

# Introducción a la Librería *survey*

## Muestreo y Planificación de Encuestas

2023-04-19

En este ejemplo seleccionaremos una muestra de personas de Bella Unión para luego realizar las estimaciones utilizando la librería **survey**. Cargamos las librerías.

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

Leemos los datos censales de Bella Unión.

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

Vamos a seleccionar primero una muestra con un diseño SI y luego una con un diseño SY. Estimaremos el total y la media de la cantidad de jubilados y pensionistas de Bella Unión con la librería **survey**. Primero creamos la variable:

```
bu$jp=ifelse(bu$pobpcoac==5,1,0)
```

El total poblacional es:

```
t=sum(bu$jp)
t
```

```
## [1] 1587
```

Selecciono una muestra de tamaño  $n = 200$  con la función *srswor()* de la librería **sampling**:

```
N=nrow(bu)
n=200
set.seed(852)
s1=srswor(n,N)
m1=getdata(bu,s1)
```

Ahora seleccionamos una muestra SY con la librería **sampling**. Para ello debemos crear el vector de probabilidades de inclusión;

```
pik=rep(n/N,N)
set.seed(852)
s2=UPsystematic(pik)
m2=getdata(bu,s2)
```

## Estimación con la librería survey.

Cualquier cálculo que se vaya a realizar con la librería **survey** parte de un objeto que contiene al diseño muestral.

El diseño muestral lo especificamos con la función `svydesign()`. Esta función tiene múltiples argumentos que iremos viendo a medida que avancemos en el curso. Hagamos `help(svydesign)`.

Por ahora los argumentos que vamos a utilizar son:

- `ids`: indica si es muestreo directo de elementos (`~1`).
- `probs`: son las probabilidades de inclusión. Alternativamente se puede usar el argumento `weights` que es el inverso.
- `fpc`: factor de corrección de poblaciones finitas.
- `data`: los datos.

Para un diseño SI pasamos los argumentos:

```
p.s1=svydesign(ids=~1, probs = rep(n/N,n),
              fpc =rep(N,n),data=m1 )
```

Calculo el total, para ello usamos la función `svytotal()` con la variable de interés (`jp`) y el diseño como argumentos.

```
svytotal(~jp,p.s1)
```

```
##      total      SE
## jp  1159 251.5
```

Verifiquemos si se cumple la fórmula para la varianza de un total para el diseño SI  $N^2(1-f)\frac{S_{ys}^2}{n}$ . Como nos proporciona el desvío, le hacemos la raíz.

```
sqrt(N^2*(1-n/N)*var(m1$jp)/n)
```

```
## [1] 251.4954
```

Calculo los intervalos de confianza:

```
confint(svytotal(~jp,p.s1))
```

```
##      2.5 %    97.5 %
## jp 666.078 1651.922
```

Calculo la media:

```
svymean(~jp,p.s1)
```

```
##      mean      SE
## jp 0.095 0.0206
```

Verifiquemos si se cumple la fórmula para la varianza de una media para el diseño SI  $(1 - f) \frac{S_y^2}{n}$ . Como nos proporciona el desvío, le hacemos la raíz.

```
sqrt((1-n/N)*var(m1$jp)/n)
```

```
## [1] 0.02061438
```

Ahora creemos un diseño para hacer la estimación con la muestra sistemática:

```
p.s2=svydesign(ids=~1, probs = rep(n/N,n),
              fpc =rep(N,n),data=m2)
```

Calculo el total y los intervalos de confianza con la función *svytotal()*.

```
svytotal(~jp,p.s2)
```

```
##      total      SE
## jp  1403 273.63
```

```
confint(svytotal(~jp,p.s2))
```

```
##      2.5 %   97.5 %
## jp 866.694 1939.306
```

Verifiquemos que sigue utilizando la fórmula de la varianza para el diseño SI.

```
sqrt(N^2*(1-n/N)*var(m2$jp)/n)
```

```
## [1] 273.6305
```

Calcula la del SI, o sea que puede estar sobreestimando.

¿Qué pasa si no usamos el fpc para la muestra obtenida con el diseño SI?

```
p.s3=svydesign(ids=~1, probs = rep(n/N,n),data=m1)
svytotal(~jp,p.s3)
```

```
##      total      SE
## jp  1159 253.58
```

La varianza da un poco más grande que la obtenida cuando utilizamos el fpc. Verificamos que asume que el diseño es con remplazo.

```
sqrt(N^2*var(m1$jp)/n)
```

```
## [1] 253.5826
```

En este caso podemos calcular el *deff*.

```
svytotal(~jp,p.s3,deff=TRUE)
```

```
##      total      SE  DEff  
## jp 1159.00 253.58 1.0167
```