

Anexo 2.0. Selección de una muestra CUIS para experimentación

Juan Carlos Martínez-Ovando

ITAM

Se presenta y comenta el código extraer una muestra de los cuestionarios CUIS para experimentación e ilustración.

Preamble

CUIS

Base de selección

La muestra corresponde a los registros en `cuis_muestra100.csv`.

Muestra base

```
# Base de la muestra

se_integrante_cuis <- read.csv("../Bases.Datos.Original/CUIS/CUIS/cuis_muestra100.csv", head=TRUE,
                                #read.spss("../Bases.Datos.Original/CUIS/CUIS/se_integrante_cuis.sav", to.

dim(se_integrante_cuis)
```

```
## [1] 1007 109
```

```
#head(se_integrante_cuis)

LLAVE_HOGAR_H.df <- as.data.frame(unique(se_integrante_cuis$LLAVE_HOGAR_H))

LLAVE_HOGAR_H.df$IND_SAM <- 1

colnames(LLAVE_HOGAR_H.df) <- c("LLAVE_HOGAR_H", "IND_SAM")

#head(LLAVE_HOGAR_H.df)

dim(LLAVE_HOGAR_H.df)
```

```
## [1] 980 2
```

```
write.csv(LLAVE_HOGAR_H.df, file="../Bases.Cuis/hogares_cuis_4_analysis.csv")
```

Importacion

Importamos los datos completos de los cuestionarios CUIS desde su formato SAV.

Domicilio

```
# DOMICILIO CUIS
domicilio_cuis <- read.spss("../Bases.Datos.Original/CUIS/CUIS/domicilio_cuis_original.sav", to.dat

## Warning in read.spss("../Bases.Datos.Original/CUIS/CUIS/
## domicilio_cuis_original.sav", : ../Bases.Datos.Original/CUIS/CUIS/
## domicilio_cuis_original.sav: Very long string record(s) found (record type
## 7, subtype 14), each will be imported in consecutive separate variables

## re-encoding from UTF-8

#read.csv("../Bases.Datos.Original/CUIS/CUIS/domicilio_cuis_original.csv", head=1)

dim(domicilio_cuis)

## [1] 363106      91

domicilio_cuis <- merge(domicilio_cuis, LLAVE_HOGAR_H.df, by = "LLAVE_HOGAR_H")

domicilio_cuis <- domicilio_cuis[which(domicilio_cuis$IND_SAM==1),]

dim(domicilio_cuis)

## [1] 980   92

write.csv(domicilio_cuis, file="../Bases.Cuis/domicilio_cuis.csv")
```

Encuesta

```
# ENCUESTA CUIS

encuesta_cuis <- read.spss("../Bases.Datos.Original/CUIS/CUIS/encuesta_cuis_original.sav", to.dat

## re-encoding from UTF-8

#read.csv("../Bases.Datos.Original/CUIS/CUIS/encuesta_cuis_original.csv", head=1)

dim(encuesta_cuis)

## [1] 456833      20
```

```
encuesta_cuis <- merge(encuesta_cuis, LLAVE_HOGAR_H.df, by = "LLAVE_HOGAR_H")

encuesta_cuis <- encuesta_cuis[which(encuesta_cuis$IND_SAM==1),]

dim(encuesta_cuis)
```

```
## [1] 980 21
```

```
write.csv(encuesta_cuis, file="../Bases.Cuis/encuesta_cuis.csv")
```

Integrante

```
# INTEGRANTE CUIS
```

```
integrante_cuis <- read.spss("../Bases.Datos.Original/CUIS/CUIS/integrante_cuis_original.sav", t
```

```
## re-encoding from UTF-8
```

```
#read.csv("../Bases.Datos.Original/CUIS/CUIS/integrante_cuis_original.csv", he
```

```
dim(integrante_cuis)
```

```
## [1] 922613 18
```

```
integrante_cuis <- merge(integrante_cuis, LLAVE_HOGAR_H.df, by = "LLAVE_HOGAR_H")
```

```
integrante_cuis <- integrante_cuis[which(integrante_cuis$IND_SAM==1),]
```

```
dim(integrante_cuis)
```

```
## [1] 3641 19
```

```
write.csv(integrante_cuis, file="../Bases.Cuis/integrante_cuis.csv")
```

Personas filtrado

```
# PERSONA CUIS FILTRADO
```

```
persona_cuis_filtrado <- read.spss("../Bases.Datos.Original/CUIS/CUIS/persona_cuis_filtrado_orig
```

```
## re-encoding from UTF-8
```

```
#read.csv("../Bases.Datos.Original/CUIS/CUIS/persona_cuis_filtrado_origi
```

```
dim(persona_cuis_filtrado)
```

```
## [1] 922613 28
```

```

persona_cuis_filtrado <- merge(persona_cuis_filtrado, LLAVE_HOGAR_H.df, by = "LLAVE_HOGAR_H")

persona_cuis_filtrado <- persona_cuis_filtrado[which(persona_cuis_filtrado$IND_SAM==1),]

dim(persona_cuis_filtrado)

## [1] 3641    29

write.csv(persona_cuis_filtrado, file="../Bases.Cuis/persona_cuis_filtrado.csv")

```

SE Integrante

```

# SE INTEGRANTE CUIS

se_integrante_cuis <- read.spss("../Bases.Datos.Original/CUIS/CUIS/se_integrante_cuis_original.sav",
                                 as.data.frame = FALSE)

## re-encoding from UTF-8

#read.csv("../Bases.Datos.Original/CUIS/CUIS/se_integrante_cuis_linux.csv")

dim(se_integrante_cuis)

## [1] 922613    109

se_integrante_cuis <- merge(se_integrante_cuis, LLAVE_HOGAR_H.df, by = "LLAVE_HOGAR_H")

se_integrante_cuis <- se_integrante_cuis[which(se_integrante_cuis$IND_SAM==1),]

dim(se_integrante_cuis)

## [1] 3641    110

write.csv(se_integrante_cuis, file="../Bases.Cuis/se_integrante_cuis.csv")

```

SE Vivienda

```

#SE VIVIENDA

se_vivienda_cuis <- read.spss("../Bases.Datos.Original/CUIS/CUIS/se_vivienda_cuis_original.sav",
                                as.data.frame = FALSE)

## re-encoding from UTF-8

#read.csv("../Bases.Datos.Original/CUIS/CUIS/se_vivienda_cuis_original.csv",
#          as.data.frame = FALSE)

dim(se_vivienda_cuis)

## [1] 456833    130

```

```
se_vivienda_cuis <- merge(se_vivienda_cuis, LLAVE_HOGAR_H.df, by = "LLAVE_HOGAR_H")

se_vivienda_cuis <- se_vivienda_cuis[which(se_vivienda_cuis$IND_SAM==1),]

dim(se_vivienda_cuis)

## [1] 980 131

write.csv(se_vivienda_cuis, file="../Bases.Cuis/se_vivienda_cuis.csv")
```

V Encuestador

```
# V ENCUESTADOR
```

```
v_encuestador_cuis <- read.spss("../Bases.Datos.Original/CUIS/CUIS/v_encuestador_cuis_original.sav",

## Warning in read.spss("../Bases.Datos.Original/CUIS/CUIS/
## v_encuestador_cuis_original.sav", : ../Bases.Datos.Original/CUIS/CUIS/
## v_encuestador_cuis_original.sav: Very long string record(s) found (record
## type 7, subtype 14), each will be imported in consecutive separate
## variables

## re-encoding from UTF-8
```

```
#read.csv("../Bases.Datos.Original/CUIS/CUIS/v_encuestador_cuis_original.csv")

dim(v_encuestador_cuis)

## [1] 456833    23
```

```
v_encuestador_cuis <- merge(v_encuestador_cuis, LLAVE_HOGAR_H.df, by = "LLAVE_HOGAR_H")

v_encuestador_cuis <- v_encuestador_cuis[which(v_encuestador_cuis$IND_SAM==1),]

dim(v_encuestador_cuis)

## [1] 980 24

write.csv(v_encuestador_cuis, file="../Bases.Cuis/v_encuestador_cuis.csv")
```

Vivienda

```
# VIVIENDA CUIS
```

```
vivienda_cuis <- read.spss("../Bases.Datos.Original/CUIS/CUIS/vivienda_cuis_original.sav", to.data.frame=TRUE)

## re-encoding from UTF-8
```

```

#read.csv("../Bases.Datos.Original/CUIS/CUIS/vivienda_cuis_original.csv", head=T)

dim(vivienda_cuis)

## [1] 456833      11

vivienda_cuis <- merge(vivienda_cuis, LLAVE_HOGAR_H.df, by = "LLAVE_HOGAR_H")

vivienda_cuis <- vivienda_cuis[which(vivienda_cuis$IND_SAM==1),]

dim(vivienda_cuis)

## [1] 980  12

write.csv(vivienda_cuis, file="../Bases.Cuis/vivienda_cuis.csv")

gc()

##           used  (Mb) gc trigger  (Mb) max used   (Mb)
## Ncells  3317747 177.2   5684620 303.6   3571460 190.8
## Vcells 26967209 205.8   88330929 674.0 137936634 1052.4

```