

ESCALAMIENTO MULTIDIMENSIONAL

Saraí Campos Varela

2022-06-07

Introducción.

Dentro de las técnicas multivariantes podemos citar al Escalamiento Multidimensional (EMD) (Multidimensional Scaling, MDS por sus siglas en inglés). El MDS es una técnica multivariante de interdependencia que trata de representar en un espacio geométrico de pocas dimensiones las proximidades existentes entre un conjunto de objetos o de estímulos. Esta técnica, aunque tiene sus raíces a principios del siglo XX, hoy día sigue siendo infrautilizada en muchas áreas.

En este trabajo se pretende dar una visión general del funcionamiento del Escalamiento Multidimensional, tomando de la web una base de datos que muestra las distancias entre ciudades de México por carretera. La información con la que contamos son 15 distintas ciudades con su respectiva distancia en kilómetros entre una y otra, es así como se obtiene una matriz cuadrada de 15 filas y 15 columnas.

Base de datos.

Lamentablemente se presentaron algunos problemas al momento de cargar el excel que descargamos, se intentó de distintas maneras, de primer intento se agregó manualmente en el excel la diagonal de 0 en la base, de segundo intento se modificó la matriz a “double” sin embargo no se nos permitió esa opción. Por último ya encontrándome en estado de desesperación opté por cargar y armar la matriz manualmente si es que es correcto llamarlo así.

1.- Crear los vectores de la matriz.

En este caso, apoyándonos con el excel descargado se crearon los vectores de cada ciudad tomando las filas y abreviamos el nombre de las ciudades para no tener problemas.

```
CDM<-c(0,531,1790,2743,540,898,1429,1306,402,484,1884,834,2576,1216,378)
GDJR<-c(532,0,1526,2215,10,787,1165,1840,328,220,1356,698,2047,687,859)
CJ<-c(1806,1535,0,1164,1543,1349,355,3100,1384,1312,847,1092,1007,1088,2181)
TJN<-c(2747,2218,1164,0,2214,2506,1354,4055,2383,2311,863,2091,175,1553,3074)
ZPP<-c(540,9,1533,2209,0,795,1173,1848,336,228,1351,706,2042,682,867)
MTR<-c(906,795,1169,2326,804,0,803,1986,512,573,1709,88,2169,1040,1281)
CHUA<-c(1443,1172,353,1353,1180,809,0,2737,1021,949,696,729,1196,736,1818)
MRD<-c(1316,1849,3099,4061,1857,1991,2738,0,1711,1793,3202,2143,3893,2533,1574)
SLP<-c(408,328,1378,2378,336,511,1017,1702,0,165,1602,447,2221,934,783)
AGCL<-c(491,219,1306,2306,228,567,945,1785,166,0,1572,478,2149,904,866)
HMSLL<-c(1885,1356,847,860,1351,1713,697,3193,1681,1572,0,1633,693,691,2212)
STLL<-c(841,705,1085,2085,714,87,724,2135,448,483,1631,0,1929,962,1216)
MXC<-c(2578,2049,1007,169,2044,2349,1198,3885,2227,2155,694,1935,0,1384,2904)
```

```
CUL<-c(1230,701,1086,1550,696,1058,736,2537,1026,917,691,978,1382,0,1557)
ACJ<-c(378,859,2172,3070,867,1280,1812,1566,784,866,2212,1217,2903,1543,0)
```

2.- Crear la matriz a partir de los vectores.

Teniendo listos los vectores creamos la matriz para que vaya teniendo dimensión, en este paso ocupamos función “matrix” en donde agregamos los vectores y el numero de filas y columnas que es 15.

```
datos <- matrix(c(CDM,GDJR,CJ,TJN,ZPP,MTR,CHUA,MRD,SLP,AGCL,HMSLL,STLL,MXC,CUL,
ACJ), nrow = 15, ncol = 15 )
```

3.- Agregar nombre a las columnas.

Por último agregamos el nombre a cada una de las 15 columnas y posteriormente a cada una de las 15 filas.

```
colnames(datos) <-c("Ciudad de México","Guadalajara","Ciudad Juárez","Tijuana",
"Zapopan","Monterrey","Chihuahua","Mérida","San Luis Potosí",
"Aguascalientes","Hermosillo","Saltillo","Mexicali","Culiacán",
"Acapulco de Juárez")
```

```
rownames(datos) <-c("Ciudad de México","Guadalajara","Ciudad Juárez","Tijuana",
"Zapopan","Monterrey","Chihuahua","Mérida",
"San Luis Potosí","Aguascalientes","Hermosillo",
"Saltillo","Mexicali","Culiacán","Acapulco de Juárez")
```

4.- Visualizamos la base.

Aquí presento estas dos opciones donde “view” nos presetará la tabla en una ventana emergente de R. Y por otro lado, al llamar “datos” nos los presetará en la consola.

```
View (datos)
```

Como observamos nuestra matriz o base de datos esta guardada en el objeto *datos*.

```
datos
```

```
##          Ciudad de México Guadalajara Ciudad Juárez Tijuana Zapopan
## Ciudad de México          0          532          1806      2747      540
## Guadalajara             531           0          1535      2218         9
## Ciudad Juárez          1790          1526           0       1164      1533
## Tijuana                 2743          2215          1164         0      2209
## Zapopan                540           10          1543      2214         0
## Monterrey               898           787          1349      2506       795
## Chihuahua              1429          1165           355      1354      1173
## Mérida                  1306          1840          3100      4055      1848
## San Luis Potosí         402           328          1384      2383       336
## Aguascalientes         484           220          1312      2311       228
## Hermosillo             1884          1356           847       863      1351
## Saltillo                834           698          1092      2091       706
```

```
## Mexicali                2576        2047        1007        175        2042
## Culiacán                1216         687        1088       1553        682
## Acapulco de Juárez      378         859        2181       3074        867
##
## Monterrey Chihuahua Mérida San Luis Potosí Aguascalientes
## Ciudad de México        906        1443        1316         408         491
## Guadalajara            795        1172        1849         328         219
## Ciudad Juárez          1169         353        3099        1378        1306
## Tijuana                2326        1353        4061        2378        2306
## Zapopan               804         1180        1857         336         228
## Monterrey              0          809        1991         511         567
## Chihuahua              803          0        2738        1017         945
## Mérida                 1986        2737          0        1702        1785
## San Luis Potosí        512        1021        1711          0         166
## Aguascalientes        573         949        1793         165          0
## Hermosillo            1709         696        3202        1602        1572
## Saltillo               88         729        2143         447         478
## Mexicali              2169        1196        3893        2221        2149
## Culiacán              1040         736        2533         934         904
## Acapulco de Juárez    1281        1818        1574         783         866
##
## Hermosillo Saltillo Mexicali Culiacán Acapulco de Juárez
## Ciudad de México      1885         841        2578        1230         378
## Guadalajara          1356         705        2049         701         859
## Ciudad Juárez         847        1085        1007        1086        2172
## Tijuana               860        2085         169        1550        3070
## Zapopan             1351         714        2044         696         867
## Monterrey            1713          87        2349        1058        1280
## Chihuahua            697         724        1198         736        1812
## Mérida               3193        2135        3885        2537        1566
## San Luis Potosí      1681         448        2227        1026         784
## Aguascalientes      1572         483        2155         917         866
## Hermosillo           0         1631         694         691        2212
## Saltillo             1633          0        1935         978        1217
## Mexicali             693        1929          0        1382        2903
## Culiacán             691         962        1384          0        1543
## Acapulco de Juárez   2212        1216        2904        1557          0
```

5.- Transformamos la base de datos.

En este para poder realizar el análisis es necesario transformar nuestra base a una matriz.

```
datos<-as.matrix(datos)
```

Exploración de la matriz.

1.- Dimensión.

Como sabemos nuestra matriz es cuadrada cuenta con el mismo numero de filas y columnas.

```
dim(datos)
```

```
## [1] 15 15
```

2.- Variables.

Muestra la estructura interna de la matriz.

```
str(datos)

##  num [1:15, 1:15] 0 531 1790 2743 540 ...
##  - attr(*, "dimnames")=List of 2
##    ..$ : chr [1:15] "Ciudad de México" "Guadalajara" "Ciudad Juárez" "Tijuana" ...
##    ..$ : chr [1:15] "Ciudad de México" "Guadalajara" "Ciudad Juárez" "Tijuana" ...
```

3.- Nombre de columnas.

Arroja el nombre de las 15 ciudades ubicadas en las columnas.

```
colnames(datos)

##  [1] "Ciudad de México"  "Guadalajara"      "Ciudad Juárez"
##  [4] "Tijuana"           "Zapopan"         "Monterrey"
##  [7] "Chihuahua"         "Mérida"           "San Luis Potosí"
## [10] "Aguascalientes"   "Hermosillo"       "Saltillo"
## [13] "Mexicali"         "Culiacán"         "Acapulco de Juárez"
```

4.- Datos perdidos.

No tenemos datos perdidos en la matriz, por tanto, continuamos con el análisis.

```
anyNA(datos)
```

```
## [1] FALSE
```

5.- Extracción de las filas de la matriz.

Contamos con 15L, lo cual es el número de ciudades.

```
n<-nrow(datos)
```

Escalado multidimensional clásico.

1.- Cálculo de autovalores.

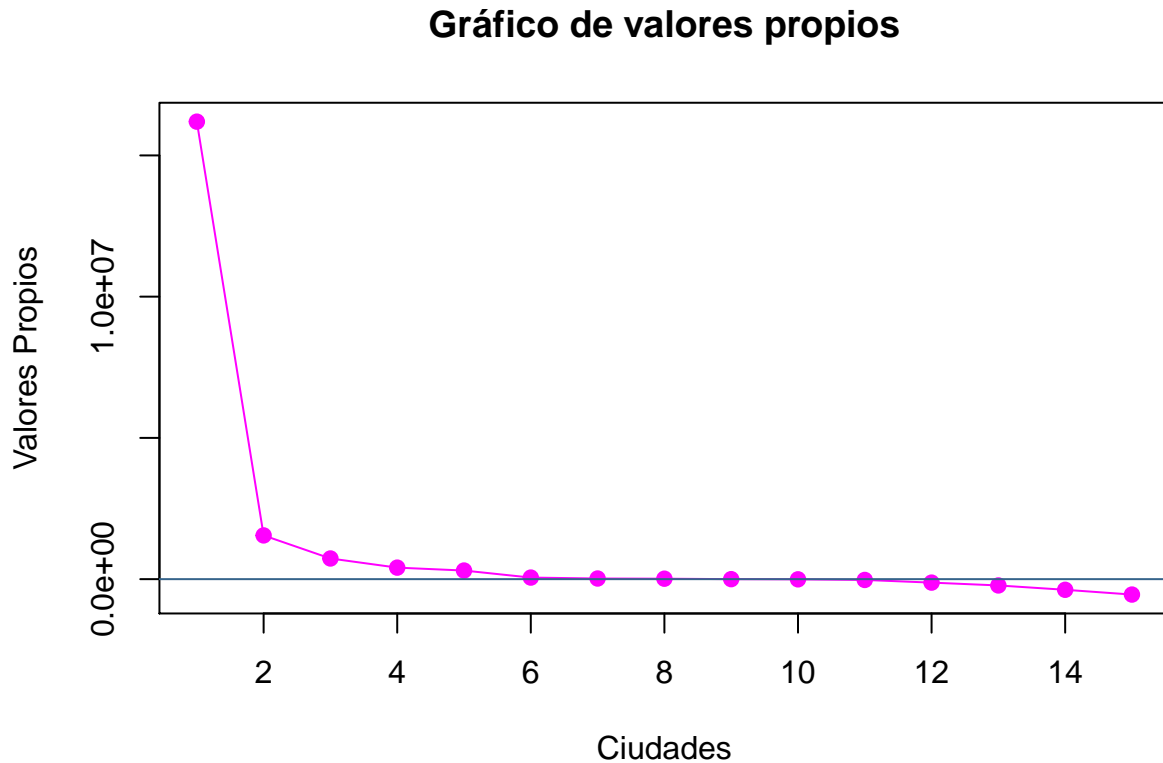
En este objeto **mds.ciudades** encontramos el cálculo del escalado multidimensional (el cual es calculado con la función **cmdscale**) junto con el cálculo de los valores propios.

```
mds.ciudades <- cmdscale(datos, eig = TRUE)
```

2.- Generación del gráfico.

Se gráficaran los valores propios del resultado del escalado multidimensional.

```
plot(mds.ciudades$eig, pch=19, col="magenta", main = "Gráfico de valores propios",  
      xlab="Ciudades", ylab="Valores Propios",  
      type="o")  
abline(a=0, b=0, col="steelblue4")
```



Interpretacion: Se identifican valores propios negativos y por lo tanto, se considera como solución el seleccionar coordenadas principales ($r=2$), lo que son las primeras dos ciudades.

3.- Medidas de precision.

```
m<-sum(abs(mds.ciudades$eig[1:2]))/sum(abs(mds.ciudades$eig))
```

4.- Obtencion de coordenadas principales.

Fijando $k=2$ que son nuestras coordenadas principales y de igual manera se realiza con los dos primeros autovalores.

```
mds.cities<-cmdscale(datos, eig=TRUE, k=2)
```

5.- Separación de columnas en X1 y X2.

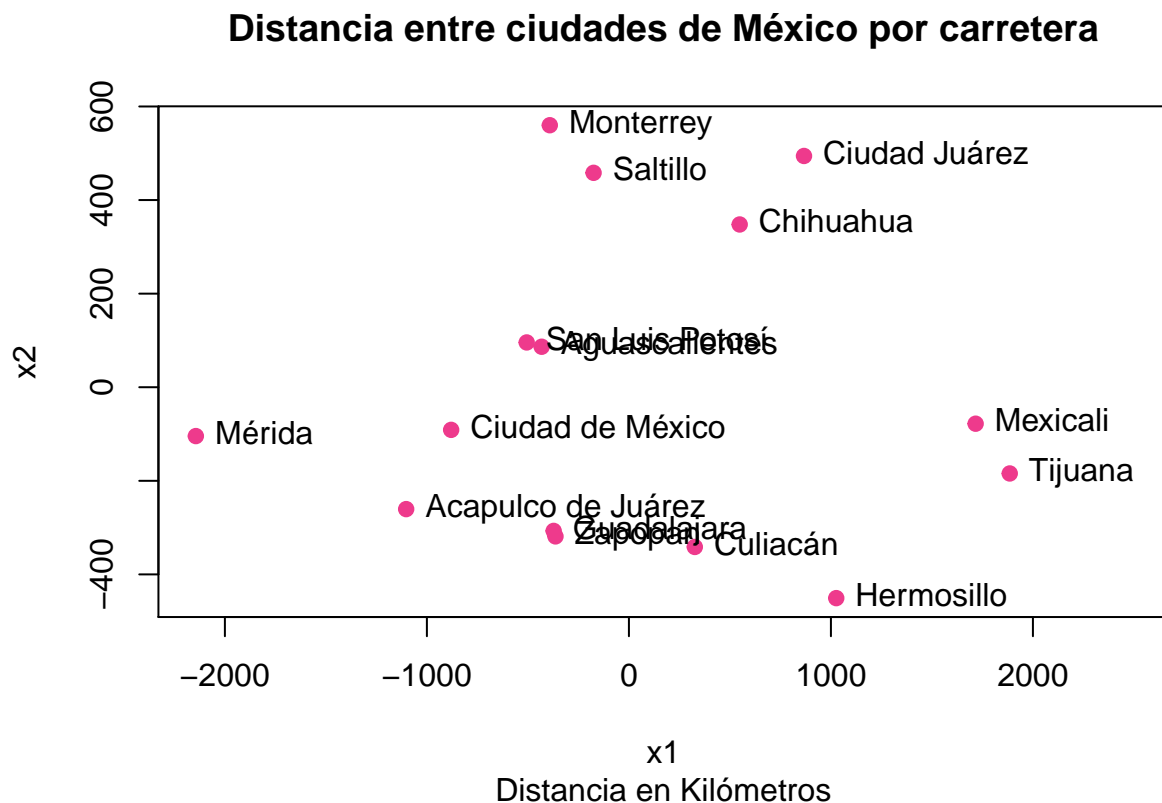
```
x1<-mds.cities$points[,1]
```

```
x2<-mds.cities$points[,2]
```

6.- Generación del gráfico en dos dimensiones.

Este paso graficará los datos con las coordenadas obtenidas.

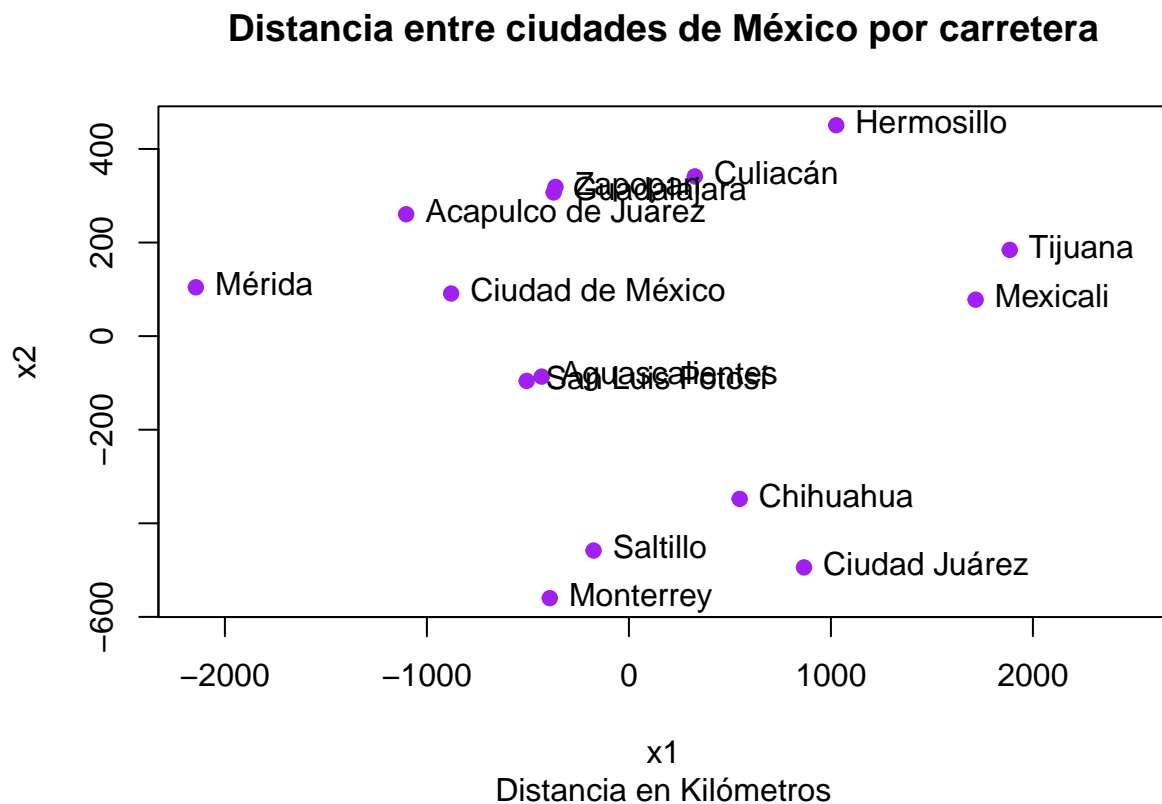
```
plot(x1,x2,pch=19, col="violetred2",main = "Distancia entre ciudades de México por carretera",
     sub = "Distancia en Kilómetros", xlim = range(x1)+c(0,600))
text(x1,x2, pos=4, labels = rownames(datos),
     col="black")
```



7.- Rotación del gráfico y visualización.

```
x2<--x2

plot<-plot(x1,x2,pch=19, col="purple", main = "Distancia entre ciudades de México por carretera",
          sub = "Distancia en Kilómetros", xlim = range(x1)+c(0,600))
text(x1,x2, pos=4, labels = rownames(datos),
     col="black")
```



Resultados.

Es fácil notar las distancias gracias al gráfico elaborado. Notamos que la ciudad San Luis Potosí está muy cerca de Aguascalientes a 165km, así como Zapopan se encuentra muy cerca de Guadalajara, a tan solo 9km de distancia. La Ciudad de México podría considerarse el centro de todas ellas y que Mexicali sería considerada la ciudad más lejana a ella con 2578km. La ciudad más lejana una con otra es Mérida y Tijuana con 4061km de distancia una entre otra. Este método es demasiado sencillo y dinámico, es bastante sencillo identificar las distancias entre las ciudades y en caso de tomar decisiones al respecto, también puede facilitarnos demasiado el trabajo.

Referencias

Base de datos tomada en: <https://www.mejoresrutas.com/tabla-de-distancias-entre-ciudades/mx/>

Información y conceptualización en: <http://eio.usc.es>