Muestreo con probabilidades Simples

Usando la libreria `TeachingSampling` accederemos a la base de datos `BigCity`

```
library(TeachingSampling)

data(BigCity)
attach(BigCity)
head(BigCity)
```

```
##          HHID PersonID   Stratum     PSU  Zone    Sex Age MaritalST Income
## 1 idHH00001  idPer01 idStrt001 PSU0001 Rural   Male  38   Married 555.00
## 2 idHH00001  idPer02 idStrt001 PSU0001 Rural Female  40   Married 555.00
## 3 idHH00001  idPer03 idStrt001 PSU0001 Rural Female  20    Single 555.00
## 4 idHH00001  idPer04 idStrt001 PSU0001 Rural   Male  19    Single 555.00
## 5 idHH00001  idPer05 idStrt001 PSU0001 Rural   Male  18    Single 555.00
## 6 idHH00002  idPer01 idStrt001 PSU0001 Rural   Male  35   Married 298.34
##   Expenditure Employment  Poverty
## 1      488.33   Employed  NotPoor
## 2      488.33   Employed  NotPoor
## 3      488.33   Inactive  NotPoor
## 4      488.33   Employed  NotPoor
## 5      488.33   Inactive  NotPoor
## 6      216.70   Employed Relative
```

Muestreo aleatorio simple

Utilizando la función `S.SI()` obtenemos una muestra aleatoria de tamaño n = 2000

```
N <- dim(BigCity)[1]
n <- 2000
sam <- S.SI(N,n)
muestra <- BigCity[sam,]
attach(muestra)
head(muestra)
```

```
##           HHID PersonID   Stratum     PSU  Zone    Sex Age MaritalST
## 31  idHH00007  idPer03 idStrt001 PSU0001 Rural Female  21    Single
## 96  idHH20653  idPer04 idStrt001 PSU0001 Rural Female  19    Single
## 140 idHH00019  idPer03 idStrt001 PSU0002 Rural Female   9      <NA>
## 151 idHH00021  idPer03 idStrt001 PSU0002 Rural Female  25    Single
## 183 idHH00029  idPer02 idStrt001 PSU0002 Rural   Male  17    Single
## 270 idHH00033  idPer02 idStrt001 PSU0003 Rural Female  19    Single
##       Income Expenditure Employment  Poverty
## 31   1607.50      637.75   Employed  NotPoor
## 96    817.50      202.33   Employed  NotPoor
## 140   209.00      164.55       <NA> Relative
## 151  1027.17      514.92   Inactive  NotPoor
## 183   350.17      267.15   Inactive  NotPoor
## 270   501.58      203.43   Inactive  NotPoor
```

Creamos una base de datos con las variables a estimar de la muestra y obtenemos los estimadores correspondientes mediante el codigo `E.SI()`

```
estima <- data.frame(Income, Expenditure)
head(estima)
```

```
##     Income Expenditure
## 1 1607.50      637.75
## 2  817.50      202.33
## 3  209.00      164.55
## 4 1027.17      514.92
## 5  350.17      267.15
## 6  501.58      203.43
```

```
E.SI(N, n, estima)
```

```
##                      N      Income  Expenditure
## Estimation      150266 8.880628e+07 5.712305e+07
## Standard Error       0 1.821837e+06 9.661676e+05
## CVE                  0 2.051473e+00 1.691380e+00
## DEFF               NaN 1.000000e+00 1.000000e+00
```

Mediante `Domains()` obtenemos variables dummys para las categorías de factores.

```
Dominios <- Domains(Sex)
head(Dominios)
```

```
##      Female Male
## [1,]      1    0
## [2,]      1    0
## [3,]      1    0
## [4,]      1    0
## [5,]      0    1
## [6,]      1    0
```

```
area.M <- Dominios[, 1] * estima
head(area.M)
```

```
##     Income Expenditure
## 1 1607.50      637.75
## 2  817.50      202.33
## 3  209.00      164.55
## 4 1027.17      514.92
## 5    0.00        0.00
## 6  501.58      203.43
```

```
area.H <- Dominios[, 2] * estima
head(area.H)
```

```
##   Income Expenditure
## 1   0.00        0.00
```

```
## 2    0.00         0.00
## 3    0.00         0.00
## 4    0.00         0.00
## 5  350.17       267.15
## 6    0.00         0.00
```

Estimamos el tamaño poblacional de cada dominio

```
E.SI(N, n, Dominios)
```

```
##                        N         Female          Male
## Estimation        150266 79565.847000 70700.153000
## Standard Error         0  1666.316702  1666.316702
## CVE                    0     2.094261     2.356878
## DEFF                 NaN     1.000000     1.000000
```

Estimamos los ingresos y gastos de cada dominio

```
E.SI(N, n, area.H)
```

```
##                        N       Income  Expenditure
## Estimation        150266 4.105684e+07 2.613495e+07
## Standard Error         0 1.504102e+06 8.778435e+05
## CVE                    0 3.663462e+00 3.358887e+00
## DEFF                 NaN 1.000000e+00 1.000000e+00
```

```
E.SI(N, n, area.M)
```

```
##                        N       Income  Expenditure
## Estimation        150266 4.774944e+07 3.098810e+07
## Standard Error         0 1.729764e+06 9.810013e+05
## CVE                    0 3.622584e+00 3.165735e+00
## DEFF                 NaN 1.000000e+00 1.000000e+00
```

Muestreo bernoulli

La idea es obtener una muestra mediante un muestreo bernoulli

```
data(BigCity)
N <- dim(BigCity)[1]
pik <- 0.025
sam <- S.BE(N,pik)
muestra <- BigCity[sam,]
head(muestra)
```

```
##          HHID PersonID   Stratum     PSU  Zone    Sex Age MaritalST Income
## 23  idHH00005   idPer05 idStrt001 PSU0001 Rural   Male  22    Single 623.75
## 66  idHH20647   idPer02 idStrt001 PSU0001 Rural Female  29   Married 298.34
## 121 idHH00014   idPer03 idStrt001 PSU0002 Rural Female  14    Single 899.30
## 135 idHH00017   idPer03 idStrt001 PSU0002 Rural   Male  13    Single 472.08
```

```
## 212 idHH20665  idPer03 idStrt001 PSU0002 Rural Female  26    Single 384.50
## 453 idHH00056  idPer01 idStrt001 PSU0005 Rural   Male  53   Partner 681.17
##      Expenditure Employment  Poverty
## 23        613.70 Unemployed  NotPoor
## 66        216.70   Employed Relative
## 121       647.58       <NA>  NotPoor
## 135       247.86       <NA>  NotPoor
## 212       142.04   Inactive Relative
## 453       310.63   Employed  NotPoor
```

Luego se construye una base de datos para estimar ingresos y gastos

```
n <- dim(muestra)[1]
n
```

```
## [1] 3766
```

```
estima <- data.frame(Income, Expenditure)
head(estima)
```

```
##      Income Expenditure
## 1 1607.50       637.75
## 2  817.50       202.33
## 3  209.00       164.55
## 4 1027.17       514.92
## 5  350.17       267.15
## 6  501.58       203.43
```

```
E.BE(estima, pik)
```

```
##                         N       Income   Expenditure
## Estimation    80000.00000 4.727951e+07 3.041170e+07
## Standard Error  1766.35217 1.420873e+06 8.439154e+05
## CVE               2.20794 3.005262e+00 2.774970e+00
## DEFF                  Inf 2.171748e+00 2.724022e+00
```

Muestreo sistemático

Escogemos el valor para el muestreo sistematico y obtenemos la muestra sistemática con la función `S.SY()`

```
data(BigCity)
attach(BigCity)
N <- dim(BigCity)[1]
a <- 40
floor(N/a)
```

```
## [1] 3756
```

```
sam <- S.SY(N, a)
muestra <- BigCity[sam,]
attach(muestra)
n <- dim(muestra)[1]
n
```

```
## [1] 3757
```

```
head(muestra)
```

```
##            HHID PersonID   Stratum     PSU  Zone    Sex Age MaritalST Income
## 4    idHH00001  idPer04 idStrt001 PSU0001 Rural   Male  19    Single 555.00
## 44   idHH00010  idPer01 idStrt001 PSU0001 Rural Female  59   Married 279.25
## 84   idHH20651  idPer02 idStrt001 PSU0001 Rural Female  60    Single 135.00
## 124  idHH00015  idPer02 idStrt001 PSU0002 Rural   Male  18    Single 430.00
## 164  idHH00024  idPer03 idStrt001 PSU0002 Rural   Male  22    Single 689.34
## 204  idHH20663  idPer01 idStrt001 PSU0002 Rural Female  49 Separated 785.00
##     Expenditure Employment  Poverty
## 4        488.33   Employed  NotPoor
## 44       179.20   Inactive Relative
## 84       219.60   Inactive  Extreme
## 124      273.01   Inactive  NotPoor
## 164      357.20   Employed  NotPoor
## 204      116.77   Employed  NotPoor
```

El calculo de las estimaciones respectivas se realizan mediante la función E.SY()

```
estima <- data.frame(Income, Expenditure)
head(estima)
```

```
##    Income Expenditure
## 1 555.00      488.33
## 2 279.25      179.20
## 3 135.00      219.60
## 4 430.00      273.01
## 5 689.34      357.20
## 6 785.00      116.77
```

```
E.SY(N, a, estima)
```

```
##                       N       Income  Expenditure
## Estimation       150280 8.757491e+07 5.629281e+07
## Standard Error        0 1.329507e+06 6.691192e+05
## CVE                   0 1.518137e+00 1.188640e+00
## DEFF                NaN 1.000000e+00 1.000000e+00
```