

# CÁLCULO DE LOS ESTIMADORES DE REGRESIÓN Y CALIBRACIÓN CON EL PAQUETE SAMPLING

El paquete *sampling* de R (Tillé y Matei, 2012) permite seleccionar muestras mediante distintos diseños muestrales complejos como estratificación, conglomerados, muestreos polietápicos o probabilidades desiguales, y hacer estimaciones. En este documento se pretende mostrar cómo se utilizan algunas de sus funciones mediante la implementación de un ejemplo básico.

Dado que *sampling* es un paquete adicional de R, antes de utilizarlo la primera vez deberemos instalarlo, si se tiene internet se puede hacer, por ejemplo, de la siguiente forma

```
install.packages("sampling")
```

siendo necesario para utilizarlo cargarlo previamente mediante la orden

```
library(sampling)
```

Para el ejemplo vamos a considerar la población MU284, referenciada en Särndal, Swensson, y Wretman (1992), incluida en el propio paquete. Esta población está formada por 284 municipios de Suecia, con las siguientes variables:

LABEL: identificador, de 1 a 284,

P85: población en 1985, en miles,

P75: población en 1975, en miles,

RMT85: recaudación de impuestos municipales en 1985, in millones de coronas,

CS82: número de escaños conservadores en el consistorio,

SS82: número de escaños social demócratas en el consistorio,

S82: número total de escaños en el consistorio,

ME84: número de empleados municipales en 1984,

REV84: valor total de las propiedades del municipio, en millones de coronas,

REG: indicador geográfico de región, de 1 a 8,

CL: indicador de conglomerado o conjunto de municipios, de 1 a 50.

El **objetivo** es: estimar la recaudación total de impuestos en 1985 (RMT85) utilizando una m.a.s de tamaño 100.

En primer lugar se hará sin utilizar información auxiliar, esto es, por medio del estimador de Horvitz-Thompson y, después, mediante los estimadores de regresión y calibración utilizando como variables auxiliares la población en 1975 (P75) y en 1985 (P85) y un modelo lineal. El código es el siguiente:

```
library(sampling)
data(MU284)

# comprobamos las variables
names(MU284)
attach(MU284)

# tamaño de la población
N=length(P75)

# tamaño de la muestra
n=100

# selección de los municipios que van a pertenecer a la muestra
s=sample(N,n)

# definición de los valores muestrales
# para la variable de interés y las variables auxiliares
y = RMT85[s]
x1 = P75[s]
x2 = P85[s]

# totales de las variables auxiliares
tx1=sum(P75)
tx2=sum(P85)
totales=c(tx1,tx2)

# vector de probabilidades de inclusión para un m.a.s
pik=rep(n/N,n)

# cálculo del estimador de Horvitz-Thompson para el total poblacional
t_ypi=HTestimator(y,pik)
```

```

# definición del modelo de superpoblación  $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ 
modelo=y~x1+x2

# matriz de probabilidades de inclusión para un m.a.s
pikl=matrix(n*(n-1)/(N*(N-1)),n,n)

# cálculo del estimador de regresión 0
# considerando el modelo lineal  $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ 
r=regest(formula=modelo, weights= 1/pik, Tx=totales, pikl, n)

# estimador de regresión
t_ygreg=r$regest

# coeficientes
coef=r$coefficients

# estimador de los pesos de calibración
Xs=cbind(x1,x2)
g1=calib(Xs,d=1/pik,totales,method="linear")

# chequeo de la calibración
checkcalibration(Xs,d=1/pik,totales,g1)

# estimador de calibración
c=calibev(y,Xs,totales,pikl,d=1/pik,g1,with=TRUE,EPS=1e-6)
t_yc=c$calest

t_ypi
t_ygreg
t_yc
#total de la variable de interés desconocido en la práctica
sum(RMT85)

```

## Referencias

Särndal, C.E., Swensson, B. y Wretman, J. (1992). *Model Assisted Survey Sampling*. Springer-Verlag. New York, Inc.

Tillé, Y. y Matei, A. (2016). *Sampling: Survey Sampling*. R package version 2.8. <http://CRAN.R-project.org/package=sampling>.