mediante los términos **modo** y **vía**

- 1 *Modo*. Cada conjunto de objetos distintos implicados en la obtención de las proximidades. Así,  $\delta_{ijr}$  del ejemplo son proximidades a dos modos.
- 2 *Vía*. Cada índice implicado en las proximidades entre objetos. Así, en  $\delta_{ijr}$  del ejemplo, las proximidades son a tres vías.

# Clasificación de los modelos de MDS de acuerdo al número de modos y al número de vías de las disimilaridades

Los diferentes modelos MDS se construyen de acuerdo al número de *modos* y *vías* de las proximidades.

## Modelos de MDS para disimilaridades unimodales a dos vías

En el ejemplo de los vinos, si solo se considera un catador, las disimilaridades serían unimodales y a dos vías. Estos modelos se conocen como *modelos MDS para disimilaridades unimodales a dos vías*.

## Modelos de MDS para disimilaridades bimodales a tres vías

Si se consideran varios catadores, las disimilaridades serán bimodales a tres vías. Los modelos de MDS que tratan con estas disimilaridades se conocen como *modelos de MDS de diferencias individuales*.

# Clasificación de los modelos de MDS de acuerdo al número de modos y al número de vías de las disimilaridades

## Modelos de MDS para disimilaridades bimodales a dos vías

Las disimilaridades bimodales a dos vías, generalmente corresponden a datos de preferencias de  $R$  sujetos sobre  $n$  objetos. El análisis de datos de preferencia, se conocen como *modelos de unfolding*.

## Modelo de Unfolding dada una matriz bimodal a dos vías (preferencias)

- $\mathbf{V} = \{v_1, \dots, v_R\}$  de  $R$  individuos y  $\mathbf{O} = \{o_1, \dots, o_n\}$  de  $n$  objetos y  $\mathbf{S} = (s_{ij})$  ( $R \times n$ ) matriz de preferencias entre  $\mathbf{V}$  y  $\mathbf{O}$ .
- Modelo de distancia:  $\mathbf{A}$  ( $R \times M$ ) y  $\mathbf{B}$  ( $n \times M$ ), donde  $\mathbf{a}_i$ , y  $\mathbf{b}_j$ , representan las coordenadas de  $v_i$  y  $o_j$ .
- **Objetivo:** Encontrar simultáneamente  $\mathbf{A}$ ,  $\mathbf{B}$  tal que para cada  $v_i$  y  $o_j$ ,  $s_{ij} \approx d_{ij}$ , donde  $d_{ij} = d(\mathbf{a}_i, \mathbf{b}_j) = [(\mathbf{a}_i - \mathbf{b}_j)'(\mathbf{a}_i - \mathbf{b}_j)]^{\frac{1}{2}}$

# Transformaciones de las disimilaridades

- En la realidad es irrelevante aproximar las disimilaridades  $\delta_{ij} \approx d(\mathbf{x}_i, \mathbf{x}_j)$ , donde  $\mathbf{x}_i$  y  $\mathbf{x}_j$  representan las filas  $i$  y  $j$  de  $\mathbf{X}$
- Entonces se considera  $f(\delta_{ij})$  y el objetivo es representar  $f(\delta_{ij}) \approx d(\mathbf{x}_i, \mathbf{x}_j)$ , donde  $\mathbf{x}_i$  y  $\mathbf{x}_j$  representan las filas  $i$  y  $j$  de  $\mathbf{X}$ .
- Las  $f(\delta)$  son llamadas *disparidades* y se denotan por  $\hat{d}_{ij}$ .

## Objetivo reformulado del MDS

Dado  $\mathbf{O} = \{o_1, \dots, o_n\}$  y  $\hat{\Delta} = \{\hat{d}_{ij}\}$  obtener una configuración  $\mathbf{X} = (\mathbf{x}'_1, \dots, \mathbf{x}'_n)'$ ,  $\mathbf{x}_i \in \mathbb{R}^M$ ,  $M \ll n$ , tal que  $d(\mathbf{x}_i, \mathbf{x}_j) \approx \hat{d}_{ij}$ .

# Transformaciones de las disimilaridades

## Transformaciones métricas

$\hat{d}_{ij}$  continua y monótona, que toma en cuenta además del orden, las magnitudes de las disimilaridades. Ejemplo: Transformación de razón,  $\hat{d}_{ij} = a\delta_{ij}$ . Transformación de intervalo,  $\hat{d}_{ij} = a + b\delta_{ij}$ .

## Transformaciones no métricas

$\hat{d}_{ij}$  considera únicamente el orden de las disimilaridades, descartando sus valores numéricos. En este caso la transformación  $\hat{d}_{ij}$  puede ser arbitraria y solo cumple la restricción de monotonicidad

$$\delta_{ij} < \delta_{rs} \quad \Rightarrow \quad \hat{d}_{ij} \leq \hat{d}_{rs}.$$

# Principales modelos de MDS de acuerdo al método de estimación de $\mathbf{X}$

- 1 *MDS clásico*. Este modelo trata las disimilaridades directamente como distancias Euclídeas mediante la relación  $d_{ij} = \delta_{ij}$ .  
Mediante el resultado del teorema clasico de MDS se obtiene la representación  $\mathbf{X}$ , usando los valores y vectores propios de la matriz  $\mathbf{\Delta}$  centrada
- 2 *MDS mínimos cuadrados*. Este tipo de modelos obtiene la configuración  $\mathbf{X}$  ajustando mediante mínimos cuadrados las distancias  $\{d_{ij}\}$  a las disimilaridades transformadas  $f(\delta_{ij}) = \hat{d}_{ij}$ .
- 3 *MDS máxima verosimilitud*. Este modelo fue desarrollado por Ramsay (1977) y bajo la hipótesis de lognormalidad de las disimilaridades permite encontrar  $\mathbf{X}$  mediante la maximización de una función de verosimilitud.



## Error de representación

- En el enfoque de mínimos cuadrados, las distancias entre los puntos de la representación  $\mathbf{X}$  se expresan como

$$d_{ij}(\mathbf{X}) = \hat{d}_{ij} + \varepsilon, \quad \forall i, j = 1, \dots, n$$

donde  $\hat{d}_{ij} = f(\delta_{ij})$  representa la función de regresión a estimar, que pueden ser lineales, polinomiales o monótonas y  $\varepsilon$  es el **error de representación**

- En los modelos de MDS, idealmente se desearía que cada disimilaridad o su transformada quedara exactamente relacionada con su correspondiente distancia, lo cual supondría que el error de representación quedaría descartado.
- Sin embargo esta situación rara vez se presenta en la realidad, debido a que los datos empíricos siempre contienen algún componente de error, producto de imprecisiones en las mediciones, efectos de muestreo, etc.

# Modelos de MDS: Mínimos cuadrados

## Error de representación

En los modelos MDS de mínimos cuadrados, se define el **error de representación** para cada par de objetos  $(i, j)$  como

$$e_{ij}^2 = [\hat{d}_{ij} - d_{ij}(\mathbf{X})]^2, \forall i, j = 1, \dots, n.$$

## Sumando el error de representación: STRESS puro

Sumando el **error de representación** sobre  $i$  y  $j$ , se obtiene una medida de *bondad de ajuste* para la representación entera de MDS, el *STRESS puro*

$$\sigma_r^2 = \sigma_r^2(\hat{d}_{ij}, \mathbf{X}) = \sum_{i < j} [\hat{d}_{ij} - d_{ij}(\mathbf{X})]^2.$$

La relación  $i < j$  indica que es suficiente, en general, sumar sobre la mitad de los datos, debido a que las disimilaridades y distancias son simétricas.

## STRESS puro

- En la práctica algunas veces se tienen *valores ausentes*, de tal forma que algunos  $\delta_{ij}$  son indefinidas.
- La existencia de valores ausentes no implica poner restricciones sobre cualquier distancia en  $\mathbf{X}$ .
- Por lo tanto, se definen ponderadores fijos  $w_{ij}$  con valor de 1 si  $\delta_{ij}$  es conocida y  $w_{ij} = 0$  si  $\delta_{ij}$  está ausente. Se pueden permitir otros valores de  $w_{ij}$  siempre y cuando  $w_{ij} \geq 0$ .

$$\sigma_r^2(\hat{d}_{ij}, \mathbf{X}) = \sum_{i < j} w_{ij} [\hat{d}_{ij} - d_{ij}(\mathbf{X})]^2$$

## Minimización del STRESS puro

El objetivo es encontrar  $\hat{d}_{ij}$  y  $\mathbf{X}$  que minimicen

$$\sigma_r^2(\hat{d}_{ij}, \mathbf{X}) = \sum_{i < j} w_{ij} [\hat{d}_{ij} - d_{ij}(\mathbf{X})]^2.$$

- La minimización de  $\sigma_r^2(\hat{d}_{ij}, \mathbf{X})$  es un problema complejo que no puede ser resuelto en forma exacta.
- Los paquetes estadístico implementan algoritmos numéricos de manera iterativa para encontrar  $\hat{d}_{ij}$  y  $\mathbf{X}$ .

## Algoritmos de optimización mas utilizados

- **SMACOF** desarrollado por De Leeuw y Heiser(1980)
- **ALSCAL** desarrollado por Takane(1977). Se aplica en los modelos de MDS no métricos

## STRESS normalizado

- Sin embargo, al optimizar el STRESS puro se pueden obtener las soluciones triviales,  $\mathbf{X} = \mathbf{0}$  y  $\hat{d}_{ij} = 0$ .
- Para evitar la solución trivial de MDS,  $\mathbf{X} = \mathbf{0}$  y  $\hat{d}_{ij} = 0$ , usualmente se imponen condiciones de longitud sobre las disparidades  $\hat{d}_{ij}$ .
- Una forma común es dividir el STRESS puro por la suma de las disparidades al cuadrado, obteniéndose el STRESS normalizado

$$\sigma_n^2(\hat{d}_{ij}, \mathbf{X}) = \frac{\sum_{i < j} w_{ij} [\hat{d}_{ij} - d_{ij}(\mathbf{X})]^2}{\sum_{i < j} w_{ij} \hat{d}_{ij}^2}.$$

## STRESS1 como una medida de ajuste alternativa

El stress normalizado no es la única medida que puede utilizarse para establecer la bondad de ajuste en MDS.

$$STRESS1 = \sigma_1(\hat{d}_{ij}, \mathbf{X}) = \sqrt{\frac{\sum_{i,j} w_{ij} (\hat{d}_{ij} - d_{ij}(\mathbf{X}))^2}{\sum_{i,j} w_{ij} d_{ij}^2(\mathbf{X})}}$$

- La función STRESS1 evita el problema de la dependencia de escalas de las disimilaridades o de sus transformaciones, es decir al dividir entre las distancias euclidianas se eliminan las escalas de las disimilaridades o de las disparidades

## STRESS2

Otra medida de bondad de ajuste es el STRESS2, que se define como

$$STRESS2 = \sigma_2(\hat{d}_{ij}, \mathbf{X}) = \sqrt{\frac{\sum_{i < j} w_{ij} [\hat{d}_{ij} - d_{ij}(\mathbf{X})]^2}{\sum_{i < j} [d_{ij}^2(\mathbf{X}) - \bar{d}]^2}},$$

donde  $\bar{d}$  es el promedio de las distancias euclidianas.

- El STRESS2 se usa para evitar situaciones donde todas las distancias son casi iguales (soluciones degeneradas)

## S-STRESS

El S-STRESS se define como

$$\sigma_S(\hat{d}_{ij}, \mathbf{X}) = \sum_{i < j} (d_{ij}^2(\mathbf{X}) - \hat{d}_{ij}^2)^2$$

- Esta medida de bondad de ajuste es utilizada por el algoritmo ALSCAL de Takane
- La desventaja de esta medida es que tiende a dar soluciones, que dan mayor peso a disparidades grandes
- Mientras que las disparidades pequeñas no son bien representadas



# Diagnóstico de la solución de MDS

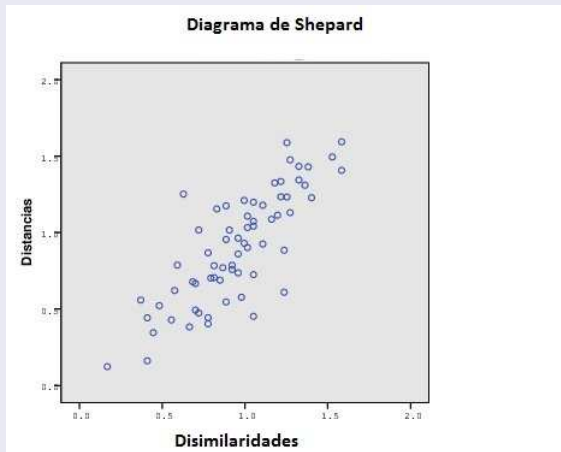
Una vez que se obtiene la configuración solución  $\mathbf{X}$  y las disparidades  $\hat{d}$ , minimizando ya se el *STRESS normalizado*, el *STRESS1* el *STRESS2* o el *S-STRESS*. El siguiente paso es ver que tan buena es la solución, es decir realizar un diagnóstico de la solución MDS.

## Diagrama de Shepard

- Para verificar la calidad de la solución de MDS es aconsejable comparar las distancias de los puntos de la configuración obtenida  $\mathbf{X}$  y las disimilaridades.
- El diagrama de Shepard grafica simultáneamente los pares  $(\delta_{ij}, d_{ij}(\mathbf{X}))$  y  $(\delta_{ij}, \hat{d}_{ij})$
- Se utiliza para inspeccionar simultáneamente los errores de la representación MDS y los errores de la transformación
- También pueden ser usados para detectar outliers.

# Diagnóstico de la solución de MDS

## Diagrama de Shepard

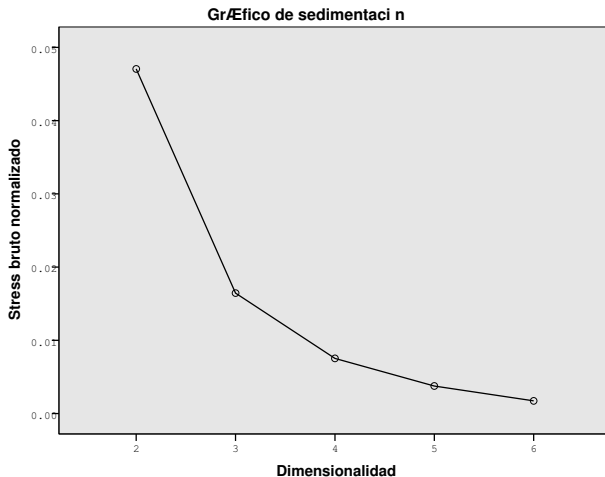


## Elección de la dimensionalidad adecuada de la representación

- Un criterio muy utilizado es calcular soluciones MDS para un rango de dimensiones, por decir entre 2 y 6 dimensiones, y graficar el STRESS vs la dimensión.
- Se utiliza el criterio del codo de la curva, es decir, se elige el número de dimensiones, donde se presenta una *torcedura* de la curva.

# Diagnóstico de la solución de MDS

## Elección de la dimensionalidad adecuada de la representación



# Diagnóstico de la solución de MDS

## Elección de la dimensionalidad adecuada de la representación

- Algunos investigadores sugieren que si el valor del STRESS1 para una dimensionalidad dada es menor de .05, la representación es *muy buena* y por tanto se puede elegir esa dimensionalidad  $M$  para la configuración solución  $X$

Sin embargo el criterio mas importante, es elegir la dimensionalidad que facilite la interpretación del mapa de representación. Este criterio es válido especialmente cuando el MDS es usado como un método exploratorio

# Una aplicación de MDS

**Objetivo:** estimar las ventas de una cadena de farmacias de acuerdo las características de su entorno

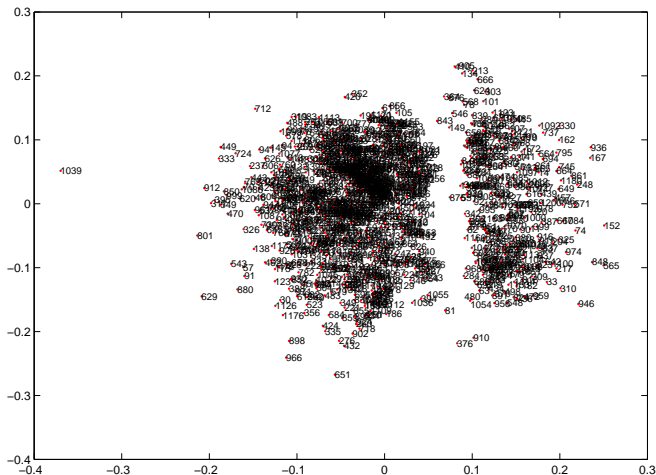
- Se tienen 3000 observaciones multivariadas, las variables consideradas son cuantitativas y cualitativas.
- Todas estas variables se consideran potencialmente como variables predictoras de ventas.
- Es complicado encontrar una relación entre las ventas y las variables cuantitativas y cualitativas.
- Se construyó una medida de disimilaridad para cada par de observaciones mediante el coeficiente de Gower

# Una aplicacion de MDS

- Entonces de  $\Delta_{Gow} \Rightarrow \mathbf{X}$  aplicando el modelo de MDS clasico.
- La dimensión óptima de la solución  $\mathbf{X}$  resultó muy alta, lo cual nos garantizaba tener la mejor aproximación de las disimilaridades a las distancias euclidianas de la configuración  $\mathbf{X}$ .
- Debido a que el objetivo no era la reducción de la dimensionalidad de los datos (2 o 3 dimensiones), sino obtener la mejor aproximación de las distancias euclidianas a las disimilaridades
- Entonces, en este caso se sacrificó la reducción de dimensionalidad por la precisión.
- De esta forma trasladamos el problema del espacio original de variables cuantitativas y cualitativas a un espacio de variables continuas,  $\mathbf{X}$ .

# Una aplicación de MDS

Configuración MDS obtenida a partir de la matriz de disimilaridades de Gower, graficada sobre las primeras dos dimensiones





# Una aplicación de MDS

- Se observa una división de las farmacias en 2 segmentos
- Dentro de cada segmento se construyó un modelo de regresión múltiple, tomando como variables predictoras las coordenadas  $X$  asociadas a las farmacias de cada grupo y como variable dependiente sus ventas correspondientes.
- En cada segmento se seleccionó el mejor modelo de regresión considerando procedimientos de stepwise con validación cruzada.

## **Para estimar las ventas de una nueva farmacia.**

- 1 Se construyen sus coordenadas en el espacio MDS generado por las 3000 farmacias
- 2 Mediante un criterio de clasificación, la farmacia se asigna a uno de los dos grupos formados
- 3 Entonces la farmacia se evalúa en el grupo asignado, con los parámetros del modelo de regresión correspondiente, obteniendo su pronóstico de ventas.