

# Homework 6

Marc Mendez & Joel Cantero

5th May, 2019

First of all, we are going to install and load all the libraries we need for this exercise. The factominer library is needed for the catdes function, which is needed on exercise 2. Then the arules function is needed to the the association rules.

```
packages <- c("FactoMineR", "arules")
for (package in packages) {
  if(!require(package, character.only=TRUE)){
    install.packages(package, repos="http://cran.rstudio.com")
    library(package, character.only=TRUE)
  }
}
```

## 1. Read the file: tic\_tt.csv. Check the “class” of every variable of tic\_tt.

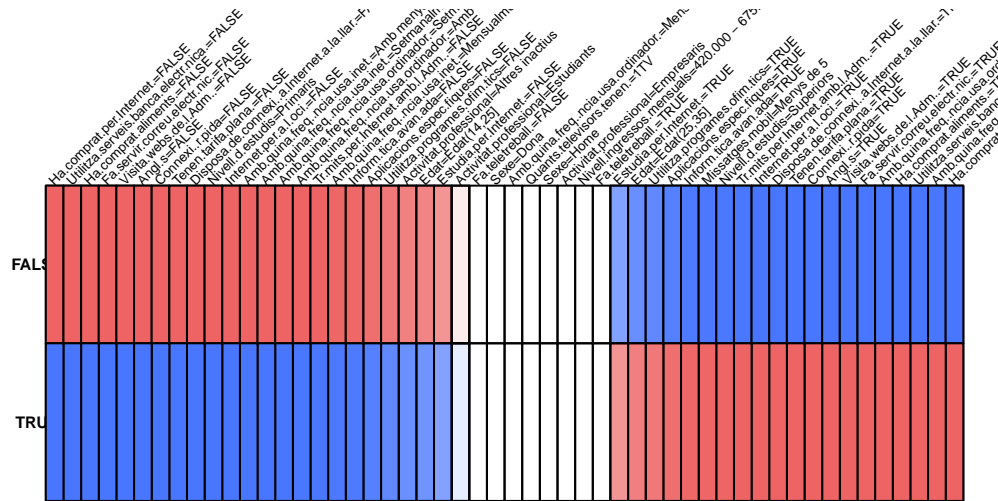
First of all we need to check the data and we can see that all the variables are booleans or factors with more than 2 levels. So, for loading the data, we will set the classes as factors and set the first one as row names because they are numbers that don't give information to the study. We will make a summary of it and analyze all the results to see that the load has been succesfull.

```
tic <- read.csv("tic_tt.csv", sep=";", colClasses = c(rep("factor",33)), row.names=1)
#summary(tic)
```

## 2. Find the profile of the people who do payments by Internet (use the function catdes of FactoMineR). (to perform the catdes of a variable, it should be categorical (factor in the R nomenclature)).

To study the profile of the people who do payments by Internet we are going to execute the catdes function over the variable “Pagament.a.trav.s.d.Internet.”. By doing this we can see the variables that have a realtion with this one.

```
a <- catdes(tic, 28)
plot(a)
```



As we can see in this plot the users that make payments through internet use electronic payments services and also they use internet every day. Most of them know English so there is a relation with the language also. We can see that, on the other hand, people that not pay through internet tend to not have electronic payment services as well as few usage of e-mail. And also they don't speak English, which makes sense as english is the most common language on the computers world. We can notice that there are differences between users that do payments on internet and people that not.

### 3. Convert the tic\_tt file to a transactions file.

To do this conversion we will use `as()` function. This conversion is to make the association rules after. To do this conversion we needed to delete the first column as it wasn't a factor and have all the other columns as factors.

```
tic.trans <- as(tic, "transactions")
```

### 4. Define the parameters: Min\_support, min\_confidence and maximum size of itemsets, and run the apriori function.

Now that we have the transaction file we are going to execute it with the function `apriori` from the package "arules". To execute this we set the min support, min confidence and maximum size of the rules. The more precise we make the variables the larger number of rules we are going to obtain and longer time will last to calculate it.

```
minSupp <- 0.3
minConf <- 0.75
maxSize <- 5
rules <- apriori (tic.trans, parameter = list
                  (support=minSupp, confidence=minConf, maxlen = maxSize))
```

```
##      lhs      rhs      support confidence
## [1] {} => {Fa.servir.xat.=FALSE} 0.7616487 0.7616487
## [2] {} => {Utilitza.programes.ofim.tics=TRUE} 0.7894265 0.7894265
## [3] {} => {Castell.=TRUE} 0.7921147 0.7921147
## [4] {} => {Ha.comprat.per.Internet.=FALSE} 0.8100358 0.8100358
## [5] {} => {Tr.mite.per.Internet.amb.l.Adm..=FALSE} 0.8127240 0.8127240
## [6] {} => {Connexi...r.pida=FALSE} 0.8189964 0.8189964
## [7] {} => {Visita.webs.sanitaris.=FALSE} 0.8351254 0.8351254
## [8] {} => {Inform.tica.avan.ada=FALSE} 0.8467742 0.8467742
## [9] {} => {Pagament.a.trav.s.d.Internet.=FALSE} 0.8629032 0.8629032
## [10] {} => {Estudia.per.Internet.=FALSE} 0.9301075 0.9301075
##      lift count
## [1] 1 850
## [2] 1 881
## [3] 1 884
## [4] 1 904
```

```
## [5] 1 907
## [6] 1 914
## [7] 1 932
## [8] 1 945
## [9] 1 963
## [10] 1 1038
```

We have set this parameters as the next questions doesn't require to have large amount of solutions and this ones are more than enough for it.

## 5. List the 10 most frequent itemsets.

To obtain the list with the most frequent itemsets we need to filter the repeated itemsets.

```
fsets <- unique(generatingItemsets(rules))
fsets.df <- as(fsets, "data.frame")
sor.fsets <- fsets.df[order(-fsets.df$support),]
sor.fsets[1:10,1]
```

```
## [1] {Tr.mits.amb.els.serveis.sanitaris.=FALSE}
## [2] {Fa.teletreball.=FALSE}
## [3] {Ha.comprat.aliments.=FALSE}
## [4] {Fa.teletreball.=FALSE,Tr.mits.amb.els.serveis.sanitaris.=FALSE}
## [5] {Ha.comprat.aliments.=FALSE,Tr.mits.amb.els.serveis.sanitaris.=FALSE}
## [6] {Fa.teletreball.=FALSE,Ha.comprat.aliments.=FALSE}
## [7] {Estudia.per.Internet.=FALSE}
## [8] {Estudia.per.Internet.=FALSE,Tr.mits.amb.els.serveis.sanitaris.=FALSE}
## [9] {Fa.teletreball.=FALSE,Ha.comprat.aliments.=FALSE,Tr.mits.amb.els.serveis.sanitaris.=FALSE}
## [10] {Estudia.per.Internet.=FALSE,Fa.teletreball.=FALSE}
## 29129 Levels: {Activitat.professional=Assalariats,Estudia.per.Internet.=FALSE,Fa.servir.correu.elect.
```

As we can see on this list, the most repeated itemsets are those ones above.

## 6. List the first 10 rules sorted by the lift.

To make this sorting we need to order by lift.

```
top10.lift <- inspect(sort(rules, by="lift")[1:10])
```

##	lhs	rhs	support	conf
## [1]	{Disposa.de.connexi..a.Internet.a.la.llar.=FALSE}	=> {Tenen.tarifa.plana.=FALSE}	0.3100358	
## [2]	{Disposa.de.connexi..a.Internet.a.la.llar.=FALSE, Connexi..r.pida=FALSE}	=> {Tenen.tarifa.plana.=FALSE}	0.3100358	
## [3]	{Disposa.de.connexi..a.Internet.a.la.llar.=FALSE, Estudia.per.Internet.=FALSE}	=> {Tenen.tarifa.plana.=FALSE}	0.3037634	
## [4]	{Disposa.de.connexi..a.Internet.a.la.llar.=FALSE, Ha.comprat.aliments.=FALSE}	=> {Tenen.tarifa.plana.=FALSE}	0.3082437	
## [5]	{Disposa.de.connexi..a.Internet.a.la.llar.=FALSE, Fa.teletreball.=FALSE}	=> {Tenen.tarifa.plana.=FALSE}	0.3073477	
## [6]	{Disposa.de.connexi..a.Internet.a.la.llar.=FALSE, Tr.mits.amb.els.serveis.sanitaris.=FALSE}	=> {Tenen.tarifa.plana.=FALSE}	0.3064516	
## [7]	{Disposa.de.connexi..a.Internet.a.la.llar.=FALSE, Connexi..r.pida=FALSE, Estudia.per.Internet.=FALSE}	=> {Tenen.tarifa.plana.=FALSE}	0.3037634	
## [8]	{Disposa.de.connexi..a.Internet.a.la.llar.=FALSE, Connexi..r.pida=FALSE, Ha.comprat.aliments.=FALSE}	=> {Tenen.tarifa.plana.=FALSE}	0.3082437	

```
## [9] {Disposa.de.connexi..a.Internet.a.la.llar.=FALSE,
##      Connexi..r.pida=FALSE,
##      Fa.teletreball.=FALSE} => {Tenen.tarifa.plana.=FALSE} 0.3073477
## [10] {Disposa.de.connexi..a.Internet.a.la.llar.=FALSE,
##      Connexi..r.pida=FALSE,
##      Tr.mits.amb.els.serveis.sanitaris.=FALSE} => {Tenen.tarifa.plana.=FALSE} 0.3064516
```

## 7. List the 10 rules according the lift, where the Consequent is “Pagament.a.traves.d.Internet.”.

On this part we first need to create a subset with only the rules that have “Pagament.a.traves.d.Internet.” as the Consequent. Then, once the subset is created we proceed on doing the same as the last exercise, order by lift and thats all.

```
rulesConseq <- subset(rules, subset = rhs %in% c("Pagament.a.trav.s.d.Internet.=FALSE") )
top10.lift.Conseq <- inspect(sort(rulesConseq, by="lift")[1:10] )
```

	lhs	rhs	support	confiden
## [1]	{Ha.comprat.per.Internet.=FALSE}	=> {Pagament.a.trav.s.d.Internet.=FALSE}	0.8100358	
## [2]	{Ha.comprat.per.Internet.=FALSE, Visita.webs.de.l.Adm..=FALSE}	=> {Pagament.a.trav.s.d.Internet.=FALSE}	0.3091398	
## [3]	{Nivell.ingressos.mensuals=NS/NC, Ha.comprat.per.Internet.=FALSE}	=> {Pagament.a.trav.s.d.Internet.=FALSE}	0.3315412	
## [4]	{Catal.=FALSE, Ha.comprat.per.Internet.=FALSE}	=> {Pagament.a.trav.s.d.Internet.=FALSE}	0.3315412	
## [5]	{Sexe=Home, Ha.comprat.per.Internet.=FALSE}	=> {Pagament.a.trav.s.d.Internet.=FALSE}	0.3306452	
## [6]	{Quants.televisors.tenen.=2TVs, Ha.comprat.per.Internet.=FALSE}	=> {Pagament.a.trav.s.d.Internet.=FALSE}	0.3548387	
## [7]	{Nivell.d.estudis=Secundaris, Ha.comprat.per.Internet.=FALSE}	=> {Pagament.a.trav.s.d.Internet.=FALSE}	0.3673835	
## [8]	{Per.oci=FALSE, Ha.comprat.per.Internet.=FALSE}	=> {Pagament.a.trav.s.d.Internet.=FALSE}	0.4112903	
## [9]	{Per.oci=TRUE, Ha.comprat.per.Internet.=FALSE}	=> {Pagament.a.trav.s.d.Internet.=FALSE}	0.3987455	
## [10]	{Activitat.professional=Assalariats, Ha.comprat.per.Internet.=FALSE}	=> {Pagament.a.trav.s.d.Internet.=FALSE}	0.4489247	