

# Repaso multivariante. Combinando CA con clustering

*00R Team*

*Abril 2016*

## Índice

1. Preliminares	1
2. Visualizar los datos del análisis	2
3. La función HCPC en correspondencias	2

## 1. Preliminares

Cargar los paquetes `FactoMineR`, `factoextra`, `rgl`, `cluster`

### 1.1. Cargar datos

- Cargar el fichero de datos `housetask` incluido en el paquete `factoextra`
- ¿Cuántas variables tenemos?
- ¿Son cuantitativas o cualitativas?
- ¿Cuál es el número máximo de dimensiones que esperamos obtener?

Aplicar la función `CA()` sobre los datos cargados para hacer un análisis de correspondencias

```
# código  
# myCA <- CA(.....)  
# summary(.....)
```

La función `CA` realiza un test  $\chi^2$  sobre la independencia de las variables. Es decir la hipótesis nula es que las variables son independientes.

- A la vista de los resultados del análisis, ¿crees que hay independencia entre las variables o están de algún modo relacionadas?

Este test se puede realizar independientemente de la función `CA`, mediante la función `chisq.test()`

```
# código  
# chisq.test(.....)
```

- ¿Es coherente este resultado de la  $\chi^2$  respecto al obtenido con la función `CA`?

## 2. Visualizar los datos del análisis

### 2.1. Análisis por filas

1. Mostrar contribuciones por filas
2. Representar las contribuciones de las filas (`fviz_ca_row`)

```
# código R  
# myCA$....$.  
# fviz_ca_row(.....)
```

### 2.2. Análisis por columnas

1. Mostrar contribuciones por columnas
2. Representar las contribuciones de las columnas (`fviz_ca_col`)

```
# código R  
# myCA$....$.  
# fviz_ca_col(.....)
```

### 2.3. Biplot

Con la función `fviz_ca_biplot()` sobre el objeto CA creado se puede visualizar un gráfico biplot.

```
# código  
#fviz_ca_biplot(.....)
```

- ¿Se observa relación entre categorías de las variables?
- ¿Entre cuáles se observa mayor correlación?

1. La función `dimdesc()` nos da la correlación de las variables con cada dimensión y facilita la interpretación de los resultados.

- ¿Se observa relación entre las filas que más contribuyen y las columnas que más contribuyen?
- Conociendo sólo estos datos, ¿podríamos agrupar las tareas en función de quién las realiza?

## 3. La función HCPC en correspondencias

La función hace una partición en  $k$  grupos de los resultados del análisis de correspondencias. El número de grupos se puede seleccionar manualmente o se puede dejar a la función decidir cuál es el mejor número de grupos.

- Los grupos creados, ¿son coherentes con lo visto e interpretado anteriormente?