

# Implementación de Análisis de Correspondencias

19 de marzo de 2021

## 1. Objetivo

Hemos visto que el paquete `ca` de R produce un `summary` con información sobre las dos primeras dimensiones, pero no más allá.

Se pretende obtener la misma información para dimensiones a partir de la tercera.

## 2. Datos

Consigue algunos datos para probar las fórmulas. Puede ser un ejemplo de clase o alguno de los distribuidos en R, como los que se presentan a continuación.

### 2.1. Color de pelo y de ojos

```
Kij = apply (HairEyeColor, 1:2, sum)
```

### 2.2. Hundimiento del *Titanic*

```
D = data.frame(Titanic)
i = function (cadena) substring(cadena,1,1) # inicial
D$bio = paste0 (i(D$Sex), i(D$Age), i(D$Survived))
Kij = table (D [rep(1:nrow(D),D$Freq), c("Class", "bio")])
class (Kij) = "matrix" # necesario sólo para MASS::corresp
```

### 2.3. Alojamiento en Copenhague

```
Kij = xtabs (Freq ~ abbreviate(paste(Sat,Infl,Cont)) + Type,
            MASS::housing)
```

## 3. Solución 1: modificar el código fuente de `ca`

1. Averiguar la clase del objeto que contiene la información del análisis de correspondencias.

```
library (ca) ; a <- ca (Kij) ; class (a)
```

2. Averiguar la función (método) que produce el `summary`. Su nombre tendrá la forma `summary.clase`.

```
methods (summary)
```

3. Obtener su código fuente.

```
getAnywhere (nombreDelMétodo)
```

4. Identificar en el código fuente dónde se limita el número de dimensiones a dos.
5. Definir una nueva función a partir de ese código fuente, modificándolo para que permita más dimensiones.

#### 4. Solución 2: implementar el análisis de correspondencias

1. Ten en cuenta las fórmulas de <https://bellman.ciencias.uniovi.es/~carleos/master/manadine/curso1/AnalisisDatos1/5-correspondencias/teoria/correspondencias.pdf>
2. Considera las capacidades de cálculo matricial de R:
  - `%%` para producto matricial.
  - `apply` para aplicar una función por filas o columnas.
  - `diag` para extraer la diagonal de una matriz o crear una matriz diagonal a partir de un vector.
  - Pre o posmultiplicar por una matriz diagonal equivale a multiplicar (elemento a elemento) filas o columnas por el vector diagonal.
  - `t` para trasponer una matriz.
3. Obtén la matriz  $T = X^t X$  (página 5).
4. La información del `summary` contiene
  - inercias principales (autovalores)
  - por cada fila y cada columna, en milésimas:
    - masa relativa
    - calidad de la reconstrucción
    - inercia relativa
    - por cada dimensión:
      - coordenadas

- contribución relativa (a la calidad)
- contribución absoluta (a la inercia)

5. Autovalores.

- Son los  $\lambda_\alpha$
- Se obtienen mediante la función `eigen` aplicada a la matriz  $T$ .
- Los acumulados pueden calcularse mediante `cumsum`.

6. Masas relativas: las frecuencias relativas marginales,  $f_i$  y  $f_{\cdot j}$ , en milésimas.

7. Calidad de la reconstrucción en los ejes representados:

- Suma de las contribuciones relativas en los ejes  $\alpha$ :  $\mathbf{qlt} = \sum_\alpha \text{Cr}_i(\alpha)$

8. Inercias relativas, en milésimas: es la parte de la variabilidad total debida al punto. Puede calcularse a partir de  $\psi$  y  $\varphi$  o de las contribuciones absolutas sumando para todo  $\alpha = 1, \dots, \min\{n, p\} - 1$ :

$$\mathbf{inr}_i = \frac{\sum f_i \cdot \psi_{\alpha i}}{\sum \lambda_\alpha} = \frac{\sum \lambda_\alpha \text{Ca}_\alpha(i)}{\sum \lambda_\alpha}$$

9. Coordenadas en el eje  $\alpha$ , en milésimas:

- para filas,  $\psi_{\alpha i}$
- para columnas,  $\varphi_{\alpha j}$

Ojo, porque pueden tener el signo cambiado.

10. Contribuciones relativas, en milésimas:

- Se calculan a partir de las coordenadas  $\psi$  (para filas) y  $\varphi$  para columnas.
- Para el punto fila  $i$  sobre el eje  $\alpha$ :

$$\text{Cr}_i = \frac{\psi_{\alpha i}^2}{\sum_\beta \psi_{\beta i}^2}$$

- Para el punto columna  $j$  sobre el eje  $\alpha$ :

$$\text{Cr}_j = \frac{\varphi_{\alpha j}^2}{\sum_\beta \varphi_{\beta j}^2}$$

11. Contribuciones absolutas, en milésimas:

- Se calculan a partir de las coordenadas  $\psi$  (para filas) y  $\varphi$  para columnas.

- Para el punto fila  $i$  sobre el eje  $\alpha$ :

$$Ca_i = \frac{f_i \cdot \psi_{\alpha i}^2}{\lambda_{\alpha}}$$

- Para el punto columna  $j$  sobre el eje  $\alpha$ :

$$Ca_j = \frac{f_j \cdot \varphi_{\alpha j}^2}{\lambda_{\alpha}}$$