

TareaInei1

Cochachin Victor

13/8/2021

Lectura de datos

```
library(survey)
```

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

```
## Loading required package: Matrix
```

```
## Loading required package: survival
```

```
##
```

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

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

```
##
```

```
##      dotchart
```

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5      v purrr  0.3.4
```

```
## v tibble  3.1.3      v dplyr  1.0.7
```

```
## v tidyr   1.1.3      v stringr 1.4.0
```

```
## v readr   2.0.0      v forcats 0.5.1
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x tidyr::expand() masks Matrix::expand()
```

```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag()    masks stats::lag()
```

```
## x tidyr::pack()   masks Matrix::pack()
```

```
## x tidyr::unpack() masks Matrix::unpack()
```

```
library(haven)
```

```
library(knitr)
```

```
library(kableExtra)
```

```
##
## Attaching package: 'kableExtra'

## The following object is masked from 'package:dplyr':
##
##     group_rows
```

```
library(readstata13)
knit_hooks$set(inline = function(x) {
  prettyNum(x, big.mark=" ")
})
```

Cargando paquetes nec

Estableciendo rutas de los archivos a importar.

```
maindir <- "D:/Ciclo V/Econometría/Semana 2"
subdir.empleo <- paste0("/Datos/687-Modulo05/Enaho01a-2019-500.sav")
subdir.sumaria <- paste0("/Datos/687-Modulo34/Sumaria-2019.sav")
#-----*
dir.deflactor.tiempo <- paste0(maindir, "/Datos/687-Modulo34/Gasto2019/Bases/deflactores_base2019_new.dta")
dir.deflactor.espacial <- paste0(maindir, "/Datos/687-Modulo34/Gasto2019/Bases/despacial_ldnew.dta")
```

Estimacion puntual

Ejemplo 1

```
empleo <- read_sav(paste0(maindir,subdir.empleo))
#empleo <- read_sav(paste0(maindir,subdir.empleo),col_select = c(1:20,323))
```

El siguiente comando nos mostrará una tabla de frecuencias de la variable **ESTRATO**, debemos notar que nos muestra un conteo del número de filas asociadas a cada valor de la variable.

```
table(empleo$P510A1)
```

```
##
##      1      2      3
## 9418 7588 45287
```

Estos valores se visualizan como números, pero cada categoría tiene un significado distinto. De acuerdo con el diccionario de datos de la ENAHO podemos etiquetar estos números con el siguiente script.

```
#-----
#Modificando etiquetas de la variable categórica (o factor varibale)
#-----
empleo$P510A1 <- factor(empleo$P510A1,
  levels = c("1","2","3"),
  labels = c("Persona juridica","Persona natural","N.A."))
#-----
table(empleo$P510A1)
```

```
##
## Persona juridica Persona natural N.A.
## 9418 7588 45287
```

Otro ejemplo de etiquetas.

```
#-----
#Modificando etiquetas de la variable categórica (o factor varibale)
#-----
empleo$ESTRATO <- factor(empleo$ESTRATO,
  levels = c("1","2","3","4",
             "5","6","7","8"),
  labels = c("1.De 500 000 a más habitantes.",
             "2.De 100 000 a 499 999 habitantes.",
             "3.De 50 000 a 99 999 habitantes.",
             "4.De 20 000 a 49 999 habitantes.",
             "5.De 2 000 a 19 999 habitantes.",
             "6.De 500 a 1 999 habitantes.",
             "7.Área de Empadronamiento Rural (AER) Compuesto.",
             "8.Área de Empadronamiento Rural (AER) Simple.))

empleo$conteo <- 1 #Esta variable serviría para contar el número de filas
svy.empleo <- svydesign(ids = ~1, strata = ~ESTRATO, weights = ~FAC500A, data = empleo)
```

Podemos estimar la cantidad de hogar estimada para cada estrato con el siguiente código.

```
est1 <- svytable(~P510A1,svy.empleo); est1
```

```
## P510A1
## Persona juridica Persona natural N.A.
## 3498415 2333098 10530033
```

El código anterior no nos proporciona el error estandar, para ello podemos utiliza el siguiente código.

```
est2 <- svyby(~conteo,~P510A1,svy.empleo,svytotal,na.rm=T);
est2.1 <- data.frame(est2,row.names = 1); est2.1
```

```
##          conteo      se
## Persona juridica 3498415 44839.46
## Persona natural  2333098 35861.43
## N.A.            10530033 58819.03
```

El total de hogares encuestados es 16 361 545 y el número de hogares estimados en áreas con más de 500 000 habitantes es 3 498 415. Ahora que tenemos los errores estandar, podemos calcular los coeficientes de variación y los intervalos de confianza al 95%.

```
#Coeficientes de variación
cv_est2 <- cv(est2); cv_est2
```

```
## Persona juridica Persona natural N.A.
## 0.012817078 0.015370739 0.005585835
```

```
#Intervalos de confianza
conf_est2 <- confint(est2); conf_est2
```

```
##                2.5 %   97.5 %
## Persona juridica 3410531 3586299
## Persona natural  2262810 2403385
## N.A.             10414750 10645316
```

Del cuadro anterior podemos afirmar que al 95 por ciento de confianza, el número de hogares en áreas con más de 500 000 habitantes, se encuentra entre 3 410 531 y 3 586 299.

En la práctica, los reportes deben presentarse de manera que pueda darse una lectura rápida y comprensiva al cuadro. Por ello, ahora construiremos un solo cuadro resumen de los cálculos realizados.

```
est0 <- as.numeric(table(empleo$P510A1))
est3 <- cbind.data.frame(est0,
  round(est2.1,0),
  cv_est2*100,
  round(conf_est2,0)
); est3
```

```
##                est0   conteo   se cv_est2 * 100   2.5 %   97.5 %
## Persona juridica 9418 3498415 44839   1.2817078 3410531 3586299
## Persona natural  7588 2333098 35861   1.5370739 2262810 2403385
## N.A.             45287 10530033 58819   0.5585835 10414750 10645316
```

```
r.names <- est2[,1]
rownames(est3) <- r.names
```

Ahora daremos formato a nuestra tabla resumen.

```
colnames(est3) <- c("Muestra","Estimacion","Error estandar","CV%","2.5%","97.5%")
kable(est3,
  booktabs = T, longtable = T, #format = 'latex',
  linesep = " ",
  caption = "EMPLEO",
  digits = 2,
  align=rep("c", 5),
  format.args = list(decimal.mark = ".", big.mark = " ")) %>%
  row_spec(c(1:1) - 1) %>%
kable_styling(font_size = 9,
  latex_options = c("repeat_header")) %>%
add_header_above(c(" " = 5, "Intervalo al 95% de confianza" = 2)) %>%
footnote(general = "ENAH0 2019",
  general_title = "Fuente: ",
footnote_as_chunk = T, title_format = c("italic"),
threeparttable = T)
```

Table 1: EMPLEO

	Muestra	Estimacion	Error estandar	CV%	Intervalo al 95% de confianza	
					2.5%	97.5%
Persona juridica	9 418	3 498 415	44 839	1.28	3 410 531	3 586 299
Persona natural	7 588	2 333 098	35 861	1.54	2 262 810	2 403 385
N.A.	45 287	10 530 033	58 819	0.56	10 414 750	10 645 316

Fuente: ENAHO 2019