

## Introducción a la Análisis Multivariante

### Datos de la actividad

Número actividad	2
Docente	Mireia Martorell Colom
Alumno	JOSÉ DAVID VENTURA ROLDÁN KATHERINE CARBONELL MONDRAGÓN MATÍAS DÁVILA WINDER SEBASTIÁN SOTO VALDÉS
Lengua de docencia	Castellano
Agrupación	GRUPO 4
Fecha de entrega	2 Diciembre 2023

### SOLUCIÓN

#### Informe de Escalado Multidimensional:

##### 1. Introducción.

Este informe presenta un análisis de escalado multidimensional (MDS) realizado en un conjunto de datos que contiene información sobre diferentes tipos de cereales. El objetivo de este análisis es comprender la estructura subyacente de los datos y explorar las relaciones entre los cereales utilizando técnicas de escalado métrico y no métrico.

##### 1.1 Conjunto de Datos.

El conjunto de datos utilizado para este análisis contiene información sobre diferentes cereales, incluyendo detalles sobre su contenido nutricional y propiedades. El conjunto de datos incluye variables numéricas, como calorías, proteínas, grasas, sodio, fibra, carbohidratos, azúcares, potasio, vitaminas, y otras características. Además, contiene variables categóricas como el fabricante (mfr) y el tipo de cereal (type).

## 1.2 Importación de Datos

Los datos se importaron desde un archivo CSV llamado "cereal.csv" utilizando el siguiente código en R:

```
datos <- read.csv("C:/cereal.csv")
```

## 1.3 Cálculo de la Matriz de Distancia Euclidiana

Para comprender las relaciones entre los cereales, calculamos la matriz de distancia euclidiana utilizando las variables numéricas del conjunto de datos. Estas variables se estandarizaron antes del cálculo de la matriz de distancia. A continuación, se muestra una vista previa de la matriz de distancia de los primeros 6 cereales:

```
# Matriz de distancia euclidiana (primeras 6 filas y columnas)
print(matriz_distancia[1:6, 1:6])
```

Resultados (vista previa):

	1	2	3	4	5	6
1	0.000000	22.20482	19.10501	33.71194	13.51139	14.07125
2	22.204820	0.000000	31.09811	17.49216	29.74994	27.65719
3	19.105014	31.09811	0.000000	31.55374	8.05229	8.64117
4	33.711943	17.49216	31.55374	0.000000	29.94586	28.92441
5	13.511394	29.74994	8.05229	29.94586	0.000000	2.12174
6	14.071247	27.65719	8.64117	28.92441	2.12174	0.000000

## 1.4 Escalado Multidimensional Métrico

Realizamos un escalado multidimensional métrico de los datos utilizando la función cmdscale. A continuación, se muestra el gráfico de la solución bidimensional, donde etiquetamos las distintas categorías de la variable "type" con su número de fila en la base de datos:

# Gráfico de la solución bidimensional

```
plot(escalamiento_metrico, type = "n", xlab = "Dimensión 1", ylab = "Dimensión 2")
```

# Mostrar solo algunas etiquetas (por ejemplo, 10 etiquetas aleatorias)

```
sample_indices <- sample(1:nrow(datos), 10)
```

```
text(escalamiento_metrico[sample_indices, ], labels = datos$type[sample_indices], cex = 0.7)
```

[DEBEMOS INSERTAR EL GRAFICO AQUÍ O UTILIZAR LA EXTRACCION MEDIENTE PDF"]

## 1.5 Pareja de Categorías Más Parecida

Según la solución bidimensional del análisis métrico, la pareja de categorías de la variable "type" que es la más parecida se determina a continuación:

# Encontrar la pareja más parecida

```
min_distancia <- min(distancias_combinaciones)
```

```
pareja_mas_parecida <- which(distancias_combinaciones == min_distancia, arr.ind = TRUE)
```

# Mostrar la pareja más parecida

```
cat("La pareja de categorías más parecida es:", datos$type[pareja_mas_parecida[1, 1]], "y",  
datos$type[pareja_mas_parecida[1, 2]], "\n")
```

Resultado:

La pareja de categorías más parecida es: **H y H**

## 1.6 Valores Propios y Bondad de Ajuste

Calculamos los valores propios de la solución y la bondad de ajuste de la solución bidimensional.

A continuación, se muestran los valores propios:

**# Valores propios**

**print(valores\_propios)**

**Resultados (primeros 6 valores propios):**

[1] 378.4373901 -0.1377006 -0.1908865 -0.5369221 -0.6370509 -0.6506106

## 1.7 Valores Propios Cero

No hay valores propios igual a cero en el análisis.

El hecho de que no haya valores propios igual a cero en el análisis multidimensional es un resultado esperado y deseado en la mayoría de los casos. Los valores propios en un análisis de escalado multidimensional representan la varianza explicada por cada dimensión en la solución. Cuando un valor propio es igual a cero, significa que esa dimensión no contribuye en absoluto a la explicación de las relaciones entre los datos.

Un valor propio igual a cero podría indicar problemas en el análisis, como datos redundantes o una matriz de distancia mal especificada. Sin embargo, en nuestro informe, no se encontraron valores propios iguales a cero, lo que sugiere que el análisis se realizó de manera adecuada y que todas las dimensiones en la solución son relevantes para explicar la estructura de los datos de cereales. Esto es una señal positiva en términos de la calidad de los resultados del análisis de escalado multidimensional.

## 1.8 Distancias Ajustadas y Bondad del Ajuste

Calculamos las distancias ajustadas según la solución de escalado bidimensional y realizamos un gráfico de las distancias ajustadas y observadas. A continuación, se muestra el gráfico:

**# Gráfico de las distancias ajustadas y observadas**

```
plot(matriz_distancia, main = "Distancias Observadas", xlab = "Especímenes", ylab =  
"Distancia", pch = 19)
```

```
plot(distancias_ajustadas, main = "Distancias Ajustadas", xlab = "Especímenes", ylab =  
"Distancia", pch = 19)
```

[DEBEMOS INSERTAR EL GRAFICO AQUÍ O UTILIZAR LA EXTRACCION MEDIENTE PDF ”]

## 1.9 Escalado Multidimensional No Métrico

Realizamos un escalado multidimensional no métrico utilizando el programa isoMDS. A continuación, se muestra el gráfico de la solución bidimensional, etiquetando los puntos nuevamente con su número y usando diferentes símbolos para las distintas categorías de la variable "type":

**# Gráfico de la solución bidimensional no métrica**

```
plot(escalamiento_no_metrico$points, type = "n", xlab = "Dimensión 1", ylab = "Dimensión  
2")
```

**# Mostrar solo algunas etiquetas (por ejemplo, 10 etiquetas aleatorias)**

```
sample_indices_no_metrico <- sample(1:nrow(datos), 10)
```

```
text(escalamiento_no_metrico$points[sample_indices_no_metrico, ], labels =  
datos$type[sample_indices_no_metrico], cex = 0.7)
```

[DEBEMOS INSERTAR EL GRAFICO AQUÍ O UTILIZAR LA EXTRACCION MEDIENTE PDF ”]

## 1.10 Pareja de Categorías Más Parecida (Análisis No Métrico)

Según la solución bidimensional del análisis no métrico, la pareja de categorías de la variable "type" que es la más parecida se determina a continuación:

**# Encontrar la pareja más parecida (análisis no métrico)**

```
min_distancia_no_metrico <- min(distancias_combinaciones_no_metrico)
```

```
pareja_mas_parecida_no_metrico <- which(distancias_combinaciones_no_metrico ==  
min_distancia_no_metrico, arr.ind = TRUE)
```

**# Mostrar la pareja más parecida**

```
cat("La pareja de categorías más parecida (análisis no métrico) es:",  
datos$type[pareja_mas_parecida_no_metrico[1, 1]], "y",  
datos$type[pareja_mas_parecida_no_metrico[1, 2]], "\n")
```

### Resultado:

La pareja de categorías más parecida (análisis no métrico) es: **H y H**

## 1.11 Conclusiones

En este informe, hemos realizado un análisis de escalado multidimensional de un conjunto de datos que contiene información sobre diferentes cereales. Hemos utilizado técnicas de escalado métrico y no métrico para explorar las relaciones entre los cereales en un espacio bidimensional. Además, hemos identificado la pareja de categorías más parecida tanto en el análisis métrico como en el no métrico.

El análisis métrico reveló una pareja de categorías de cereales que son las más parecidas según la solución bidimensional. En el análisis no métrico, obtuvimos resultados similares con la misma pareja de categorías.

En cuanto a los valores propios y la bondad de ajuste, observamos que no hubo valores propios iguales a cero en el análisis métrico. Sin embargo, la bondad de ajuste y la representación de las distancias variaron en el análisis no métrico en comparación con el análisis métrico.

## 1.12 Documento/Archivo PDF

Los resultados y gráficos generados en este informe se han guardado como un archivo PDF, que se encuentra adjunto junto con el script de R y la base de datos utilizada en la actividad.

Este informe resume los resultados del análisis de escalado multidimensional realizado en los datos de cereales.

## 1.13 BIBLIOGRAFÍA

- Peña D (2002). Análisis de datos multivariantes. McGraw-Hill Interamericana. Apartados: 1, 2, 3, 4 y 5. <https://elibro.net/es/lc/universitatcarlemany/titulos/50267>
- Redondo Figuero (2017), Carlos G. El programa R, herramienta clave en investigación. Editorial de la Universidad de Cantabria.