

Ejercicios Tema 3 - MDS MÉTRICO Y NO MÉTRICO

Cristina Ortega Cartelle

Ejercicio 3.1.

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

MDS clásico

Extraemos primero la dimensión del MDS clásico:

```
data("eurodist")

# MDS clásico con diferentes dimensiones
mds1 <- cmdscale(eurodist, k = 1, eig = TRUE)
mds2 <- cmdscale(eurodist, k = 2, eig = TRUE)
mds3 <- cmdscale(eurodist, k = 3, eig = TRUE)

# Calculo el Goodness of Fit
#
GOF1 <- mds1$GOF[2]
GOF2 <- mds2$GOF[2]
GOF3 <- mds3$GOF[2]

GOF <- data.frame(
  Dimensiones = 1:3,
  GOF = c(GOF1, GOF2, GOF3)
)
GOF
```

```
##   Dimensiones      GOF
## 1           1 0.5401388
## 2           2 0.8679134
## 3           3 0.9101784
```

La medida GOF (Goodness of Fit) indica qué proporción de la variabilidad (o información de las distancias originales) es capturada por la solución en k dimensiones. En este caso, la solución en una dimensión ($K = 1$) explica el 54% de la variación (bajo), mientras que la solución en dos dimensiones ($K = 2$) aumenta la explicación al 87%, y con tres dimensiones ($K = 3$) alcanza el 91%.

Aunque el modelo tridimensional mejora el ajuste, el incremento (de 87% a 91%) es relativamente pequeño comparado con la ganancia entre una y dos dimensiones (de 54% a 87%).

Elegimos $K = 2$ dimensiones porque explica aproximadamente en 87%, conseguimos simplicidad e interpretación visual sin una pérdida sustancial de información.

MDS métrico por SMACOF

Vamos a elegir primero la dimensión en base a los valores de Stress.

```
library(smacof)

dimensiones <- 1:20
estres <- numeric(length(dimensiones))

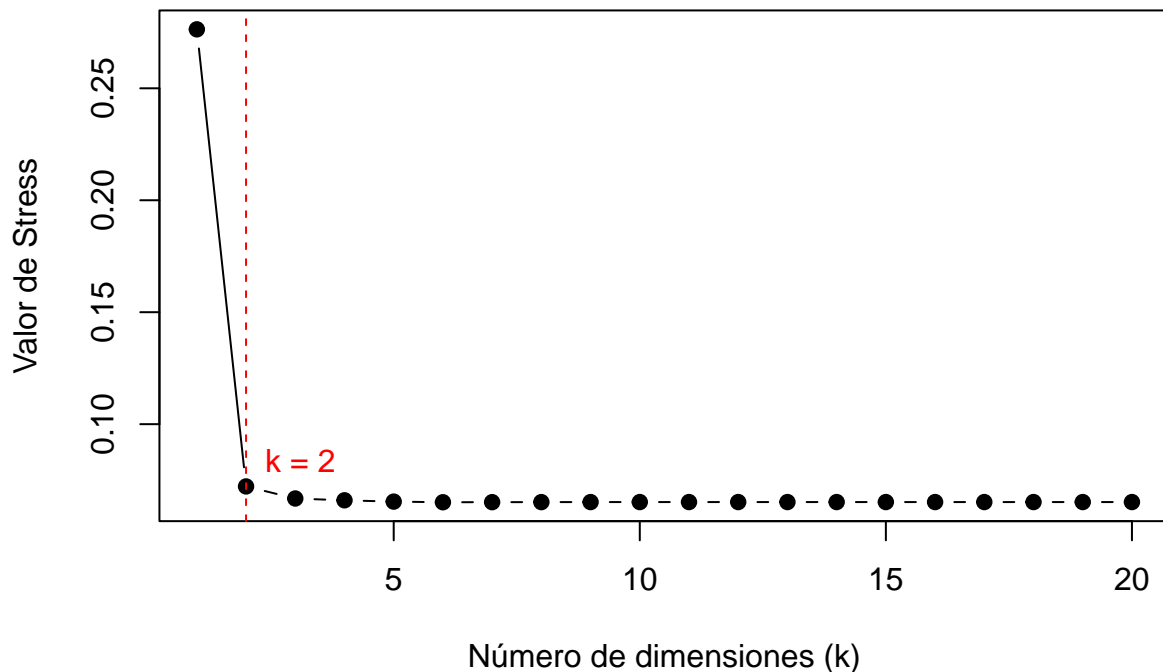
# Calcular el stress para cada número de dimensiones
for (k in dimensiones) {
  mds_smacof <- smacofSym(as.dist(eurodist), ndim = k)
  estres[k] <- mds_smacof$stress
}

data.frame(Dimensiones = dimensiones, Stress = round(estres, 4))
```

##	Dimensiones	Stress
## 1	1	0.2763
## 2	2	0.0722
## 3	3	0.0668
## 4	4	0.0660
## 5	5	0.0654
## 6	6	0.0651
## 7	7	0.0652
## 8	8	0.0652
## 9	9	0.0652
## 10	10	0.0652
## 11	11	0.0652
## 12	12	0.0652
## 13	13	0.0652
## 14	14	0.0652
## 15	15	0.0652
## 16	16	0.0652
## 17	17	0.0652
## 18	18	0.0652
## 19	19	0.0652
## 20	20	0.0652

```
plot(dimensiones, estres, type = "b", pch = 19,
      xlab = "Número de dimensiones (k)",
      ylab = "Valor de Stress",
      main = "Curva de Stress - MDS Métrico (SMACOF)")
abline(v = 2, col = "red", lty = 2)
text(2, estres[2] + 0.01, "k = 2", col = "red", pos = 4)
```

Curva de Stress – MDS Métrico (SMACOF)



Se aplicó el procedimiento SMACOF del escalamiento multidimensional métrico (MDS) para valores de K entre 1 y 10. El valor de stress disminuyó desde 0.273 ($K = 1$) hasta 0.072 ($K = 2$), observándose un cambio pronunciado entre una y dos dimensiones, mientras que a partir de $K = 3$ la reducción se vuelve marginal. Con gráfica, se evidencia un punto de inflexión en $K = 2$.

Se selecciona dimensión $K = 2$.

```
eurodist_smacof <- smacofSym(as.dist(eurodist), ndim = 2)
#Resumen numérico
summary(eurodist_smacof)
```

```
##
## 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
```

Si miramos el stress por punto, podemos observar:

- Los mayores valores de stress por punto corresponden a Atenas(13.59), Roma(12.47) y Ginebra(11.22), lo que sugiere que las relaciones de estas ciudades con el resto presentan mayor distorsión en el espacio bidimensional. Esto es esperable debido a su ubicación periférica respecto al conjunto.
- Las ciudades centrales como Bruselas(0.45), París(0.42), Munich(1.43) o Hamburgo(1.32) presentan unos valores de stress muy bajos, lo cual indica que el modelo bidimensional representa bien sus relaciones espaciales.

Comparativa MDS Clásico con MDS métrico (SMACOF)

Con el análisis de Procrustes podemos comparar dos configuraciones de puntos en el espacio (por ejemplo, los resultados de un MDS métrico y clásico) ajustando rotación, traslación y escala para superponerlas lo mejor posible. Su principal objetivo es medir cuán similares son las estructuras espaciales entre ambas configuraciones. Es útil para evaluar si dos representaciones multidimensionales capturan de manera equivalente las relaciones entre objetos, identificando diferencias globales o locales y cuantificando el ajuste mediante medidas como el error cuadrático medio.

```
library(vegan)
# Ajuste de Procrustes
pro <- procrustes(mds2$points, eurodist_smacof$conf)

#Resumen numérico
summary(pro)

##
## Call:
## procrustes(X = mds2$points, Y = eurodist_smacof$conf)
##
## Number of objects: 21      Number of dimensions: 2
##
## Procrustes sum of squares:
## 161783.4
```

```
## Procrustes root mean squared error:
## 87.77226
## Quantiles of Procrustes errors:
##      Min      1Q      Median      3Q      Max
## 13.55868 39.81629 75.53757 82.66620 224.13872
##
## Rotation matrix:
##      [,1]      [,2]
## [1,] 0.999999852 -0.000544258
## [2,] 0.000544258 0.999999852
##
## Translation of averages:
##      [,1]      [,2]
## [1,] -3.835051e-13 -6.446124e-13
##
## Scaling of target:
## [1] 1771.914
```

El análisis de Procrustes entre las configuraciones del MDS clásico y del MDS métrico (SMACOF) revela una rotación casi nula (matriz de rotación \approx identidad).

La traslación media fue prácticamente cero indica una correspondencia casi perfecta entre ambas soluciones. Por lo que, ambas representaciones describen la misma estructura espacial de distancias.

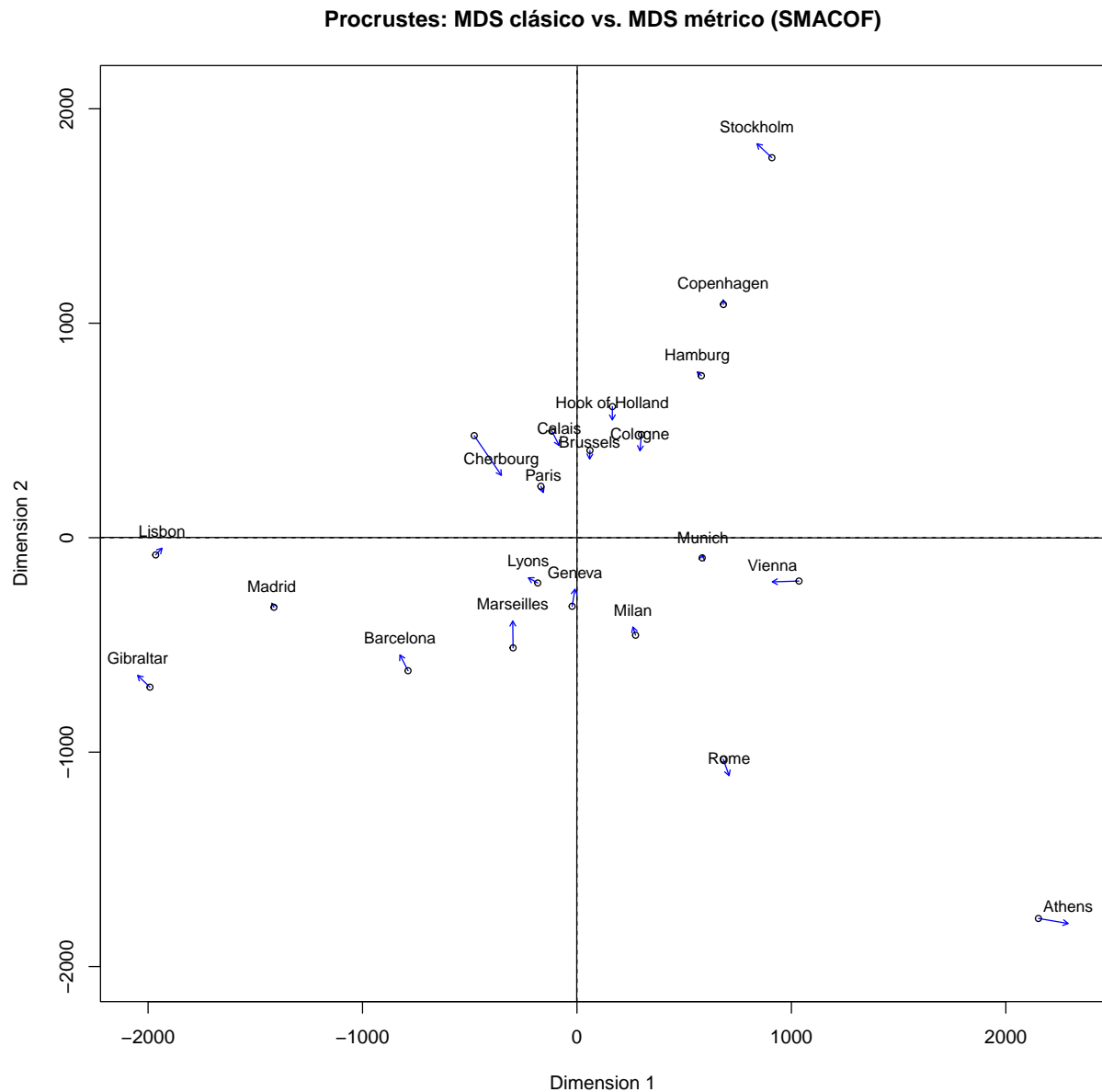
Representación comparativa en 2 dimensiones

Con el siguiente gráfico, se dibuja los puntos del primer conjunto, en este caso el MDS clásico. Las flechas, apuntan hacia las posiciones correspondientes del segundo conjunto, el MDS métrico (SMACOF), después del alineamiento.

Así podemos ver cuánto se movió cada punto entre configuraciones.

```
#rotacion del eje y para la representación
pro$X[,2] <- -pro$X[,2]
pro$Yrot[,2] <- - pro$Yrot[,2]
```

```
plot(pro, kind = 1, main = "Procrustes: MDS clásico vs. MDS métrico (SMACOF)")
text(pro$X[,1], pro$X[,2], labels = rownames(pro$X), cex = 0.8, pos = 3, col = "black")
```



Las flechas son muy cortas y orientadas en direcciones aleatorias, ambas configuraciones son muy parecidas.

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.

Análisis No Métrico usando SMACOF

Como hicimos para el análisis métrico anterior, vamos a elegir la dimensión en base a los valores de Stress.

```
ciudades <- matrix(c(
0,569,667,530,141,140,357,396,570,190,
569,0,1212,1043,617,446,325,423,787,648,
```

```

667,1212,0,201,596,768,923,882,714,714,
530,1043,201,0,431,608,740,690,516,622,
141,617,596,431,0,177,340,337,436,320,
140,446,768,608,177,0,218,272,519,302,
357,325,923,740,340,218,0,114,472,514,
396,423,882,690,337,272,114,0,364,573,
569,787,714,516,436,519,472,364,0,755,
190,648,714,622,320,302,514,573,755,0),
nrow = 10, byrow = TRUE
)
rownames(ciudades) <- colnames(ciudades) <- c("Londres","Estocolmo","Lisboa","Madrid",
"Paris","Amsterdam","Berlin","Praga",
"Roma","Dublin")

eurociti <- as.dist(ciudades) #Lo ponemos en la clase dist.

dimensiones <- 1:9
estres <- numeric(length(dimensiones))

# Calcular el stress para cada número de dimensiones
for (k in dimensiones) {
  mds_smacof_nometrico <- smacofSym(eurociti, ndim = k, type = "ordinal")
  estres[k] <- mds_smacof_nometrico$stress
}

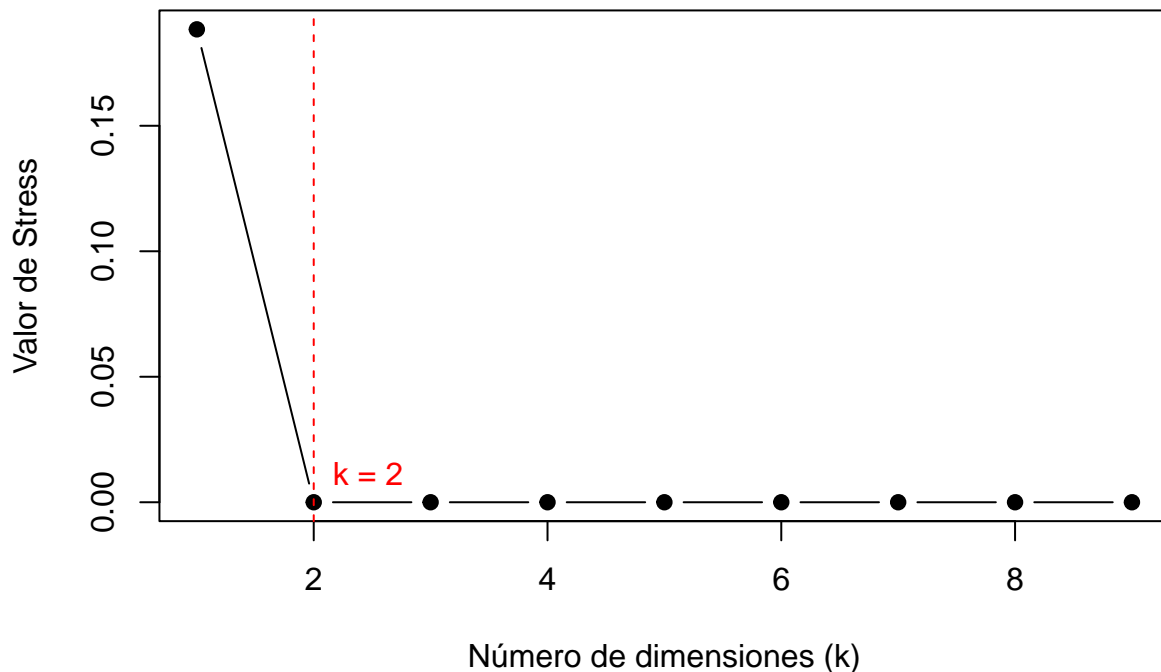
data.frame(Dimensiones = dimensiones, Stress = round(estres,8))

##   Dimensiones   Stress
## 1           1 0.18841001
## 2           2 0.00000009
## 3           3 0.00000011
## 4           4 0.00000016
## 5           5 0.00000020
## 6           6 0.00000020
## 7           7 0.00000020
## 8           8 0.00000020
## 9           9 0.00000020

plot(dimensiones, estres, type = "b", pch = 19,
     xlab = "Número de dimensiones (k)",
     ylab = "Valor de Stress",
     main = "Curva de Stress - MDS No Métrico (SMACOF)")
abline(v = 2, col = "red", lty = 2)
text(2, estres[2] + 0.01, "k = 2", col = "red", pos = 4)

```

Curva de Stress – MDS No Métrico (SMACOF)



El valor de stress disminuyó desde 0.188 para $K = 1$ hasta prácticamente 0 para $K = 2$, observándose un cambio pronunciado entre una y dos dimensiones, mientras que a partir de $K = 3$, el stress aumenta ligerísimamente.

Por ello, elegimos la solución para dos dimensiones ($K = 2$).

```
mds_nometrico <- smacofSym(eurociti, ndim=2, type = "ordinal") # ordinal = no métrico
#Resumen numérico
summary(mds_nometrico)
```

```
##
## Configurations:
##           D1      D2
## Londres    0.0347 -0.2863
## Estocolmo  1.0085 -0.0695
## Lisboa    -1.1181 -0.0847
## Madrid    -0.8135  0.0927
## Paris     -0.0735 -0.0647
## Amsterdam  0.2286 -0.1353
## Berlin     0.4832  0.1508
## Praga      0.3756  0.3190
## Roma      -0.1395  0.6968
## Dublin     0.0139 -0.6189
##
##
## Stress per point (in %):
##   Londres Estocolmo   Lisboa   Madrid   Paris Amsterdam   Berlin   Praga
```



```
##      5.92      16.35      18.97      12.45      5.11      5.78      7.85      7.70
##      Roma      Dublin
##     10.61      9.26
```

Análisis métrico con SMACOF

```
dimensiones <- 1:9
estres <- numeric(length(dimensiones))

# Calcular el stress para cada número de dimensiones
for (k in dimensiones) {
  mds_smacof_metrico <- smacofSym(eurociti, ndim = k)
  estres[k] <- mds_smacof_metrico$stress
}

data.frame(Dimensiones = dimensiones, Stress = round(estres,4))
```

```
##   Dimensiones Stress
## 1             1 0.2569
## 2             2 0.0004
## 3             3 0.0005
## 4             4 0.0005
## 5             5 0.0006
## 6             6 0.0006
## 7             7 0.0006
## 8             8 0.0006
## 9             9 0.0006
```

El valor de stress disminuyó desde 0.257 para $K = 1$ hasta 0.0004 para $K = 2$, observándose un cambio pronunciado entre una y dos dimensiones, mientras que a partir de $K = 3$, el stress aumenta ligerisimamente.

Por ello, volvemos a elegir la solución para dos dimensiones ($K = 2$).

```
mds_metrico <- smacofSym(eurociti, ndim=2)
#Resumen numérico
summary(mds_metrico)
```

```
##
## Configurations:
##           D1      D2
## Londres    0.0347 -0.2863
## Estocolmo  1.0085 -0.0695
## Lisboa    -1.1181 -0.0847
## Madrid    -0.8135  0.0927
## Paris     -0.0735 -0.0647
## Amsterdam  0.2286 -0.1353
## Berlin     0.4832  0.1508
## Praga      0.3756  0.3190
## Roma      -0.1395  0.6968
## Dublin     0.0139 -0.6189
##
##
## Stress per point (in %):
##   Londres Estocolmo   Lisboa   Madrid   Paris Amsterdam   Berlin   Praga
```

##	9.44	11.56	14.95	7.21	14.80	11.34	9.30	7.25
##	Roma	Dublin						
##	6.27	7.88						

Comparativa MDS Métrico vs No Métrico

El MDS métrico y el no métrico se diferencian en el tipo de información que intentan preservar y, por tanto, en su interpretación. El MDS métrico busca reproducir lo mejor posible las distancias originales, manteniendo sus magnitudes relativas; es adecuado cuando las distancias tienen un significado cuantitativo claro, como los kilómetros entre ciudades. El MDS no métrico, en cambio, sólo intenta conservar el orden de las distancias, permitiendo una transformación monótona que hace el modelo más flexible y capaz de adaptarse a datos con estructuras no lineales.

Stress Global

En primer lugar, el Stress global en no métrico era menor que en métrico. Lo cual es de esperar, por su mayor flexibilidad.

Stress por punto

Si comparamos el Stress por ciudad:

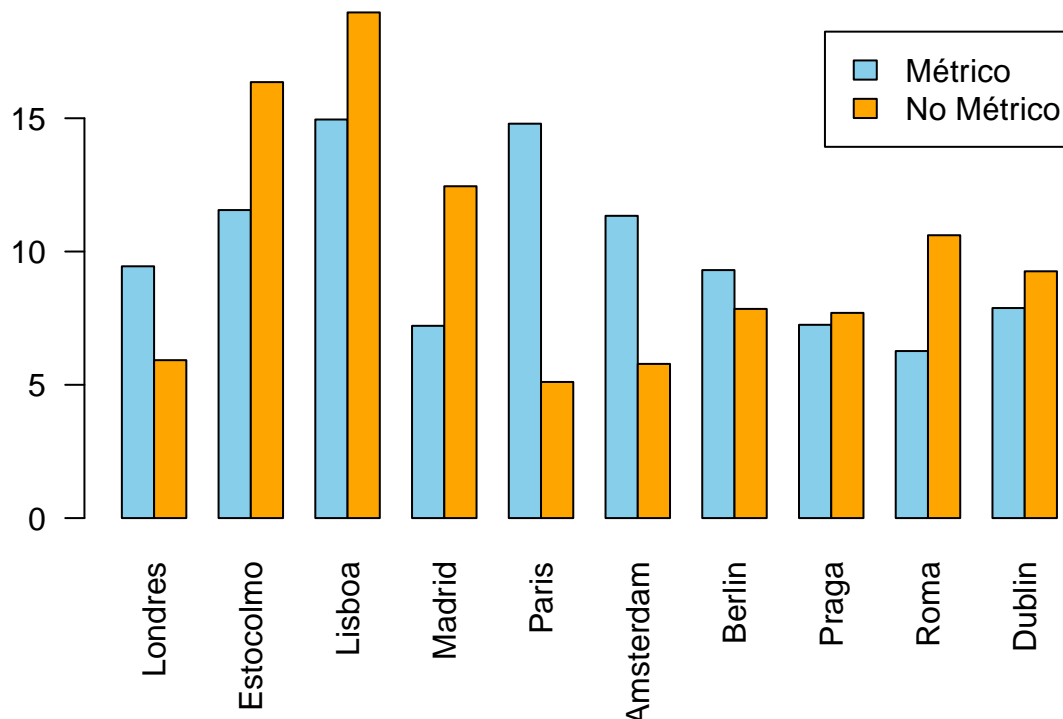
```
# Extraer stress por punto
stress_metrico <- mds_metrico$spp
stress_nometrico <- mds_nometrico$spp

# comparar
stress_comparacion <- data.frame(
  Stress_Metrico = stress_metrico,
  Stress_NoMetrico = stress_nometrico
)
print(stress_comparacion)
```

##	Stress_Metrico	Stress_NoMetrico
## Londres	9.444833	5.923984
## Estocolmo	11.555536	16.354338
## Lisboa	14.952710	18.969049
## Madrid	7.212556	12.448288
## Paris	14.795558	5.106493
## Amsterdam	11.338564	5.784196
## Berlin	9.303427	7.846975
## Praga	7.250955	7.698002
## Roma	6.265980	10.610525
## Dublin	7.879880	9.258152

```
barplot(
  t(as.matrix(stress_comparacion)),
  beside = TRUE,
  names.arg = stress_comparacion$Ciudad,
  col = c("skyblue", "orange"),
  legend.text = c("Métrico", "No Métrico"),
  main = "Comparación de Stress por Ciudad",
  las = 2
)
```

Comparación de Stress por Ciudad



Ciudades bien representadas en ambos modelos

Berlin, Praga, Dublin: el stress es similar en métrico y no métrico ($\approx 7-9$). Esto indica que la representación de estas ciudades es robusta, independientemente del tipo de MDS.

Ciudades que mejoran en MDS no métrico

- París: Stress cae de 14.8 (métrico) a 5.1 (no métrico)
- Ámsterdam: Stress cae de 11.3 a 5.8
- Londres: Stress cae de 9.4 a 5.9

Esto indica que estas ciudades se representan mucho mejor si solo se preserva el orden de las distancias, es decir, la estructura global es no lineal y el MDS métrico no logra reproducirla correctamente.

Ciudades que empeoran en MDS no métrico

- Estocolmo: Stress sube de 11.5 a 16.3
- Lisboa: Stress sube de 15 a 19
- Madrid y Roma: también aumentan también de 7.2 y 6.26 a 12.4 y 10.6 respectivamente

Estas ciudades son periféricas o extremas, y el MDS no métrico las fuerza a ajustarse al orden de proximidades, lo que peor reproduce sus distancias absolutas.

```
# Ajuste de Procrustes
pro2 <- procrustes(mds_métrico$conf, mds_nométrico$conf)

#Resumen numérico
summary(pro2)
```

```
##
```

```
## Call:
## procrustes(X = mds_metrico$conf, Y = mds_nometrico$conf)
##
## Number of objects: 10      Number of dimensions: 2
##
## Procrustes sum of squares:
## 1.776357e-15
## Procrustes root mean squared error:
## 1.3328e-08
## Quantiles of Procrustes errors:
##      Min      1Q  Median      3Q      Max
##       0       0       0       0       0
##
## Rotation matrix:
##      [,1]      [,2]
## [1,] 1 8.470329e-22
## [2,] 0 1.000000e+00
##
## Translation of averages:
##      [,1] [,2]
## [1,] 0 0
##
## Scaling of target:
## [1] 1
```

```
residuals(pro2)
```

```
##      Londres Estocolmo      Lisboa      Madrid      Paris Amsterdam      Berlin      Praga
##           0           0           0           0           0           0           0           0
##           Roma      Dublin
##           0           0
```

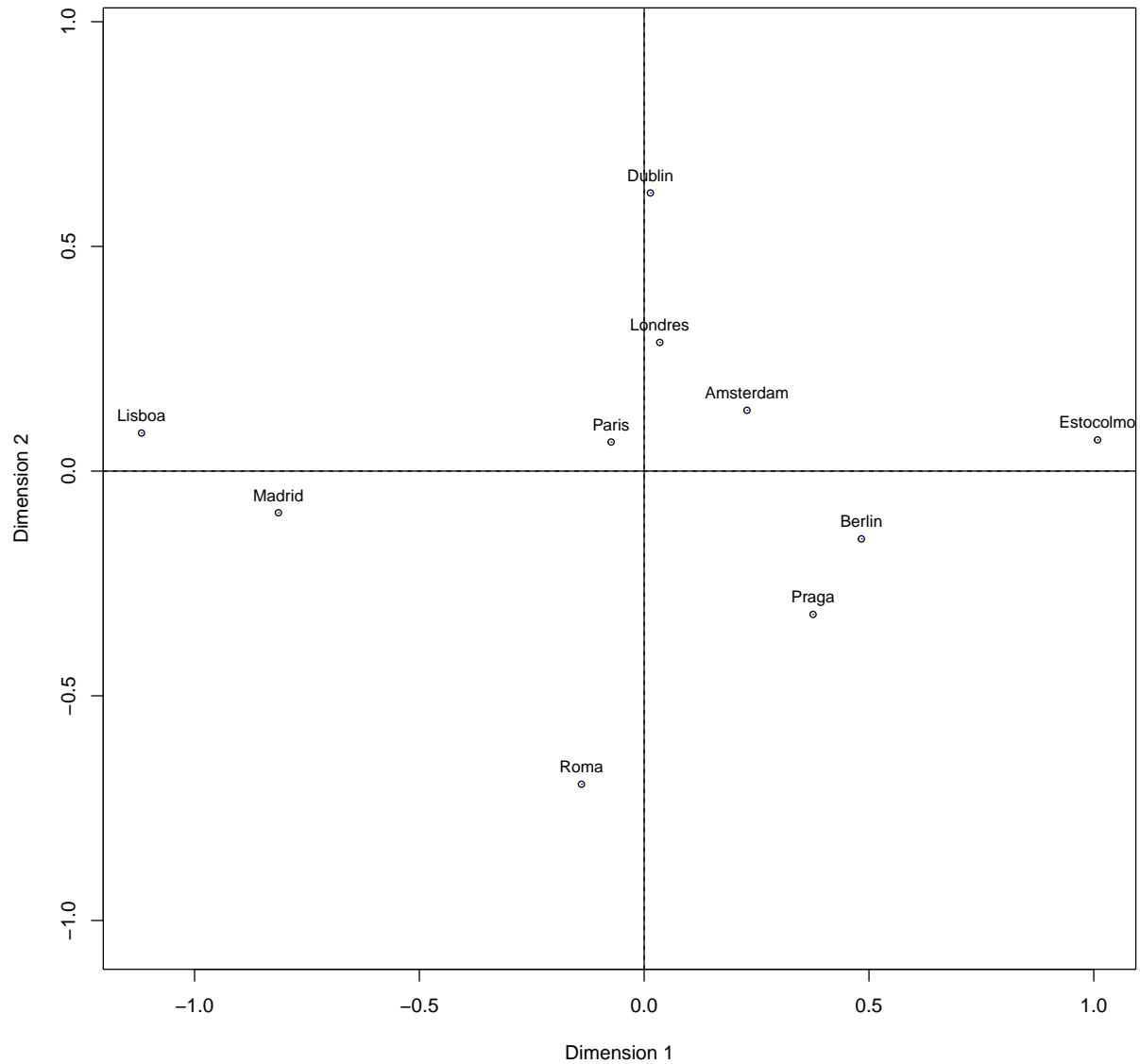
El análisis Procrustes muestra que las configuraciones de coordenadas del MDS métrico y no métrico son prácticamente idénticas en 2 dimensiones. Los valores de sum of squares y RMSE son prácticamente cero, y todos los errores por ciudad son nulos, lo que indica que no fue necesaria ninguna rotación, traslación o escalado significativo. Esto significa que, aunque el stress por punto varía para algunas ciudades, la estructura global del mapa se mantiene, y las diferencias de stress reflejan cómo cada modelo ajusta localmente las distancias, pero no altera la disposición general de las ciudades.

```
#rotacion del eje y para la representación
```

```
pro2$X[,2] <- -pro2$X[,2]
pro2$Yrot[,2] <- - pro2$Yrot[,2]
```

```
plot(pro2, kind = 1, main = "Procrustes: MDS Métrico vs. No métrico (SMACOF)")
text(pro2$X[,1], pro2$X[,2], labels = rownames(pro2$X), cex = 0.8, pos = 3, col = "black")
```

Procrustes: MDS Métrico vs. No métrico (SMACOF)



Los warnings que aparecían en la salida R **Warning: zero-length arrow is of indeterminate angle and so skipped** (y que ahora no se porqué no aparecen en el pdf) no indican un error real; reflejan que las configuraciones de MDS métrico y no métrico son prácticamente idénticas y muchas flechas tienen longitud cero, por lo que no se pueden dibujar.

Como dijimos, las configuraciones de MDS métrico y no métrico son prácticamente superponibles, lo que indica que la disposición relativa de las ciudades se conserva.