

Escalado multidimensional no-métrico.



Introducción.

El objetivo del escalado multidimensional es analizar una matriz de **proximidades** entre los casos en estudio, obtenida a partir de la percepción de diversos encuestados; para obtener gráficamente la estructura interna de dicha matriz. Este análisis incluye tanto la determinación del **número de variables o dimensiones** que los individuos han tenido en cuenta en sus procesos mentales; como lo cerca o lejos que se hallan los casos a partir de las opiniones de tales encuestados (coordenadas). Preferentemente, se busca que el número de las dimensiones elegidas sea 2, para poder hacer así un gráfico bidimensional de sencilla interpretación.

Cuando las proximidades entre casos no son muy precisas (no están en escala métrica), el escalado multidimensional será de tipo no-métrico. Este tipo de escalado suele aplicarse a las ciencias sociales. A su vez, en el escalado multidimensional no-métrico, es habitual que la matriz de proximidades no se obtenga directamente; sino a partir de la valoración que hacen los encuestados de cada caso dentro de una escala tipo *Likert*. Tras tener todas las valoraciones de todos los casos realizadas por todos los individuos en esta escala, se calculará la matriz de proximidades aplicando algún tipo de medida de distancia, como la *distancia euclídea*.

En la práctica, hay 6 expertos que opinan sobre la salud económico-financiera de 20 empresas de generación eléctrica mediante tecnología eólica. Para ello, puntúan a cada empresa en una escala Likert de 7 puntos, en respuesta a la cuestión: “La salud económico-financiera de la empresa es buena” (1: totalmente en desacuerdo”; 2: “bastante en desacuerdo”; 3: “ligeramente en desacuerdo”; 4: “ni de acuerdo ni en desacuerdo”; 5: “ligeramente de acuerdo”; 6: “bastante de acuerdo”; 7: “totalmente de acuerdo”). A partir de esta información, se calculará la matriz de

proximidades y se realizará el escalado multidimensional de 2 dimensiones, con el objetivo de que se puedan representar las empresas con facilidad. El “stress” nos informará de la bondad del análisis cuando se utilizan dos dimensiones.

Preparando Datos.

Abriremos **R-Studio** y crearemos nuestro **proyecto** siguiendo la instrucción **File → New Project**. Nos preguntará si crea el proyecto en una nueva carpeta o en una ya existente. Vamos a crearlo, por ejemplo, en el disco extraíble D, carpeta R, subcarpeta “**emd**”, que ya existe. Aparecerá una ventana para buscar la carpeta y, tras localizarla, pulsaremos **Open y Create Project**.

Vamos a ir a la carpeta del proyecto y vamos a guardar en ella los dos archivos de esta práctica: un archivo de **Microsoft® Excel®** llamado “eolica_20_emd.xlsx” y un *script* denominado “emd_eolica.R”. Si abrimos el archivo de **Microsoft® Excel®**, comprobaremos que se compone de dos hojas. La primera muestra la descripción de las variables consideradas, y la segunda (hoja “Datos”) guarda los datos que debemos importar desde **R-Studio**. Estos datos se corresponden con las puntuaciones en la escala Likert que han otorgado cada uno de los 6 expertos (columnas) a las empresas productoras de electricidad mediante generación eólica (filas).

Luego vamos a cerrar el archivo de **Microsoft® Excel®** y volveremos a **R-Studio**. Vamos a abrir nuestro script “emd_eolica.R” con **File → Open File...** Este script contiene el programa que vamos a ir ejecutando en la práctica.

La primera línea / instrucción en el script es:

```
rm(list = ls())
```

La instrucción tiene como objeto limpiar el **Environment** de objetos de anteriores sesiones de trabajo. Para importar los datos, ejecutaremos el código:

```
# Importando
library(readxl)
datos <- read_excel("eolica_20_emd.xlsx", sheet = "Datos")
```

Podemos observar cómo, en el *Environment*, ya aparece un objeto. Este objeto es una estructura de datos tipo *data frame*, se llama “datos” y contiene 7 columnas.

R ha considerado la primera columna (NOMBRE) como una variable de tipo cualitativo. En realidad, no es una variable, sino el nombre de los casos (empresas). Para evitar que R tome los nombres de los individuos como una variable, podemos redefinir nuestro *data frame* diciéndole que considere esa primera columna como los nombres de los individuos o casos (filas):

```
datos <- data.frame(datos, row.names = 1)
summary (datos)
```

Vemos que ya no aparece NOMBRE como variable, y en el *Environment* el *data frame* “datos” cuenta con 6 variables (una menos). Cada variable o columna es la valoración de uno de los expertos otorgada a cada una de las empresas de la muestra, en la escala *Likert*. Con `summary()` comprobamos las variables que contiene el *data frame*, y sus características.

Matriz de proximidades.

El siguiente paso consiste en el cálculo de la matriz de proximidades a partir de las valoraciones de los expertos. Para ello, se utilizará la distancia euclídea, mediante la función `dist()`:

```
# distancias
proximidades <- dist(datos)
```

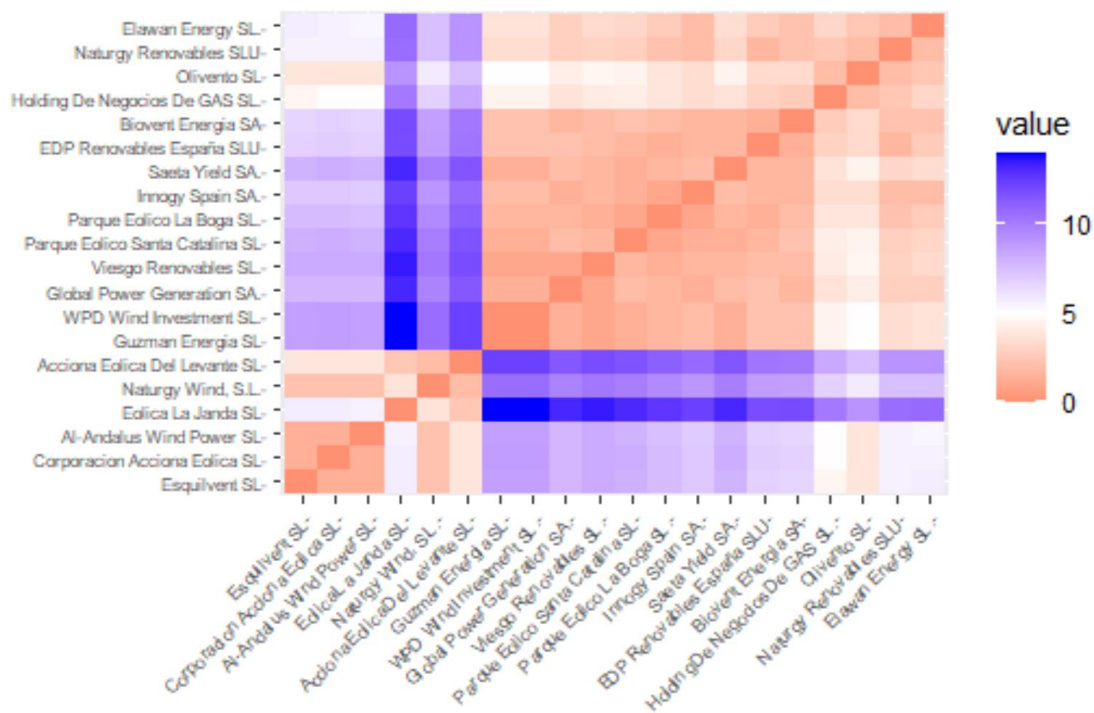
De este modo, se ha creado la matriz, a la que hemos asignado el nombre “proximidades”. Es importante hacer hincapié en que, **aunque los elementos de la matriz estén en escala métrica, no se trata de un análisis de escalado multidimensional métrico, sino no-métrico**; ya que los elementos de la matriz provienen de las percepciones de individuos (expertos) sobre los casos (empresas), expresadas en una escala no-métrica (*Likert*, ordinal).

Un modo de visualizar la matriz de proximidades, para extraer alguna conclusión previa, es mediante un gráfico de “temperatura”, que puede construirse fácilmente a partir de la función `fviz_dist()` del paquete **factoextra**:

```
library (factoextra)
```

```
fviz_dist(proximidades, lab_size = 6)
```

El gráfico obtenido es el siguiente:



Las intersecciones entre empresas en tonos tostados o anaranjados indican que las empresas implicadas han sido percibidas (en cuanto a su salud económico-financiera) como muy próximas, a juzgar por los expertos. En cambio, los tonos azulados indican que han sido percibidas como muy distantes.

Escalado multidimensional.

Una vez se tiene construida la matriz de proximidades, se ha de proceder al escalado multidimensional no-métrico, propiamente dicho. Existen diversos algoritmos para llevar a cabo el escalado. En esta práctica se utilizará uno de los más habituales: el algoritmo **Smacof**. Para ello, se ha de activar el paquete **smacof**, y utilizar su función **mds()**:

```
# Escalado Multidimensional No-M (Algoritmo: Smacof. Defecto: 2 dim.)
library (smacof)
solucion = mds(proximidades, type= "ordinal")
solucion
```

La solución obtenida se asigna al objeto “solucion”. Como argumentos, la función **mds()** tiene la matriz de proximidades, y el tipo de análisis a realizar

(si es un **escalado no-métrico**, un tipo muy apropiado es el **“ordinal”**). Por defecto, esta función considera dos dimensiones, a fin de poder representar los casos con facilidad. En concreto, calcula las coordenadas para cada caso (empresa), a partir de las “proximidades” entre estas captadas en la matriz de proximidades, tales que hacen mínimo el **“stress”**. El “stress” es una **medida de bondad** del ajuste: a menor “stress”, mejor se representa la matriz de proximidades a partir de las dimensiones consideradas, y las coordenadas en estas dimensiones calculadas. **Si se quiere modificar el número de dimensiones** a considerar, se deberá incluir el argumento “ndim= “.

Tras ejecutar el código anterior, en pantalla se mostrará, al nombrar a la “solucion”, la información básica de la aplicación del algoritmo:

```
Call:
mds(delta = proximidades, type = "ordinal")

Model: Symmetric SMACOF
Number of objects: 20
Stress-1 value: 0.013
Number of iterations: 34
```

La información más importante se refiere al valor del “stress”, que es de 0,013. Este valor, según la interpretación de Kruskal, viene a decirnos que la representación de la realidad propuesta por esta solución de dos dimensiones es prácticamente perfecta.

El output más importante que nos ofrece el escalado multidimensional es la **representación gráfica de los casos, de modo que, a menor distancia entre ellos, mayor será la similitud que han observado los encuestados**, en general, a partir de sus propias percepciones globales. Por otro lado, hay que tener en cuenta que, en general, las dimensiones no tienen una interpretación económica clara.

Para realizar el gráfico que sitúe a las empresas de la práctica según se perciben más o menos próximas (parecidas), podremos ejecutar este código:

```
library (ggplot2)
library (ggrepel)

dimensiones <- as.data.frame(solucion$conf)
ggplot(data = dimensiones, map = aes(x = D1, y = D2, label =
row.names(dimensiones))) +
  geom_point(colour = "red", alpha = 0.8) +
  geom_label_repel(size=3, color="black", alpha = 0.5) +
```

```
ggtitle("EMPRESAS EÓLICAS", subtitle = "Configuración de opiniones de
expertos")+
  xlab("Dimensión 1") +
  ylab("Dimensión 2")
```

En primer lugar, se activan el paquete gráfico `ggplot2`, y el paquete `ggrepel`, que complementa al primero en cuanto a las posibilidades de etiquetación de los elementos del gráfico.

Posteriormente, se crea un *data frame* con las coordenadas, para cada caso o empresa, en cada una de las dos dimensiones consideradas en la solución. Este data frame se ha llamado, por ejemplo, “dimensiones”, y tendrá dos columnas (las dos dimensiones), y 20 filas (las empresas de la muestra). Estas coordenadas estaban almacenadas en el elemento de “solucion” llamado **\$conf**.

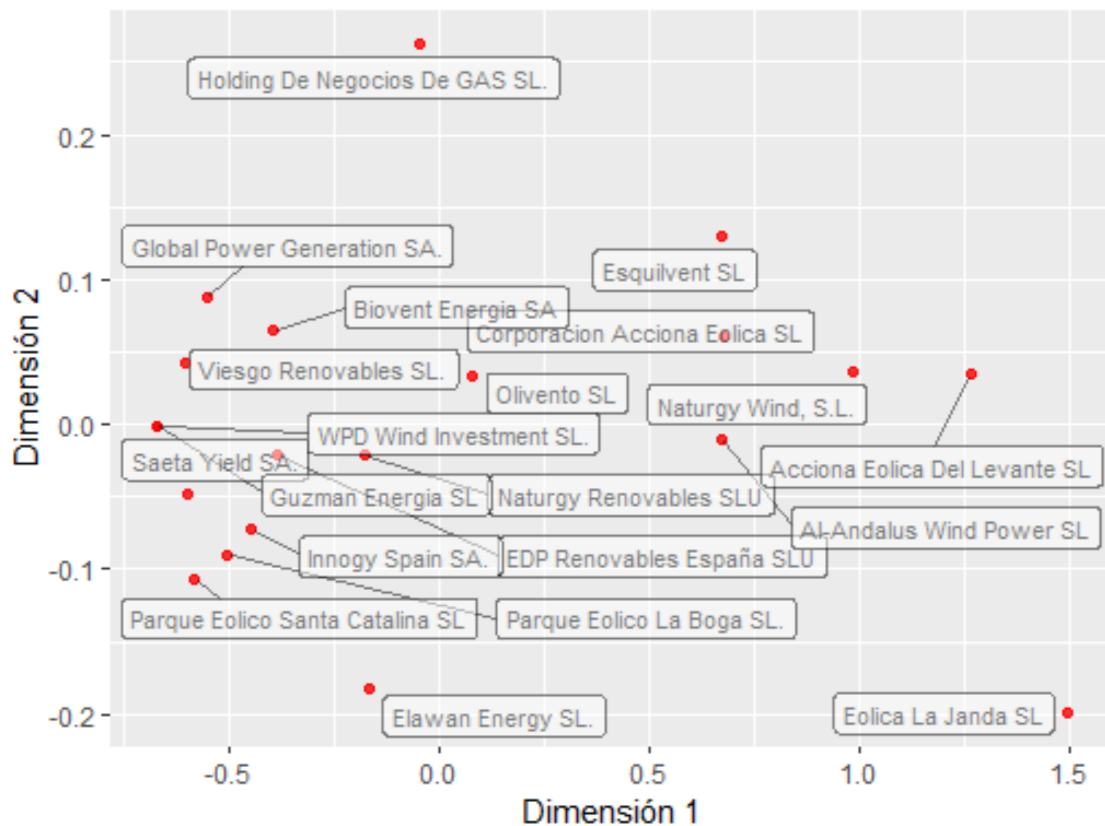
A continuación, se añadirá el código correspondiente al gráfico, a partir de la función `ggplot()`. Hay que tener en cuenta que, por defecto, las variables que recogen las coordenadas de las dimensiones 1 y 2, y que se corresponden con las columnas del *data frame* “dimensiones”, se denominan D1 y D2, respectivamente.

Las etiquetas de las empresas se configuran a partir de la función `geom_label_reppel()`, perteneciente al paquete `ggrepel`.


Como resultado, se obtiene el siguiente gráfico bidimensional:

EMPRESAS EÓLICAS

Configuración de opiniones de expertos



Como puede observarse, los expertos perciben como muy diferentes del resto, en cuanto a la salud económico-financiera, a las empresas “Eólica La Janda” y “Holding de Negocios de Gas SL”. Si observamos las encuestas directamente, se aprecia cómo la primera de estas empresas posee una valoración especialmente positiva de los evaluadores, lo que la podría hacer tan diferente, en términos globales, al resto. Por otro lado, existe un grupo de empresas, situadas en la parte inferior – izquierda del gráfico (Santa Catalina, La Boga, Innogy, Saeta, Guzmán, WPD...) que están relativamente concentradas, lo que significa que los expertos las perciben como bastante parecidas. En efecto, si observamos las encuestas, en general, los expertos coinciden al valorar estas empresas en un grado de desacuerdo elevado en cuanto a que posean una buena salud económico-financiera. Lo contrario ocurre con el grupo formado por Corporación Acciona Eólica, Naturgy Wind, Acciona Eólica del Levante, Al-Ándalus Wind Power y Esquilvent.

This work © 2023 by [Miguel Ángel Tarancón](#) and [Consolación Quintana](#) is licensed under [Attribution-NonCommercial-NoDerivatives 4.0 International](#) 

Updated: 03/03/2023