Muestras probabilísticas

Muestras probabilisticas

Cargamos la libreria a utilizar

```r
library(TeachingSampling)
```

Creamos el vector de caracteres y definimos su tamaño poblacional y muestral

```r
U <- c("Yves", "Ken", "Erik", "Sharon", "Leslie")
U[1]
```

```
## [1] "Yves"
```

```r
N <- length(U)
N
```

```
## [1] 5
```

```r
n <- 2
n
```

```
## [1] 2
```

Con la función `support` observamos las posibles muestras de tamaño `n`

```r
Support(N,n)
```

```
##       [,1] [,2]
##  [1,]    1    2
##  [2,]    1    3
##  [3,]    1    4
##  [4,]    1    5
##  [5,]    2    3
##  [6,]    2    4
##  [7,]    2    5
##  [8,]    3    4
##  [9,]    3    5
## [10,]    4    5
```

```r
Support(N,n,U)
```

```
##       [,1]     [,2]
##  [1,] "Yves"   "Ken"
##  [2,] "Yves"   "Erik"
##  [3,] "Yves"   "Sharon"
##  [4,] "Yves"   "Leslie"
##  [5,] "Ken"    "Erik"
##  [6,] "Ken"    "Sharon"
##  [7,] "Ken"    "Leslie"
##  [8,] "Erik"   "Sharon"
##  [9,] "Erik"   "Leslie"
## [10,] "Sharon" "Leslie"
```

Considere el siguiente diseño de muestreo p() tal que asigna las siguientes probabilidades de selección a cada una de las 10 posibles muestras de tamaño 2 del soporte Q de la población U.

```
p <- c(0.13,0.2,0.15,0.1,0.15,0.04,0.02,0.06,0.07,0.08)
p
```

```
##  [1] 0.13 0.20 0.15 0.10 0.15 0.04 0.02 0.06 0.07 0.08
```

Con las siguientes instrucciones verificamos que las propiedades de diseño muestral sean satisfechas.

```
sum(p)
```

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

```
p<0
```

```
##  [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

La función `Ik(N,n)` entrega una matriz que indica la pertenencia o no del elemento en la muestra.

```
Ind <- Ik(N, n)
Q <- Support(N, n, U)
data.frame(Q, p, Ind)
```

```
##        X1     X2    p X1.1 X2.1 X3 X4 X5
## 1    Yves    Ken 0.13    1    1  0  0  0
## 2    Yves   Erik 0.20    1    0  1  0  0
## 3    Yves Sharon 0.15    1    0  0  1  0
## 4    Yves Leslie 0.10    1    0  0  0  1
## 5     Ken   Erik 0.15    0    1  1  0  0
## 6     Ken Sharon 0.04    0    1  0  1  0
## 7     Ken Leslie 0.02    0    1  0  0  1
## 8    Erik Sharon 0.06    0    0  1  1  0
## 9    Erik Leslie 0.07    0    0  1  0  1
## 10 Sharon Leslie 0.08    0    0  0  1  1
```

El cálculo de las probabilidades de inclusión se hace muy sencillo al multiplicar las probabilidades de selección con cada una de las variables indicadoras. El resultado se suma por columnas y la salida es un vector de tamaño $N = 5$ de probabilidades de inclusión.

```
multip <- p * Ind
colSums(multip)
```

```
## [1] 0.58 0.34 0.48 0.33 0.27
```

Nótese que la suma de probabilidades de inclusión es el tamaño de muestra esperado, en este caso igual a 2.

```
pik <- Pik(p,Ind)
pik
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,] 0.58 0.34 0.48 0.33 0.27
```

```
sum(pik)
```

```
## [1] 2
```

Las probabilidades de inclusión de segundo orden se obtienen de la siguiente manera:

```
pikl <- Pikl(N,n,p)
pikl
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,] 0.58 0.13 0.20 0.15 0.10
## [2,] 0.13 0.34 0.15 0.04 0.02
## [3,] 0.20 0.15 0.48 0.06 0.07
## [4,] 0.15 0.04 0.06 0.33 0.08
## [5,] 0.10 0.02 0.07 0.08 0.27
```

Parámetros y estimadores

Suponga que en nuestra población de ejemplo se quiere estimar el total de la variable y. El valor para cada uno de los elementos de la población es el siguiente:

```r
y <- c(32, 34, 46, 89, 35)
ty <- sum(y)
ty
```

```
## [1] 236
```

```r
ybar <- ty / N
ybar
```

```
## [1] 47.2
```

En el caso del estimador horvitz thompson, si la primera muestra (cuyos elementos son Yves y Ken) hubiese sido seleccionada, el estimador de Horvitz-Thompson arrojaría la siguiente estimación:

```r
y.s <- c(32, 34)
pik.s <- c(0.58, 0.34)
HT(y.s, pik.s)
```

```
##          [,1]
## [1,] 155.1724
```

Calcularemos el estimador para todas las muestras posibles

```r
all.pik <- Support(N, n, pik)
all.y <- Support(N, n, y)
all.HT <- rep(0, 10)

for(k in 1:10){
all.HT[k] <- HT(all.y[k,], all.pik[k,])
}

AllSamples = data.frame(Q, p, all.pik, all.y, all.HT)
AllSamples
```

```
##        X1     X2    p X1.1 X2.1 X1.2 X2.2   all.HT
## 1    Yves    Ken 0.13 0.58 0.34   32   34 155.1724
## 2    Yves   Erik 0.20 0.58 0.48   32   46 151.0057
## 3    Yves Sharon 0.15 0.58 0.33   32   89 324.8694
## 4    Yves Leslie 0.10 0.58 0.27   32   35 184.8020
## 5     Ken   Erik 0.15 0.34 0.48   34   46 195.8333
## 6     Ken Sharon 0.04 0.34 0.33   34   89 369.6970
## 7     Ken Leslie 0.02 0.34 0.27   34   35 229.6296
## 8    Erik Sharon 0.06 0.48 0.33   46   89 365.5303
## 9    Erik Leslie 0.07 0.48 0.27   46   35 225.4630
```

```
## 10 Sharon Leslie 0.08 0.33 0.27   89   35 399.3266
```

El vector all.HT contiene las estimaciones Horvitz-Thompson para cada una de las 10 posibles muestras, su esperanza se calcula en R como

```
sum(p * all.HT)
```

```
## [1] 236
```

Muestreo con reemplazo

Mediante el siguiente codigo obtenemos todas las muestras posibles de tamaño m = 2

```
m <- 2
SupportWR(N, m,ID=FALSE)
```

```
##       [,1] [,2]
##  [1,]    1    1
##  [2,]    1    2
##  [3,]    1    3
##  [4,]    1    4
##  [5,]    1    5
##  [6,]    2    2
##  [7,]    2    3
##  [8,]    2    4
##  [9,]    2    5
## [10,]    3    3
## [11,]    3    4
## [12,]    3    5
## [13,]    4    4
## [14,]    4    5
## [15,]    5    5
```

```
SupportWR(N,m,ID=U)
```

```
##       [,1]     [,2]
##  [1,] "Yves"   "Yves"
##  [2,] "Yves"   "Ken"
##  [3,] "Yves"   "Erik"
##  [4,] "Yves"   "Sharon"
##  [5,] "Yves"   "Leslie"
##  [6,] "Ken"    "Ken"
##  [7,] "Ken"    "Erik"
##  [8,] "Ken"    "Sharon"
##  [9,] "Ken"    "Leslie"
## [10,] "Erik"   "Erik"
## [11,] "Erik"   "Sharon"
## [12,] "Erik"   "Leslie"
## [13,] "Sharon" "Sharon"
## [14,] "Sharon" "Leslie"
## [15,] "Leslie" "Leslie"
```

```
pk <- c(0.25, 0.25, 0.125, 0.125, 0.25)
QWR <- SupportWR(N,m,ID=U)
pWR <- p.WR(N, m, pk)
```

```r
nkWR <- nk(N, m)
SamplesWR <- data.frame(QWR, pWR, nkWR)
```

Para el estimador de hansen-Hurwitz de cada muestra se tiene

```r
all.y <- SupportWR(N, n, y)
all.pk <- SupportWR(N, n, pk)
all.HH <- rep(0, 15)

for(k in 1:15){
  all.HH[k] <- HH(all.y[k,], all.pk[k,])[1,1]
}

AllSamplesWR <- data.frame(QWR, all.pk, pWR, all.y, all.HH)
AllSamplesWR
```

```
##        X1     X2  X1.1  X2.1      pWR X1.2 X2.2 all.HH
## 1    Yves   Yves 0.250 0.250 0.062500   32   32    128
## 2    Yves    Ken 0.250 0.250 0.125000   32   34    132
## 3    Yves   Erik 0.250 0.125 0.062500   32   46    248
## 4    Yves Sharon 0.250 0.125 0.062500   32   89    420
## 5    Yves Leslie 0.250 0.250 0.125000   32   35    134
## 6     Ken    Ken 0.250 0.250 0.062500   34   34    136
## 7     Ken   Erik 0.250 0.125 0.062500   34   46    252
## 8     Ken Sharon 0.250 0.125 0.062500   34   89    424
## 9     Ken Leslie 0.250 0.250 0.125000   34   35    138
## 10   Erik   Erik 0.125 0.125 0.015625   46   46    368
## 11   Erik Sharon 0.125 0.125 0.031250   46   89    540
## 12   Erik Leslie 0.125 0.250 0.062500   46   35    254
## 13 Sharon Sharon 0.125 0.125 0.015625   89   89    712
## 14 Sharon Leslie 0.125 0.250 0.062500   89   35    426
## 15 Leslie Leslie 0.250 0.250 0.062500   35   35    140
```