### A picture containing text Description automatically generatedUniversidad de Granada

### Escuela Internacional de Posgrado

### Máster en Estadística Aplicada

### Materia: Análisis de datos. Técnicas aplicadas a datos de proximidad.

### Alumno: Francisco Javier Márquez Rosales

# **Tema 3: MDS métrico y no métrico:**

# **Ejercicios:**

Noviembre, 2022

Ejercicios de Tema 3: MDS métrico y no métrico.

Ejercicio 3.1

Efectuar MDS métrico usando SMACOF par los datos eurodist. Compara los resultados con los obtenidos mediante el procedimiento clásico.

Ejercicio 3.2:

Efectuar un análisis no métrico de los datos eurocitis de la Tabla 1 usando SMACOF. Compara los resultados obtenidos con los de la solución métrica.

**Ejercicio 3.1**

**Efectuar MDS métrico usando SMACOF par los datos eurodist. Compara los resultados con los obtenidos mediante el procedimiento clásico.**

**Solución**

En primer lugar

str(eurodist)

## 'dist' num [1:210] 3313 2963 3175 3339 2762 ...

## - attr(\*, "Size")= num 21

## - attr(\*, "Labels")= chr [1:21] "Athens" "Barcelona" "Brussels" "Calais" ...

summary(eurodist)

## Min. 1st Qu. Median Mean 3rd Qu. Max.

## 158 808 1312 1505 2064 4532

Luego obtenemos las matriz de distancias a partir de los valores iniciales. Por propósitos prácticos se muestra una fracción de los datos incluidos en la matriz (6 ciudades de 21).

eurodist2=as.dist(eurodist)

eurodist2

## Athens Barcelona Brussels Calais Cherbourg Cologne

## Barcelona 3313

## Brussels 2963 1318

## Calais 3175 1326 204

## Cherbourg 3339 1294 583 460

## Cologne 2762 1498 206 409 785

Para poder realizar la comparación, primero se calculan los valores propios mediante la solución clásica .

cmdscale(eurodist2, k = 20, eig = TRUE, add = FALSE, x.ret = FALSE)$eig

## [1] 1.953838e+07 1.185656e+07 1.528844e+06 1.118742e+06 7.893472e+05

## [6] 5.816552e+05 2.623192e+05 1.925976e+05 1.450845e+05 1.079673e+05

## [11] 5.139484e+04 -3.259629e-09 -9.496124e+03 -5.305820e+04 -1.322166e+05

## [16] -2.573360e+05 -3.326719e+05 -5.162523e+05 -9.191491e+05 -1.006504e+06

## [21] -2.251844e+06

Se aprecia que los datos no son distancias Euclídeas.

Ahora, realizamos un análisis de escalamiento multidimensional métrico usando el paquete ‘smacof’.

**library**("smacof")

resm.eurodist2 <- smacofSym(eurodist2,2,)

resm.eurodist2

##

## Call:

## smacofSym(delta = eurodist2, ndim = 2)

##

## Model: Symmetric SMACOF

## Number of objects: 21

## Stress-1 value: 0.072

## Number of iterations: 17

Obtenemos los valores de las configuraciones obtenidas por cada ciudad y de los valores de puntos de estrés .

## Configurations:

## D1 D2

## Athens 1.2139 1.0027

## Barcelona -0.4448 0.3499

## Brussels 0.0343 -0.2296

## Calais -0.0661 -0.2806

## Cherbourg -0.2703 -0.2687

## Cologne 0.1698 -0.2713

## Copenhagen 0.3858 -0.6136

## Geneva -0.0128 0.1806

## Gibraltar -1.1240 0.3928

## Hamburg 0.3276 -0.4260

## Hook of Holland 0.0934 -0.3450

## Lisbon -1.1087 0.0448

## Lyons -0.1034 0.1192

## Madrid -0.7976 0.1827

## Marseilles -0.1683 0.2898

## Milan 0.1541 0.2566

## Munich 0.3295 0.0534

## Paris -0.0949 -0.1354

## Rome 0.3846 0.5830

## Stockholm 0.5136 -0.9996

## Vienna 0.5843 0.1144

## Stress per point (in %):

## Athens Barcelona Brussels Calais Cherbourg

## 13.59 1.57 0.45 1.58 5.31

## Cologne Copenhagen Geneva Gibraltar Hamburg

## 11.15 5.50 11.22 2.14 1.32

## Hook of Holland Lisbon Lyons Madrid Marseilles

## 3.86 7.60 6.87 1.60 1.75

## Milan Munich Paris Rome Stockholm

## 1.66 1.43 0.42 12.47 5.66

## Vienna

## 2.85

Finalmente, obtenemos el gráfico con las dos primeras dimensiones obtenidas a partir del análisis y podemos comprobar como la dispersión resultante encaja con la realidad, sólo bastaría hacer una rotación al superponerla en un mapa.

Chart, scatter chart

Description automatically generated

**Ejercicio 3.2:**

**Efectuar un análisis no métrico de los datos eurocitis de la Tabla 1 usando SMACOF. Compara los resultados obtenidos con los de la solución métrica.**

**Solución**

En primer lugar hacemos lectura de los datos a analizar y sobre los mismos aplicamos el análisis MDS no métrico. Lo primero a tomar en cuenta es que no es necesario la transformación sim2diss() porque ya los datos son disimilaridades. Aplicamos la función ‘mds’ con la opción type=”ordinal” indicando que el análisis es no métrico.

tabla1<-read.csv("tabla1.csv",header = T)

resnm.tabla1=mds(tabla1, ndim=2, type = "ordinal")

resnm.tabla1

##

## Call:

## mds(delta = tabla1, ndim = 2, type = "ordinal")

##

## Model: Symmetric SMACOF

## Number of objects: 10

## Stress-1 value: 0

## Number of iterations: 1

Obtenemos los valores de las configuraciones obtenidas por cada ciudad y de los valores de puntos de estrés.

## Configurations:

## D1 D2

## 1 0.0347 -0.2863

## 2 1.0085 -0.0695

## 3 -1.1181 -0.0847

## 4 -0.8135 0.0927

## 5 -0.0735 -0.0647

## 6 0.2286 -0.1353

## 7 0.4832 0.1508

## 8 0.3756 0.3190

## 9 -0.1395 0.6968

## 10 0.0139 -0.6189

##

##

## Stress per point (in %):

## 1 2 3 4 5 6 7 8 9 10

## 5.92 16.35 18.97 12.45 5.11 5.78 7.85 7.70 10.61 9.26

Como resultado obtenemos la siguiente gráfica.

Chart, scatter chart

Description automatically generated

La cual en forma vemos como es muy parecida al análisis métrico obtenido en el ejercicio 3.1