

Anexo 2.4. Predicción de los datos CUIS

Juan Carlos Martinez-Ovando

ITAM

Se presenta y comenta el código para implementar el procedimiento de identificación del estado de los hogares empleando el modelo de ingreso (MDI) propuesto.

Preambulo

=====

Datos CUIS

```
# Datos CUIS
hogares_cuis_agr <- read.csv("../Bases.Cuis/hogares_cuis_agr.csv", header=TRUE)

colnames(hogares_cuis_agr)
```

```
## [1] "X" "LLAVE_HOGAR_H"
## [3] "tot_per" "muj12a49"
## [5] "int0a12" "int12a64"
## [7] "int65a98" "trabsub"
## [9] "trabind" "trab_s_pago"
## [11] "segsoc1" "seg_pop"
## [13] "es_jefe" "j_esc"
## [15] "jtrab_ind" "es_cony"
## [17] "c_esc" "CVE_ENTIDAD_FEDERATIVA"
## [19] "CVE_MUNICIPIO" "CVE_LOCALIDAD"
## [21] "depdemog" "ltot_per"
## [23] "p_esc" "p_esc3"
## [25] "p_esc4" "p_esc5"
## [27] "p_esc5b" "trab_sub"
## [29] "trab_ind" "trab_s_pag"
## [31] "seg_alim2" "seg_alim3"
## [33] "seg_alim_a" "ss"
## [35] "ssjtrabind" "con_remesas"
## [37] "viv_prop" "viv_rent"
## [39] "tot_cuar" "bao1"
## [41] "bao2" "bao3"
## [43] "bao13" "piso_fir"
## [45] "piso_rec" "combustible"
## [47] "sin_refri" "sin_vehi"
```

```
## [49] "sin_compu"          "sin_vidvd"
## [51] "sin_telef"          "sin_horno"
```

```
hogares_cuis_agr$ENTIDAD <- hogares_cuis_agr$CVE_ENTIDAD_FEDERATIVA
hogares_cuis_agr$MUN <- hogares_cuis_agr$CVE_MUNICIPIO - 1000*hogares_cuis_agr$ENTIDAD
hogares_cuis_agr$LOC <- hogares_cuis_agr$CVE_LOCALIDAD - 10000000*hogares_cuis_agr$ENTIDAD - 10000000

colnames(hogares_cuis_agr)
```

```
## [1] "X"                  "LLAVE_HOGAR_H"
## [3] "tot_per"            "muj12a49"
## [5] "int0a12"            "int12a64"
## [7] "int65a98"           "trabsub"
## [9] "trabind"            "trab_s_pago"
## [11] "segsoc1"            "seg_pop"
## [13] "es_jefe"             "j_esc"
## [15] "jtrab_ind"           "es_cony"
## [17] "c_esc"               "CVE_ENTIDAD_FEDERATIVA"
## [19] "CVE_MUNICIPIO"       "CVE_LOCALIDAD"
## [21] "depdemog"           "ltot_per"
## [23] "p_esc"               "p_esc3"
## [25] "p_esc4"              "p_esc5"
## [27] "p_esc5b"             "trab_sub"
## [29] "trab_ind"            "trab_s_pag"
## [31] "seg_alim2"           "seg_alim3"
## [33] "seg_alim_a"          "ss"
## [35] "ssjtrabind"          "con_remesas"
## [37] "viv_prop"            "viv_rent"
## [39] "tot_cuar"            "bao1"
## [41] "bao2"                "bao3"
## [43] "bao13"               "piso_fir"
## [45] "piso_rec"            "combustible"
## [47] "sin_refri"           "sin_vehi"
## [49] "sin_compu"           "sin_vidvd"
## [51] "sin_telef"           "sin_horno"
## [53] "ENTIDAD"             "MUN"
## [55] "LOC"
```

Datos MDI, localidades y líneas de bienestar

```
# Datos modelo
load("../Datos.Modelo/mdi_segmentacion.RData")
load("../Datos.Modelo/mdi_regresion.RData")

localidades <- read.csv("../Datos.Modelo/inegi_localidades.csv",
                        header=TRUE)
```

```
lineas_bienestar <- read.csv("../Datos.Modelo/coneval_lineas_bienestar.csv",
                             header=TRUE)
```

```
hogares_cuis_agr <- as.data.frame(hogares_cuis_agr)
dim.hog.agr <- dim(hogares_cuis_agr)
dim.hog.agr
```

```
## [1] 980 55
```

Definición de tipos de variables

```
# Numeric
```

```
#hogares_cuis_agr[,var_enighcuis_num] <- apply(hogares_cuis_agr[,var_enighcuis_num], as.numeric)
```

```
# Categorías
```

```
hogares_cuis_agr[,var_enighcuis_cat] <- lapply(hogares_cuis_agr[,var_enighcuis_cat], factor)
```

```
summary(hogares_cuis_agr)
```

```
##          X                                     LLAVE_HOGAR_H
## Min.      : 1.0   000000-12719111-00000000020130319131318-000000000001: 1
## 1st Qu.:245.8   000000-12798200-00000000020130411102355-000000000001: 1
## Median :490.5   000000-12815335-00000000020130417114930-000000000001: 1
## Mean    :490.5   000000-12847948-00000000020130426093731-000000000001: 1
## 3rd Qu.:735.2   000000-3349948-00000000020120113110333-000000000001: 1
## Max.     :980.0   000000-3350092-00000000020120113112251-000000000001: 1
##                                     (Other)                               :974
##      tot_per      muj12a49      int0a12      int12a64
## Min.      : 1.000   Min.      :0.000   Min.      :0.000   Min.      : 0.000
## 1st Qu.: 2.000   1st Qu.:1.000   1st Qu.:0.000   1st Qu.: 1.000
## Median : 3.000   Median :1.000   Median :1.000   Median : 2.000
## Mean     : 3.715   Mean     :1.119   Mean     :1.251   Mean     : 2.271
## 3rd Qu.: 5.000   3rd Qu.:1.000   3rd Qu.:2.000   3rd Qu.: 3.000
## Max.     :11.000   Max.      :5.000   Max.      :6.000   Max.     :11.000
##
##      int65a98      trabsub      trabind      trab_s_pago      segsoc1
## Min.      :0.0000   Min.      :0   Min.      :0   Min.      :0   Min.      :0.0000
## 1st Qu.:0.0000   1st Qu.:0   1st Qu.:0   1st Qu.:0   1st Qu.:0.0000
## Median :0.0000   Median :0   Median :0   Median :0   Median :0.0000
## Mean     :0.1929   Mean      :0   Mean      :0   Mean      :0   Mean      :0.2133
## 3rd Qu.:0.0000   3rd Qu.:0   3rd Qu.:0   3rd Qu.:0   3rd Qu.:0.0000
## Max.     :3.0000   Max.      :0   Max.      :0   Max.      :0   Max.      :7.0000
##
##      seg_pop      es_jefe      j_esc      jtrab_ind      es_cony
## Min.      : 0.000   Min.      :1   Min.      : 0.000   0:980   Min.      :1
## 1st Qu.: 0.000   1st Qu.:1   1st Qu.: 4.000           1st Qu.:1
```

```

## Median : 1.000 Median :1 Median : 7.000 Median :1
## Mean : 1.873 Mean :1 Mean : 7.134 Mean :1
## 3rd Qu.: 3.000 3rd Qu.:1 3rd Qu.: 9.000 3rd Qu.:1
## Max. :11.000 Max. :1 Max. :19.000 Max. :1
##
## c_esc CVE_ENTIDAD_FEDERATIVA CVE_MUNICIPIO CVE_LOCALIDAD
## Min. :0 Min. : 1.00 Min. : 1001 Min. : 10010001
## 1st Qu.:0 1st Qu.:11.00 1st Qu.:11006 1st Qu.:110105011
## Median :0 Median :16.00 Median :16024 Median :160435001
## Mean :0 Mean :17.45 Mean :17500 Mean :176328722
## 3rd Qu.:0 3rd Qu.:25.00 3rd Qu.:25012 3rd Qu.:250120001
## Max. :0 Max. :32.00 Max. :32056 Max. :320560001
## NA's :24
## depdemog ltot_per p_esc p_esc3
## Min. :0.0000 Min. :0.0000 Min. :0.000 0:807
## 1st Qu.:0.0000 1st Qu.:0.6931 1st Qu.:2.000 1:173
## Median :0.5000 Median :1.0986 Median :3.500
## Mean :0.6977 Mean :1.1984 Mean :3.567
## 3rd Qu.:1.0000 3rd Qu.:1.6094 3rd Qu.:4.500
## Max. :7.0000 Max. :2.3979 Max. :9.500
##
## p_esc4 p_esc5 p_esc5b trab_sub trab_ind
## Min. :0.000000 Min. :0 0:975 Min. :0 Min. :0
## 1st Qu.:0.000000 1st Qu.:0 1: 5 1st Qu.:0 1st Qu.:0
## Median :0.000000 Median :0 Median :0 Median :0
## Mean :0.005102 Mean :0 Mean :0 Mean :0
## 3rd Qu.:0.000000 3rd Qu.:0 3rd Qu.:0 3rd Qu.:0
## Max. :1.000000 Max. :0 Max. :0 Max. :0
##
## trab_s_pag seg_alim2 seg_alim3 seg_alim_a ss ssjtrabind
## Min. :0 0:720 0:795 0:715 Min. :0.0000 0:980
## 1st Qu.:0 1:260 1:185 1:265 1st Qu.:0.0000
## Median :0 Median :0.0000
## Mean :0 Mean :0.1571
## 3rd Qu.:0 3rd Qu.:0.0000
## Max. :0 Max. :1.0000
##
## con_remasas viv_prop viv_rent tot_cuar bao1
## 0:980 0:363 0:816 Min. :0.0000 Min. :0.00000
## 1:617 1:164 1st Qu.:0.0000 1st Qu.:0.00000
## Median :0.0000 Median :0.00000
## Mean :0.3622 Mean :0.07449
## 3rd Qu.:0.0000 3rd Qu.:0.00000
## Max. :5.0000 Max. :1.00000
##
## bao2 bao3 bao13 piso_fir piso_rec combustible
## Min. :0.0000 Min. :0.0000 0:542 0:405 0:808 0:700
## 1st Qu.:0.0000 1st Qu.:0.0000 1:438 1:575 1:172 1:280

```

```
## Median :0.0000 Median :0.0000
## Mean :0.4173 Mean :0.4469
## 3rd Qu.:1.0000 3rd Qu.:1.0000
## Max. :1.0000 Max. :1.0000
##
## sin_refri sin_vehi sin_compu sin_vidvd sin_telef sin_horno
## 0:980 0:980 0:980 0:980 0:980 0:980
##
##
##
##
##
##
## ENTIDAD MUN LOC
## Min. : 1.00 Min. : 1.0 Min. : 1.00
## 1st Qu.:11.00 1st Qu.: 10.0 1st Qu.: 1.00
## Median :16.00 Median : 30.0 Median : 1.00
## Mean :17.45 Mean : 48.2 Mean : 60.71
## 3rd Qu.:25.00 3rd Qu.: 61.0 3rd Qu.: 21.25
## Max. :32.00 Max. :570.0 Max. :2005.00
##
## NA's :24
```

```
# Caso
j <- 1
caso <- as.data.frame(hogares_cuis_agr[j,])

# Prediccion (un caso a la vez)
caso.pred <- mdi.pred(caso,localidades,lineas_bienestar,
                      modelo_actual_rur,modelo_actual_urb,
                      modelo_lm_rur,modelo_lm_urb,
                      modelo_qr_rur_lb,modelo_qr_urb_lb,
                      modelo_qr_rur_lbm,modelo_qr_urb_lbm,
                      modelo_seg_rur,modelo_seg_rur_tab,
                      modelo_seg_urb,modelo_seg_urb_tab)
caso.pred
```

```
## ENTIDAD MUN LOC X.x LLAVE_HOGAR_H
## 1 12 46 1 1 000000-12719111-00000000020130319131318-000000000001
## tot_per muj12a49 int0a12 int12a64 int65a98 trabsub trabind trab_s_pago
## 1 2 0 0 0 2 0 0
## segsoc1 seg_pop es_jefe j_esc jtrab_ind es_cony c_esc
## 1 0 0 1 0 0 1 0
## CVE_ENTIDAD_FEDERATIVA CVE_MUNICIPIO CVE_LOCALIDAD depdemog ltot_per
## 1 12 12046 120460001 0 0.6931472
## p_esc p_esc3 p_esc4 p_esc5 p_esc5b trab_sub trab_ind trab_s_pag
## 1 0 0 0 0 0 0 0 0
## seg_alim2 seg_alim3 seg_alim_a ss ssjtrabind con_remesas viv_prop
## 1 0 0 0 0 0 0 0 1
```

```
##   viv_rent tot_cuar bao1 bao2 bao3 bao13 piso_fir piso_rec combustible
## 1      0      1      0      1      0      0      1      0      1
##   sin_refri sin_vehi sin_compu sin_vidvd sin_telef sin_horno   X.y P_TOTAL
## 1      0      0      0      0      0      0      0 64679   24120
##   TAM_LOC RURURB pred.segm pred.mactual.lev pred.mactual.lb
## 1      2      2      3      5576.332      0
##   pred.mactual.lbm
## 1      0
```

```
#
# -----
#

# Iteraciones
j <- 2
#for(j in 2:dim.hog.agr[1]){
  for(j in 2:8){
    print(j)
    # Seleccion del caso
    caso <- as.data.frame(hogares_cuis_agr[j,])
    # Prediccion
    caso.pred.new <- mdi.pred(caso,localidades,lineas_bienestar,
                             modelo_actual_rur,modelo_actual_urb,
                             modelo_lm_rur,modelo_lm_urb,
                             modelo_qr_rur_lb,modelo_qr_urb_lb,
                             modelo_qr_rur_lbm,modelo_qr_urb_lbm,
                             modelo_seg_rur,modelo_seg_rur_tab,
                             modelo_seg_urb,modelo_seg_urb_tab)

    # Agrupacion
    caso.pred <- rbind(caso.pred,caso.pred.new)
  }
}
```

```
## [1] 2
## [1] 3
## [1] 4
## [1] 5
## [1] 6
## [1] 7
## [1] 8
```