

# Trabajo Reglas Asociación

*Juanjo Sierra*

*27 de enero de 2019*

En este trabajo final sobre **reglas de asociación** se ha seleccionado un dataset sobre el que trabajar obteniendo reglas que resulten de interés y que aporten información a los datos que ya se poseen. A continuación se estudiarán los itemsets frecuentes, maximales y cerrados, se obtendrán las reglas correspondientes al mínimo soporte establecido, y en general se hará uso de las técnicas aprendidas durante el curso para así alcanzar el objetivo de la práctica.

## Carga de librerías

En primer lugar es necesario cargar las librerías necesarias para trabajar con reglas de asociación. En el caso de este trabajo se van a utilizar las siguientes.

```
library(arules)

## Loading required package: Matrix
##
## Attaching package: 'arules'
## The following objects are masked from 'package:base':
##
##      abbreviate, write

library(arulesViz)

## Loading required package: grid

library(pmml)

## Loading required package: XML

library(mlbench)
```

## Lectura de los datos

El dataset que se ha escogido para desarrollar este trabajo ha sido *Contraceptive Method Choice*, obtenido del repositorio de datasets de la Universidad de California-Irvine. Este conjunto de datos determina el método de anticoncepción elegido por una pareja, que puede ser ninguno, a corto plazo o a largo plazo.

```
Contraceptive = read.csv("./Datos/cmc.data", header = FALSE)
head(Contraceptive)
```

```
##   V1 V2 V3 V4 V5 V6 V7 V8 V9 V10
## 1 24  2  3  3  1  1  2  3  0  1
## 2 45  1  3 10  1  1  3  4  0  1
## 3 43  2  3  7  1  1  3  4  0  1
## 4 42  3  2  9  1  1  3  3  0  1
## 5 36  3  3  8  1  1  3  2  0  1
## 6 19  4  4  0  1  1  3  3  0  1
```

Para dar un nombre adecuado a cada variable se van a utilizar la información extraída del archivo `cmc.names`. De esta forma también se le dará una nomenclatura adecuada a los valores que puede tomar cada una de las características, convirtiéndolas en un factor cuando sea necesario.

```

colnames(Contraceptive) = c("Wife's age", "Wife's education", "Husband's education",
                             "Children", "Wife's religion", "Wife working",
                             "Husband's occupation", "Standard-of-living",
                             "Media exposure", "Contraceptive method")

# Wife's education
Contraceptive[,2] = ordered(Contraceptive[,2], levels = 1:4,
                             labels = c("Low", "Mid-low", "Mid-high", "High"))

# Husband's education
Contraceptive[,3] = ordered(Contraceptive[,3], levels = 1:4,
                             labels = c("Low", "Mid-low", "Mid-high", "High"))

# Wife's religion
Contraceptive[,5] = factor(Contraceptive[,5], levels = 0:1,
                             labels = c("Non-Islam", "Islam"))

# Wife working
Contraceptive[,6] = factor(Contraceptive[,6], levels = 0:1,
                             labels = c("Yes", "No"))
# ¡No es un error! "0" significa "sÍ" en esta variable...

# Standard-of-living
Contraceptive[,8] = ordered(Contraceptive[,8], levels = 1:4,
                             labels = c("Low", "Mid-low", "Mid-high", "High"))

# Media exposure
Contraceptive[,9] = factor(Contraceptive[,9], levels = 0:1,
                             labels = c("Good", "Not good"))

# Contraceptive method
Contraceptive[,10] = factor(Contraceptive[,10], levels = 1:3,
                             labels = c("No-use", "Long-term", "Short-term"))

head(Contraceptive)

```

```

##   Wife's age Wife's education Husband's education Children Wife's religion
## 1         24         Mid-low         Mid-high         3         Islam
## 2         45          Low         Mid-high        10         Islam
## 3         43         Mid-low         Mid-high         7         Islam
## 4         42        Mid-high         Mid-low         9         Islam
## 5         36        Mid-high         Mid-high         8         Islam
## 6         19          High          High         0         Islam
##   Wife working Husband's occupation Standard-of-living Media exposure
## 1          No                2         Mid-high         Good
## 2          No                3          High         Good
## 3          No                3          High         Good
## 4          No                3        Mid-high         Good
## 5          No                3        Mid-low         Good
## 6          No                3        Mid-high         Good
##   Contraceptive method
## 1          No-use
## 2          No-use
## 3          No-use

```

```
## 4          No-use
## 5          No-use
## 6          No-use
```

Cabe destacar que para la variable “Husband’s occupation” no se especifica en la información del dataset el significado de los distintos valores, por lo que si aparece en futuras reglas o itemsets se tratarán estos valores de forma similar al resto de variables, siendo 1 el de valor más bajo y 4 el más alto.

Ahora que los datos son más legibles e interpretables se puede proceder a extraer itemsets frecuentes.