Análisis de encuestas con datos provenientes de teléfonos fijos y teléfonos móviles

A. Arcos[1] M. Rueda[1] M. G. Ranalli[2] D. Molina[1]

[1]Departamento de Estadística e I. O.
Universidad de Granada

[2]Departamento de Ciencias Políticas
Universidad de Perugia

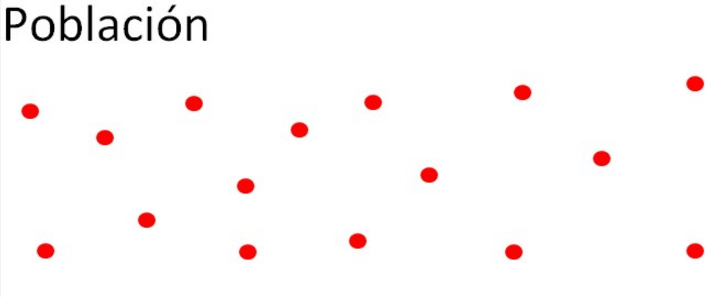
Salamanca, 5 Noviembre 2015

Índice

Introducción

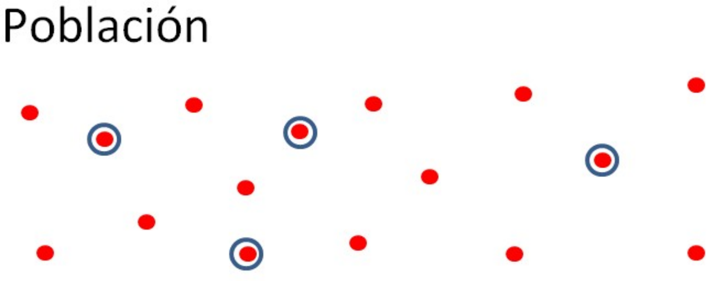
Una breve noción de muestreo

- En muestreo, el objetivo es estimar el valor de un parámetro en toda una **población** de individuos basándonos únicamente en los valores del parámetro observados en un subconjunto de ésta, al que se denomina **muestra**.



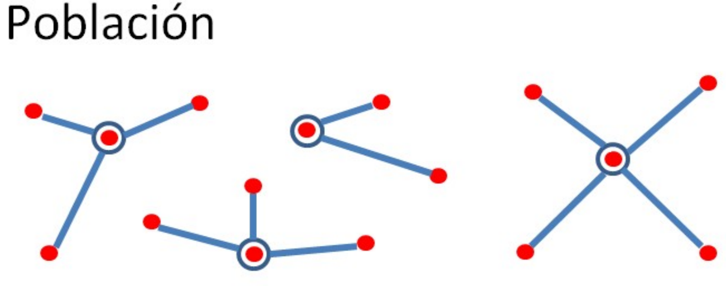
Una breve noción de muestreo

- En muestreo, el objetivo es estimar el valor de un parámetro en toda una población de individuos basándonos únicamente en los valores del parámetro observados en un subconjunto de ésta, al que se denomina **muestra** y se suele representar por la letra s.



Una breve noción de muestreo

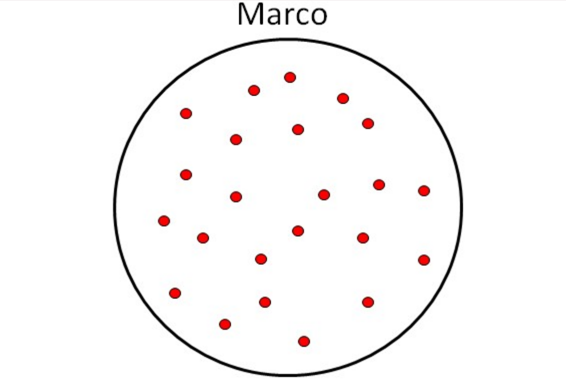
- Cada unidad de la población tiene una probabilidad de ser seleccionada en la muestra. A esta probabilidad se le conoce como **probabilidad de inclusión** y se representa por π_i. El valor inverso de la probabilidad de inclusión se conoce como **peso del diseño**, $d_i = 1/\pi_i$ y se interpreta como el número de individuos de la población a los que representa cada unidad de la muestra.



- Por tanto, si estamos interesados en estimar el total en la población de una característica Y a partir de los valores y_i observados en la muestra, parece lógico hacerlo así: $\hat{Y} = \sum_{i \in s} y_i d_i$.

Las encuestas con dos marcos: ¿qué son y por qué se utilizan?

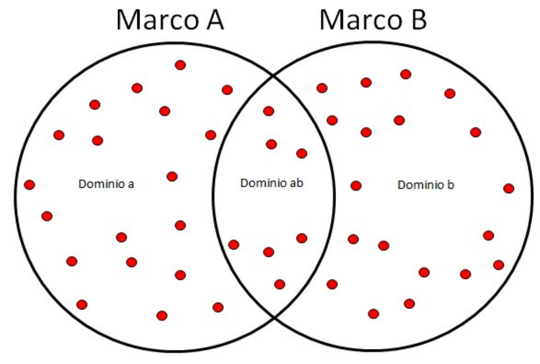
- La teoría clásica del muestreo asume la existencia de un único marco muestral que contiene **todas** las unidades de la población.
- Esta hipótesis es muy severa: las poblaciones cambian constantemente.



- **Consecuencia:** Resultados sesgados debido a la falta de representatividad de las muestras que se extraen del marco.

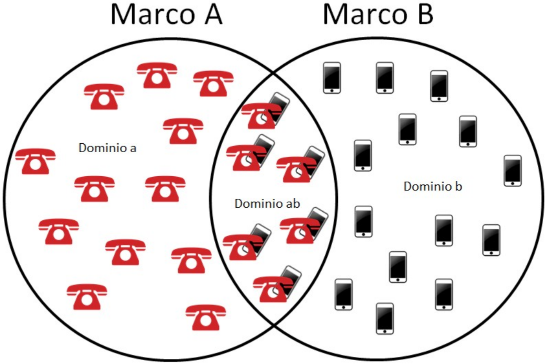
Las encuestas con dos marcos: ¿qué son y por qué se utilizan?

- La metodología de marcos duales asume que se dispone de dos marcos muestrales que, conjuntamente, cubren toda la población objeto de estudio.



Las encuestas con dos marcos: ¿qué son y por qué se utilizan?

- La metodología de marcos duales asume que se dispone de dos marcos muestrales que, conjuntamente, cubren toda la población objeto de estudio.



Introducción

Las encuestas con dos marcos: ¿qué son y por qué se utilizan?

Supongamos que la población se compone de N unidades y que queremos estimar el total poblacional de una variable Y:

$$Y = \sum_{k=1}^{N} y_k$$

Las encuestas con dos marcos: ¿qué son y por qué se utilizan?

Supongamos que la población se compone de N unidades y que queremos estimar el total poblacional de una variable Y:

$$Y = \sum_{k=1}^{N} y_k = \sum_{k=1}^{N} y_k \delta_k(a) + \sum_{k=1}^{N} y_k \delta_k(ab) + \sum_{k=1}^{N} y_k \delta_k(b)$$

Las encuestas con dos marcos: ¿qué son y por qué se utilizan?

Supongamos que la población se compone de N unidades y que queremos estimar el total poblacional de una variable Y:

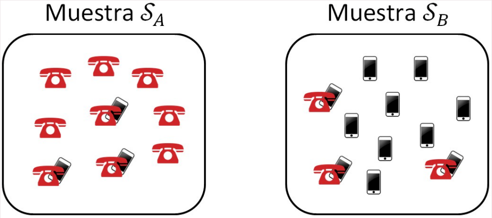
$$Y = \sum_{k=1}^{N} y_k = \sum_{k=1}^{N} y_k \delta_k(a) + \sum_{k=1}^{N} y_k \delta_k(ab) + \sum_{k=1}^{N} y_k \delta_k(b) = Y_a + Y_{ab} + Y_b,$$

Las encuestas con dos marcos: ¿qué son y por qué se utilizan?

Supongamos que la población se compone de N unidades y que queremos estimar el total poblacional de una variable Y:

$$Y = \sum_{k=1}^{N} y_k = \sum_{k=1}^{N} y_k \delta_k(a) + \sum_{k=1}^{N} y_k \delta_k(ab) + \sum_{k=1}^{N} y_k \delta_k(b) = Y_a + Y_{ab} + Y_b,$$

con y_k el valor de Y para el $k-esimo$ individuo de la población y $\delta(a), \delta(ab), \delta(b)$ las variables indicadoras de los dominios a, ab y b, respectivamente.
Extraemos dos muestras, una de cada marco:



Introducción

Muestra $\mathcal{S}_A$

$$d_k^A = \frac{1}{\pi_k^A}, \qquad k = 1, ..., n_A,$$

donde π_k^A son las probabilidades de inclusión de primer orden inducidas por el diseño muestral del marco A.

$$\hat{Y}_a^A = \sum_{k \in \mathcal{S}_A} \delta_k(a) d_k^A y_k = \sum_{k \in \mathcal{S}_a} d_k^A y_k$$

$$\hat{Y}_{ab}^A = \sum_{k \in \mathcal{S}_A} \delta_k(ab) d_k^A y_k = \sum_{k \in \mathcal{S}_{ab}^A} d_k^A y_k$$

con $\delta(a)$ y $\delta(ab)$ las variables indicadoras para los dominios a y ab.

Muestra $\mathcal{S}_B$

$$d_k^B = \frac{1}{\pi_k^B}, \qquad k = 1, ..., n_B,$$

donde π_k^B son las probabilidades de inclusión de primer orden inducidas por el diseño muestral del marco B.

$$\hat{Y}_b^B = \sum_{k \in \mathcal{S}_B} \delta_k(b) d_k^B y_k = \sum_{k \in \mathcal{S}_b} d_k^B y_k$$

$$\hat{Y}_{ab}^B = \sum_{k \in \mathcal{S}_B} \delta_k(ab) d_k^B y_k = \sum_{k \in \mathcal{S}_{ab}^B} d_k^B y_k$$

con $\delta(b)$ y $\delta(ab)$ las variables indicadoras para los dominios b y ab.

¿Y por qué no...?

$$\hat{Y} = \hat{Y}_a^A + \hat{Y}_{ab}^A + \hat{Y}_{ab}^B + \hat{Y}_b^B$$

Introducción

¿Y por qué no...?

$$\hat{Y} = \hat{Y}_a^A + \hat{Y}_{ab}^A + \hat{Y}_{ab}^B + \hat{Y}_b^B >> Y$$

De hecho: $E(Y) = Y + Y_{ab}$.

¿Y por qué no...?

$$\hat{Y} = \hat{Y}_a^A + \hat{Y}_{ab}^A + \hat{Y}_{ab}^B + \hat{Y}_b^B >> Y$$

De hecho: $E(Y) = Y + Y_{ab}$. Dos posibles soluciones:

- **Solución "dual frame":** ponderar las dos estimaciones relativas al dominio de intersección mediante un parámetro, $\eta \in [0, 1]$.
- **Solución "single frame":** considerar todas las unidades procedentes de un mismo marco y ajustar los pesos de las unidades del dominio de intersección, de manera que los pesos originales, d_k, se sustituyen por un nuevo conjunto de pesos, $\tilde{d}_k$.

Dual frame versus single frame

- **Solución "dual frame":** ponderar las dos estimaciones relativas al dominio de intersección mediante un parámetro, $\eta \in [0, 1]$.

$$\hat{Y} = \hat{Y}_a^A + \quad \hat{Y}_{ab}^A + \qquad\qquad \hat{Y}_{ab}^B + \hat{Y}_b^B$$

Dual frame versus single frame

- **Solución "dual frame":** ponderar las dos estimaciones relativas al dominio de intersección mediante un parámetro, $\eta \in [0, 1]$.

$$\hat{Y} = \hat{Y}_a^A + \eta \hat{Y}_{ab}^A + (1 - \eta) \hat{Y}_{ab}^B + \hat{Y}_b^B$$

Por ejemplo, una solución muy simple sería tomar $\eta = 1/2$, de manera que

$$\hat{Y} = \hat{Y}_a^A + \frac{1}{2} \hat{Y}_{ab}^A + \frac{1}{2} \hat{Y}_{ab}^B + \hat{Y}_b^B$$

Dual frame versus single frame

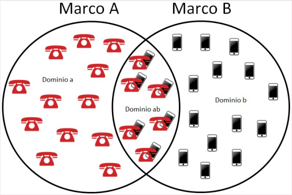
- **Solución "dual frame":** ponderar las dos estimaciones relativas al dominio de intersección mediante un parámetro, $\eta \in [0,1]$.

$$\hat{Y} = \hat{Y}_a^A + \eta \hat{Y}_{ab}^A + (1-\eta)\hat{Y}_{ab}^B + \hat{Y}_b^B$$

Por ejemplo, una solución muy simple sería tomar $\eta = 1/2$, de manera que

$$\hat{Y} = \hat{Y}_a^A + \frac{1}{2}\hat{Y}_{ab}^A + \frac{1}{2}\hat{Y}_{ab}^B + \hat{Y}_b^B$$

- **Solución "single frame":** considerar todas las unidades procedentes de un mismo marco y ajustar los pesos de las unidades del dominio de intersección, de manera que los pesos originales, d_k, se sustituyen por un nuevo conjunto de pesos, $\tilde{d}_k$.



Hartley (1962, 1974)

$$\hat{Y}_H = \left(\hat{Y}_a + \widehat{\eta}\hat{Y}_{ab}^A + (1-\widehat{\eta})\hat{Y}_{ab}^B + \hat{Y}_b\right), \text{ con } \widehat{\eta} \in (0,1)$$

Hartley (1962, 1974)

$$\hat{Y}_H = \left(\hat{Y}_a + \widehat{\eta}\hat{Y}_{ab}^A + (1-\widehat{\eta})\hat{Y}_{ab}^B + \hat{Y}_b\right), \text{ con } \widehat{\eta} \in (0,1)$$

Bankier (1986) y Kalton y Anderson (1986)

$$\hat{Y}_{BKA} = \left(\sum_{k \in s_A \cup s_B} \tilde{d}_k y_k\right),$$

$$\tilde{d}_k = \begin{cases} d_k^A, k \in a \\ (1/d_k^A + 1/d_k^B)^{-1}, k \in ab^A \cup ab^B \\ d_k^B, k \in b \end{cases}$$

El paquete de R Frames2

- Creado por los 4 autores de este taller.
- Es el único paquete estadístico para la estimación con datos provenientes de encuestas con dos marcos.
- Implementa la mayoría de los estimadores para datos provenientes de encuestas en dos marcos disponibles en la literatura.
- La primera versión se subió a CRAN en septiembre de 2014.
- La versión más reciente es la 0.1.2, de octubre de 2015.
- A pesar de ser una versión temprana, es bastante estable, ya que ha sido ampliamente testada.

El paquete de R Frames2

Estimadores

- BKA
- SFRR
- CalSF
- Hartley
- FB
- PML
- PEL
- CalDF

Jackknife

- JackBKA
- JackSFRR
- JackCalSF
- JackHartley
- JackFB
- JackPML
- JackPEL
- JackCalDF

Datos

- Dat
- DatA
- DatB
- PikIA
- PikIB

Auxiliares

- Compare
- HT
- VarHT
- CovHT
- WeightsCalSF
- WeightsCalDF

Estimación sin información auxiliar

Mínima información necesaria

- Valores de la variable respuesta en cada una de las dos muestras.
- Valores de las probabilidades de inclusión de cada individuo de cada una de las dos muestras.
- Dominio al que pertenece cada uno de los individuos dentro de cada una de las dos muestras. Por ejemplo, para los individuos de la muestra $\mathcal{S}_A$ necesitamos saber si pertenecen al dominio a o al dominio ab.

Estimadores que pueden calcularse en este contexto

- Hartley
- FB
- PEL
- CalDF

Estimación sin información auxiliar

```
> library(Frames2)
> data(DatA)
> head(DatA)
  Domain   Feed   Clo   Lei     Inc    Tax     M2 Size      ProbA     ProbB Stra
1      a 194.48 38.79 23.66 2452.07 112.90   0.00    0 0.02063274 0.0000000    1
2      a 250.23 16.92 22.68 2052.37 106.99   0.00    0 0.02063274 0.0000000    1
3     ab 199.95 24.50 23.24 2138.24 121.16 127.41    2 0.02063274 0.1133501    1
4     ab 231.29 42.60 26.76 3368.15 138.19 181.41    4 0.02063274 0.1133501    1
5      a 219.92 45.43 18.43 2897.60 128.53   0.00    0 0.02063274 0.0000000    1
6      a 186.36 34.31 24.67 2276.39 107.45   0.00    0 0.02063274 0.0000000    1
> data(DatB)
> head(DatB)
  Domain   Feed   Clo   Lei     Inc    Tax     M2 Size      ProbA     ProbB
1     ba 332.42 38.42 21.12 3109.75 148.07 186.46    3 0.02063274 0.1133501
2      b 222.47 19.94 19.74    0.00   0.00 126.79    2 0.00000000 0.1133501
3      b 215.43 35.13 20.17    0.00   0.00 148.67    3 0.00000000 0.1133501
4     ba 297.53 29.67 17.82 2523.98 133.89 154.18    3 0.02063274 0.1133501
5      b 298.69 21.37 20.19    0.00   0.00 140.58    2 0.00000000 0.1133501
6      b 267.65 27.50 10.11    0.00   0.00  90.39    2 0.00000000 0.1133501
```

Estimación sin información auxiliar

Estimador de Hartley

Hartley(ysA, ysB, pi_A, pi_B, domains_A, domains_B)

Estimación sin información auxiliar

Estimador de Hartley

Hartley(ysA, ysB, pi_A, pi_B, domains_A, domains_B)

Estimación sin información auxiliar

Estimador de Hartley

Hartley(ysA, ysB, pi_A, pi_B, domains_A, domains_B)

Estimación sin información auxiliar

Estimador de Hartley

Hartley(ysA, ysB, pi_A, pi_B, domains_A, domains_B)

Estimador de Hartley

Hartley(ysA, ysB, pi_A, pi_B, domains_A, domains_B)

```
> head(DatA)
  Domain   Feed   Clo   Lei     Inc    Tax     M2 Size      ProbA     ProbB Stra
1      a 194.48 38.79 23.66 2452.07 112.90   0.00    0 0.02063274 0.0000000    1
2      a 250.23 16.92 22.68 2052.37 106.99   0.00    0 0.02063274 0.0000000    1
3     ab 199.95 24.50 23.24 2138.24 121.16 127.41    2 0.02063274 0.1133501    1
4     ab 231.29 42.60 26.76 3368.15 138.19 181.41    4 0.02063274 0.1133501    1
5      a 219.92 45.43 18.43 2897.60 128.53   0.00    0 0.02063274 0.0000000    1
6      a 186.36 34.31 24.67 2276.39 107.45   0.00    0 0.02063274 0.0000000    1
>
> Est <- Hartley(DatA$Feed, DatB$Feed, DatA$ProbA, DatB$ProbB, DatA$Domain,
+ DatB$Domain)
> Est

Estimation:
             [,1]
Total 570867.8042
Mean     247.9484
```

Estimación sin información auxiliar

```
> summary(Est)

Call:
Hartley(ysA = DatA$Feed, ysB = DatB$Feed, pi_A = DatA$ProbA,
    pi_B = DatB$ProbB, domains_A = DatA$Domain, domains_B = DatB$Domain)

Estimation:
             [,1]
Total 570867.8042
Mean     247.9484

Variance Estimation:
                   [,1]
Var. Total 9.050344e+08
Var. Mean  1.707326e+02
```

Estimación sin información auxiliar

Pero... ¿son realmente necesarias las encuestas con dos marcos?

Vamos a comprobarlo con nuestros datos: vamos a calcular esta misma estimación suponiendo que los datos provienen de un único marco muestral y vamos a comparar los resultados que obtengamos con los del estimador de Hartley.

Estimación sin información auxiliar

Pero... ¿son realmente necesarias las encuestas con dos marcos?

Vamos a comprobarlo con nuestros datos: vamos a calcular esta misma estimación suponiendo que los datos provienen de un único marco muestral y vamos a comparar los resultados que obtengamos con los del estimador de Hartley.

```
> y <- c(DatA$Feed, DatB$Feed)
> prob <- c(DatA$ProbA, DatB$ProbB)
> HT(y, prob)
[1] 723148.7
>
> Hartley(DatA$Feed, DatB$Feed, DatA$ProbA, DatB$ProbB,
+ DatA$Domain, DatB$Domain)

Estimation:
              [,1]
Total 570867.8042
Mean     247.9484
```

Estimación sin información auxiliar

```
Total Domain Estimations:
                  [,1]
Total dom. a  263233.1
Total dom. ab 166651.7
Total dom. b  164559.2
Total dom. ba 128704.7

Mean Domain Estimations:
                 [,1]
Mean dom. a  251.8133
Mean dom. ab 241.6468
Mean dom. b  242.2443
Mean dom. ba 251.5291

Parameters:
           [,1]
theta 0.3787075
```

Estimación sin información auxiliar

Estimadores FB, PEL y CalDF

FB(ysA, ysB, pi_A, pi_B, domains_A, domains_B)
PEL(ysA, ysB, pi_A, pi_B, domains_A, domains_B)
CalDF(ysA, ysB, pi_A, pi_B, domains_A, domains_B)

```
> FB(DatA$Feed, DatB$Feed, DatA$ProbA, DatB$ProbB, DatA$Domain, DatB$Domain)

Estimation:
             [,1]
Total 571971.9511
Mean     248.4279
> CalDF(DatA$Feed, DatB$Feed, DatA$ProbA, DatB$ProbB, DatA$Domain, DatB$Domain)

Estimation:
             [,1]
Total 595162.2604
Mean     248.8422
```

Estimación sin información auxiliar

Mínima información necesaria

- Valores de la variable respuesta en cada una de las dos muestras.
- Valores de las probabilidades de inclusión de cada individuo de cada una de las dos muestras.
- Dominio al que pertenece cada uno de los individuos dentro de cada una de las dos muestras. Por ejemplo, para los individuos de la muestra $\mathcal{S}_A$ necesitamos saber si pertenecen al dominio a o al dominio ab.

Estimadores que pueden calcularse en este contexto

- Hartley
- FB
- PEL
- CalDF

Estimación sin información auxiliar

Mínima información necesaria

- Valores de la variable respuesta en cada una de las dos muestras.
- Valores de las probabilidades de inclusión de cada individuo de cada una de las dos muestras.
- Dominio al que pertenece cada uno de los individuos dentro de cada una de las dos muestras. Por ejemplo, para los individuos de la muestra $\mathcal{S}_A$ necesitamos saber si pertenecen al dominio a o al dominio ab.

Si, adicionalmente, se conocen

- Las probabilidades de inclusión bajo los diseños **de ambos marcos** de las unidades contenidas en el dominio de intersección.

Estimadores que pueden calcularse en este contexto

- Hartley
- FB
- PEL
- CalDF

Estimación sin información auxiliar

Mínima información necesaria

- Valores de la variable respuesta en cada una de las dos muestras.
- Valores de las probabilidades de inclusión de cada individuo de cada una de las dos muestras.
- Dominio al que pertenece cada uno de los individuos dentro de cada una de las dos muestras. Por ejemplo, para los individuos de la muestra $\mathcal{S}_A$ necesitamos saber si pertenecen al dominio a o al dominio ab.

Si, adicionalmente, se conocen

- Las probabilidades de inclusión bajo los diseños **de ambos marcos** de las unidades contenidas en el dominio de intersección.

Estimadores que pueden calcularse en este contexto

- Hartley
- FB
- PEL
- CalDF
- BKA
- CalSF

Estimación sin información auxiliar

Estimador BKA

BKA(ysA, ysB, pi_A, pi_B, pik_ab_B, pik_ba_A, domains_A, domains_B)

Estimación sin información auxiliar

Estimador BKA

BKA(ysA, ysB, pi_A, pi_B, pik_ab_B, pik_ba_A, domains_A, domains_B)

Estimación sin información auxiliar

Estimador BKA

BKA(ysA, ysB, pi_A, pi_B, pik_ab_B, pik_ba_A, domains_A, domains_B)

Estimación sin información auxiliar

Estimador BKA

BKA(ysA, ysB, pi_A, pi_B, pik_ab_B, pik_ba_A, domains_A, domains_B)

```
> head(DatA)
  Domain   Feed   Clo   Lei     Inc    Tax     M2 Size      ProbA     ProbB Stra
1      a 194.48 38.79 23.66 2452.07 112.90   0.00    0 0.02063274 0.0000000    1
2      a 250.23 16.92 22.68 2052.37 106.99   0.00    0 0.02063274 0.0000000    1
3     ab 199.95 24.50 23.24 2138.24 121.16 127.41    2 0.02063274 0.1133501    1
4     ab 231.29 42.60 26.76 3368.15 138.19 181.41    4 0.02063274 0.1133501    1
5      a 219.92 45.43 18.43 2897.60 128.53   0.00    0 0.02063274 0.0000000    1
6      a 186.36 34.31 24.67 2276.39 107.45   0.00    0 0.02063274 0.0000000    1
>
> Est <- BKA(DatA$Feed, DatB$Feed, DatA$ProbA, DatB$ProbB, DatA$ProbB,
+ DatB$ProbA, DatA$Domain, DatB$Domain)
> Est

Estimation:
             [,1]
Total 566434.3200
Mean     247.8845
```

Estimación sin información auxiliar

Estimador CalSF

CalSF(ysA, ysB, pi_A, pi_B, pik_ab_B, pik_ba_A, domains_A, domains_B)

```
> CalSF(DatA$Feed, DatB$Feed, DatA$ProbA, DatB$ProbB, DatA$ProbB, DatB$ProbA,
+ DatA$Domain, DatB$Domain)

Estimation:
             [,1]
Total 587175.1891
Mean     248.2153
```

Estimación sin información auxiliar

Estimador CalSF

CalSF(ysA, ysB, pi_A, pi_B, pik_ab_B, pik_ba_A, domains_A, domains_B)

```
> CalSF(DatA$Feed, DatB$Feed, DatA$ProbA, DatB$ProbB, DatA$ProbB, DatB$ProbA,
+ DatA$Domain, DatB$Domain)

Estimation:
              [,1]
Total 587175.1891
Mean     248.2153
```

Nota: La función `summary` también puede aplicarse sobre estos estimadores para obtener información adicional sobre la estimación.

Estimación sin información auxiliar

Intervalos de confianza

- Para cualquiera de los estimadores presentados hasta el momento es posible obtener un intervalo de confianza según el método propuesto por cada autor para el nivel de confianza que se desee.
- Para ello, basta con indicar el nivel de confianza mediante el argumento `conf_level`.

```
> Hartley(DatA$Feed, DatB$Feed, DatA$ProbA, DatB$ProbB, DatA$Domain,
+ DatB$Domain, conf_level = 0.95)

Estimation and  95 % Confidence Intervals:
                   [,1]
Total       570867.8042
Lower Bound 511904.6588
Upper Bound 629830.9496
Mean           247.9484
Lower Bound    222.3386
Upper Bound    273.5582
```

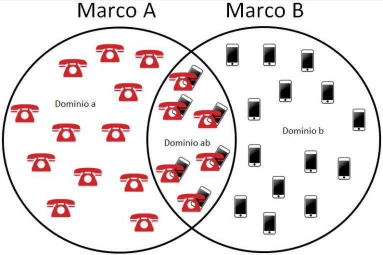
Estimación sin información auxiliar

```
> CalDF(DatA$Feed, DatB$Feed, DatA$ProbA, DatB$ProbB, DatA$Domain,
+ DatB$Domain, conf_level = 0.90)

Estimation and  90 % Confidence Intervals:
                   [,1]
Total       595162.2604
Lower Bound 577020.4135
Upper Bound 613304.1073
Mean           248.8422
Lower Bound    241.2569
Upper Bound    256.4274
>
> BKA(DatA$Feed, DatB$Feed, DatA$ProbA, DatB$ProbB, DatA$ProbB,
+ DatB$ProbA, DatA$Domain, DatB$Domain, conf_level = 0.95)

Estimation and  95 % Confidence Intervals:
                   [,1]
Total       566434.3200
Lower Bound 508299.4262
Upper Bound 624569.2139
Mean           247.8845
Lower Bound    222.4434
Upper Bound    273.3257
```

Estimación con información del tamaño de los marcos

Información conocida

- Supongamos que ahora, además del conjunto de información mínima, conocemos el número de inviduos total que hay en cada marco.



- Denotaremos a estas cantidades N_A para el conjunto de individuos del marco A y N_B para el del marco B.
- En este caso, podemos incorporar esta información al proceso de estimación para obtener estimadores más precisos.

Estimación con información del tamaño de los marcos

Información conocida

- Supongamos que ahora, además del conjunto de información mínima, conocemos el número de inviduos total que hay en cada marco.
- Denotaremos a estas cantidades N_A para el conjunto de individuos del marco A y N_B para el del marco B.
- En este caso, podemos incorporar esta información al proceso de estimación para obtener estimadores más precisos.

Estimación con información del tamaño de los marcos

Información conocida

- Supongamos que ahora, además del conjunto de información mínima, conocemos el número de inviduos total que hay en cada marco.
- Denotaremos a estas cantidades N_A para el conjunto de individuos del marco A y N_B para el del marco B.
- En este caso, podemos incorporar esta información al proceso de estimación para obtener estimadores más precisos.

Estimadores que pueden calcularse en este contexto

- Hartley
- FB

Estimación con información del tamaño de los marcos

Información conocida

- Supongamos que ahora, además del conjunto de información mínima, conocemos el número de inviduos total que hay en cada marco.
- Denotaremos a estas cantidades N_A para el conjunto de individuos del marco A y N_B para el del marco B.
- En este caso, podemos incorporar esta información al proceso de estimación para obtener estimadores más precisos.

Estimadores que pueden calcularse en este contexto

- Hartley
- FB
- PEL
- CalDF
- PML

Estimación con información del tamaño de los marcos

Estimador PML

PML(ysA, ysB, pi_A, pi_B, domains_A, domains_B, N_A, N_B)

Estimación con información del tamaño de los marcos

Estimador PML

PML(ysA, ysB, pi_A, pi_B, domains_A, domains_B, N_A, N_B)

Estimación con información del tamaño de los marcos

Estimador PML

PML(ysA, ysB, pi_A, pi_B, domains_A, domains_B, N_A, N_B)

Estimación con información del tamaño de los marcos

Estimador PML

PML(ysA, ysB, pi_A, pi_B, domains_A, domains_B, N_A, N_B)

```
> PML(DatA$Feed, DatB$Feed, DatA$ProbA, DatB$ProbB, DatA$Domain, DatB$Domain
+ N_A = 1735, N_B = 1191)

Estimation:
              [,1]
Total 594400.6320
Mean     248.0934
```

Estimación con información del tamaño de los marcos

Estimador PML

PML(ysA, ysB, pi_A, pi_B, domains_A, domains_B, N_A, N_B)

```
> PML(DatA$Feed, DatB$Feed, DatA$ProbA, DatB$ProbB, DatA$Domain, DatB$Domain
+ N_A = 1735, N_B = 1191)

Estimation:
             [,1]
Total 594400.6320
Mean     248.0934
```

Estimador CalDF

CalDF(ysA, ysB, pi_A, pi_B, domains_A, domains_B, N_A, N_B)

```
> CalDF(DatA$Feed, DatB$Feed, DatA$ProbA, DatB$ProbB, DatA$Domain, DatB$Domain
+ N_A = 1735, N_B = 1191)

Estimation:
             [,1]
Total 588416.4644
Mean     248.8131
```

Estimación con información del tamaño de los marcos

Información conocida

- Mínima información necesaria.
- Tamaño de los marcos, N_A y N_B.
- Probabilidades de inclusión bajo los diseños **de ambos marcos** de las unidades contenidas en el dominio de intersección.

Estimación con información del tamaño de los marcos

Información conocida

- Mínima información necesaria.
- Tamaño de los marcos, N_A y N_B.
- Probabilidades de inclusión bajo los diseños **de ambos marcos** de las unidades contenidas en el dominio de intersección.

Estimadores que pueden calcularse en este contexto

- Hartley
- FB
- PEL
- CalDF
- PML
- BKA

Estimación con información del tamaño de los marcos

Información conocida

- Mínima información necesaria.
- Tamaño de los marcos, N_A y N_B.
- Probabilidades de inclusión bajo los diseños **de ambos marcos** de las unidades contenidas en el dominio de intersección.

Estimadores que pueden calcularse en este contexto

- Hartley
- FB
- PEL
- CalDF
- PML
- BKA
- SFRR
- CalSF

Estimación con información del tamaño de los marcos

Estimador SFRR

SFRR(ysA, ysB, pi_A, pi_B, pik_ab_B, pik_ba_A, domains_A, domains_B, N_A, N_B)

Estimación con información del tamaño de los marcos

Estimador SFRR

SFRR(ysA, ysB, pi_A, pi_B, pik_ab_B, pik_ba_A, domains_A, domains_B, N_A, N_B)

Estimación con información del tamaño de los marcos

Estimador SFRR

SFRR(ysA, ysB, pi_A, pi_B, pik_ab_B, pik_ba_A, domains_A, domains_B, N_A, N_B)

Estimación con información del tamaño de los marcos

Estimador SFRR

SFRR(ysA, ysB, pi_A, pi_B, pik_ab_B, pik_ba_A, domains_A, domains_B, N_A, N_B)

Estimación con información del tamaño de los marcos

Estimador SFRR

SFRR(ysA, ysB, pi_A, pi_B, pik_ab_B, pik_ba_A, domains_A, domains_B, N_A, N_B)

```
> SFRR(DatA$Feed, DatB$Feed, DatA$ProbA, DatB$ProbB, DatA$ProbB,
+ DatB$ProbA, DatA$Domain, DatB$Domain, N_A = 1735, N_B = 1191)

Estimation:
             [,1]
Total 584713.4070
Mean     248.2219
```

Estimación con información del tamaño de los marcos

Estimador SFRR

SFRR(ysA, ysB, pi_A, pi_B, pik_ab_B, pik_ba_A, domains_A, domains_B, N_A, N_B)

```
> SFRR(DatA$Feed, DatB$Feed, DatA$ProbA, DatB$ProbB, DatA$ProbB,
+ DatB$ProbA, DatA$Domain, DatB$Domain, N_A = 1735, N_B = 1191)

Estimation:
             [,1]
Total 584713.4070
Mean     248.2219
```

Estimador CalSF

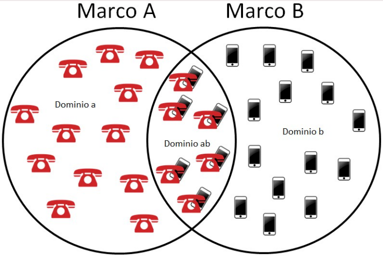
CalSF(ysA, ysB, pi_A, pi_B, pik_ab_B, pik_ba_A, domains_A, domains_B, N_A, N_B)

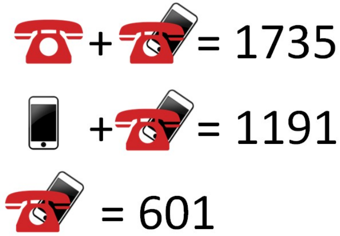
```
> CalSF(DatA$Feed, DatB$Feed, DatA$ProbA, DatB$ProbB, DatA$ProbB,
+ DatB$ProbA, DatA$Domain, DatB$Domain, N_A = 1735, N_B = 1191)

Estimation:
             [,1]
Total 584136.2083
Mean     248.2235
```

Estimación con información del tamaño de los marcos y del dominio de solapamiento

Información conocida

- Mínima información necesaria.
- Tamaño de los marcos, N_A y N_B.
- Tamaño del dominio de solapamiento, al que denotaremos por N_{ab}.

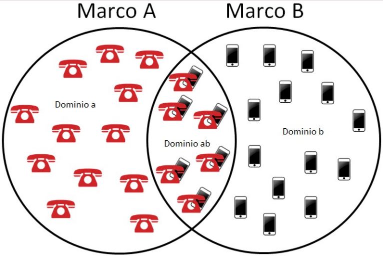


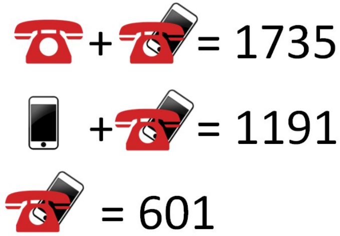


Estimación con información del tamaño de los marcos y del dominio de solapamiento

Información conocida

- Mínima información necesaria.
- Tamaño de los marcos, N_A y N_B.
- Tamaño del dominio de solapamiento, al que denotaremos por N_{ab}.





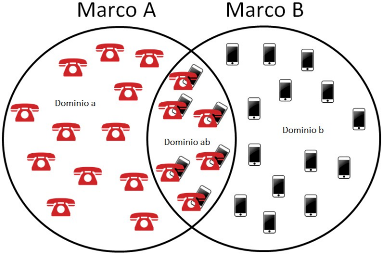
Estimadores que pueden calcularse en este contexto

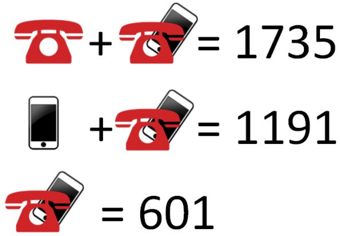
- Hartley
- FB
- PML

Estimación con información del tamaño de los marcos y del dominio de solapamiento

Información conocida

- Mínima información necesaria.
- Tamaño de los marcos, N_A y N_B.
- Tamaño del dominio de solapamiento, al que denotaremos por N_{ab}.





Estimadores que pueden calcularse en este contexto

- Hartley
- FB
- PML
- PEL
- CalDF

Est. con info. del tamaño de los marcos y del solapamiento

Estimador PEL

PEL(ysA, ysB, pi_A, pi_B, domains_A, domains_B, N_A, N_B, N_ab)

Est. con info. del tamaño de los marcos y del solapamiento

Estimador PEL

PEL(ysA, ysB, pi_A, pi_B, domains_A, domains_B, N_A, N_B, N_ab)

Est. con info. del tamaño de los marcos y del solapamiento

Estimador PEL

PEL(ysA, ysB, pi_A, pi_B, domains_A, domains_B, N_A, N_B, N_ab)

Est. con info. del tamaño de los marcos y del solapamiento

Estimador PEL

PEL(ysA, ysB, pi_A, pi_B, domains_A, domains_B, N_A, N_B, N_ab)

Est. con info. del tamaño de los marcos y del solapamiento

Estimador PEL

PEL(ysA, ysB, pi_A, pi_B, domains_A, domains_B, N_A, N_B, N_ab)

```
> PEL(DatA$Feed, DatB$Feed, DatA$ProbA, DatB$ProbB,DatA$Domain,
+ DatB$Domain, N_A = 1735, N_B = 1191, N_ab = 601)

Estimation:
              [,1]
Total 575289.2187
Mean     247.4362
```

Est. con info. del tamaño de los marcos y del solapamiento

Estimador PEL

PEL(ysA, ysB, pi_A, pi_B, domains_A, domains_B, N_A, N_B, N_ab)

```
> PEL(DatA$Feed, DatB$Feed, DatA$ProbA, DatB$ProbB,DatA$Domain,
+ DatB$Domain, N_A = 1735, N_B = 1191, N_ab = 601)

Estimation:
             [,1]
Total 575289.2187
Mean     247.4362
```

Estimador CalDF

CalDF(ysA, ysB, pi_A, pi_B, domains_A, domains_B, N_A, N_B, N_ab)

```
> CalDF(DatA$Feed, DatB$Feed, DatA$ProbA, DatB$ProbB,DatA$Domain,
+ DatB$Domain, N_A = 1735, N_B = 1191, N_ab = 601)

Estimation:
             [,1]
Total 578895.6961
Mean     248.9874
```

Estimación con información del tamaño de los marcos y del dominio de solapamiento

Información conocida

- Mínima información necesaria.
- Tamaño de los marcos, N_A y N_B.
- Tamaño del dominio de solapamiento, al que denotaremos por N_{ab}.
- Probabilidades de inclusión bajo los diseños **de ambos marcos** de las unidades contenidas en el dominio de intersección.

Estimadores que pueden calcularse en este contexto

- Hartley
- FB
- PEL
- CalDF
- PML
- BKA
- SFRR

Estimación con información del tamaño de los marcos y del dominio de solapamiento

Información conocida

- Mínima información necesaria.
- Tamaño de los marcos, N_A y N_B.
- Tamaño del dominio de solapamiento, al que denotaremos por N_{ab}.
- Probabilidades de inclusión bajo los diseños **de ambos marcos** de las unidades contenidas en el dominio de intersección.

Estimadores que pueden calcularse en este contexto

- Hartley
- FB
- PEL
- CalDF
- PML
- BKA
- SFRR
- CalSF

Estimación con información del tamaño de los marcos y del dominio de solapamiento

Estimador CalSF

CalSF(ysA, ysB, pi_A, pi_B, pik_ab_B, pik_ba_A, domains_A, domains_B, N_A, N_B, N_ab)

Estimación con información del tamaño de los marcos y del dominio de solapamiento

Estimador CalSF

CalSF(ysA, ysB, pi_A, pi_B, pik_ab_B, pik_ba_A, domains_A, domains_B, N_A, N_B, N_ab)

Estimación con información del tamaño de los marcos y del dominio de solapamiento

Estimador CalSF

CalSF(ysA, ysB, pi_A, pi_B, pik_ab_B, pik_ba_A, domains_A, domains_B, N_A, N_B, N_ab)

Estimación con información del tamaño de los marcos y del dominio de solapamiento

Estimador CalSF

CalSF(ysA, ysB, pi_A, pi_B, pik_ab_B, pik_ba_A, domains_A, domains_B, N_A, N_B, N_ab)

Estimación con información del tamaño de los marcos y del dominio de solapamiento

Estimador CalSF

CalSF(ysA, ysB, pi_A, pi_B, pik_ab_B, pik_ba_A, domains_A, domains_B, N_A, N_B, N_ab)

Estimación con información del tamaño de los marcos y del dominio de solapamiento

Estimador CalSF

CalSF(ysA, ysB, pi_A, pi_B, pik_ab_B, pik_ba_A, domains_A, domains_B, N_A, N_B, N_ab)

```
> CalSF(DatA$Feed, DatB$Feed, DatA$ProbA, DatB$ProbB, DatA$ProbB,
+ DatB$ProbB, DatA$Domain, DatB$Domain, N_A = 1735, N_B = 1191,
+ N_ab = 601)

Estimation:
             [,1]
Total 577071.959
Mean     248.203
```

Estimación con información de variables auxiliares

Información conocida

- Mínima información necesaria.
- Tamaño de los marcos, N_A y N_B, (y del dominio de solapamiento, N_{ab}).
- Información sobre una (o varias) variable(s) auxiliar(es) en el marco A y/o información sobre una (o varias) variable(s) auxiliar(es) en el marco B.

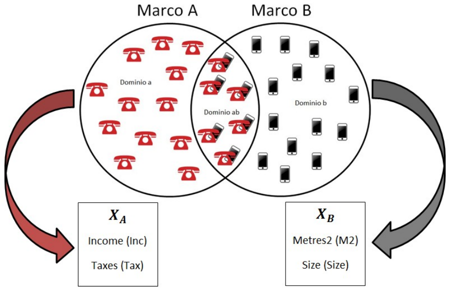
Estimación con información de variables auxiliares

```
> head(DatA)
  Domain   Feed   Clo   Lei     Inc    Tax     M2 Size      ProbA     ProbB Stra
1      a 194.48 38.79 23.66 2452.07 112.90   0.00    0 0.02063274 0.0000000    1
2      a 250.23 16.92 22.68 2052.37 106.99   0.00    0 0.02063274 0.0000000    1
3     ab 199.95 24.50 23.24 2138.24 121.16 127.41    2 0.02063274 0.1133501    1
4     ab 231.29 42.60 26.76 3368.15 138.19 181.41    4 0.02063274 0.1133501    1
5      a 219.92 45.43 18.43 2897.60 128.53   0.00    0 0.02063274 0.0000000    1
6      a 186.36 34.31 24.67 2276.39 107.45   0.00    0 0.02063274 0.0000000    1
> head(DatB)
  Domain   Feed   Clo   Lei     Inc    Tax     M2 Size      ProbA     ProbB
1     ba 332.42 38.42 21.12 3109.75 148.07 186.46    3 0.02063274 0.1133501
2      b 222.47 19.94 19.74    0.00   0.00 126.79    2 0.00000000 0.1133501
3      b 215.43 35.13 20.17    0.00   0.00 148.67    3 0.00000000 0.1133501
4     ba 297.53 29.67 17.82 2523.98 133.89 154.18    3 0.02063274 0.1133501
5      b 298.69 21.37 20.19    0.00   0.00 140.58    2 0.00000000 0.1133501
6      b 267.65 27.50 10.11    0.00   0.00  90.39    2 0.00000000 0.1133501
```

Estimación con información de variables auxiliares

Información conocida

- Mínima información necesaria.
- Tamaño de los marcos, N_A y N_B, (y del dominio de solapamiento, N_{ab}).
- Información sobre una (o varias) variable(s) auxiliar(es) en el marco A y/o información sobre una (o varias) variable(s) auxiliar(es) en el marco B.



Estimación con información de variables auxiliares

Información conocida

- Mínima información necesaria.
- Tamaño de los marcos, N_A y N_B, (y del dominio de solapamiento, N_{ab}).
- Información sobre una (o varias) variable(s) auxiliar(es) en el marco A y/o información sobre una (o varias) variable(s) auxiliar(es) en el marco B junto a sus respectivos totales poblacionales.

Estimadores que pueden calcularse en este contexto

- Hartley
- FB
- PML

Estimación con información de variables auxiliares

Información conocida

- Mínima información necesaria.
- Tamaño de los marcos, N_A y N_B, (y del dominio de solapamiento, N_{ab}).
- Información sobre una (o varias) variable(s) auxiliar(es) en el marco A y/o información sobre una (o varias) variable(s) auxiliar(es) en el marco B junto a sus respectivos totales poblacionales.

Estimadores que pueden calcularse en este contexto

- Hartley
- FB
- PML
- PEL
- CalDF

Nota: Por simplicidad vamos a considerar, como máximo, una variable auxiliar en cada marco (*Income* en el marco A y *Metres2* en el marco B), aunque podrían utilizarse tantas como se quisieran.

Estimación con información de variables auxiliares

Estimador PEL

PEL(ysA, ysB, pi_A, pi_B, domains_A, domains_B, N_A, N_B, xsAFrameA, xsBFrameA, xsAFrameB, xsBFrameB, XA, XB)

Estimador PEL

PEL(ysA, ysB, pi_A, pi_B, domains_A, domains_B, N_A, N_B, xsAFrameA, xsBFrameA, xsAFrameB, xsBFrameB, XA, XB)

Estimación con información de variables auxiliares

Estimador PEL

PEL(ysA, ysB, pi_A, pi_B, domains_A, domains_B, N_A, N_B, xsAFrameA, xsBFrameA, xsAFrameB, xsBFrameB, XA, XB)

Estimación con información de variables auxiliares

Estimador PEL

PEL(ysA, ysB, pi_A, pi_B, domains_A, domains_B, N_A, N_B, xsAFrameA, xsBFrameA, xsAFrameB, xsBFrameB, XA, XB)

Estimador PEL

PEL(ysA, ysB, pi_A, pi_B, domains_A, domains_B, N_A, N_B, xsAFrameA, xsBFrameA, xsAFrameB, xsBFrameB, XA, XB)

```
> PEL(DatA$Feed, DatB$Feed, DatA$ProbA, DatB$ProbB, DatA$Domain, DatB$Domain,
+ 1735, 1191, xsAFrameA = DatA$Inc, xsBFrameA = DatB$Inc, XA = 4300260)

Estimation:
             [,1]
Total 590478.1484
Mean     246.8837
>
> PEL(DatA$Feed, DatB$Feed, DatA$ProbA, DatB$ProbB, DatA$Domain, DatB$Domain,
1735, 1191, 601, xsAFrameA = DatA$Inc, xsBFrameA = DatB$Inc, XA = 4300260)

Estimation:
             [,1]
Total 573507.2840
Mean     246.6698
```

```
> PEL(DatA$Feed, DatB$Feed, DatA$ProbA, DatB$ProbB, DatA$Domain, DatB$Domain,
+ 1735, 1191, xsAFrameB = DatA$M2, xsBFrameB = DatB$M2, XB = 176553)

Estimation:
             [,1]
Total 590674.6280
Mean     246.9659
>
> PEL(DatA$Feed, DatB$Feed, DatA$ProbA, DatB$ProbB, DatA$Domain, DatB$Domain,
+ 1735, 1191, xsAFrameA = DatA$Inc, xsBFrameA = DatB$Inc, xsAFrameB = DatA$M2,
+ xsBFrameB = DatB$M2, XA = 4300260, XB = 176553)

Estimation:
             [,1]
Total 589273.9262
Mean     246.3802
```

Estimación con información de variables auxiliares

Información conocida

- Mínima información necesaria.
- Tamaño de los marcos, N_A y N_B, (y del dominio de solapamiento, N_{ab}).
- Información sobre una (o varias) variable(s) auxiliar(es) en el marco A y/o información sobre una (o varias) variable(s) auxiliar(es) en el marco B junto a sus respectivos totales poblacionales.
- Probabilidades de inclusión bajo los diseños **de ambos marcos** de las unidades contenidas en el dominio de intersección.

Estimadores que pueden calcularse en este contexto

- Hartley
- FB
- PML
- PEL
- CalDF
- BKA
- SFRR

Estimación con información de variables auxiliares

Información conocida

- Mínima información necesaria.
- Tamaño de los marcos, N_A y N_B, (y del dominio de solapamiento, N_{ab}).
- Información sobre una (o varias) variable(s) auxiliar(es) en el marco A y/o información sobre una (o varias) variable(s) auxiliar(es) en el marco B junto a sus respectivos totales poblacionales.
- Probabilidades de inclusión bajo los diseños **de ambos marcos** de las unidades contenidas en el dominio de intersección.

Estimadores que pueden calcularse en este contexto

- Hartley
- FB
- PML
- PEL
- CalDF
- BKA
- SFRR
- CalSF

Estimación con información de variables auxiliares

Estimador CalSF

CalSF(ysA, ysB, pi_A, pi_B, pik_ab_B, pik_ba_A, domains_A, domains_B, N_A, N_B, xsAFrameA, xsBFrameA, xsAFrameB, xsBFrameB, XA, XB)

Estimación con información de variables auxiliares

Estimador CalSF

CalSF(ysA, ysB, pi_A, pi_B, pik_ab_B, pik_ba_A, domains_A, domains_B, N_A, N_B, xsAFrameA, xsBFrameA, xsAFrameB, xsBFrameB, XA, XB)

Estimación con información de variables auxiliares

Estimador CalSF

CalSF(ysA, ysB, pi_A, pi_B, pik_ab_B, pik_ba_A, domains_A, domains_B, N_A, N_B, xsAFrameA, xsBFrameA, xsAFrameB, xsBFrameB, XA, XB)

Estimación con información de variables auxiliares

Estimador CalSF

CalSF(ysA, ysB, pi_A, pi_B, pik_ab_B, pik_ba_A, domains_A, domains_B, N_A, N_B, xsAFrameA, xsBFrameA, xsAFrameB, xsBFrameB, XA, XB)

Estimación con información de variables auxiliares

Estimador CalSF

CalSF(ysA, ysB, pi_A, pi_B, pik_ab_B, pik_ba_A, domains_A, domains_B, N_A, N_B, xsAFrameA, xsBFrameA, xsAFrameB, xsBFrameB, XA, XB)

Estimación con información de variables auxiliares

Estimador CalSF

CalSF(ysA, ysB, pi_A, pi_B, pik_ab_B, pik_ba_A, domains_A, domains_B, N_A, N_B, xsAFrameA, xsBFrameA, xsAFrameB, xsBFrameB, XA, XB)

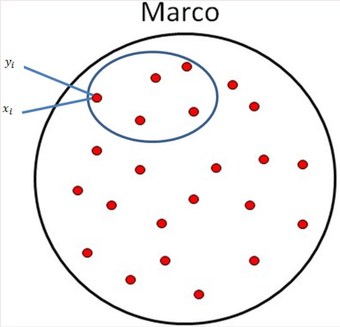
```
> CalSF(DatA$Feed, DatB$Feed, DatA$ProbA, DatB$ProbB, DatA$ProbB, DatB$ProbA,
+ DatA$Domain, DatB$Domain, 1735, 1191, xsAFrameA = DatA$Inc,
+ xsBFrameA = DatB$Inc, XA = 4300260)

Estimation:
              [,1]
Total 584700.7448
Mean     248.4826
>
> CalSF(DatA$Feed, DatB$Feed, DatA$ProbA, DatB$ProbB, DatA$ProbB, DatB$ProbA,
+ DatA$Domain, DatB$Domain, 1735, 1191, 601, xsAFrameA = DatA$Inc,
+ xsBFrameA = DatB$Inc, xsAFrameB = DatA$M2, xsBFrameB = DatB$M2, XA = 4300260,
+ XB = 176553)

Estimation:
              [,1]
Total 575636.7877
Mean     247.5857
```

Estimación con información de variables auxiliares

Unas notas sobre calibración

- En muestreo, es habitual disponer de variables auxiliares relacionadas con la variable de estudio.
- Esta información se suele incorporar a la estimación para obtener mejores resultados mediante diversos procedimientos.
- Uno de estos procedimientos es la **calibración**.

Unas notas sobre calibración



La idea de la calibración es modificar los pesos del diseño originales, d_i, para obtener unos pesos nuevos, w_i, de manera que:

- Los nuevos pesos, w_i, estén tan cerca como sea posible de los pesos originales, d_i, para que así "hereden" sus buenas propiedades. Para ello: $w_i = d_i * g_i$
- Los nuevos pesos, w_i, sean capaces de estimar de forma exacta el total de la variable auxiliar, es decir: $\sum_{i \in s} w_i x_i = X$. De esta forma, si los nuevos pesos w_i funcionan bien para la variable auxiliar, es de esperar que también funcionen bien para la variable de estudio.

De esta forma, la calibración propone estimar el total de la variable de estudio Y de esta forma:

$$\hat{Y} = \sum_{i \in s} w_i y_i$$

Estimación con información de variables auxiliares

Unas notas sobre calibración: un ejemplo sencillo

1. Vamos a utilizar el estimador de calibración single frame sobre la variable Feeding, suponiendo que conocemos el tamaño de los marcos de fijos y móviles ($N_A = 1735$ y $N_B = 1191$) y el tamaño del dominio de solapamiento ($N_{ab} = 601$).

```
> CalSF(DatA$Feed, DatB$Feed, DatA$ProbA, DatB$ProbB, DatA$ProbB,
+ DatB$ProbA, DatA$Domain, DatB$Domain, N_A = 1735,  N_B = 1191,
+ N_ab = 601)

Estimation:
             [,1]
Total 577163.6066
Mean     248.2424
```

Estimación con información de variables auxiliares

Unas notas sobre calibración: un ejemplo sencillo

2. Ahora vamos a calcular de nuevo el estimador del total, sabiendo que éste puede obtenerse como $\hat{Y} = \sum_{k \in s} \tilde{d}_k g_k y_k$. Necesitamos, por tanto:
   - Los pesos $\tilde{d}_k$, que son los pesos "single frame"

$$\tilde{d}_k = \begin{cases} d_k^A, k \in a \\ (1/d_k^A + 1/d_k^B)^{-1}, k \in ab^A \cup ab^B \\ d_k^B, k \in b \end{cases}$$

   - Los valores g_k, que son las modificaciones de los pesos originales, que hacen que se cumplan las condiciones de la calibración.
   - Los valores y_k, que son los valores de la variable a estimar.

Unas notas sobre calibración: un ejemplo sencillo

2. Ahora vamos a calcular de nuevo el estimador del total, sabiendo que éste puede obtenerse como $\hat{Y} = \sum_{k \in s} \tilde{d}_k g_k y_k$. Necesitamos, por tanto:
   - Los pesos $\tilde{d}_k$, que son los pesos "single frame"

$$\tilde{d}_k = \begin{cases} d_k^A, k \in a \\ (1/d_k^A + 1/d_k^B)^{-1}, k \in ab^A \cup ab^B \\ d_k^B, k \in b \end{cases}$$

```
> dA <- 1/DatA$ProbA
> dB <- 1/DatB$ProbB
> dA
> dA[DatA$Domain == "ab"] <- 1/(DatA$ProbA[DatA$Domain == "ab"] +
+ DatA$ProbB[DatA$Domain == "ab"])
> dA
> dB[DatB$Domain == "ba"] <- 1/(DatB$ProbB[DatB$Domain == "ba"] +
+ DatB$ProbB[DatB$Domain == "ba"])
> dtilde <- c(dA, dB)
```

Estimación con información de variables auxiliares

Unas notas sobre calibración: un ejemplo sencillo

2. Ahora vamos a calcular de nuevo el estimador del total, sabiendo que éste puede obtenerse como $\hat{Y} = \sum_{k \in s} \tilde{d}_k g_k y_k$. Necesitamos, por tanto:
   - Los valores g_k, que son las modificaciones de los pesos originales, que hacen que se cumplan las condiciones de la calibración.

```
> g <- WeightsCalSF(DatA$Feed, DatB$Feed, DatA$ProbA, DatB$ProbB,
+ DatA$ProbB, DatB$ProbA, DatA$Domain, DatB$Domain, N_A = 1735,
+ N_B = 1191, N_ab = 601)
> g
          [,1]
 [1,] 1.0848041
 [2,] 1.0848041
 [3,] 1.0724249
 [4,] 1.0724249
 [5,] 1.0848041
 [6,] 1.0848041
 [7,] 1.0724249
```

Unas notas sobre calibración: un ejemplo sencillo

2. Ahora vamos a calcular de nuevo el estimador del total, sabiendo que éste puede obtenerse como $\hat{Y} = \sum_{k \in s} \tilde{d}_k g_k y_k$. Necesitamos, por tanto:
   - Los valores y_k, que son los valores de la variable a estimar.

```
> y <- c(DatA$Feed, DatB$Feed)
```

Unas notas sobre calibración: un ejemplo sencillo

3. Ya tenemos todo lo que necesitamos. Ahora podemos calcular el estimador según la fórmula $\hat{Y} = \sum_{k \in s} \tilde{d}_k g_k y_k$.

```
> sum(dtilde * g * y)
```

Unas notas sobre calibración: un ejemplo sencillo

Pero...¿qué tienen de especial los nuevos pesos $w_k = \tilde{d}_k * g_k$?

```
> y <- c(as.numeric(DatA$Domain == "a"),
+ as.numeric(DatB$Domain == "a"))
> sum(dtilde * g * y)
> y <- c(as.numeric(DatA$Domain == "ab"),
+ as.numeric(DatB$Domain == "ba"))
> sum(dtilde * g * y)
> y <- c(as.numeric(DatA$Domain == "b"),
+ as.numeric(DatB$Domain == "b"))
> sum(dtilde * g * y)
```

¿Qué estimadores puedo calcular?

- Hasta ahora, hemos presentado en cada situación los posibles estimadores que podíamos calcular de acuerdo a la información conocida.
- Pero... ¿y si esto no es así? Es decir, ¿y si no conocemos de antemos qué estimadores se pueden calcular a partir de una determinada información?
- En este tipo de situaciones, la función `Compare` incluida en el paquete, podría ser útil.

¿Qué estimadores puedo calcular?

La función Compare

Compare (ysA, ysB, pi_A, pi_B, domains_A, domains_B, ...)

```
> Compare(DatA$Feed, DatB$Feed, DatA$ProbA, DatB$ProbB, DatA$Domain,
+ DatB$Domain)
$Hartley

Estimation:
             [,1]
Total 570867.8042
Mean     247.9484

$FullerBurmeister

Estimation:
             [,1]
Total 571971.9511
Mean     248.4279

$PEL

Estimation:
                   [,1]
Total 1.791588e+08
Mean  2.479314e+02

$Calibration_DF

Estimation:
                 [,1]
Total 595162.2604
Mean     248.8422
```

¿Qué estimadores puedo calcular?

```
> Compare(DatA$Feed, DatB$Feed, DatA$ProbA, DatB$ProbB, DatA$Domain,
+ DatB$Domain, N_A = 1735, N_B = 1191)
$PML

Estimation:
             [,1]
Total 594400.6320
Mean     248.0934
```

```
$PEL

Estimation:
             [,1]
Total 591956.1900
Mean     247.5017

$Calibration_DF

Estimation:
             [,1]
Total 588416.4644
Mean     248.8131
```

¿Por qué no aparecen los estimadores de Hartley o de FB en los resultados anteriores? La respuesta es que la función `Compare` únicamente muestra aquellos estimadores que utilizan **TODA** la información que se le suministra.

¿Qué estimadores puedo calcular?

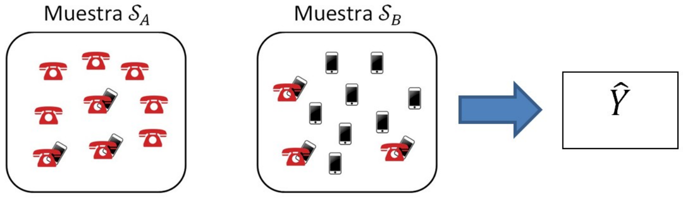
¿Qué estimador es el mejor?

- La respuesta a esta pregunta depende de la situación en la que nos encontremos: no existe un estimador que sea mejor que el resto en todas las situaciones
- No necesariamente el estimador que utiliza máyor cantidad de información auxiliar tiene por qué ser mejor que el resto (dependerá, entre otras cosas de la calidad de la información auxiliar)
- Existen muchos criterios en los cuales basar la elección del **mejor** estimador como, por ejemplo, la minimización de la varianza.
- Es arriesgado comparar los estimadores utilizando las varianzas que proporcionan los métodos analíticos.

Estimación confidencial mediante jackknife

Estimación jackknife

- El jackknife es una técnica de exploración intensiva de la muestra propuesta por Quenouille (1949, 1956)
- La técnica se basa en las estimaciones resultantes de omitir sucesivamente una unidad de la muestra de observaciones.



- El paquete `Frames2` incorpora funciones para el cálculo de intervalos de confianza mediante este método.

Estimación confidencial mediante jackknife

Estimadores

- BKA
- SFRR
- CalSF
- Hartley
- FB
- PML
- PEL
- CalDF

Jackknife

- JackBKA
- JackSFRR
- JackCalSF
- JackHartley
- JackFB
- JackPML
- JackPEL
- JackCalDF

Datos

- Dat
- DatA
- DatB
- PikIA
- PikIB

Auxiliares

- Compare
- HT
- VarHT
- CovHT
- WeightsCalSF
- WeightsCalDF

Estimación confidencial mediante jackknife

La función JackHartley

JackHartley(ysA, ysB, piA, piB, domainsA, domainsB, conf_level, sdA = "srs", sdB = "srs", strA = NULL, strB = NULL, clusA = NULL, clusB = NULL, fcpA = FALSE, fcpB = FALSE)

Estimación confidencial mediante jackknife

La función JackHartley

JackHartley(ysA, ysB, piA, piB, domainsA, domainsB, conf_level, sdA = "srs", sdB = "srs", strA = NULL, strB = NULL, clusA = NULL, clusB = NULL, fcpA = FALSE, fcpB = FALSE)

- Información necesaria para calcular el estimador

Estimación confidencial mediante jackknife

La función JackHartley

JackHartley(ysA, ysB, piA, piB, domainsA, domainsB, conf_level, sdA = "srs", sdB = "srs", strA = NULL, strB = NULL, clusA = NULL, clusB = NULL, fcpA = FALSE, fcpB = FALSE)

- Información necesaria para calcular el estimador
- Nivel de confianza (ahora obligatorio)

Estimación confidencial mediante jackknife

La función JackHartley

JackHartley(ysA, ysB, piA, piB, domainsA, domainsB, conf_level, sdA = "srs", sdB = "srs", strA = NULL, strB = NULL, clusA = NULL, clusB = NULL, fcpA = FALSE, fcpB = FALSE)

- Información necesaria para calcular el estimador
- Nivel de confianza (ahora obligatorio)
- Información relativa a los diseños muestrales utilizados en cada uno de los marcos. Los posibles valores para los diseños muestrales son:
  - "srs": muestreo aleatorio simple (por defecto)
  - "str": muestreo estratificado (en este caso, indicar la variable de estratificación en cada marco mediante `strA` y/o `strB`)
  - "pps": muestreo con probabilidades proporcionales al tamaño
  - "clu": muestreo por conglomerados (en este caso, indicar la variable de agrupación de conglomerados en cada marco mediante `cluA` y/o `cluB`)
  - "strclu": muestreo estratificado por conglomerados (en este caso, indicar tanto la variable de estratificación como la de agrupación por conglomerados)

Estimación confidencial mediante jackknife

La función JackHartley

JackHartley(ysA, ysB, piA, piB, domainsA, domainsB, conf_level, sdA = "srs", sdB = "srs", strA = NULL, strB = NULL, clusA = NULL, clusB = NULL, fcpA = FALSE, fcpB = FALSE)

- Información necesaria para calcular el estimador
- Nivel de confianza (ahora obligatorio)
- Información relativa a los diseños muestrales utilizados en cada uno de los marcos. Los posibles valores para los diseños muestrales son:
  - "srs": muestreo aleatorio simple (por defecto)
  - "str": muestreo estratificado (en este caso, indicar la variable de estratificación en cada marco mediante `strA` y/o `strB`)
  - "pps": muestreo con probabilidades proporcionales al tamaño
  - “clu”: muestreo por conglomerados (en este caso, indicar la variable de agrupación de conglomerados en cada marco mediante `cluA` y/o `cluB`)
  - "strclu": muestreo estratificado por conglomerados (en este caso, indicar tanto la variable de estratificación como la de agrupación por conglomerados)
- Factor de corrección por finitud de la población (opcional)

Estimación confidencial mediante jackknife

```
> JackHartley(DatA$Feed, DatB$Feed, DatA$ProbA, DatB$ProbB, DatA$Domain,
+ DatB$Domain, 0.95, sdA = "str", sdB = "srs", strA = DatA$Stratum)
                       [,1]
Total           570867.8042
Jack Upper End 611330.1794
Jack Lower End 530405.4290
Mean               247.9484
Jack Upper End     265.5226
Jack Lower End     230.3741
>
> Hartley(DatA$Feed, DatB$Feed, DatA$ProbA, DatB$ProbB, DatA$Domain,
+ DatB$Domain)

Estimation:
             [,1]
Total 570867.8042
Mean     247.9484
```

Estimación confidencial mediante jackknife

```
> JackHartley(DatA$Feed, DatB$Feed, DatA$ProbA, DatB$ProbB, DatA$Domain,
+ DatB$Domain, 0.95, sdA = "str", sdB = "srs", strA = DatA$Stratum,
+ fcpA = TRUE)
                      [,1]
Total          570867.8042
Jack Upper End 610664.7131
Jack Lower End 531070.8954
Mean              247.9484
Jack Upper End    265.2336
Jack Lower End    230.6631
>
> JackHartley(DatA$Feed, DatB$Feed, DatA$ProbA, DatB$ProbB, DatA$Domain,
+ DatB$Domain, 0.95, sdA = "str", sdB = "srs", strA = DatA$Stratum,
+ fcpA = TRUE, fcpB = TRUE)
                      [,1]
Total          570867.8042
Jack Upper End 610359.4803
Jack Lower End 531376.1281
Mean              247.9484
Jack Upper End    265.1010
Jack Lower End    230.7957
```

Estimación confidencial mediante jackknife

La función JackCalSF

JackCalSF(ysA, ysB, piA, piB, pik_ab_B, pik_ba_A, domainsA, domainsB, N_A = NULL, N_B = NULL, N_ab = NULL, xsAFrameA = NULL, xsBFrameA = NULL, xsAFrameB = NULL, xsBFrameB = NULL, XA = NULL, XB = NULL, conf_level, sdA = "srs", sdB = "srs", strA = NULL, strB = NULL, clusA = NULL, clusB = NULL, fcpA = FALSE, fcpB = FALSE)

Estimación confidencial mediante jackknife

La función JackCalSF

JackCalSF(ysA, ysB, piA, piB, pik_ab_B, pik_ba_A, domainsA, domainsB, N_A = NULL, N_B = NULL, N_ab = NULL, xsAFrameA = NULL, xsBFrameA = NULL, xsAFrameB = NULL, xsBFrameB = NULL, XA = NULL, XB = NULL, conf_level, sdA = "srs", sdB = "srs", strA = NULL, strB = NULL, clusA = NULL, clusB = NULL, fcpA = FALSE, fcpB = FALSE)

Estimación confidencial mediante jackknife

La función JackCalSF

JackCalSF(ysA, ysB, piA, piB, pik_ab_B, pik_ba_A, domainsA, domainsB, N_A = NULL, N_B = NULL, N_ab = NULL, xsAFrameA = NULL, xsBFrameA = NULL, xsAFrameB = NULL, xsBFrameB = NULL, XA = NULL, XB = NULL, conf_level, sdA = "srs", sdB = "srs", strA = NULL, strB = NULL, clusA = NULL, clusB = NULL, fcpA = FALSE, fcpB = FALSE)

Estimación confidencial mediante jackknife

La función JackCalSF

JackCalSF(ysA, ysB, piA, piB, pik_ab_B, pik_ba_A, domainsA, domainsB, N_A = NULL, N_B = NULL, N_ab = NULL, xsAFrameA = NULL, xsBFrameA = NULL, xsAFrameB = NULL, xsBFrameB = NULL, XA = NULL, XB = NULL, conf_level, sdA = "srs", sdB = "srs", strA = NULL, strB = NULL, clusA = NULL, clusB = NULL, fcpA = FALSE, fcpB = FALSE)

Estimación confidencial mediante jackknife

La función JackCalSF

JackCalSF(ysA, ysB, piA, piB, pik_ab_B, pik_ba_A, domainsA, domainsB, N_A = NULL, N_B = NULL, N_ab = NULL, xsAFrameA = NULL, xsBFrameA = NULL, xsAFrameB = NULL, xsBFrameB = NULL, XA = NULL, XB = NULL, conf_level, sdA = "srs", sdB = "srs", strA = NULL, strB = NULL, clusA = NULL, clusB = NULL, fcpA = FALSE, fcpB = FALSE)

Estimación confidencial mediante jackknife

La función JackCalSF

JackCalSF(ysA, ysB, piA, piB, pik_ab_B, pik_ba_A, domainsA, domainsB, N_A = NULL, N_B = NULL, N_ab = NULL, xsAFrameA = NULL, xsBFrameA = NULL, xsAFrameB = NULL, xsBFrameB = NULL, XA = NULL, XB = NULL, conf_level, sdA = "srs", sdB = "srs", strA = NULL, strB = NULL, clusA = NULL, clusB = NULL, fcpA = FALSE, fcpB = FALSE)

Estimación confidencial mediante jackknife

La función JackCalSF

JackCalSF(ysA, ysB, piA, piB, pik_ab_B, pik_ba_A, domainsA, domainsB, N_A = NULL, N_B = NULL, N_ab = NULL, xsAFrameA = NULL, xsBFrameA = NULL, xsAFrameB = NULL, xsBFrameB = NULL, XA = NULL, XB = NULL, conf_level, sdA = "srs", sdB = "srs", strA = NULL, strB = NULL, clusA = NULL, clusB = NULL, fcpA = FALSE, fcpB = FALSE)

Estimación confidencial mediante jackknife

La función JackCalSF

JackCalSF(ysA, ysB, piA, piB, pik_ab_B, pik_ba_A, domainsA, domainsB, N_A = NULL, N_B = NULL, N_ab = NULL, xsAFrameA = NULL, xsBFrameA = NULL, xsAFrameB = NULL, xsBFrameB = NULL, XA = NULL, XB = NULL, conf_level, sdA = "srs", sdB = "srs", strA = NULL, strB = NULL, clusA = NULL, clusB = NULL, fcpA = FALSE, fcpB = FALSE)

```
> JackCalSF(DatA$Feed, DatB$Feed, DatA$ProbA, DatB$ProbB, DatA$ProbB,
+ DatB$ProbA, DatA$Domain, DatB$Domain, 1735, 1191, 601, conf_level = 0.95,
+ sdA = "str", sdB = "srs", strA = DatA$Stratum, fcpA = TRUE, fcpB = TRUE)
                      [,1]
Total           577163.6066
Jack Upper End  598762.9871
Jack Lower End  555564.2261
Mean               248.2424
Jack Upper End     257.5325
Jack Lower End     238.9524
```

Preparación de un conjunto de datos

- Hasta ahora hemos trabajado con los conjuntos de datos `DatA` y `DatB`, incluidos en el paquete.
- Estos conjuntos de datos ya están adecuadamente formateados, de manera que podemos trabajar con ellos directamente.
- Pero, ¿y si esto no es así?
- En la mayoría de las ocasiones, los datos de las dos muestras están mezclados y hay que separarlos y formatearlos adecuadamente para poder usar las funciones del paquete.

```
> data(Dat)
> head(Dat)
  Drawnby Stratum Opinion Landline Cell ProbLandline ProbCell  Income
1       1       2       0        1    1 0.0006736230 8.49e-05 1629.31
2       1       5       1        1    1 0.0021932970 5.86e-05 2084.03
3       1       1       0        1    1 0.0018314890 7.81e-05 1718.65
4       1       4       0        1    0 0.0005665875 0.00e+00 1881.01
5       1       7       1        1    0 0.0002934890 0.00e+00 2185.26
6       1       6       0        1    0 0.0015381470 0.00e+00 1939.93
```

Preparación de un conjunto de datos

Esquema a seguir

1. Crear 4 subconjuntos de datos (uno para cada dominio), basándonos en los valores de las variables `Landline`, `Cell` y `Drawnby`.
2. Reconstruir las muestras de los dos marcos a partir de los subconjuntos creados en el paso anterior (de manera que obtengamos dos conjuntos de datos similares a `DatA` y `DatB`).
3. Etiquetar adecuadamente cada unidad mediante "a", "b", "ab" o "ba"para indicar el dominio al que pertenece.

Preparación de un conjunto de datos

Esquema a seguir

1. Crear 4 subconjuntos de datos (uno para cada dominio), basándonos en los valores de las variables `Landline`, `Cell` y `Drawnby`.

```
> DomainOnlyLandline <- Dat[Landline == 1 & Cell == 0,] #a
> DomainBothLandline <- Dat[Landline == 1 & Cell == 1  #ab
+ & Drawnby == 1,]
> DomainOnlyCell <- Dat[Landline == 0 & Cell == 1,]     #b
> DomainBothCell <- Dat[Landline == 1 & Cell == 1      #ba
+ & Drawnby == 2,]
```

Preparación de un conjunto de datos

Esquema a seguir

2. Reconstruir las muestras de los dos marcos a partir de los subconjuntos creados en el paso anterior (de manera que obtengamos dos conjuntos de datos similares a `DatA` y `DatB`).

```
> sLandline <- rbind(DomainOnlyLandline, DomainBothLandline) #sA
> sCell <- rbind(DomainOnlyCell, DomainBothCell)             #sB
```

Preparación de un conjunto de datos

Esquema a seguir

3. Etiquetar adecuadamente cada unidad mediante “a”, ”b”, “ab” o ”ba”para indicar el dominio al que pertenece.

```
> Domain <- c(rep("a", nrow(DomainOnlyLandline)), rep("ab",
+ nrow(DomainBothLandline)))
> sLandline <- cbind(sLandline, Domain)
> Domain <- c(rep("b", nrow(DomainOnlyCell)), rep("ba",
+ nrow(DomainBothCell)))
> sCell <- cbind(sCell, Domain)
```

Esquema a seguir

4. Ahora ya podemos utilizar cualquiera de los estimadores que se han visto hasta el momento.

```
> Hartley(sLandline$Opinion, sCell$Opinion,
+ sLandline$ProbLandline, sCell$ProbCell,
+ sLandline$Domain, sCell$Domain)

Estimation:
             [,1]
Total 3.46686e+06
Mean  4.93861e-01
```

Interfaz

Una interfaz más amigable

- La consultora estadística granadina **p-evalúa** está elaborando una interfaz de usuario para acercar el análisis de encuestas con datos provenientes de dos marcos a usuarios que no están familiarizados con R.
- La interfaz está basada en shiny.
- La interfaz cubre todas las situaciones que se han descrito durante el taller.

Interfaz

DESIGN AN ANALYSIS OF COMPLEX DUAL SURVEYS DATA | Home | Data | Estimation | Technical file | Tools | Help

Frames2

Data from complex survey designs require special consideration with regard to estimation of finite population parameters and corresponding variance estimation procedures, as a consequence of significant departures from the simple random sampling assumption. In the past decade a number of statistical software packages have been developed to facilitate the analysis of complex survey data. All these statistical software packages are able to treat samples selected from one sampling frame containing all population units. Dual frame surveys are very useful when it is not possible to guarantee a complete coverage of the target population and may result in considerable cost savings over a single frame design with comparable precision. There are several estimators available in the statistical literature but no existing software covers dual frame estimation procedures. This gap is now filled by package **Frames2**. The package includes the main estimators in dual frame surveys and also provides interval confidence estimation. Point and interval estimation in dual frame surveys. In contrast to classic sampling theory, where only one sampling frame is considered, dual frame methodology assumes that there are two frames available for sampling and that, overall, they cover the entire target population. Then, two probability samples (one from each frame) are drawn and information collected is suitably combined to get estimators of the parameter of interest.

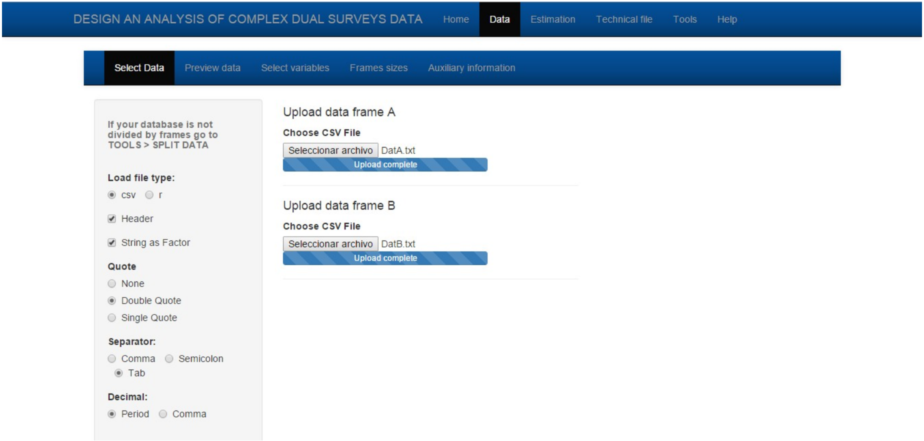
ugr Universidad de Granada

Unión Europea

Fondo Europeo
de Desarrollo Regional
"Una manera de hacer Europa"

p-evalua
Asesores Estadísticos

Interfaz



DESIGN AN ANALYSIS OF COMPLEX DUAL SURVEYS DATA
Home
Data
Estimation
Technical file
Tools
Help
Select Data
Preview data
Select variables
Frames sizes
Auxiliary information
If your database is not divided by frames go to TOOLS > SPLIT DATA
Load file type:
csv
r
Header
String as Factor
Quote
None
Double Quote
Single Quote
Separator:
Comma
Semicolon
Tab
Decimal:
Period
Comma
Upload data frame A
Choose CSV File
Seleccionar archivo
DatA.txt
Upload complete
Upload data frame B
Choose CSV File
Seleccionar archivo
DatB.txt
Upload complete

Interfaz

DESIGN AN ANALYSIS OF COMPLEX DUAL SURVEYS DATA | Home | Data | Estimation | Technical file | Tools | Help

Select Data | Preview data | Select variables | Frames sizes | Auxiliary information

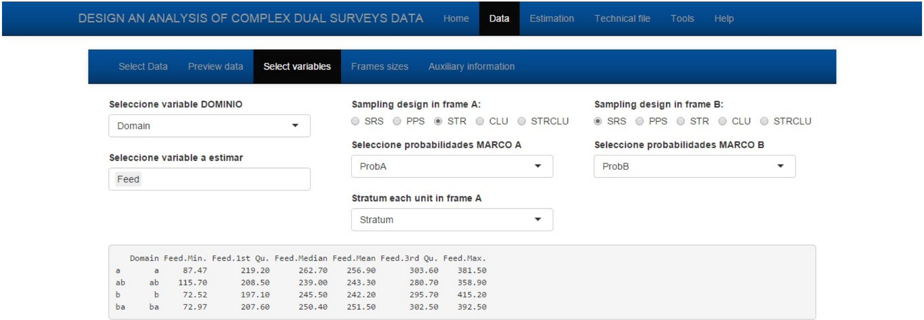
Preview observed Data A

	Domain	Feed	Clo	Lei	Inc	Tax	M2	Size	ProbA	ProbB	Stratum
1	a	194.48	38.79	23.66	2452.07	112.90	0.00	0	0.02	0.00	1
2	a	250.23	16.92	22.68	2052.37	106.99	0.00	0	0.02	0.00	1
3	ab	199.95	24.50	23.24	2138.24	121.16	127.41	2	0.02	0.11	1
4	ab	231.29	42.60	26.76	3368.15	138.19	181.41	4	0.02	0.11	1
5	a	219.92	45.43	18.43	2897.60	128.53	0.00	0	0.02	0.00	1
6	a	186.36	34.31	24.67	2276.39	107.45	0.00	0	0.02	0.00	1

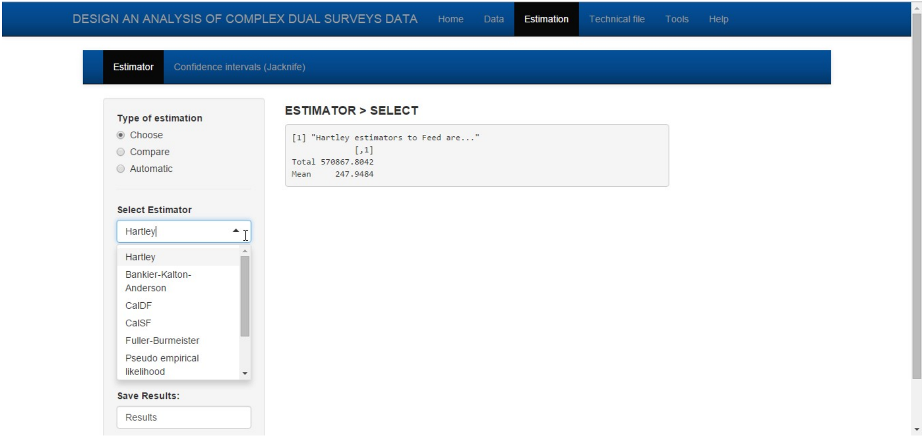
Preview observed Data B

	Domain	Feed	Clo	Lei	Inc	Tax	M2	Size	ProbA	ProbB
1	ba	332.42	38.42	21.12	3109.75	148.07	186.46	3	0.02	0.11
2	b	222.47	19.94	19.74	0.00	0.00	126.79	2	0.00	0.11
3	b	215.43	35.13	20.17	0.00	0.00	148.67	3	0.00	0.11
4	ba	297.53	29.67	17.82	2523.98	133.89	154.18	3	0.02	0.11
5	b	298.69	21.37	20.19	0.00	0.00	140.58	2	0.00	0.11

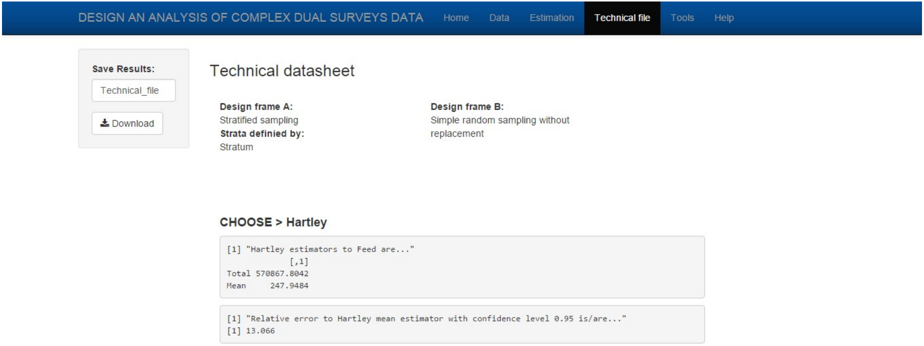
Interfaz


DESIGN AN ANALYSIS OF COMPLEX DUAL SURVEYS DATA
Home
Data
Estimation
Technical file
Tools
Help
Select Data
Preview data
Select variables
Frames sizes
Auxiliary information
Seleccione variable DOMINIO
Domain
Seleccione variable a estimar
Feed
Sampling design in frame A:
SRS PPS STR CLU STRCLU
Seleccione probabilidades MARCO A
ProbA
Stratum each unit in frame A
Stratum
Sampling design in frame B:
SRS PPS STR CLU STRCLU
Seleccione probabilidades MARCO B
ProbB
Domain Feed.Min. Feed.1st Qu. Feed.Median Feed.Mean Feed.3rd Qu. Feed.Max.
a a 87.47 219.20 262.70 256.90 303.60 381.50
ab ab 115.70 208.50 239.00 243.30 280.70 358.90
b b 72.52 197.10 245.50 242.20 295.70 415.20
ba ba 72.97 207.60 250.40 251.50 302.50 392.50

Interfaz


DESIGN AN ANALYSIS OF COMPLEX DUAL SURVEYS DATA
Home
Data
Estimation
Technical file
Tools
Help
Estimator
Confidence intervals (Jacknife)
Type of estimation
Choose
Compare
Automatic
Select Estimator
Hartley
Hartley
Bankier-Kalton-Anderson
CalDF
CalSF
Fuller-Burmeister
Pseudo empirical likelihood
Save Results:
Results
ESTIMATOR > SELECT
[1] "Hartley estimators to Feed are..."
[,1]
Total 570867.8042
Mean 247.9484

Interfaz


DESIGN AN ANALYSIS OF COMPLEX DUAL SURVEYS DATA
Home
Data
Estimation
Technical file
Tools
Help
Save Results:
Technical_file
Download
Technical datasheet
Design frame A:
Stratified sampling
Strata definied by:
Stratum
Design frame B:
Simple random sampling without replacement
CHOOSE > Hartley
[1] "Hartley estimators to Feed are..."
[,1]
Total 570867.8042
Mean 247.9484
[1] "Relative error to Hartley mean estimator with confidence level 0.95 is/are..."
[1] 13.066

Algunos ejemplos más

¿Y si...

- ...queremos realizar estimaciones para más de una variable principal?

```
> var_sA <- data.frame(DatA$Feed, DatA$Clo, DatA$Lei)
> var_sB <- data.frame(DatB$Feed, DatB$Clo, DatB$Lei)
> names(var_sA) <- c("Feed", "Clo", "Lei")
> names(var_sB) <- c("Feed", "Clo", "Lei")
> Hartley(var_sA, var_sB, DatA$ProbA, DatB$ProbB,
+ DatA$Domain, DatB$Domain)

Estimation:
             Feed         Clo        Lei
Total 570867.8042 69473.86532 51284.2727
Mean     247.9484    30.17499    22.2746
```

¿Y si...

- ...disponemos de las probabilidades de inclusión de segundo orden, π_{ij}?

```
> data(PiklA)
> data(PiklB)
> PiklA[1:6, 1:6]
            [,1]        [,2]        [,3]        [,4]        [,5]        [,6]
[1,] 0.020632737 0.000397876 0.000397876 0.000397876 0.000397876 0.000397876
[2,] 0.000397876 0.020632737 0.000397876 0.000397876 0.000397876 0.000397876
[3,] 0.000397876 0.000397876 0.020632737 0.000397876 0.000397876 0.000397876
[4,] 0.000397876 0.000397876 0.000397876 0.020632737 0.000397876 0.000397876
[5,] 0.000397876 0.000397876 0.000397876 0.000397876 0.020632737 0.000397876
[6,] 0.000397876 0.000397876 0.000397876 0.000397876 0.000397876 0.020632737
> dim(PiklA)
[1] 105 105
> Hartley(var_sA, var_sB, PiklA, PiklB, DatA$Domain, DatB$Domain)
```

Algunos ejemplos más

¿Y si...

- ...disponemos de más de una variable auxiliar?

```
> auxsAFrameA <- data.frame(DatA$Inc, DatA$Tax)
> auxsBFrameA <- data.frame(DatB$Inc, DatB$Tax)
> auxsAFrameB <- data.frame(DatA$M2, DatA$Size)
> auxsBFrameB <- data.frame(DatB$M2, DatB$Size)
> totFrameA <- c(4300260, 215577)
> totFrameB <- c(176553, 3529)
> CalDF (DatA$Feed, DatB$Feed, DatA$ProbA, DatB$ProbB, DatA$Domain,
+ DatB$Domain, N_A = 1735, N_B = 1191, N_ab = 601, xsAFrameA = auxsAFrameA,
+ xsBFrameA = auxsBFrameA, xsAFrameB = auxsAFrameB, xsBFrameB = auxsBFrameB,
+ XA = totFrameA, XB = totFrameB)
```

Algunos ejemplos más

¿Y si...

- ...la información auxiliar viene dada para toda la población, y no por marcos, como hasta ahora?
- Esta sería la situación en el conjunto de datos que hemos dividido en dos, donde la variable *Income* es común a ambos marcos.

```
> xstotal <- c(sLandline$Income, sCell$Income)
> Xtotal <- 49646576
> CalDF (sLandline$Opinion, sCell$Opinion, sLandline$ProbLandline,
+ sCell$ProbCell, sLandline$Domain, sCell$Domain, xsT = xstotal,
+ X = Xtotal)
```

Otros ejemplos de utilización

Aunque este taller ha girado en torno a una encuesta con teléfonos fijos y móviles, las encuestas con dos marcos podrían aplicarse en muchas otras situaciones como, por ejemplo:

- Datos procedentes de una encuesta con teléfonos fijos y una encuesta web.
- Datos procedentes de una encuesta con teléfonos móviles y una encuesta web.
- Datos procedentes de una encuesta general de una población y una encuesta específica con alta tasa de individuos con una característica de interés.
  - **Los individuos de interés son las personas sin techo**: Encuesta de registrados en albergues y comedores sociales y encuesta general de la población.
  - **Los individuos de interés son personas con una determinada enfermedad**: Encuesta de registrados en centros especializados y encuesta general de la población.
- Datos procedentes de una encuesta a comerciantes al por mayor y una encuesta de pymes.

Algunos ejemplos más

¿Usando los datos del fichero que se acaba de dividir, podrías...?

- ...calcular el estimador de calibración dual frame sin ningún tipo de información adicional?
- ...incorporar al estimador anterior la información relativa al tamaño de los marcos? ($N_A = 4982920$ y $N_B = 5707655$)
- ...incorporar al estimador anterior la información relativa al tamaño del dominio de intersección? ($N_a b = 4339659$)
- ...incorporar al estimador anterior la información sobre la variable auxiliar Income? ($X = 12693191467$)
- ...calcular un intervalo de confianza al 95 % para las estimaciones anteriores usando el método jackknife?

References I

- Arcos, A., Molina, D., Ranalli, M. G. and Rueda, M. *Frames2: A package for estimation in dual frame surveys*, The R Journal (2015), **7(1)**, 52 – 72.
- Bankier, M. D., *Estimators based on several stratified samples with applications to multiple frame surveys*, Journal of the American Statistical Association (1986) 1074–1079.
- Hartley, H. O., *Multiple frame surveys*, Proceedings of the American Statistical Association, Social Statistics Section (1962) 203–206.
- Hartley, H. O., *Multiple Frame Methodology and Selected Applications*, Sankhyā C. (1974), **36**, 99–118.
- Kalton G. and Anderson, D. W., *Sampling rare populations*, Journal of the Royal Statistical Society. Series A (1986) 65–82.

References II

- Quenouille, M. H., *Problems in plane sampling*, The Annals of Mathematical Statistics, **20(3)** (1949) 355–375.
- Quenouille, M. H., *Notes on bias in estimation*, Biometrika, **43(3/4)** (1956) 353–360.
- Ranalli, M. G., Arcos, A., Rueda, M. and Teodoro, A., *Calibration estimation in dual frame surveys*, arXiv preprint arXiv:1312.0761v2