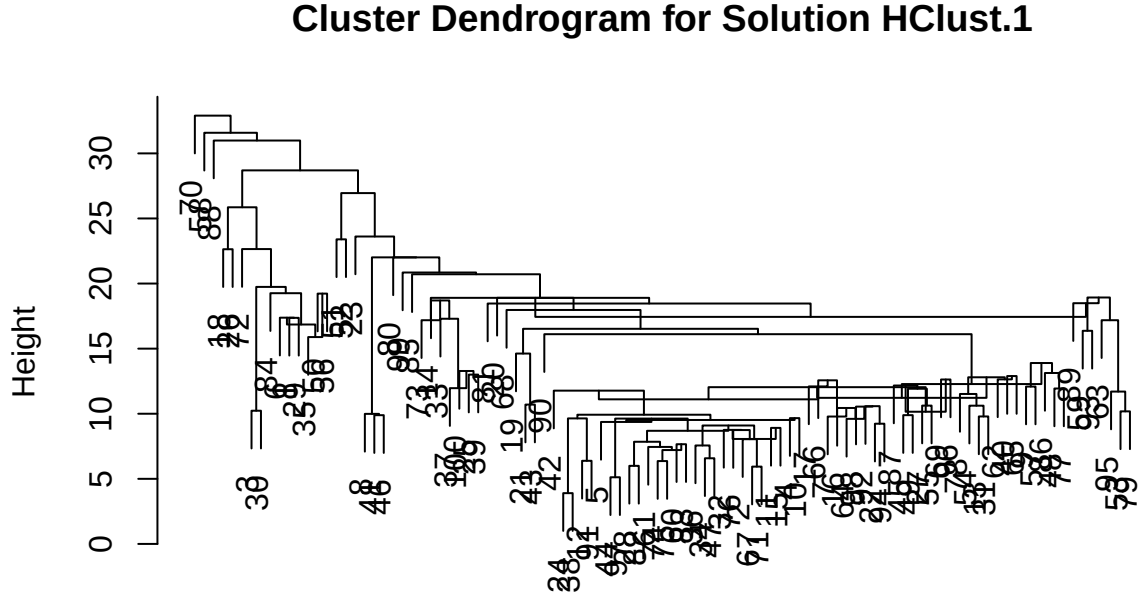
Análisis Multivariado - Maestría en Estadística Aplicada

Dra. Maria Ines Stimolo

Còrdoba- 2019

Agrupamiento Jerarquico

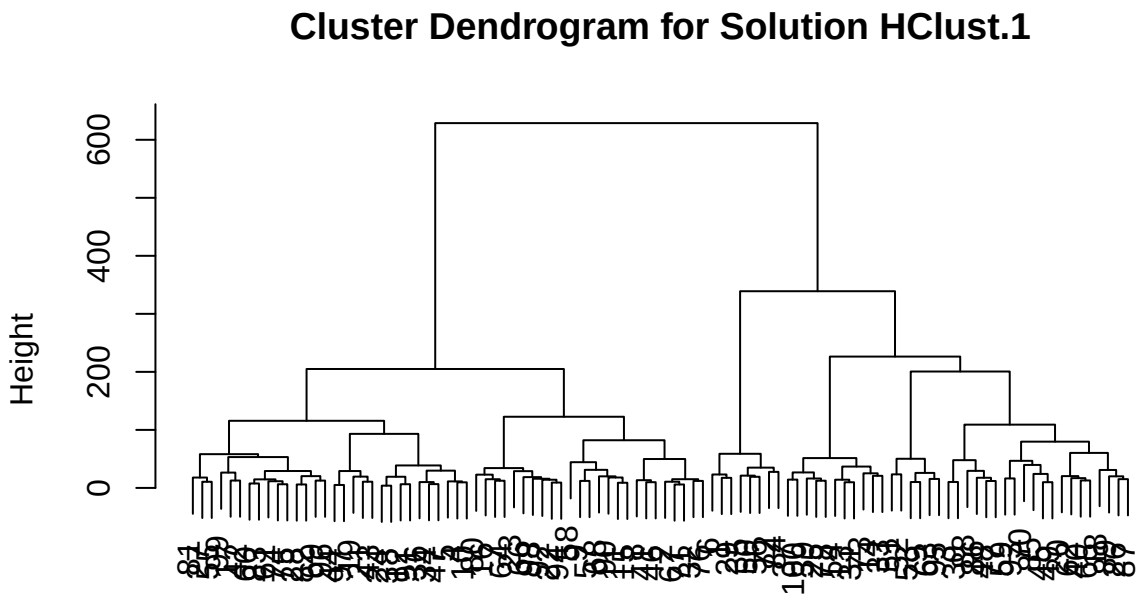
```
knitr::opts_chunk$set(echo = TRUE)

#Agrupar los países utilizando el método Jerárquico.
#Utilizar distancias euclídeas y metodo de Ward.e

load("satif.rda")

## Agrupación jerárquica
#?hclust
### ver los metodos de aglomeración que utiliza.
##utilizando el metodo delcentroide
HClust.1<-hclust(dist(model.matrix(~-1+ANTEMP+ANTIG+EDUCAC+HORASEM+EDAD+INGRFLIA+INGRPER,Datos)),
method="centroid")
plot(HClust.1,main="Cluster Dendrogram for Solution HClust.1",
     xlab="Observation Number in Data Set Datos",sub="Method=single;Distance=euclidian")
```



```
##utilizando el metodo de ward
HClust.1<-hclust(dist(model.matrix(~-1+ANTEMP+ANTIG+EDAD+EDUCAC+HORASEM+INGRFLIA+INGRPER,Datos)),
method="ward")
```

```
## The "ward" method has been renamed to "ward.D"; note new "ward.D2"
```

```
plot(HClust.1,main="Cluster Dendrogram for Solution HClust.1",
     xlab="Observation Number in Data Set Datos",sub="Method=ward;Distance=euclidian")
```




```
names(HClust.1)
```

```
## [1] "merge"       "height"      "order"       "labels"      "method"
## [6] "call"        "dist.method"
```

```
#detalle del agrupamiento
detalle=cbind(HClust.1$merge, HClust.1$height); detalle[1:20,]
```

```
##       [,1] [,2]      [,3]
##  [1,]  -24  -38  3.899756
##  [2,]  -44  -97  5.120439
##  [3,]  -67  -71  5.937171
##  [4,]  -28  -82  5.997499
##  [5,]  -74  -75  6.363961
##  [6,]  -12  -91  6.383573
##  [7,]  -34  -47  6.578754
##  [8,]  -60  -83  7.658982
##  [9,]  -11  -15  8.911229
## [10,]  -22  -94  9.234479
```

```
## [11,]   -4  -10  9.662815
## [12,]  -13  -31  9.790204
## [13,]  -41  -46  9.889388
## [14,]  -45  -49  9.891916
## [15,]  -36    7 10.083916
## [16,]  -53  -79 10.181753
## [17,]   -3  -30 10.232302
## [18,]  -27  -55 10.625272
## [19,]  -21  -43 10.702336
## [20,]   -2    3 10.725312
```

```
d1<- dist(Datos[,2:8])

d2 <-cophenetic(HClust.1)
(as.matrix(d2))[1:10,1:10]
```

```
##           1         2        3          4        5        6        7
## 1    0.0000 628.68994 338.9685 628.689936 628.6899  23.1701 628.6899
## 2  628.6899   0.00000 628.6899 204.955672 204.9557 628.6899 122.6748
## 3  338.9685 628.68994   0.0000 628.689936 628.6899 338.9685 628.6899
## 4  628.6899 204.95567 628.6899   0.000000  11.2713 628.6899 204.9557
## 5  628.6899 204.95567 628.6899  11.271300   0.0000 628.6899 204.9557
## 6   23.1701 628.68994 338.9685 628.689936 628.6899   0.0000 628.6899
## 7  628.6899 122.67477 628.6899 204.955672 204.9557 628.6899   0.0000
## 8  628.6899  49.84697 628.6899 204.955672 204.9557 628.6899 122.6748
## 9  628.6899  82.18701 628.6899 204.955672 204.9557 628.6899 122.6748
## 10 628.6899 204.95567 628.6899   9.662815  11.2713 628.6899 204.9557
##            8         9         10
## 1  628.68994 628.68994 628.689936
## 2   49.84697  82.18701 204.955672
## 3  628.68994 628.68994 628.689936
## 4  204.95567 204.95567   9.662815
## 5  204.95567 204.95567  11.271300
## 6  628.68994 628.68994 628.689936
## 7  122.67477 122.67477 204.955672
## 8    0.00000  82.18701 204.955672
## 9   82.18701   0.00000 204.955672
## 10 204.95567 204.95567   0.000000
```

```
cor(d1,d2)
```

```
## [1] 0.4705713
```

```
apply (Datos[,2:8],2,mean)
```

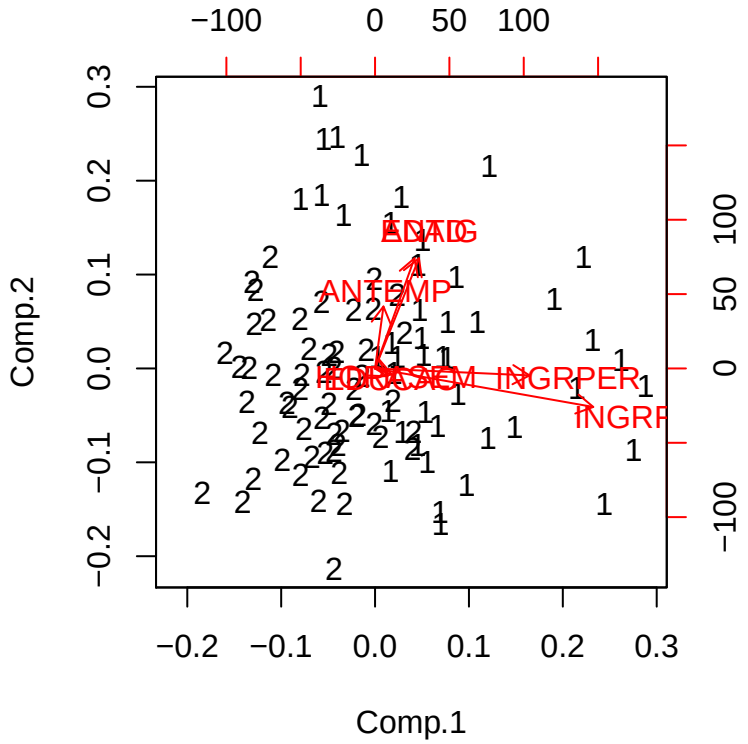
```
##  HORASEM     EDAD   EDUCAC  INGRPER INGRFLIA    ANTIG   ANTEMP
##  44.9000  40.0300  14.2900  32.4490  44.9010  21.2200   7.9136
```

```
#Resumen de los 2 grupos
summary(as.factor(cutree(HClust.1, k = 2)))# Cluster Sizes
```

```
##  1  2
## 45 55
```

```
## Generar una variable grupo(con dos grupos)  h2
Datos$h2 <- c(as.factor(cutree(HClust.1, k = 2)))
biplot(princomp(model.matrix(~-1+ANTEMP+ANTIG+EDAD+EDUCAC+HORASEM+INGRFLIA+INGRPER,Datos)),
```

```
       xlabs=as.character(cutree(HClust.1,k=2)))
```




```
# suma de cuadrados dentro de 2 grupos
G1=as.matrix(subset(Datos[,2:8],Datos$h2==1))
apply(G1,2,mean);dim(G1)
```

```
##  HORASEM     EDAD   EDUCAC  INGRPER INGRFLIA    ANTIG   ANTEMP
## 46.62222 45.55556 14.62222 39.73556 58.97333 27.60000 10.61778
```

```
## [1] 45  7
```

```
G2=as.matrix(subset(Datos[,2:8],Datos$h2==2))
apply(G2,2,mean);dim(G2)
```

```
##   HORASEM      EDAD    EDUCAC   INGRPER  INGRFLIA     ANTIG    ANTEMP
## 43.490909 35.509091 14.018182 26.487273 33.387273 16.000000  5.701091
```

```
## [1] 55  7
```

```
SCD1= sum(diag(var(G1))*(nrow(G1)-1))
SCD2= sum(diag(var(G2))*(nrow(G2)-1))
SCDG2=SCD1+SCD2;SCDG2
```

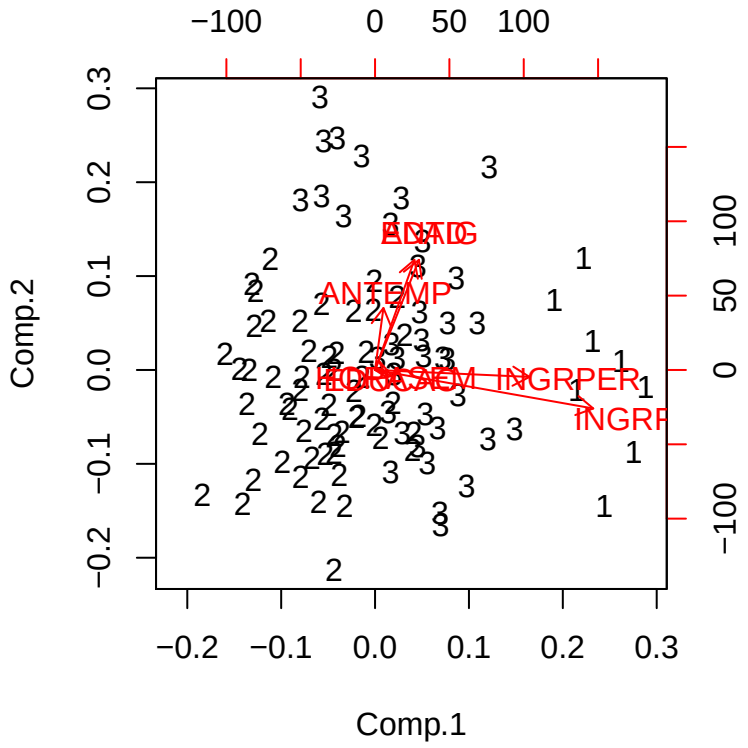
```
## [1] 73296.58
```

```
#Resumen de los 3 grupos
#Generar las variable de los grupos 3 (h3) y 4 (h4) variables
summary(as.factor(cutree(HClust.1, k = 3)))# Cluster Sizes
```

```
##  1  2  3
```

```
##  8 55 37
```

```
## Generar una variable grupo(con tres grupos)  h3
Datos$h3 <- c(as.factor(cutree(HClust.1, k = 3)))
biplot(princomp(model.matrix(~-1+ANTEMP+ANTIG+EDAD+EDUCAC+HORASEM+INGRFLIA+INGRPER,Datos)),
       xlabs=as.character(cutree(HClust.1,k=3)))
```




```
# suma de cuadrados dentro de 3 grupos
G1=as.matrix(subset(Datos[,2:8],Datos$h3==1))
apply(G1,2,mean)
```

```
##  HORASEM     EDAD   EDUCAC  INGRPER INGRFLIA    ANTIG   ANTEMP
##  47.5000  47.7500  16.7500  65.4500  88.1875  30.0000   7.9775
```

```
dim(G1)
```

```
## [1] 8 7
```

```
G2=as.matrix(subset(Datos[,2:8],Datos$h3==2))
apply(G2,2,mean)
```

```
##   HORASEM      EDAD    EDUCAC   INGRPER  INGRFLIA     ANTIG    ANTEMP
## 43.490909 35.509091 14.018182 26.487273 33.387273 16.000000  5.701091
```

```
dim(G2)
```

```
## [1] 55  7
```

```
G3=as.matrix(subset(Datos[,2:8],Datos$h3==3))
apply(G3,2,mean);dim(G3)
```

```
##  HORASEM     EDAD   EDUCAC  INGRPER INGRFLIA    ANTIG   ANTEMP
## 46.43243 45.08108 14.16216 34.17568 52.65676 27.08108 11.18865
```

```
## [1] 37  7
```

```
SCD1= sum(diag(var(G1))*(nrow(G1)-1))
SCD2= sum(diag(var(G2))*(nrow(G2)-1))
SCD3= sum(diag(var(G3))*(nrow(G3)-1))
SCD1;SCD2;SCD3
```

```
## [1] 3089.698
```

```
## [1] 24454.75
```

```
## [1] 30792.24
```

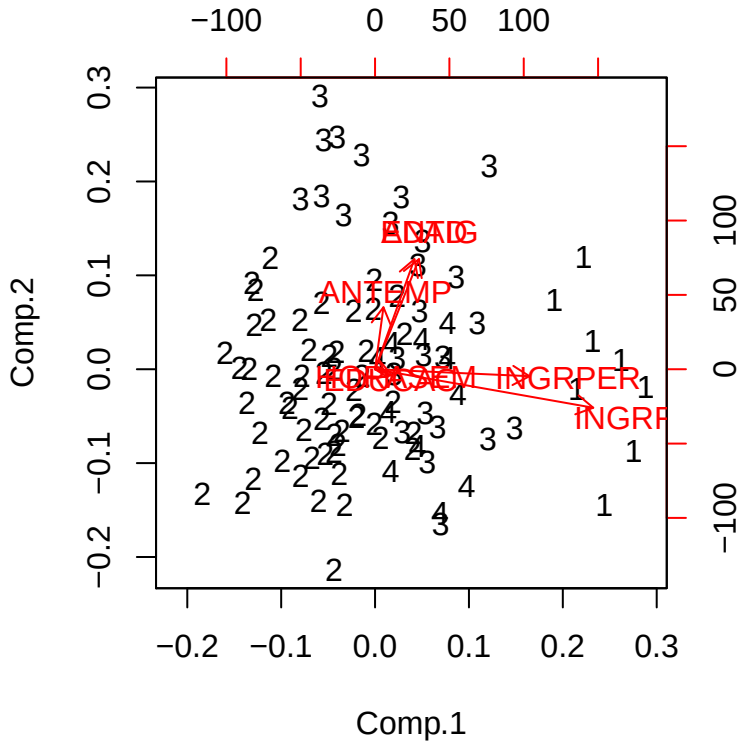
```
SCDG3=SCD1+SCD2+SCD3; SCDG3
```

```
## [1] 58336.69
```

```
#Resumen de los 4 grupos
summary(as.factor(cutree(HClust.1, k = 4)))# Cluster Sizes
```

```
##  1  2  3  4
##  8 55 26 11
```

```
## Generar una variable grupo(con 4 grupos)  h4
Datos$h4 <- c(as.factor(cutree(HClust.1, k = 4)))
biplot(princomp(model.matrix(~-1+ANTEMP+ANTIG+EDAD+EDUCAC+HORASEM+INGRFLIA+INGRPER,Datos)),
       xlabs=as.character(cutree(HClust.1,k=4)))
```



```r
# suma de cuadrados dentre de 4 grupos
G1=as.matrix(subset(Datos[,2:8],Datos$h4==1))
dim(G1)
```

```
## [1] 8 7
```

```r
G2=as.matrix(subset(Datos[,2:8],Datos$h4==2))
dim(G2)
```

```
## [1] 55  7
```

```r
G3=as.matrix(subset(Datos[,2:8],Datos$h4==3))
dim(G3)
```

```
## [1] 26  7
```

```r
G4=as.matrix(subset(Datos[,2:8],Datos$h4==4))
dim(G4)
```

```
## [1] 11  7
```

```r
SCD1= sum(diag(var(G1))*(nrow(G1))-1)
SCD2= sum(diag(var(G2))*(nrow(G2))-1)
SCD3= sum(diag(var(G3))*(nrow(G3))-1)
SCD4= sum(diag(var(G4))*(nrow(G4))-1)
SCD1;SCD2;SCD3;SCD4
```

```
## [1] 3524.084
```

```
## [1] 24900.62
```

```
## [1] 21260.78
```

```
## [1] 3063.879
```

```r
SCDG4=SCD1+SCD2+SCD3+SCD4; SCDG4
```

```
## [1] 52749.36
```

```r
#PARA COMPARAR SUMAS DE CUADRADOS

#calculamos el valor F =(SCDT(k grupos)-SCDT (k+1 grupos))/(SCDT(k+1 grupos)/n-k-1)
# de 2 a 3 grupos
F=(SCDG2-SCDG3)/(SCDG3/(100-2-1))
F
```

```
## [1] 24.87474
```

```r
# de 3 a 4 grupos
F=(SCDG3-SCDG4)/(SCDG4/(100-3-1))
F
```

```
## [1] 10.16853
```

```r
###actividad
#seleccionar la cantidad de conglomerados con otra medida diferente a la de Hartigan.

#Descripción de los grupos
M=apply(Datos[,2:8],2,mean)
S=diag(var(Datos[,2:8]))
D=nrow(Datos[,2:8])
```

```r
# 2 grupos
G1=as.matrix(subset(Datos[,2:8],Datos$h2==1))
dim(G1)
```

```
## [1] 45  7
```

```r
D1=nrow(G1)
M1=apply(G1,2,mean)
S1=diag(var(G1))
Sp=sqrt((S/D1)*((D-D1)/(D-1)));Sp
```

```
##   HORASEM      EDAD    EDUCAC   INGRPER  INGRFLIA     ANTIG    ANTEMP
## 1.0871146 1.1543176 0.3095902 1.7347267 2.1851370 1.1550871 0.8983221
```

```r
T1=(M1-M)/Sp;T1
```

```
##  HORASEM     EDAD   EDUCAC  INGRPER INGRFLIA    ANTIG   ANTEMP
## 1.584214 4.786859 1.073103 4.200406 6.440023 5.523393 3.010254
```

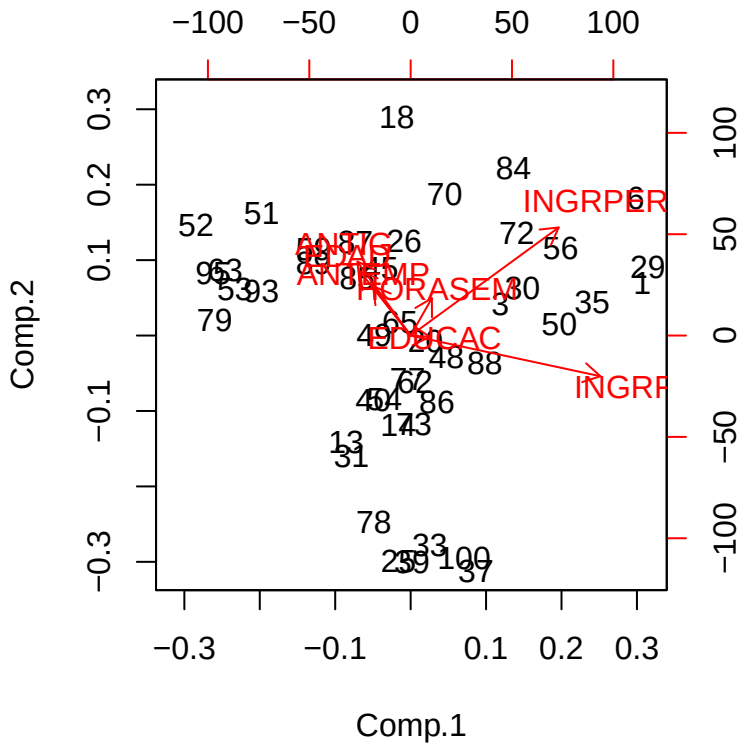
```r
gl=D1-1;gl
```

```
## [1] 44
```

```r
Pt1=pt(abs(T1),df=gl, lower.tail=FALSE)
round(cbind(T1,Pt1),4)
```

```
##              T1    Pt1
## HORASEM  1.5842 0.0602
## EDAD     4.7869 0.0000
## EDUCAC   1.0731 0.1445
## INGRPER  4.2004 0.0001
## INGRFLIA 6.4400 0.0000
## ANTIG    5.5234 0.0000
## ANTEMP   3.0103 0.0022
```

```r
P1=princomp(G1)
biplot(P1)
```




```
G2=as.matrix(subset(Datos[,2:8],Datos$h2==2))
dim(G2)
```

```
## [1] 55  7
```

```
D2=nrow(G2)
M2=apply(G2,2,mean)
S2=diag(var(G2))
Sp=sqrt((S/D2)*(D-D2)/(D-1));Sp
```

```
##   HORASEM      EDAD    EDUCAC   INGRPER  INGRFLIA     ANTIG    ANTEMP
## 0.8894574 0.9444417 0.2533011 1.4193218 1.7878393 0.9450713 0.7349908
```

```
T2=(M2-M)/Sp
gl=D2-1
Pt2=pt(abs(T2), df=gl, lower.tail=FALSE)
round(cbind(T2,Pt2),4)
```

```
##               T2    Pt2
## HORASEM  -1.5842 0.0595
## EDAD     -4.7869 0.0000
## EDUCAC   -1.0731 0.1440
## INGRPER  -4.2004 0.0001
## INGRFLIA -6.4400 0.0000
## ANTIG    -5.5234 0.0000
## ANTEMP   -3.0103 0.0020
```

```r
# 3 grupos
G1=as.matrix(subset(Datos[,2:8],Datos$h3==1))
dim(G1)
```

```
## [1] 8 7
```

```r
D1=nrow(G1)
M1=apply(G1,2,mean)
S1=diag(var(G1))
Sp=sqrt((S/D1)*((D-D1)/(D-1)));Sp
```

```
##   HORASEM      EDAD    EDUCAC   INGRPER  INGRFLIA     ANTIG    ANTEMP
## 3.3346416 3.5407818 0.9496446 5.3211426 6.7027419 3.5431421 2.7555350
```

```r
T1=(M1-M)/Sp;T1
```

```
##    HORASEM       EDAD     EDUCAC    INGRPER   INGRFLIA      ANTIG
## 0.77969398 2.18030947 2.59044286 6.20186354 6.45802879 2.47802652
##     ANTEMP
## 0.02318969
```

```r
gl=D1-1;gl
```

```
## [1] 7
```

```r
Pt1=pt(abs(T1),df=gl, lower.tail=FALSE)
round(cbind(T1,Pt1),4)
```

```
##              T1    Pt1
## HORASEM  0.7797 0.2306
## EDAD     2.1803 0.0328
## EDUCAC   2.5904 0.0180
## INGRPER  6.2019 0.0002
## INGRFLIA 6.4580 0.0002
## ANTIG    2.4780 0.0212
## ANTEMP   0.0232 0.4911
```

```r
G2=as.matrix(subset(Datos[,2:8],Datos$h3==2))
dim(G2)
```

```
## [1] 55  7
```

```r
D2=nrow(G2)
M2=apply(G2,2,mean)
S2=diag(var(G2))
Sp=sqrt((S/D2)*(D-D2)/(D-1));Sp
```

```
##   HORASEM      EDAD    EDUCAC   INGRPER  INGRFLIA     ANTIG    ANTEMP
## 0.8894574 0.9444417 0.2533011 1.4193218 1.7878393 0.9450713 0.7349908
```

```r
T2=(M2-M)/Sp
gl=D2-1
Pt2=pt(abs(T2), df=gl, lower.tail=FALSE)
round(cbind(T2,Pt2),4)
```

```
##               T2    Pt2
## HORASEM  -1.5842 0.0595
## EDAD     -4.7869 0.0000
## EDUCAC   -1.0731 0.1440
```

```
## INGRPER  -4.2004 0.0001
## INGRFLIA -6.4400 0.0000
## ANTIG    -5.5234 0.0000
## ANTEMP   -3.0103 0.0020
```

```
G3=as.matrix(subset(Datos[,2:8],Datos$h3==3))
dim(G3)
```

```
## [1] 37  7
```

```
D3=nrow(G3)
M3=apply(G3,2,mean)
S3=diag(var(G3))
Sp=sqrt((S/D3)*((D-D3)/(D-1)));Sp
```

```
##  HORASEM     EDAD   EDUCAC  INGRPER INGRFLIA    ANTIG   ANTEMP
## 1.283127 1.362447 0.365411 2.047507 2.579129 1.363355 1.060294
```

```
T3=(M3-M)/Sp;T3
```

```
##    HORASEM       EDAD     EDUCAC    INGRPER   INGRFLIA      ANTIG
##  1.1942952  3.7073594 -0.3498467  0.8433064  3.0071230  4.2990121
##     ANTEMP
##  3.0888109
```

```
gl=D3-1;gl
```

```
## [1] 36
```

```
Pt3=pt(abs(T3),df=gl, lower.tail=FALSE)
round(cbind(T3,Pt3),4)
```

```
##               T3    Pt3
## HORASEM   1.1943 0.1201
## EDAD      3.7074 0.0004
## EDUCAC   -0.3498 0.3642
## INGRPER   0.8433 0.2023
## INGRFLIA  3.0071 0.0024
## ANTIG     4.2990 0.0001
## ANTEMP    3.0888 0.0019
```

Agrupamiento K MEDIAS

```
# DOS GRUPOS
.cluster<-kmeans(model.matrix(~-1+ANTEMP+ANTIG+EDAD+EDUCAC+HORASEM+INGRFLIA+INGRPER,Datos),
                 centers=2,iter.max=10,nstart=10)
names(.cluster)
```

```
## [1] "cluster"      "centers"      "totss"        "withinss"
## [5] "tot.withinss" "betweenss"    "size"         "iter"
## [9] "ifault"
```

```
.cluster$size # Cluster Sizes
```

```
## [1] 32 68
```

```
.cluster$centers # Cluster Centroids
```

```
##      ANTEMP    ANTIG     EDAD   EDUCAC  HORASEM INGRFLIA  INGRPER
## 1 8.924062 24.34375 42.56250 15.62500 47.84375 66.55625 46.64688
## 2 7.438088 19.75000 38.83824 13.66176 43.51471 34.71029 25.76765
```

```
.cluster$withinss # Within Cluster Sum of Squares
```

```
## [1] 27680.9 39985.6
```

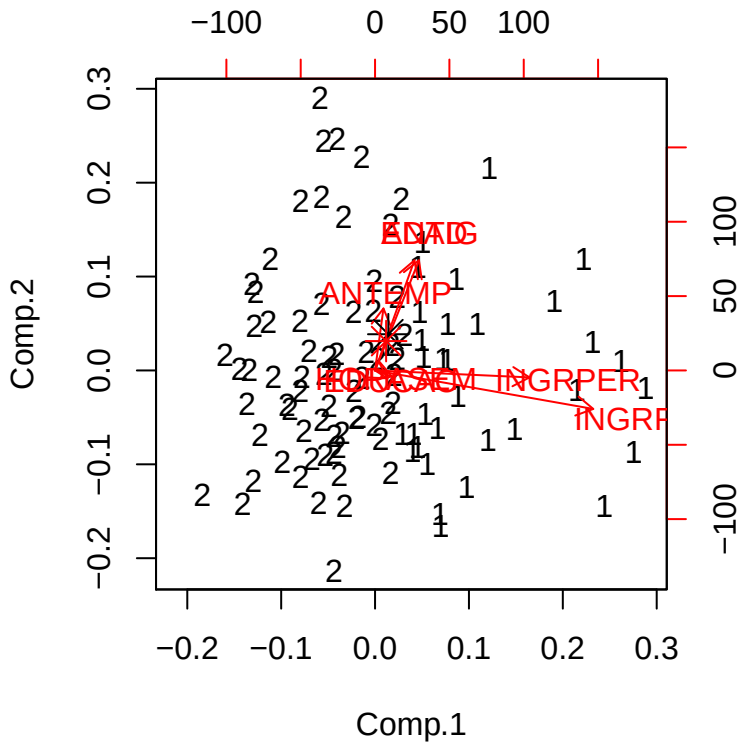
```
.cluster$tot.withinss # Total Within Sum of Squares
```

```
## [1] 67666.5
```

```
.cluster$betweenss # Between Cluster Sum of Squares
```

```
## [1] 32855.05
```

```
biplot(princomp(model.matrix(~-1+ANTEMP+ANTIG+EDAD+EDUCAC+HORASEM+INGRFLIA+INGRPER,Datos)),
       xlabs=as.character(.cluster$cluster))
points(.cluster$centers,col=1:4,pch=8,cex=2)
```




```
Datos$k2<-c(.cluster$cluster)

#TRES GRUPOS
.cluster<-kmeans(model.matrix(~-1+ANTEMP+ANTIG+EDAD+EDUCAC+HORASEM+INGRFLIA+INGRPER,Datos),
                 centers=3,iter.max=10,nstart=1)
.cluster$size # Cluster Sizes
```

```
## [1] 40 48 12
```

```
.cluster$centers # Cluster Centroids
```

```
##     ANTEMP    ANTIG     EDAD   EDUCAC  HORASEM INGRFLIA  INGRPER
## 1 8.491250 22.65000 41.22500 14.70000 44.12500 52.24000 35.35750
## 2 6.956042 18.06250 37.27083 13.35417 44.20833 30.33125 22.40208
## 3 9.818333 29.08333 47.08333 16.66667 50.25000 78.71667 62.94167
```

```
.cluster$withinss # Within Cluster Sum of Squares
```

```
## [1] 21786.236 25360.110  8038.533
```

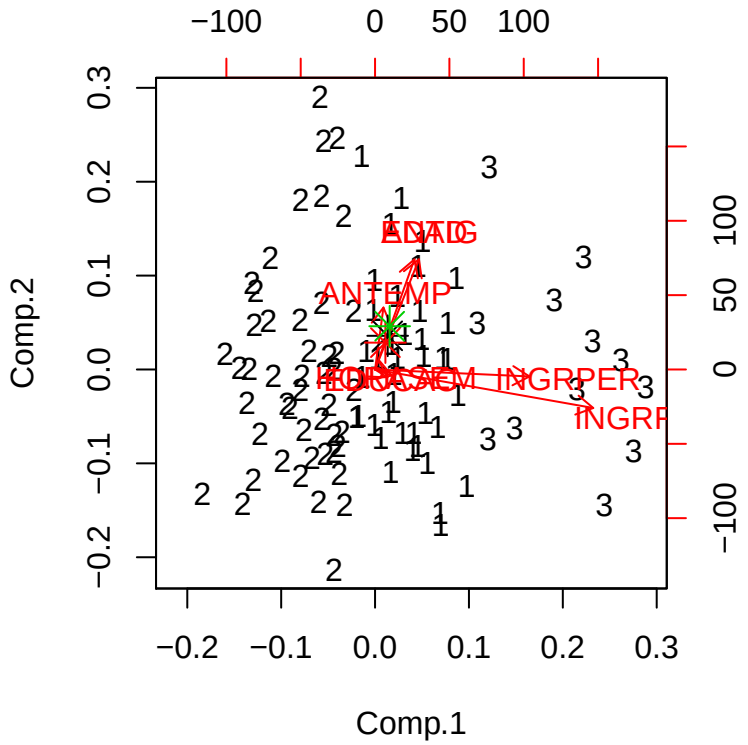
```
.cluster$tot.withinss # Total Within Sum of Squares
```

```
## [1] 55184.88
```

```
.cluster$betweenss # Between Cluster Sum of Squares
```

```
## [1] 45336.68
```

```
biplot(princomp(model.matrix(~-1+ANTEMP+ANTIG+EDAD+EDUCAC+HORASEM+INGRFLIA+INGRPER,Datos)),
       xlabs=as.character(.cluster$cluster))
points(.cluster$centers,col=1:4,pch=8,cex=2)
```




```
Datos$K3 <- c(.cluster$cluster)
```

K means recortado

```
####OTRO EJEMPLO CON KMEANS Y KMEANS recortado
load("indice.RDA")
n=nrow(indice)
names(indice)
```

```
##  [1] "EMPRESA"  "CONDICIO" "LIQACID"  "SOLVENC"  "PROPACT"  "PNOCOR"
##  [7] "AUTOFIN"  "INMACT"   "INMPN"    "RENTECO"  "MAREXP"   "REXP_INT"
```

```
# K MEDIAS
X=as.matrix(indice[1:n,2:8])

cl2 <- kmeans(X,2)
cl2
```

```
## K-means clustering with 2 clusters of sizes 3, 47
##
## Cluster means:
##   CONDICIO   LIQACID   SOLVENC  PROPACT    PNOCOR   AUTOFIN   INMACT
## 1 1.666667 742.16433 1279.0784 91.90774  2.821241  6.167939 51.10217
## 2 1.489362  78.35591  230.6136 48.53629 12.374457 -2.987081 62.28457
##
## Clustering vector:
##  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
##  2  1  2  2  1  2  2  2  2  2  2  1  2  2  2  2  2  2  2  2  2  2  2  2  2
## 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50
##  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2
##
## Within cluster sum of squares by cluster:
## [1] 174169.9 870779.3
##  (between_SS / total_SS =  80.6 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"
## [5] "tot.withinss" "betweenss"    "size"         "iter"
## [9] "ifault"
```

```
s2=sum(cl2$withinss);s2
```

```
## [1] 1044949
```

```
cl2$size
```

```
## [1]  3 47
```

```
cl3 <-kmeans(X,3)
cl3
```

```
## K-means clustering with 3 clusters of sizes 8, 3, 39
##
## Cluster means:
##   CONDICIO   LIQACID   SOLVENC  PROPACT    PNOCOR    AUTOFIN   INMACT
## 1 1.875000 160.27661  436.8398 75.19077  6.192613  0.2309873 56.25505
## 2 1.666667 742.16433 1279.0784 91.90774  2.821241  6.1679395 51.10217
## 3 1.410256  61.55167  188.3108 43.06870 13.642528 -3.6471977 63.52139
```

```
##
## Clustering vector:
##  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
##  3  2  3  3  2  3  3  3  3  1  3  2  3  3  3  3  3  1  3  3  1  3  3  1  3
## 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50
##  3  3  1  3  3  3  1  3  3  3  3  3  3  3  3  3  1  3  3  3  3  3  3  3  1
##
## Within cluster sum of squares by cluster:
## [1] 192506.0 174169.9 195877.1
##  (between_SS / total_SS =  89.6 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"
## [5] "tot.withinss" "betweenss"    "size"         "iter"
## [9] "ifault"
```

```
s3=sum(cl3$withinss);s3
```

```
## [1] 562553
```

```
cl3$size
```

```
## [1]  8  3 39
```

```
testF3= (s2-s3)/(s3/(n-2-1))
testF3
```

```
## [1] 40.30308
```

```
cl4 <-kmeans(X,4)
cl4
```

```
## K-means clustering with 4 clusters of sizes 3, 34, 11, 2
##
## Cluster means:
##   CONDICIO   LIQACID   SOLVENC  PROPACT     PNOCOR    AUTOFIN   INMACT
## 1 1.666667 742.16433 1279.0784 91.90774  2.8212408   6.167939 51.10217
## 2 1.382353  52.34859  176.8410 40.24455 13.7056421  -2.858011 63.67491
## 3 1.727273 140.11211  322.5720 67.65583 10.4612777 -10.307035 60.03730
## 4 2.000000 180.82124  638.9758 84.33828  0.2668003  35.078465 51.00858
##
## Clustering vector:
##  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
##  2  1  2  2  1  3  3  2  2  4  2  1  2  2  2  2  2  3  2  3  3  2  2  3  2
## 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50
##  2  2  4  2  2  2  3  2  3  3  2  2  2  2  2  2  3  2  2  2  2  2  2  2  3
##
## Within cluster sum of squares by cluster:
## [1] 174169.88 117288.53 104596.33  24916.41
##  (between_SS / total_SS =  92.2 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"
## [5] "tot.withinss" "betweenss"    "size"         "iter"
## [9] "ifault"
```

```r
s4=sum(cl4$withinss);s4
```

```
## [1] 420971.2
```

```r
cl4$size
```

```
## [1]  3 34 11  2
```

```r
testF4= (s3-s4)/(s4/(n-3-1))
testF4
```

```
## [1] 15.47081
```

```r
cl5 <-kmeans(X,5)
cl5
```

```
## K-means clustering with 5 clusters of sizes 7, 3, 17, 2, 21
##
## Cluster means:
##   CONDICIO   LIQACID   SOLVENC  PROPACT     PNOCOR    AUTOFIN   INMACT
## 1 1.714286 161.88446  352.1346 70.36508 10.3976444 -12.032387 58.30404
## 2 1.666667 742.16433 1279.0784 91.90774  2.8212408   6.167939 51.10217
## 3 1.294118  47.11992  141.5077 28.47016 16.9631575 -12.630218 60.66515
## 4 2.000000 180.82124  638.9758 84.33828  0.2668003  35.078465 51.00858
## 5 1.523810  66.04074  223.3482 54.09432 10.4718428   4.209080 65.99627
##
## Clustering vector:
##  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
##  5  2  3  3  2  5  1  5  3  4  3  2  5  5  3  5  3  1  3  5  1  3  5  1  5
## 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50
##  5  5  4  3  5  3  1  5  5  5  5  5  5  3  3  3  1  5  3  5  5  3  3  3  1
##
## Within cluster sum of squares by cluster:
## [1]  69853.59 174169.88  33632.18  24916.41  58804.24
##  (between_SS / total_SS =  93.3 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"
## [5] "tot.withinss" "betweenss"    "size"         "iter"
## [9] "ifault"
```

```r
s5=sum(cl5$withinss);s5
```

```
## [1] 361376.3
```

```r
cl5$size
```

```
## [1]  7  3 17  2 21
```

```r
testF5= (s4-s5)/(s5/(n-4-1))
testF5
```

```
## [1] 7.420986
```

```r
#tabla valores F y SCREs#
g=c()
s=c()
f=c()
```

```r
for (i in 2:6)
{cl <- kmeans(X,i)
w=c((cl$withinss))
s[i]=sum(cl$withinss)
f[i]=(s[(i-1)]-s[i])/(s[i]/(n-i))
g[i]=i}
ta=cbind(g,s,f)
w=c(w)

#SE DECIDE TRABAJAR CON 4 GRUPOS
cl4gru <-kmeans(X,4)
g1=cl4gru$cluster
cl4$size
```

```
## [1]  3 34 11  2
```

```r
#CLUSTER RECORTADO
library(RSKC)
```

```
## Loading required package: flexclust
```

```
## Loading required package: grid
```

```
## Loading required package: lattice
```

```
## Loading required package: modeltools
```

```
## Loading required package: stats4
```

```r
#?(RSKC)
clrec <- RSKC(X, ncl = 4, alpha = 0.10)
clrec
```

```
##
## Input:
## #obs= 50  #feature= 7
## L1= 12  alpha= 0.1
##
## Result:
## wbss: 270758.1
## trimmed cases: 2 5 10 12 28
## #non-zero weights: 7
##  4 clusters of sizes 17, 10, 16, 7
## Cluster labels: 1 2 3 3 2 4 4 1 3 2 3 2 1 1 3 4 3 2 1 4 4 3 1 2 1 1 1 2 3 1 3 2 1 4 4 1 1 1 3 3 3 2
```

```r
table(clrec$labels )
```

```
##
##  1  2  3  4
## 17 10 16  7
```

```r
#(el grupo de pertenencia)
clrec$oW
```

```
## [1]  2  5 10 12 28
```

```r
#(observaciones eliminadas)


dim(indice)
```

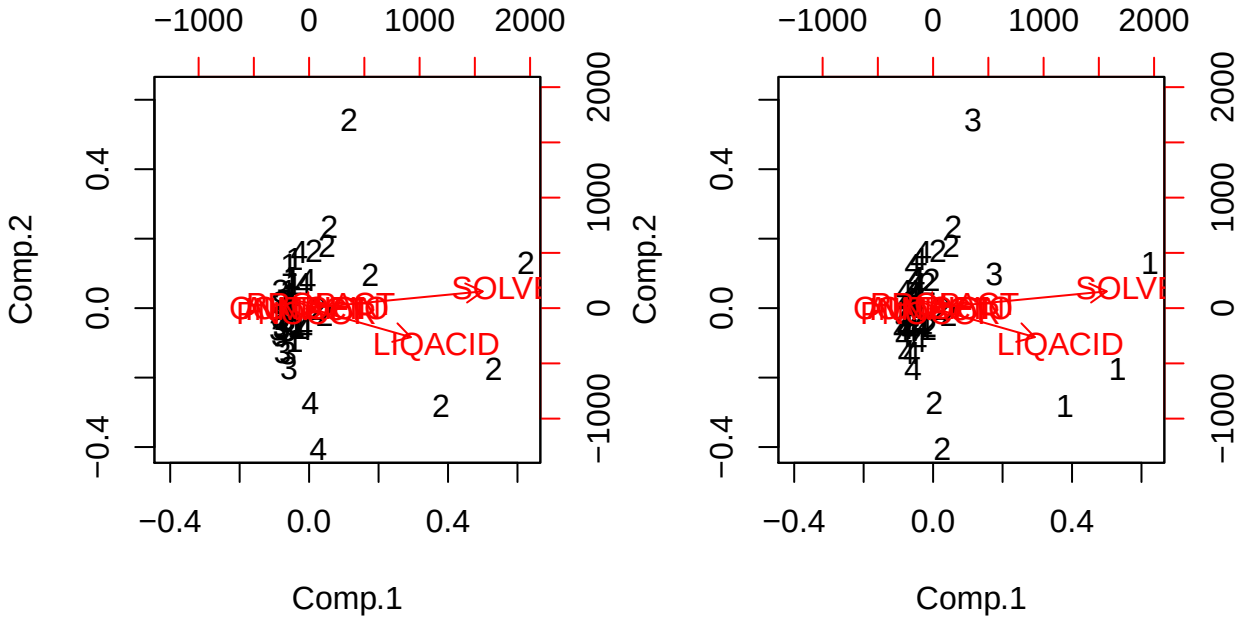
```
## [1] 50 12
```

```
g2=clrec$labels

indice=cbind(indice,g1,g2)
par(mfrow=c(1,2))
biplot(princomp(indice[,2:8]), xlabs=as.character(clrec$labels))
```

```
## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], length
## = arrow.len): zero-length arrow is of indeterminate angle and so skipped
```

```
biplot(princomp(indice[,2:8]), xlabs=as.character(cl4gru$cluster))
```

```
## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], length
## = arrow.len): zero-length arrow is of indeterminate angle and so skipped
```



K modas

```
####K modes

#install.packages"klaR")
require("klaR")
```

```
## Loading required package: klaR
```

```
## Loading required package: MASS
```

```
#edit(kmodes)
load("alumnos_2018.Rdata")
x_acat=alumnos_2018[,c("Sexo","Edad_cat","IMC_cat","Vive_en","vivienda","con_quien",
"procedencia","trabaja","Fuma","actividad física","calificacion_primer_año",
"aprobadas_cat","aplazos_cat","matematicas","Nota Mate I_cat",
"Nota Mate II_cat","Libros_cat","horas_mate_cat","horas_adm_cat",
"autocalificación")]
names(x_acat)
```

```
##  [1] "Sexo"                    "Edad_cat"
##  [3] "IMC_cat"                 "Vive_en"
##  [5] "vivienda"                "con_quien"
##  [7] "procedencia"             "trabaja"
##  [9] "Fuma"                    "actividad física"
## [11] "calificacion_primer_año" "aprobadas_cat"
## [13] "aplazos_cat"             "matematicas"
## [15] "Nota Mate I_cat"         "Nota Mate II_cat"
## [17] "Libros_cat"              "horas_mate_cat"
## [19] "horas_adm_cat"           "autocalificación"
```

```
dim(x_acat)
```

```
## [1] 173  20
```

```
## run algorithm on x:
cl <- kmodes(x_acat,3)
cl
```

```
## K-modes clustering with 3 clusters of sizes 78, 37, 58
##
## Cluster modes:
##    Sexo Edad_cat     IMC_cat      Vive_en           vivienda     con_quien
## 1 Varón  20 a 21 Peso Normal         Casa             Propia        Padres
## 2 Mujer  22 a 23   Sobrepeso         Casa             Propia        Padres
## 3 Varón  20 a 21 Peso Normal Departamento Prestada/Alquilada Otro familiar
##             procedencia    trabaja Fuma actividad física
## 1       Córdoba Capital No trabaja   No               Si
## 2       Córdoba Capital No trabaja   No               Si
## 3 Otras Pcias. / Países No trabaja   No               Si
##   calificacion_primer_año aprobadas_cat aplazos_cat matematicas
## 1                 Regular        6 a 10     Hasta 5   Regulares
## 2                 Regular       11 a 15     Hasta 5     Fáciles
## 3                 Regular        6 a 10     Hasta 5   Regulares
##   Nota Mate I_cat Nota Mate II_cat  Libros_cat horas_mate_cat
## 1           Bueno            Bueno No solicita        16 a 30
## 2     Distinguido      Distinguido No solicita       Hasta 15
## 3           Bueno      Distinguido No solicita       Hasta 15
##   horas_adm_cat autocalificación
## 1       16 a 30            Bueno
## 2      Hasta 15          Regular
## 3      Hasta 15          Regular
##
## Clustering vector:
##   [1] 3 2 2 3 2 1 3 2 1 2 2 3 3 3 3 1 2 3 2 1 1 2 1 1 1 3 1 1 3 1 1 3 2 2 1
##  [36] 3 3 3 2 1 1 1 1 2 2 1 1 2 1 3 2 3 1 1 1 2 3 1 3 1 1 1 1 1 2 1 1 3 1 3
```

```
##  [71] 3 1 3 1 1 1 2 3 1 1 1 3 2 1 1 1 3 3 2 3 1 1 2 1 1 3 1 1 3 3 3 3 1 2 3
## [106] 1 2 3 1 1 1 3 1 2 1 1 3 1 2 1 1 1 3 3 3 3 3 3 3 2 3 2 1 1 1 1 2 3 2 1
## [141] 1 3 3 2 3 1 1 1 1 1 2 2 1 3 3 3 1 3 2 1 1 3 2 3 3 2 1 3 3 3 1 1 2
##
## Within cluster simple-matching distance by cluster:
## [1] 538 270 427
##
## Available components:
## [1] "cluster"    "size"       "modes"      "withindiff" "iterations"
## [6] "weighted"
```

```
clasif_2018=cbind(alumnos_2018,cl$cluster)
write.csv(clasif_2018,"clasif_2018")
```

K prototipos

```
####K Proptotipes

#install.packages("clustMixType")
require("clustMixType")
```

```
## Loading required package: clustMixType
```

```
x_p=alumnos_2018
dim(x_p)
```

```
## [1] 173  36
```

```
names(x_p)
```

```
##  [1] "año"                          "Ultimas tres cifras del DNI"
##  [3] "Sexo"                         "Edad"
##  [5] "Edad_cat"                     "Peso"
##  [7] "Estatura"                     "IMC"
##  [9] "IMC_cat"                      "Vive_en"
## [11] "vivienda"                     "con_quien"
## [13] "procedencia"                  "trabaja"
## [15] "trabaja_1"                    "calificacion_primer_año"
## [17] "aprobadas"                    "aprobadas_cat"
## [19] "aplazos"                      "aplazos_cat"
## [21] "matematicas"                  "Nota_MatemáticaI"
## [23] "Nota Mate I_cat"              "Nota_MatemáticaII"
## [25] "Nota Mate II_cat"             "Cuántas_libro_año"
## [27] "Libros_cat"                   "gasto promedio libros"
## [29] "gasto promedio Movilidad"     "horas_matemáticas"
## [31] "horas_mate_cat"               "horas_administrativa contable"
## [33] "horas_adm_cat"                "autocalificación"
## [35] "Fuma"                         "actividad física"
```

```
Sexo=as.factor(x_p$Sexo)
Vive_en= as.factor(x_p$Vive_en)
vivienda= as.factor(x_p$vivienda)
con_quien= as.factor(x_p$con_quien)
procedencia= as.factor(x_p$procedencia)
```

```r
trabaja= as.factor(x_p$trabaja)
calificacion_primer_anio= as.factor(x_p$"calificacion_primer_año")
matematicas= as.factor(x_p$matematicas)
autocalificacion= as.factor(x_p$"autocalificación")
Fuma= as.factor(x_p$Fuma)
Nota_MatI= as.factor(x_p$"Nota Mate I_cat")
Nota_MatII= as.factor(x_p$"Nota Mate II_cat")
actividad_fisica=as.factor(x_p$"actividad física")


x_proto=cbind.data.frame(x_p$Edad,x_p$IMC,x_p$aprobadas,
x_p$aplazos,x_p$"Cuántas_libro_año",x_p$"horas_matemáticas",
x_p$"horas_administrativa contable",Sexo,
Vive_en,vivienda,con_quien,procedencia,trabaja,
calificacion_primer_anio,matematicas,autocalificacion,
Fuma,actividad_fisica,Nota_MatI,Nota_MatII)
dim(x_proto)
```

```
## [1] 173  20
```

```r
names(x_proto)
```

```
##  [1] "x_p$Edad"
##  [2] "x_p$IMC"
##  [3] "x_p$aprobadas"
##  [4] "x_p$aplazos"
##  [5] "x_p$Cuántas_libro_año"
##  [6] "x_p$horas_matemáticas"
##  [7] "x_p$\"horas_administrativa contable\""
##  [8] "Sexo"
##  [9] "Vive_en"
## [10] "vivienda"
## [11] "con_quien"
## [12] "procedencia"
## [13] "trabaja"
## [14] "calificacion_primer_anio"
## [15] "matematicas"
## [16] "autocalificacion"
## [17] "Fuma"
## [18] "actividad_fisica"
## [19] "Nota_MatI"
## [20] "Nota_MatII"
```

```r
summary(x_proto)
```

```
##     x_p$Edad        x_p$IMC      x_p$aprobadas     x_p$aplazos
##  Min.   :18.00   Min.   :16.14   Min.   : 1.000   Min.   : 0.000
##  1st Qu.:20.00   1st Qu.:20.82   1st Qu.: 7.000   1st Qu.: 2.000
##  Median :21.00   Median :23.08   Median : 8.000   Median : 4.000
##  Mean   :21.76   Mean   :23.37   Mean   : 9.312   Mean   : 4.445
##  3rd Qu.:23.00   3rd Qu.:25.39   3rd Qu.:11.000   3rd Qu.: 5.000
##  Max.   :47.00   Max.   :38.45   Max.   :25.000   Max.   :25.000
##  x_p$Cuántas_libro_año x_p$horas_matemáticas
##  Min.   : 0.000        Min.   :  0.01
##  1st Qu.: 0.000        1st Qu.: 12.00
```

```
##  Median : 1.500        Median : 24.00
##  Mean   : 2.012        Mean   : 31.47
##  3rd Qu.: 2.000        3rd Qu.: 42.00
##  Max.   :30.000        Max.   :120.00
##  x_p$"horas_administrativa contable"    Sexo             Vive_en
##  Min.   :  0.00                      Mujer: 64   Casa        :98
##  1st Qu.: 10.00                      Varón:109   Departamento:75
##  Median : 24.00
##  Mean   : 30.86
##  3rd Qu.: 40.00
##  Max.   :320.00
##                vivienda          con_quien                  procedencia
##  Prestada/Alquilada:79   Otro familiar:39   Córdoba Capital      :80
##  Propia            :94   Otros        :14   Córdoba Interior     :45
##                          Padres       :92   Otras Pcias. / Países:48
##                          Solo         :28
##
##
##         trabaja    calificacion_primer_anio    matematicas
##  Hasta 4 Hs : 27   Difícil: 20              Difíciles: 14
##  Más de 4 Hs: 33   Fácil  : 22              Fáciles  : 53
##  No trabaja :113   Regular:131              Regulares:106
##
##
##
##  autocalificacion Fuma     actividad_fisica         Nota_MatI
##  Bueno  :90       No:151   No: 42           Bueno        :71
##  Malo   : 4       Si: 22   Si:131           Distinguido  :77
##  Regular:79                                 Insuficente  : 1
##                                             No contesta  : 3
##                                             Sobresaliente: 7
##                                             Suficiente   :14
##          Nota_MatII
##  Bueno        :53
##  Distinguido  :67
##  Insuficiente : 3
##  No rindiò    :35
##  Sobresaliente: 3
##  Suficiente   :12
```

```
kproto <- kproto(x_proto, 3)
```

```
## # NAs in variables:
##                             x_p$Edad                              x_p$IMC
##                                    0                                    0
##                         x_p$aprobadas                          x_p$aplazos
##                                    0                                    0
##                 x_p$Cuántas_libro_año                x_p$horas_matemáticas
##                                    0                                    0
## x_p$"horas_administrativa contable"                                 Sexo
##                                    0                                    0
##                               Vive_en                              vivienda
##                                    0                                    0
##                             con_quien                            procedencia
##                                    0                                    0
```

```
##                              trabaja            calificacion_primer_anio
##                                    0                                   0
##                           matematicas                    autocalificacion
##                                    0                                   0
##                                 Fuma                    actividad_fisica
##                                    0                                   0
##                             Nota_MatI                          Nota_MatII
##                                    0                                   0
## 0 observation(s) with NAs.
##
## Estimated lambda: 541.853
```

```
kproto
```

```
## Numeric predictors: 7
## Categorical predictors: 13
## Lambda: 541.853
##
## Number of Clusters: 3
## Cluster sizes: 21 90 62
## Within cluster error: 129826.1 242409.6 186292.6
##
## Cluster prototypes:
##    x_p$Edad  x_p$IMC x_p$aprobadas x_p$aplazos x_p$Cuántas_libro_año
## 67 21.38095 22.83575      8.142857    4.761905              2.238095
## 97 21.66667 23.71401      9.744444    4.426667              2.155556
## 35 22.03226 23.04240      9.080645    4.364516              1.725806
##    x_p$horas_matemáticas x_p$"horas_administrativa contable"  Sexo
## 67              85.04762                            93.14286 Varón
## 97              23.51133                            21.62244 Varón
## 35              24.88774                            23.17806 Varón
##         Vive_en           vivienda     con_quien           procedencia
## 67         Casa             Propia        Padres       Córdoba Capital
## 97         Casa             Propia        Padres       Córdoba Capital
## 35 Departamento Prestada/Alquilada Otro familiar Otras Pcias. / Países
##       trabaja calificacion_primer_anio matematicas autocalificacion Fuma
## 67 No trabaja                  Regular   Regulares            Bueno   No
## 97 No trabaja                  Regular   Regulares            Bueno   No
## 35 No trabaja                  Regular   Regulares          Regular   No
##    actividad_fisica   Nota_MatI  Nota_MatII
## 67               Si       Bueno       Bueno
## 97               Si Distinguido Distinguido
## 35               Si       Bueno Distinguido
```

```
summary(kproto)
```

```
## x_p$Edad
##   Min. 1st Qu. Median     Mean 3rd Qu. Max.
## 1   18      19     20 21.38095      22   31
## 2   18      20     21 21.66667      23   47
## 3   18      20     21 22.03226      23   42
##
## ------------------------------------------------------------
## x_p$IMC
##       Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
```

```
## 1 16.13539 18.42404 21.93635 22.83575 25.10388 38.44675
## 2 17.63085 21.49175 23.51861 23.71401 25.45704 33.95062
## 3 18.49650 20.70139 22.75831 23.04240 25.02344 32.50852
##
## ------------------------------------------------------------------
## x_p$aprobadas
##   Min. 1st Qu. Median     Mean 3rd Qu. Max.
## 1    5    7.00      7 8.142857       8   14
## 2    1    8.00      9 9.744444      12   24
## 3    3    6.25      8 9.080645      10   25
##
## ------------------------------------------------------------------
## x_p$aplazos
##   Min. 1st Qu. Median     Mean 3rd Qu. Max.
## 1    1    2.00      4 4.761905       5   17
## 2    0    2.25      4 4.426667       5   20
## 3    0    2.00      4 4.364516       5   25
##
## ------------------------------------------------------------------
## x_p$Cuántas_libro_año
##   Min. 1st Qu. Median     Mean 3rd Qu. Max.
## 1    0       0    1.5 2.238095       3   10
## 2    0       0    1.0 2.155556       2   30
## 3    0       0    1.5 1.725806       2   10
##
## ------------------------------------------------------------------
## x_p$horas_matemáticas
##    Min. 1st Qu. Median     Mean 3rd Qu. Max.
## 1 50.00    70.0   72.0 85.04762     120  120
## 2  0.01    10.5   20.0 23.51133      30   60
## 3  0.01    10.0   20.5 24.88774      36   70
##
## ------------------------------------------------------------------
## x_p$"horas_administrativa contable"
##    Min. 1st Qu. Median     Mean 3rd Qu. Max.
## 1 48.00      70     80 93.14286   90.00  320
## 2  0.01      10     20 21.62244   30.00   60
## 3  0.00      10     20 23.17806   33.75   72
##
## ------------------------------------------------------------------
## Sexo
##
## cluster Mujer Varón
##       1 0.095 0.905
##       2 0.367 0.633
##       3 0.468 0.532
##
## ------------------------------------------------------------------
## Vive_en
##
## cluster  Casa Departamento
##       1 0.619        0.381
##       2 0.933        0.067
##       3 0.016        0.984
```

```
##
## ------------------------------------------------------------------
## vivienda
##
## cluster Prestada/Alquilada Propia
##       1              0.476  0.524
##       2              0.133  0.867
##       3              0.919  0.081
##
## ------------------------------------------------------------------
## con_quien
##
## cluster Otro familiar Otros Padres  Solo
##       1         0.143 0.095  0.524 0.238
##       2         0.033 0.022  0.878 0.067
##       3         0.532 0.161  0.032 0.274
##
## ------------------------------------------------------------------
## procedencia
##
## cluster Córdoba Capital Córdoba Interior Otras Pcias. / Países
##       1           0.524            0.143                 0.333
##       2           0.733            0.200                 0.067
##       3           0.048            0.387                 0.565
##
## ------------------------------------------------------------------
## trabaja
##
## cluster Hasta 4 Hs Más de 4 Hs No trabaja
##       1      0.048       0.095      0.857
##       2      0.211       0.300      0.489
##       3      0.113       0.065      0.823
##
## ------------------------------------------------------------------
## calificacion_primer_anio
##
## cluster Difícil Fácil Regular
##       1   0.095 0.048   0.857
##       2   0.100 0.200   0.700
##       3   0.145 0.048   0.806
##
## ------------------------------------------------------------------
## matematicas
##
## cluster Difíciles Fáciles Regulares
##       1     0.095   0.143     0.762
##       2     0.067   0.322     0.611
##       3     0.097   0.339     0.565
##
## ------------------------------------------------------------------
## autocalificacion
##
## cluster Bueno  Malo Regular
##       1 0.571 0.000   0.429
```

```
##       2 0.578 0.022   0.400
##       3 0.419 0.032   0.548
##
## ------------------------------------------------------------------
## Fuma
##
## cluster    No    Si
##       1 0.810 0.190
##       2 0.911 0.089
##       3 0.839 0.161
##
## ------------------------------------------------------------------
## actividad_fisica
##
## cluster    No    Si
##       1 0.190 0.810
##       2 0.267 0.733
##       3 0.226 0.774
##
## ------------------------------------------------------------------
## Nota_MatI
##
## cluster Bueno Distinguido Insuficente No contesta Sobresaliente Suficiente
##       1 0.571       0.381       0.000       0.000         0.000      0.048
##       2 0.333       0.511       0.011       0.022         0.044      0.078
##       3 0.468       0.371       0.000       0.016         0.048      0.097
##
## ------------------------------------------------------------------
## Nota_MatII
##
## cluster Bueno Distinguido Insuficiente No rindiò Sobresaliente Suficiente
##       1 0.333       0.333        0.048     0.143         0.048      0.095
##       2 0.333       0.389        0.022     0.178         0.011      0.067
##       3 0.258       0.403        0.000     0.258         0.016      0.065
##
## ------------------------------------------------------------------
```

```
names(kproto)
```

```
##  [1] "cluster"      "centers"      "lambda"       "size"
##  [5] "withinss"     "tot.withinss" "dists"        "iter"
##  [9] "trace"        "data"
```

```
kproto$cluster
```

```
##   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18
##   3   2   3   3   2   2   3   2   1   3   2   3   3   3   3   2   2   3
##  19  20  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35  36
##   2   2   2   2   1   2   2   3   2   1   3   2   1   1   2   2   2   3
##  37  38  39  40  41  42  43  44  45  46  47  48  49  50  51  52  53  54
##   1   3   2   3   2   3   2   2   1   2   2   2   2   3   3   1   3   2
##  55  56  57  58  59  60  61  62  63  64  65  66  67  68  69  70  71  72
##   1   2   3   1   3   2   3   2   2   1   2   2   2   3   2   3   3   2
##  73  74  75  76  77  78  79  80  81  82  83  84  85  86  87  88  89  90
##   3   2   2   2   3   3   2   2   2   2   2   2   2   3   3   3   3   3
```

```
##  91  92  93  94  95  96  97  98  99 100 101 102 103 104 105 106 107 108
##   2   2   2   2   2   1   2   2   3   3   3   3   2   2   3   2   2   3
## 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126
##   1   2   2   1   2   2   1   2   3   2   2   2   2   2   3   3   3   3
## 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144
##   3   3   2   2   3   2   2   1   2   2   2   3   3   2   1   3   3   2
## 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162
##   1   2   2   2   2   2   2   2   2   3   1   1   2   3   2   2   1   3
## 163 164 165 166 167 168 169 170 171 172 173
##   3   3   3   3   2   3   3   3   2   2   3
```

```
clasif_2018 <- read.csv("clasif_2018")
dim(clasif_2018)
```

```
## [1] 173  38
```

```
clasif_2018=cbind(clasif_2018,kproto$cluster)
write.csv(clasif_2018,"clasif_2018")
```

‘‘‘