# Estrategias transversales

## Muestreo en dos etapas estratificado

```
library(dplyr)
library(TeachingSampling)
data('BigCity')
FrameI <- BigCity %>% group_by(PSU) %>%
            summarise(Stratum = unique(Stratum),
                      Persons = n(),
                      Income = sum(Income),
                      Expenditure = sum(Expenditure))
attach(FrameI)
head(FrameI)
```

```
## # A tibble: 6 x 5
##   PSU     Stratum   Persons Income Expenditure
##   <chr>   <chr>       <int>  <dbl>       <dbl>
## 1 PSU0001 idStrt001     118 70912.      44232.
## 2 PSU0002 idStrt001     136 68887.      38382.
## 3 PSU0003 idStrt001      96 37213.      19495.
## 4 PSU0004 idStrt001      88 36926.      24031.
## 5 PSU0005 idStrt001     110 57494.      31142.
## 6 PSU0006 idStrt001     116 75272.      43473.
```

```
sizes = FrameI %>% group_by(Stratum) %>%
          summarise(NIh = n(),
                    nIh = 2,
                    dI = NIh/nIh)
NIh <- sizes$NIh
nIh <- sizes$nIh
head(sizes)
```

```
## # A tibble: 6 x 4
##   Stratum     NIh   nIh    dI
##   <chr>     <int> <dbl> <dbl>
## 1 idStrt001     9     2   4.5
## 2 idStrt002    11     2   5.5
## 3 idStrt003     7     2   3.5
## 4 idStrt004    13     2   6.5
## 5 idStrt005    11     2   5.5
## 6 idStrt006     5     2   2.5
```

```
samI <- S.STSI(Stratum, NIh, nIh)
UI <- levels(as.factor(FrameI$PSU))
sampleI <- UI[samI]
FrameII <- left_join(sizes, BigCity[which(BigCity$PSU %in% sampleI), ], by = "Stratum")
attach(FrameII)
head(FrameII)
```

```
## # A tibble: 6 x 15
##   Stratum   NIh   nIh    dI HHID  PersonID PSU   Zone  Sex     Age MaritalST
##   <chr>   <int> <dbl> <dbl> <chr> <chr>    <chr> <chr> <chr> <int> <fct>
## 1 idStrt~     9     2   4.5 idHH~ idPer01  PSU0~ Rural Male     57 Married
## 2 idStrt~     9     2   4.5 idHH~ idPer02  PSU0~ Rural Fema~    48 Married
## 3 idStrt~     9     2   4.5 idHH~ idPer03  PSU0~ Rural Male     20 Single
## 4 idStrt~     9     2   4.5 idHH~ idPer04  PSU0~ Rural Fema~    18 Single
## 5 idStrt~     9     2   4.5 idHH~ idPer05  PSU0~ Rural Male     16 Single
## 6 idStrt~     9     2   4.5 idHH~ idPer06  PSU0~ Rural Fema~     0 <NA>
## # ... with 4 more variables: Income <dbl>, Expenditure <dbl>, Employment <fct>,
## #   Poverty <fct>
```

```
HHdb <- FrameII %>%
        group_by(PSU) %>%
        summarise(Ni = length(unique(HHID)))
Ni <- as.numeric(HHdb$Ni)
ni <- ceiling(Ni * 0.1)
sum(ni)
```

```
## [1] 702
```

```
sam = S.SI(Ni[1], ni[1])
clusterII = FrameII[which(FrameII$PSU == sampleI[1]), ]
sam.HH <- data.frame(HHID = unique(clusterII$HHID)[sam])
clusterHH <- left_join(sam.HH, clusterII, by = "HHID")
clusterHH$dki <- Ni[1]/ni[1]
clusterHH$dk <- clusterHH$dI * clusterHH$dki
data = clusterHH
```

```
for (i in 2:length(Ni)) {
      sam = S.SI(Ni[i], ni[i])
      clusterII = FrameII[which(FrameII$PSU == sampleI[i]), ]
      sam.HH <- data.frame(HHID = unique(clusterII$HHID)[sam])
      clusterHH <- left_join(sam.HH, clusterII, by = "HHID")
      clusterHH$dki <- Ni[i]/ni[i]
      clusterHH$dk <- clusterHH$dI * clusterHH$dki
      data1 = clusterHH
      data = rbind(data, data1)
}
```

```
dim(data)
```

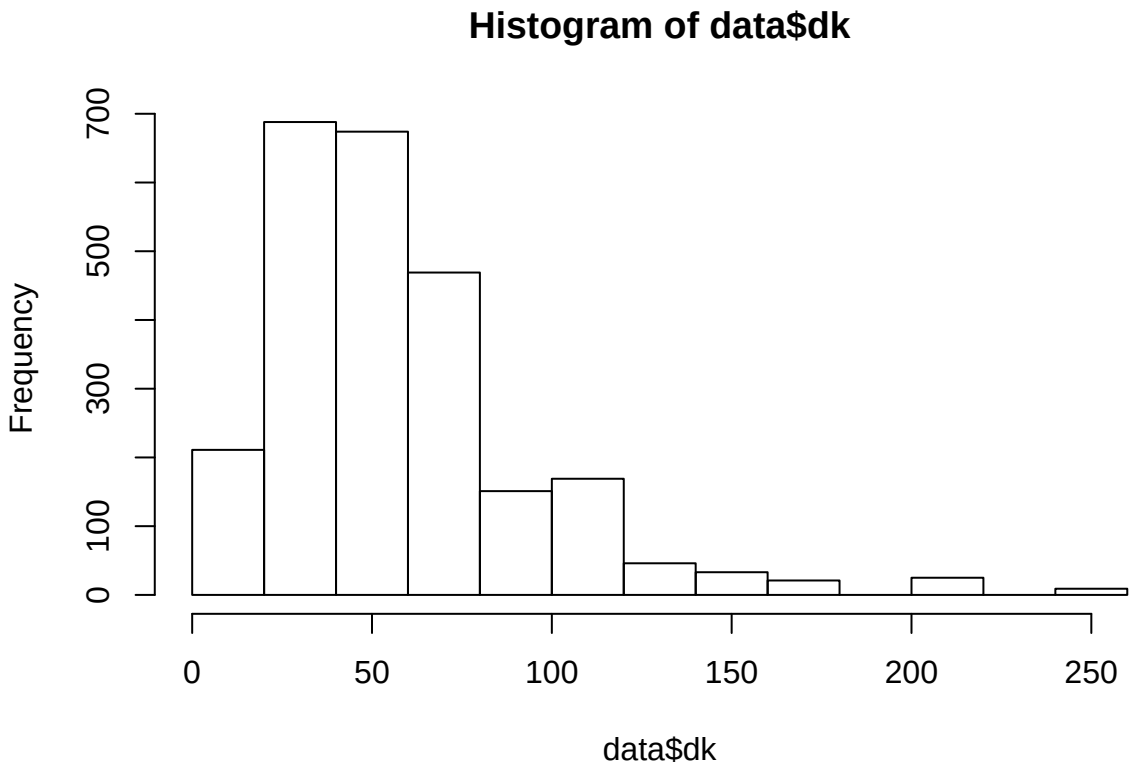
```
## [1] 2496   17
```

```
sum(data$dk)
```

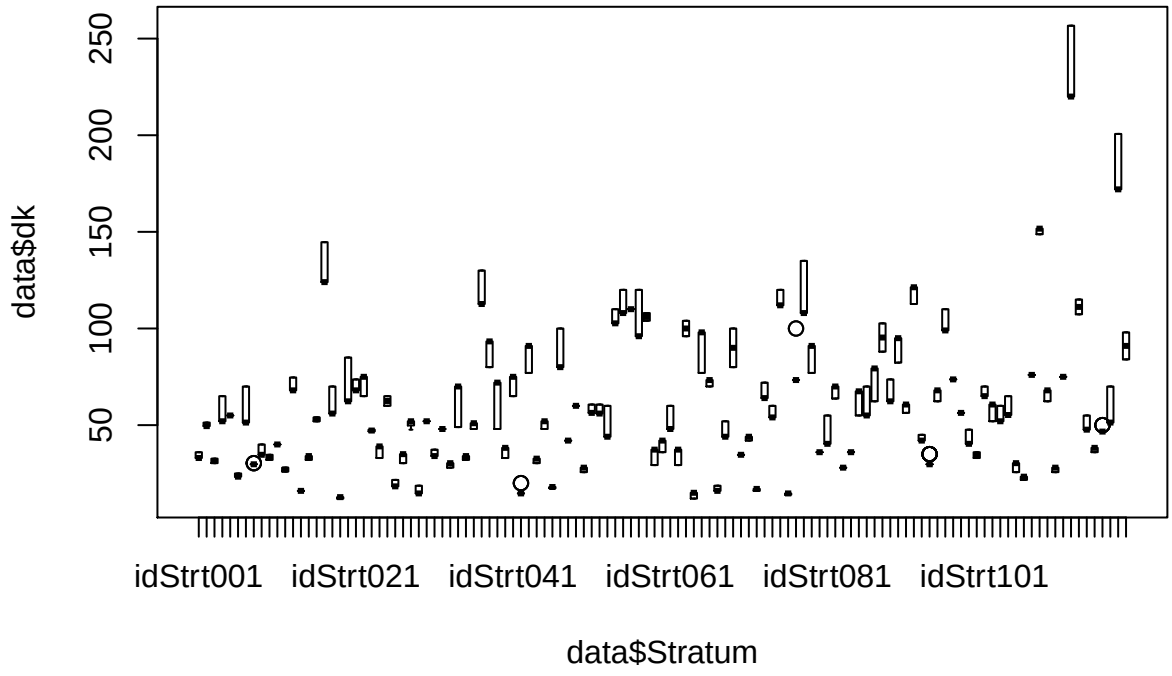
```
## [1] 146880.9
```

```
attach(data)
estima <- data.frame(Income, Expenditure)
area <- as.factor(PSU)
stratum <- as.factor(Stratum)
```

```
hist(data$dk)
```




```
boxplot(data$dk ~ data$Stratum)
```




```
E.UC(stratum, area, dk, estima)
```

```
##                            N       Income  Expenditure
## Estimation     1.468809e+05 8.290466e+07 5.427982e+07
## Standard Error 3.351918e+03 4.465049e+06 2.620232e+06
## CVE            2.282065e+00 5.385764e+00 4.827267e+00
## DEFF                    Inf 1.145389e+01 1.355435e+01
```

## Muestreo autoponderado en dos etapas estratificado

```
data('BigCity')
FrameI <- BigCity %>% group_by(PSU) %>%
          summarise(Stratum = unique(Stratum),
                    Households = length(unique(HHID)),
                    Income = sum(Income),
                    Expenditure = sum(Expenditure))
attach(FrameI)
head(FrameI)
```

```
## # A tibble: 6 x 5
##   PSU     Stratum   Households Income Expenditure
##   <chr>   <chr>          <int>  <dbl>       <dbl>
## 1 PSU0001 idStrt001         26 70912.      44232.
## 2 PSU0002 idStrt001         32 68887.      38382.
## 3 PSU0003 idStrt001         24 37213.      19495.
## 4 PSU0004 idStrt001         22 36926.      24031.
## 5 PSU0005 idStrt001         28 57494.      31142.
## 6 PSU0006 idStrt001         30 75272.      43473.
```

```
sizes = FrameI %>% group_by(Stratum) %>%
        summarise(NIh = n(), nIh = 2)
NIh <- sizes$NIh
nIh <- sizes$nIh
head(sizes)
```

```
## # A tibble: 6 x 3
##   Stratum     NIh   nIh
##   <chr>     <int> <dbl>
## 1 idStrt001     9     2
## 2 idStrt002    11     2
## 3 idStrt003     7     2
## 4 idStrt004    13     2
## 5 idStrt005    11     2
## 6 idStrt006     5     2
```

```
resI <- S.STpiPS(Stratum, Households, nIh)
head(resI)
```

```
##      [,1]      [,2]
## [1,]    4 0.1774194
## [2,]    6 0.2419355
## [3,]   12 0.1717791
## [4,]   20 0.1717791
## [5,]   23 0.2553191
## [6,]   26 0.2978723
```

```
samI <- resI[, 1]
piI <- resI[, 2]
UI <- levels(as.factor(FrameI$PSU))
sampleI <- data.frame(PSU = UI[samI], dI = 1/piI)
FrameII <- left_join(sampleI, BigCity[which(BigCity$PSU %in% sampleI[,1]), ])
attach(FrameII)
head(FrameII)
```

```
##       PSU       dI      HHID PersonID   Stratum  Zone    Sex Age MaritalST
## 1 PSU0004 5.636364 idHH00042  idPer01 idStrt001 Rural   Male  57   Married
## 2 PSU0004 5.636364 idHH00042  idPer02 idStrt001 Rural Female  48   Married
## 3 PSU0004 5.636364 idHH00042  idPer03 idStrt001 Rural   Male  20    Single
## 4 PSU0004 5.636364 idHH00042  idPer04 idStrt001 Rural Female  18    Single
## 5 PSU0004 5.636364 idHH00042  idPer05 idStrt001 Rural   Male  16    Single
## 6 PSU0004 5.636364 idHH00042  idPer06 idStrt001 Rural Female   0      <NA>
##   Income Expenditure Employment Poverty
## 1 527.84      309.09   Inactive NotPoor
## 2 527.84      309.09   Inactive NotPoor
## 3 527.84      309.09   Employed NotPoor
## 4 527.84      309.09   Inactive NotPoor
## 5 527.84      309.09   Inactive NotPoor
## 6 527.84      309.09       <NA> NotPoor
```

```
HHdb <- FrameII %>%
        group_by(PSU) %>%
        summarise(Ni = length(unique(HHID)),
                  ni = 5)
Ni <- as.numeric(HHdb$Ni)
ni <- 3
head(HHdb)
```

```
## # A tibble: 6 x 3
##   PSU        Ni    ni
##   <chr>   <int> <dbl>
## 1 PSU0004    22     5
## 2 PSU0006    30     5
## 3 PSU0012    28     5
## 4 PSU0020    28     5
## 5 PSU0023    24     5
## 6 PSU0026    28     5
```

```
sam = S.SI(Ni[1], ni)
clusterII = FrameII[which(FrameII$PSU == sampleI$PSU[1]), ]
sam.HH <- data.frame(HHID = unique(clusterII$HHID)[sam])
clusterHH <- left_join(sam.HH, clusterII, by = "HHID")
clusterHH$dki <- Ni[1]/ni
clusterHH$dk <- clusterHH$dI * clusterHH$dki
data = clusterHH
head(data)
```

```
##        HHID     PSU       dI PersonID   Stratum  Zone    Sex Age MaritalST
## 1 idHH00050 PSU0004 5.636364  idPer01 idStrt001 Rural Female  41 Separated
## 2 idHH00050 PSU0004 5.636364  idPer02 idStrt001 Rural   Male  19    Single
## 3 idHH00050 PSU0004 5.636364  idPer03 idStrt001 Rural Female  16    Single
## 4 idHH20687 PSU0004 5.636364  idPer01 idStrt001 Rural   Male  57   Married
## 5 idHH20687 PSU0004 5.636364  idPer02 idStrt001 Rural Female  48   Married
## 6 idHH20687 PSU0004 5.636364  idPer03 idStrt001 Rural   Male  20    Single
##   Income Expenditure Employment Poverty      dki       dk
## 1 503.92      331.92   Employed NotPoor 7.333333 41.33333
## 2 503.92      331.92   Inactive NotPoor 7.333333 41.33333
## 3 503.92      331.92   Inactive NotPoor 7.333333 41.33333
## 4 527.84      309.09   Inactive NotPoor 7.333333 41.33333
## 5 527.84      309.09   Inactive NotPoor 7.333333 41.33333
## 6 527.84      309.09   Employed NotPoor 7.333333 41.33333
```

```
for (i in 2:length(Ni)) {
      sam = S.SI(Ni[i], ni)
      clusterII = FrameII[which(FrameII$PSU == sampleI$PSU[i]), ]
      sam.HH <- data.frame(HHID = unique(clusterII$HHID)[sam])
      clusterHH <- left_join(sam.HH, clusterII, by = "HHID")
      clusterHH$dki <- Ni[i]/ni
      clusterHH$dk <- clusterHH$dI * clusterHH$dki
      data1 = clusterHH
      data = rbind(data, data1)
}
```

```
sum(data$dk)
```

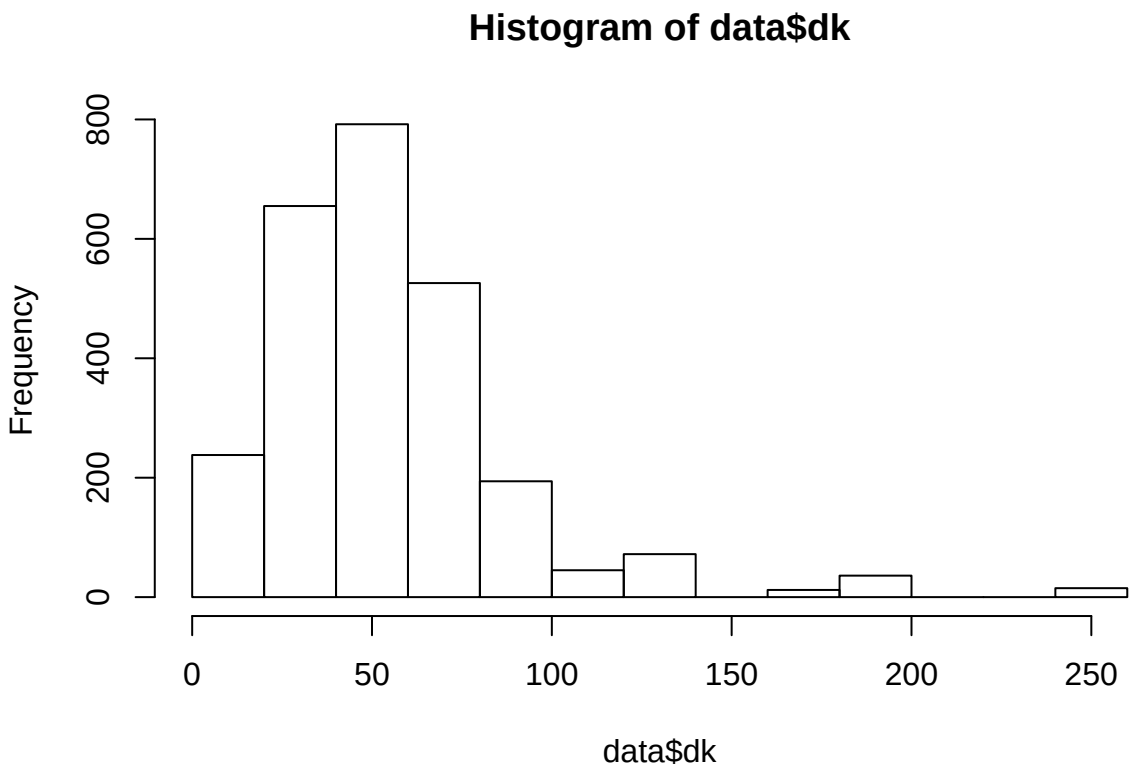
```
## [1] 148253
```

```
dim(data)
```

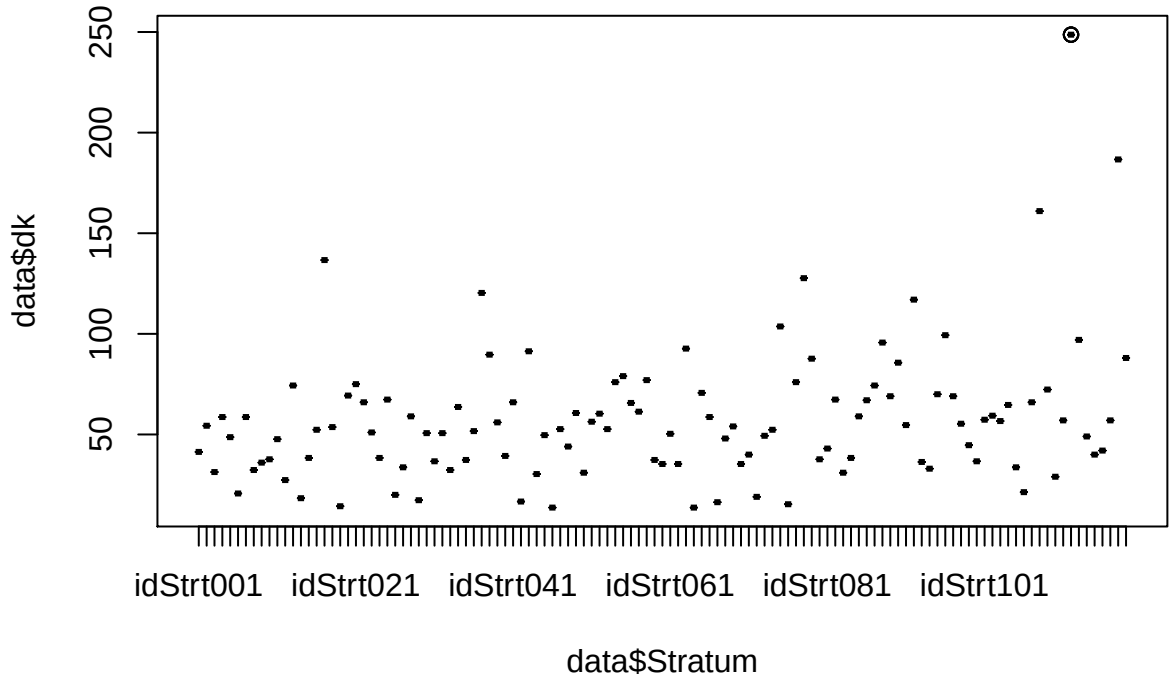
```
## [1] 2585   15
```

```
attach(data)
estima <- data.frame(Income, Expenditure)
area <- as.factor(PSU)
stratum <- as.factor(Stratum)
```

```
hist(data$dk)
```




```
boxplot(data$dk ~ data$Stratum)
```




```
E.UC(stratum, area, dk, estima)
```

```
##                             N       Income  Expenditure
## Estimation     1.482530e+05 8.382537e+07 5.363328e+07
## Standard Error 4.056097e+03 4.551031e+06 2.262998e+06
## CVE            2.735929e+00 5.429181e+00 4.219391e+00
## DEFF                    Inf 7.373944e+00 7.260026e+00
```