Anexo 2.2. Segmentacion de hogares ENIGH

Juan Carlos Martinez-Ovando
ITAM

Se presenta y comenta el codigo para segmentar los hogares de la ENIGH 2016

Preambulo

Leemos los paquetes necesarios para realizar la segmentacion de los datos generados de la ENIGH.

```
source("./Code/load.R.packages.R")
```

```
## Loading required package: digest

## Loading required package: devtools

## Loading required package: abind

## Loading required package: aplpack

## Loading required package: tcltk

## Loading required package: bayesQR

## Loading required package: car

## Loading required package: carData

## Loading required package: class

## Loading required package: colorspace

## Loading required package: data.table

## Loading required package: doBy

## Loading required package: DPpackage

##

## DPpackage 1.1-7.4

##

## Copyright (C) 2006 - 2012, Alejandro Jara
```

```
## Department of Statistics

## P.U. Catolica de Chile

##

## Support provided by Fondecyt

## 11100144 grant.

##

## Loading required package: GB2

## Loading required package: effects

## lattice theme set by effectsTheme()
## See ?effectsTheme for details.

## Loading required package: foreign

## Loading required package: GPDPQuantReg

## Loading required package: pscl

## Classes and Methods for R developed in the
## Political Science Computational Laboratory
## Department of Political Science
## Stanford University
## Simon Jackman
## hurdle and zeroinfl functions by Achim Zeileis

## Loading required package: tmvtnorm

## Loading required package: mvtnorm

## Loading required package: Matrix

## Loading required package: stats4

## Loading required package: gmm

## Loading required package: sandwich

## Loading required package: coda

## Loading required package: reshape

##
## Attaching package: 'reshape'
```

```
## The following object is masked from 'package:Matrix':
##
##     expand

## The following object is masked from 'package:data.table':
##
##     melt

## The following object is masked from 'package:class':
##
##     condense

## Loading required package: lattice

## Loading required package: doParallel

## Loading required package: foreach

## Loading required package: iterators

## Loading required package: parallel

## Loading required package: Hmisc

## Loading required package: survival

## Loading required package: Formula

## Loading required package: ggplot2

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':
##
##     format.pval, units

## Loading required package: kamila

## Loading required package: leaps

## Loading required package: lmtest

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
## Loading required package: memisc

## Loading required package: MASS

##
## Attaching package: 'memisc'

## The following objects are masked from 'package:Hmisc':
##
##     %nin%, html

## The following object is masked from 'package:foreach':
##
##     foreach

## The following object is masked from 'package:reshape':
##
##     rename

## The following object is masked from 'package:Matrix':
##
##     as.array

## The following object is masked from 'package:car':
##
##     recode

## The following objects are masked from 'package:stats':
##
##     contr.sum, contr.treatment, contrasts

## The following object is masked from 'package:base':
##
##     as.array

## Loading required package: monomvn

## Loading required package: pls

##
## Attaching package: 'pls'

## The following object is masked from 'package:stats':
##
##     loadings

## Loading required package: lars

## Loaded lars 1.2

## Loading required package: multcomp
```

```
## Loading required package: TH.data

##
## Attaching package: 'TH.data'

## The following object is masked from 'package:MASS':
##
##     geyser

## Loading required package: quantreg

## Loading required package: SparseM

##
## Attaching package: 'SparseM'

## The following object is masked from 'package:base':
##
##     backsolve

##
## Attaching package: 'quantreg'

## The following object is masked from 'package:Hmisc':
##
##     latex

## The following object is masked from 'package:survival':
##
##     untangle.specials

## Loading required package: Rcpp

## Loading required package: relimp

## Loading required package: sampling

##
## Attaching package: 'sampling'

## The following objects are masked from 'package:survival':
##
##     cluster, strata

## Loading required package: survey

## Loading required package: grid

##
## Attaching package: 'survey'
```

```
## The following object is masked from 'package:Hmisc':
##
##     deff

## The following object is masked from 'package:graphics':
##
##     dotchart

## Loading required package: yaml
```

```
load("../Datos.Modelo/mdi_variables_modelo.RData")
```

Los datos con los que trabajaremos son:

- `hogares_enigh_agr.csv` – Tabla de hogares de la ENIGH 2016

Preparacion

```
hogares_agr <- read.csv("../Bases.Enigh/hogares_enigh_agr.csv", header=TRUE)
colnames(hogares_agr)
```

```
##  [1] "X"           "FOLIOVIV"    "FOLIOHOG"    "tam_loc"     "sexo_jefe"
##  [6] "edad_jefe"   "educa_jefe"  "tot_integ"   "p12_64"      "p65mas"
## [11] "remesas"     "ing_cor"     "int0a12"     "int12a64"    "int65a98"
## [16] "depdemog"    "muj12a49"    "tot_per"     "ltot_per"    "p_esc3"
## [21] "p_esc5b"     "trab_sub"    "trab_ind"    "trab_s_pago" "seg_alim2"
## [26] "seg_alim3"   "seg_alim_a"  "seg_pop"     "ss"          "jtrab_ind"
## [31] "ssjtrabind"  "con_remesas" "viv_prop"    "viv_rent"    "tot_cuar"
## [36] "bao13"       "piso_fir"    "piso_rec"    "combustible" "sin_refri"
## [41] "sin_vehi"    "sin_compu"   "sin_vidvd"   "sin_telef"   "sin_horno"
## [46] "factor"      "rururb"      "tamhogesc"   "ict"         "ictpc"
```

```
dim(hogares_agr)
```

```
## [1] 43609    50
```

Ingreso corriente (dos versiones)

Variables ENIGH-CUIS

```
# Numeric
hogares_agr[,var_enighcuis_num] <- lapply(hogares_agr[,var_enighcuis_num],
                                          as.numeric)

# Categoric
hogares_agr[,var_enighcuis_cat] <- lapply(hogares_agr[,var_enighcuis_cat],
```

```
                                              as.numeric)

hogares_agr <- as.data.frame(hogares_agr)

table(hogares_agr[,c("rururb","tam_loc")])
```

```
##       tam_loc
## rururb     1     2     3     4
##      0 14364  6018  6960     0
##      1     0     0     0 16267
```

```
table(hogares_agr$rururb)
```

```
##
##     0     1
## 27342 16267
```

Parte I - Segmentacion Rural

```
# Numerical
hogares_rur_num <- hogares_agr[which(hogares_agr$rururb==1),
                                              var_enighcuis_seg_num]
hogares_rur_num <- lapply(hogares_rur_num,as.numeric)
hogares_rur_num <- as.data.frame(hogares_rur_num)

# Categorical
hogares_rur_cat <- hogares_agr[which(hogares_agr$rururb==1),
                                              var_enighcuis_seg_cat]
hogares_rur_cat <- lapply(hogares_rur_cat,factor)
hogares_rur_cat <- as.data.frame(hogares_rur_cat)

# Segmentacion rural
modelo_seg_rur <- kamila(hogares_rur_num,
                         hogares_rur_cat,
                         numClust=5,
                         numInit=10)

summary(modelo_seg_rur)
```

```
##              Length Class  Mode
## finalMemb    16267  -none- numeric
## numIter         10  -none- numeric
## finalLogLik      1  -none- numeric
## finalObj         1  -none- numeric
## finalCenters    60  -none- numeric
## finalProbs       8  -none- list
```

```
## input           10  -none- list
## verbose          0  -none- list
## nClust           0  -none- list
```

```
table(modelo_seg_rur$finalMemb)
```

```
##
##    1    2    3    4    5
## 3399 1807 3624 2453 4984
```

Unimos los clasificaciones

```
hogares_agr_rur <- hogares_agr[which(hogares_agr$rururb==1),]
dim(hogares_agr_rur)
```

```
## [1] 16267    50
```

```
finalMemb_rur <- as.data.frame(modelo_seg_rur$finalMemb)
dim(finalMemb_rur)
```

```
## [1] 16267     1
```

```
colnames(finalMemb_rur) <- c("finalMemb_rur")

hogares_agr_rur <- cbind(hogares_agr_rur,finalMemb_rur)
colnames(hogares_agr_rur)
```

```
##  [1] "X"             "FOLIOVIV"      "FOLIOHOG"      "tam_loc"
##  [5] "sexo_jefe"     "edad_jefe"     "educa_jefe"    "tot_integ"
##  [9] "p12_64"        "p65mas"        "remesas"       "ing_cor"
## [13] "int0a12"       "int12a64"      "int65a98"      "depdemog"
## [17] "muj12a49"      "tot_per"       "ltot_per"      "p_esc3"
## [21] "p_esc5b"       "trab_sub"      "trab_ind"      "trab_s_pago"
## [25] "seg_alim2"     "seg_alim3"     "seg_alim_a"    "seg_pop"
## [29] "ss"            "jtrab_ind"     "ssjtrabind"    "con_remesas"
## [33] "viv_prop"      "viv_rent"      "tot_cuar"      "bao13"
## [37] "piso_fir"      "piso_rec"      "combustible"   "sin_refri"
## [41] "sin_vehi"      "sin_compu"     "sin_vidvd"     "sin_telef"
## [45] "sin_horno"     "factor"        "rururb"        "tamhogesc"
## [49] "ict"           "ictpc"         "finalMemb_rur"
```

```
100 * table(hogares_agr_rur$finalMemb_rur) /
  sum(table(hogares_agr_rur$finalMemb_rur))
```

```
##
##        1        2        3        4        5
## 20.89506 11.10838 22.27823 15.07961 30.63872
```

```
write.csv(hogares_agr_rur,file="../Datos.Modelo/hogares_agr_rur.csv")
```

Parte II - Segmentacion Urbano

```
# Numerical
hogares_urb_num <- hogares_agr[which(hogares_agr$rururb==0),
                               var_enighcuis_seg_num]
hogares_urb_num <- lapply(hogares_urb_num,as.numeric)
hogares_urb_num <- as.data.frame(hogares_urb_num)

# Categorical
hogares_urb_cat <- hogares_agr[which(hogares_agr$rururb==0),
                               var_enighcuis_seg_cat]
hogares_urb_cat <- lapply(hogares_urb_cat,factor)
hogares_urb_cat <- as.data.frame(hogares_urb_cat)

# Segmentacion urbano
modelo_seg_urb <- kamila(hogares_urb_num,
                         hogares_urb_cat,
                         numClust=3,
                         numInit=10,
                         verbose=TRUE)

summary(modelo_seg_urb)
```

```
##              Length Class  Mode
## finalMemb    27342  -none- numeric
## numIter         10  -none- numeric
## finalLogLik      1  -none- numeric
## finalObj         1  -none- numeric
## finalCenters    36  -none- numeric
## finalProbs       8  -none- list
## input           10  -none- list
## verbose          6  -none- list
## nClust           0  -none- list
```

```
table(modelo_seg_urb$finalMemb)
```

```
##
##     1     2     3
##  8993  6237 12112
```

Unimos los clasificaciones

```r
hogares_agr_urb <- hogares_agr[which(hogares_agr$rururb==0),]
dim(hogares_agr_urb)
```

```
## [1] 27342    50
```

```r
finalMemb_urb <- as.data.frame(modelo_seg_urb$finalMemb)
dim(finalMemb_urb)
```

```
## [1] 27342     1
```

```r
colnames(finalMemb_urb) <- c("finalMemb_urb")

hogares_agr_urb <- cbind(hogares_agr_urb,finalMemb_urb)
colnames(hogares_agr_urb)
```

```
##  [1] "X"             "FOLIOVIV"      "FOLIOHOG"      "tam_loc"
##  [5] "sexo_jefe"     "edad_jefe"     "educa_jefe"    "tot_integ"
##  [9] "p12_64"        "p65mas"        "remesas"       "ing_cor"
## [13] "int0a12"       "int12a64"      "int65a98"      "depdemog"
## [17] "muj12a49"      "tot_per"       "ltot_per"      "p_esc3"
## [21] "p_esc5b"       "trab_sub"      "trab_ind"      "trab_s_pago"
## [25] "seg_alim2"     "seg_alim3"     "seg_alim_a"    "seg_pop"
## [29] "ss"            "jtrab_ind"     "ssjtrabind"    "con_remesas"
## [33] "viv_prop"      "viv_rent"      "tot_cuar"      "bao13"
## [37] "piso_fir"      "piso_rec"      "combustible"   "sin_refri"
## [41] "sin_vehi"      "sin_compu"     "sin_vidvd"     "sin_telef"
## [45] "sin_horno"     "factor"        "rururb"        "tamhogesc"
## [49] "ict"           "ictpc"         "finalMemb_urb"
```

```r
100 * table(hogares_agr_urb$finalMemb_urb) /
  sum(table(hogares_agr_urb$finalMemb_urb))
```

```
##
##        1        2        3
## 32.89079 22.81106 44.29815
```

```r
write.csv(hogares_agr_urb,file="../Datos.Modelo/hogares_agr_urb.csv")
```

Descripcion de salida

- `finalMemb` - Vector numerico con etiquetas numericas de asignacion
- `numIter`
- `finalLogLik` - Pseudo log-likelihood de la clasificacion final
- `finalObj`

- `finalCenters`
- `finalProbs`
- `input` - Vector con los parametros de inicio
- `nClust` - Descripcion de los resultados de la seleccion de segmentos
- `verbose` - Informacion compementaria

Exportacion

```
save( finalMemb_rur,finalMemb_urb,
      hogares_agr,
      hogares_agr_rur,hogares_agr_urb,
      hogares_rur_cat,hogares_rur_num,
      hogares_urb_cat,hogares_urb_num,
      modelo_seg_rur,modelo_seg_urb,
      var_enighcuis_cat,var_enighcuis_num,
      var_enighcuis_reg_cat,var_enighcuis_reg_num,
      var_enighcuis_seg_cat,var_enighcuis_seg_num,
     file = "../Datos.Modelo/modelo_segmentacion.RData")
ls()
```

```
##  [1] "finalMemb_rur"         "finalMemb_urb"
##  [3] "hogares_agr"           "hogares_agr_rur"
##  [5] "hogares_agr_urb"       "hogares_rur_cat"
##  [7] "hogares_rur_num"       "hogares_urb_cat"
##  [9] "hogares_urb_num"       "modelo_seg_rur"
## [11] "modelo_seg_urb"        "var_enighcuis_cat"
## [13] "var_enighcuis_num"     "var_enighcuis_reg_cat"
## [15] "var_enighcuis_reg_num" "var_enighcuis_seg_cat"
## [17] "var_enighcuis_seg_num"
```

```
gc()
```

```
##            used  (Mb) gc trigger  (Mb) max used  (Mb)
## Ncells  1968519 105.2    3205452 171.2  3205055 171.2
## Vcells 13339715 101.8   21625148 165.0 21625075 165.0
```