

Ejercicio Diseño STSI

Muestreo y Planificación de Encuestas

2023-04-26

En este ejemplo seleccionaremos una muestra de hogares de Bella Unión con un diseño STSI, con el fin de estimar la cantidad de desocupados, utilizando como estrato al segmento censal. Cargamos las librerías:

```
library(sampling)
library(survey)
library(haven)
library(tidyverse)
library(here)
```

Leemos los datos censales de Bella Unión.

```
bu=read_spss(here("Datos","bella_union.sav"))
```

Creamos la variable de interés:

```
bu$desocupado=ifelse(bu$pobpcoac==3|bu$pobpcoac==4,1,0)
```

Visualizamos a los segmentos como estratos geográficos, para ello debemos leer el archivo *shp* con la función *st_read()* de la librería **sf** y luego visualizarlo con **tmap**:

```
bu.map=st_read("zonas_bu.shp")
bu.map$SEGMENTO=as.character(bu.map$SEGMENTO)

tmap_mode("view")
tm_shape(bu.map)+
  tm_polygons("SEGMENTO",alpha = 0.3,id="SEGMENTO")+
  tm_layout(legend.outside = TRUE)+
  tm_basemap("OpenStreetMap")
```

Definimos el tamaño de muestra y creamos un objeto que agregue por estrato la cantidad de hogares total y el desvío estándar de la variable de interés.

```
n=200

tam=bu %>% group_by(SEGM) %>% summarise(Nh=n(),sh=sd(desocupado))
tam
```

```
## # A tibble: 11 x 3
##   SEGM      Nh      sh
```

```
##      <chr> <int> <dbl>
## 1 008      1 NA
## 2 013     1261 0.147
## 3 104      525 0.137
## 4 105     1335 0.141
## 5 106      561 0.194
## 6 204     1285 0.138
## 7 205     1478 0.164
## 8 206     1343 0.143
## 9 304     1508 0.146
## 10 305     1685 0.141
## 11 306     1218 0.107
```

Observamos que hay un segmento con una sola persona. Para el fin de este ejercicio no contaremos con ese segmento, ya que puede perjudicar la asignación entre estratos. Volvemos a crear el objeto anterior, descartando a este segmento y lo descartamos del marco.

```
tam=bu %>% filter(!SEGM%in%"008") %>% group_by(SEGM) %>% summarise(Nh=n(),sh=sd(desocupado))
tam
```

```
## # A tibble: 10 x 3
##   SEGM      Nh    sh
##   <chr> <int> <dbl>
## 1 013     1261 0.147
## 2 104      525 0.137
## 3 105     1335 0.141
## 4 106      561 0.194
## 5 204     1285 0.138
## 6 205     1478 0.164
## 7 206     1343 0.143
## 8 304     1508 0.146
## 9 305     1685 0.141
## 10 306     1218 0.107
```

```
bu=bu %>% filter(!SEGM%in%"008")
```

Creamos una columna con la asignación proporcional de la muestra y otra con la asignación óptima asumiendo costos constantes. Recordar que la muestra siempre se redondea hacia arriba (por eso la función *ceiling()*), y que tenemos un estrato con una sola observación, el que tendrá que ser censado. Es celilling porque el tamaño de muestra aproxima se aproxima para arriba

```
tam=tam %>% mutate(n1=ceiling(n*tam$Nh/sum(tam$Nh)),
                  n2=ceiling(n*Nh*sh/sum(Nh*sh,na.rm=TRUE)))
tam
```

```
## # A tibble: 10 x 5
##   SEGM      Nh    sh    n1    n2
##   <chr> <int> <dbl> <dbl> <dbl>
## 1 013     1261 0.147    21    22
## 2 104      525 0.137     9     9
## 3 105     1335 0.141    22    22
## 4 106      561 0.194    10    13
```

```
## 5 204 1285 0.138 22 21
## 6 205 1478 0.164 25 28
## 7 206 1343 0.143 23 22
## 8 304 1508 0.146 25 26
## 9 305 1685 0.141 28 28
## 10 306 1218 0.107 20 15
```

En vistas de que usaremos la librería **survey** para obtener las estimaciones, conviene agregar las variables creadas en el objeto *tam* al marco de muestreo. hay que decirle el fps pasarle cada tamaño, lo jineo segun segmento

```
bu=left_join(bu,tam,by="SEGM")
```

Seleccionamos una muestra estratificada con la función *strata()* de la librería **sampling**. La base tiene que estar ordenada por el estrato. tiene que estar ordenado

```
bu=bu %>% arrange(SEGM)
```

Selecciono la muestra con la asignación proporcional. Indico el nombre del estrato (*stratanames*) y el tamaño en un vector. El método por defecto es el STSI. hay que indicarle que agarre de la libreria sampling

```
set.seed(12345)
s1=sampling::strata(bu,stratanames = "SEGM",size=tam$n1)
```

```
## Warning in sampling::strata(bu, stratanames = "SEGM", size = tam$n1): the method
## is not specified; by default, the method is srswor
```

```
m1=getdata(bu,s1)
```

Selecciono la muestra con la asignación óptima. En este caso especifico el *method* para eliminar el warning. por defecto es srswor

```
set.seed(12345)
s2=sampling::strata(bu,stratanames = "SEGM",size=tam$n2,method = "srswor")
m2=getdata(bu,s2)
```

Ahora vamos a procesar con la librería **survey**. Primero la muestra obtenida bajo asignación proporcional. ids con *vivorita 1* significa que es seleccion directo de elementos sin etapas cuando no le pasamos prob entienede que es con prob cte

```
p.s1=svydesign(ids = ~1,strata = ~SEGM, data=m1,fpc =~Nh)
summary(p.s1)
```

Realizamos un *summary()* a los pesos muestrales, y le calculamos es desvío.

```
summary(weights(p.s1))
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    56.10  58.41   60.05   59.51  60.32   60.90
```

```
sd(weights(p.s1))
```

```
## [1] 1.196052
```

$\pi_k = (nNh/N)/Nh = n/N$ es auto ponderado
Calculamos el total de desocupados con su *deff*:

```
svytotal(~desocupado,p.s1,deff=TRUE)
```

```
##           total      SE  DEff  
## desocupado 178.42 100.73 0.9817
```

Repetimos el ejercicio para la muestra obtenida bajo asignación óptima.

```
p.s2=svydesign(ids = ~1,strata = ~SEGM, data=m2,fpc =~Nh)  
summary(p.s2)
```

Realizamos un *summary()* a los pesos muestrales, y le calculamos es desvío.

```
summary(weights(p.s2))
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.  
##    43.15  57.32   60.18   59.22  61.05   81.20
```

```
sd(weights(p.s2))
```

```
## [1] 7.687341
```

si tengo un rango grande es un problema Calculamos el total de desocupados con su *deff*:

```
svytotal(~desocupado,p.s2,deff=TRUE)
```

```
##           total      SE  DEff  
## desocupado 226.80 111.76 0.9593
```

Ejercicio

Verificar si está utilizando la fórmula de la varianza del diseño STSI para calcular la varianza de la estimación.