

Package ‘TeachingSampling’

January 2, 2012

Type Package

Title Sampling designs and parameter estimation in finite population

License GPL (≥ 2)

Version 2.0.1

Date 2011-04-01

Author Hugo Andres Gutierrez Rojas <hugogutierrez@usantotomas.edu.co>

Maintainer Hugo Andres Gutierrez Rojas <hugogutierrez@usantotomas.edu.co>

Depends R ($\geq 2.6.0$)

Description Foundations of inference in survey sampling

URL <http://www.gutierrezandres.com/software/the-teachingsampling-package>

Encoding latin1

Repository CRAN

Date/Publication 2011-04-03 07:42:38

R topics documented:

Deltakl	2
Domains	4
E.2SI	5
E.BE	8
E.Beta	10
E.piPS	13
E.PO	14
E.PPS	15
E.Quantile	17
E.SI	19
E.STPPS	20

E.STSI	22
E.SY	24
E.WR	25
GREG.SI	26
HH	30
HT	33
Ik	39
IkRS	41
IkWR	42
IPFP	43
Lucy	44
Marco	46
nk	47
OrderWR	48
p.WR	49
Pik	51
PikHol	52
Pikl	54
PikPPS	55
S.BE	57
S.piPS	58
S.PO	60
S.PPS	61
S.SI	63
S.STPPS	64
S.STSI	66
S.SY	68
S.WR	69
Support	71
SupportRS	72
SupportWR	73
T.SIC	75
VarHT	76
Wk	78

Index	82
--------------	-----------

Deltak1	<i>Variance-Covariance Matrix of the Sample Membership Indicators for Fixed Size Without Replacement Sampling Designs</i>
---------	-------------------------------------------------------------------------------------------------------------------------------

Description

Computes the Variance-Covariance matrix of the sample membership indicators in the population given a fixed sample size design

Usage

```
Deltakl(N, n, p)
```

Arguments

N	Population size
n	Sample size
p	A vector containing the selection probabilities of a fixed size without replacement sampling design. The sum of the values of this vector must be one

Details

The k lth unit of the Variance-Covariance matrix of the sample membership indicators is defined as $\Delta_{kl} = \pi_{kl} - \pi_k \pi_l$

Value

The function returns a symmetric matrix of size $N \times N$ containing the variances-covariances among the sample membership indicators for each pair of units in the finite population.

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

Särndal, C-E. and Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*. Springer.
 Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*. Editorial Universidad Santo Tomás.

See Also

[VarHT](#), [Pikl](#), [Pik](#)

Examples

```
# Vector U contains the label of a population of size N=5
U <- c("Yves", "Ken", "Erik", "Sharon", "Leslie")
N <- length(U)
# The sample size is n=2
n <- 2
# p is the probability of selection of every sample.
p <- c(0.13, 0.2, 0.15, 0.1, 0.15, 0.04, 0.02, 0.06, 0.07, 0.08)
# Note that the sum of the elements of this vector is one
sum(p)
# Computation of the Variance-Covariance matrix of the sample membership indicators
Deltakl(N, n, p)
```

 Domains

Domains Indicator Matrix

Description

Creates a matrix of domain indicator variables for every single unit in the selected sample or in the entire population

Usage

```
Domains(y)
```

Arguments

`y` Vector of the domain of interest containing the membership of each unit to a specified category of the domain

Details

Each value of `y` represents the domain which a specified unit belongs

Value

The function returns a $n \times p$ matrix, where n is the number of units in the selected sample and p is the number of categories of the domain of interest. The values of this matrix are zero, if the unit does not belong to a specified category and one, otherwise.

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

Särndal, C-E. and Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*. Springer.
 Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*. Editorial Universidad Santo Tomás.

See Also

[E.SI](#)

Examples

```
#####
## Example 1
#####
# This domain contains only two categories: "yes" and "no"
x <- as.factor(c("yes", "yes", "yes", "no", "no", "no", "no", "yes", "yes"))
Domains(x)
```

```
#####
## Example 2
#####
# Uses the Marco and Lucy data to draw a random sample of units according
# to a SI design
data(Marco)
data(Lucy)

N <- dim(Marco)[1]
n <- 400
sam <- sample(N,n)
# The information about the units in the sample is stored in an object called data
data <- Lucy[sam,]
attach(data)
names(data)
# The variable SPAM is a domain of interest
Doma <- Domains(SPAM)
Doma
# HT estimation of the absolute domain size for every category in the domain
# of interest
E.SI(N,n,Doma)

#####
## Example 3
#####
# Following with Example 2...
# The variables of interest are: Income, Employees and Taxes
# This function allows to estimate the population total of this variables for every
# category in the domain of interest SPAM
estima <- data.frame(Income, Employees, Taxes)
SPAM.no <- estima*Doma[,1]
SPAM.yes <- estima*Doma[,2]
E.SI(N,n,SPAM.no)
E.SI(N,n,SPAM.yes)
```

E.2SI

Estimation of the Population Total under Two Stage Simple Random Sampling Without Replacement

Description

Computes the Horvitz-Thompson estimator of the population total according to a 2SI sampling design

Usage

E.2SI(NI, nI, Ni, ni, y, PSU)

Arguments

NI	Population size of Primary Sampling Units
nI	Sample size of Primary Sampling Units
Ni	Vector of population sizes of Secondary Sampling Units selected in the first draw
ni	Vector of sample sizes of Secondary Sampling Units
y	Vector, matrix or data frame containig the recollected information of the variables of interest for every unit in the selected sample
PSU	Vector identifying the membership to the strata of each unit in the population

Details

Returns the estimation of the population total of every single variable of interest, its estimated variance and its estimated coefficient of variation

Value

The function returns a data matrix whose columns correspond to the estimated parameters of the variables of interest

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

Särndal, C-E. and Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*. Springer.
 Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*. Editorial Universidad Santo Tomás.

See Also

[S.SI](#)

Examples

```
#####
## Example 1
#####
# Uses the Marco and Lucy data to draw a twostage simple random sample
# accordind to a 2SI design. Zone is the clustering variable
data(Lucy)
data(Marco)
attach(Marco)
summary(Zone)
# The population of clusters or Primary Sampling Units
UI<-c("A", "B", "C", "D", "E")
NI <- length(UI)
# The sample size is nI=3
```

```

nI <- 3
# Selects the sample of PSUs
samI<-S.SI(NI,nI)
dataI<-UI[samI]
dataI
# The sampling frame of Secondary Sampling Unit is saved in Lucy1 ... Lucy3
Lucy1<-Lucy[which(Zone==dataI[1]),]
Lucy2<-Lucy[which(Zone==dataI[2]),]
Lucy3<-Lucy[which(Zone==dataI[3]),]
# The size of every single PSU
N1<-dim(Lucy1)[1]
N2<-dim(Lucy2)[1]
N3<-dim(Lucy3)[1]
Ni<-c(N1,N2,N3)
# The sample size in every PSI is 135 Secondary Sampling Units
n1<-135
n2<-135
n3<-135
ni<-c(n1,n2,n3)
# Selects a sample of Secondary Sampling Units inside the PSUs
sam1<-S.SI(N1,n1)
sam2<-S.SI(N2,n2)
sam3<-S.SI(N3,n3)
# The information about each Secondary Sampling Unit in the PSUs
# is saved in data1 ... data3
data1<-Lucy1[sam1,]
data2<-Lucy2[sam2,]
data3<-Lucy3[sam3,]
# The information about each unit in the final selected sample is saved in data
data<-rbind(data1, data2, data3)
attach(data)
# The clustering variable is Zone
Cluster <- as.factor(as.integer(Zone))
# The variables of interest are: Income, Employees and Taxes
# This information is stored in a data frame called estima
estima <- data.frame(Income, Employees, Taxes)
# Estimation of the Population total
E.2SI(NI,nI,Ni,ni,estima,Cluster)

#####
## Example 2 Total Census to the entire population
#####
# Uses the Marco and Lucy data to draw a cluster random sample
# accordind to a SI design ...
# Zone is the clustering variable
data(Marco)
attach(Marco)
summary(Zone)
# The population of clusters
UI<-c("A","B","C","D","E")
NI <- length(UI)
# The sample size equals to the population size of PSU
nI <- NI

```

```

# Selects every single PSU
samI<-S.SI(NI,nI)
dataI<-UI[samI]
dataI
# The sampling frame of Secondary Sampling Unit is saved in Lucy1 ... Lucy5
Lucy1<-Lucy[which(Zone==dataI[1]),]
Lucy2<-Lucy[which(Zone==dataI[2]),]
Lucy3<-Lucy[which(Zone==dataI[3]),]
Lucy4<-Lucy[which(Zone==dataI[4]),]
Lucy5<-Lucy[which(Zone==dataI[5]),]
# The size of every single PSU
N1<-dim(Lucy1)[1]
N2<-dim(Lucy2)[1]
N3<-dim(Lucy3)[1]
N4<-dim(Lucy4)[1]
N5<-dim(Lucy5)[1]
Ni<-c(N1,N2,N3,N4,N5)
# The sample size of Secondary Sampling Units equals to the size of each PSU
n1<-N1
n2<-N2
n3<-N3
n4<-N4
n5<-N5
ni<-c(n1,n2,n3,n4,n5)
# Selects every single Secondary Sampling Unit inside the PSU
sam1<-S.SI(N1,n1)
sam2<-S.SI(N2,n2)
sam3<-S.SI(N3,n3)
sam4<-S.SI(N4,n4)
sam5<-S.SI(N5,n5)
# The information about each unit in the cluster is saved in Lucy1 ... Lucy5
data1<-Lucy1[sam1,]
data2<-Lucy2[sam2,]
data3<-Lucy3[sam3,]
data4<-Lucy4[sam4,]
data5<-Lucy5[sam5,]
# The information about each Secondary Sampling Unit
# in the sample (census) is saved in data
data<-rbind(data1, data2, data3, data4, data5)
attach(data)
# The clustering variable is Zone
Cluster <- as.factor(as.integer(Zone))
# The variables of interest are: Income, Employees and Taxes
# This information is stored in a data frame called estima
estima <- data.frame(Income, Employees, Taxes)
# Estimation of the Population total
E.2SI(NI,nI,Ni,ni,estima,Cluster)
# Sampling error is null

```

E.BE

Estimation of the Population Total under Bernoulli Sampling Without Replacement

Description

Computes the Horvitz-Thompson estimator of the population total according to a BE sampling design

Usage

E.BE(y, prob)

Arguments

y	Vector, matrix or data frame containig the recollected information of the variables of interest for every unit in the selected sample
prob	Inclusion probability for each unit in the population

Details

Returns the estimation of the population total of every single variable of interest, its estimated variance and its estimated coefficient of variation under an SI sampling design

Value

The function returns a data matrix whose columns correspond to the estimated parameters of the variables of interest

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

Särndal, C-E. and Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*. Springer.
Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*. Editorial Universidad Santo Tomás.

See Also

[S.BE](#)

Examples

```
# Uses the Marco and Lucy data to draw a Bernoulli sample
data(Lucy)
data(Marco)

N <- dim(Marco)[1]
# The population size is 2396. If the expected sample size is 400,
# then, the inclusion probability must be 400/2396=0.1669
sam <- S.BE(N,0.1669)
# The information about the units in the sample is stored in an object called data
data <- Lucy[sam,]
attach(data)
names(data)
# The variables of interest are: Income, Employees and Taxes
# This information is stored in a data frame called estima
estima <- data.frame(Income, Employees, Taxes)
E.BE(estima,0.1669)
```

E.Beta

Estimation of the population regression coefficients

Description

Computes the estimation of regression coefficients using the principles of the Horvitz-Thompson estimator

Usage

```
E.Beta(y, x, Pik, ck=1, b0=FALSE)
```

Arguments

y	Vector, matrix or data frame containig the recollected information of the variables of interest for every unit in the selected sample
x	Vector, matrix or data frame containig the recollected auxiliary information for every unit in the selected sample
Pik	A vector containing the inclusion probabilities for each unit in the selected sample
ck	By default equals to one. It is a vector of weights induced by the structure of variance of the supposed model
b0	By default FALSE. The intercept of the regression model

Details

Returns the estimation of the population regression coefficients in a supposed linear model

Value

The function returns a vector whose entries correspond to the estimated parameters of the regression coefficients

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

Särndal, C-E. and Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*. Springer.
 Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*. Editorial Universidad Santo Tomás.

See Also

[GREG.SI](#)

Examples

```
#####
## Example 1: Linear models involving continuous auxiliary information
#####

# Draws a simple random sample without replacement
data(Lucy)
data(Marco)
N <- dim(Marco)[1]
n <- 400
sam <- S.SI(N,n)
# The information about the units in the sample is stored in an object called data
data <- Lucy[sam,]
attach(data)
names(data)
# Vector of inclusion probabilities for the units in the sample
Pik<-rep(n/N,n)

##### common mean model #####

estima<-data.frame(Income, Employees, Taxes)
x <- rep(1,n)
E.Beta(estima,x,Pik,ck=1,b0=FALSE)

##### common ratio model #####

estima<-data.frame(Income)
x <- data.frame(Employees)
E.Beta(estima,x,Pik,ck=x,b0=FALSE)
```

```
##### Simple regression model without intercept #####

estima<-data.frame(Income, Employees)
x <- data.frame(Taxes)
E.Beta(estima,x,Pik,ck=1,b0=FALSE)

##### Multiple regression model without intercept #####

estima<-data.frame(Income)
x <- data.frame(Employees, Taxes)
E.Beta(estima,x,Pik,ck=1,b0=FALSE)

##### Simple regression model with intercept #####

estima<-data.frame(Income, Employees)
x <- data.frame(Taxes)
E.Beta(estima,x,Pik,ck=1,b0=TRUE)

##### Multiple regression model with intercept #####

estima<-data.frame(Income)
x <- data.frame(Employees, Taxes)
E.Beta(estima,x,Pik,ck=1,b0=TRUE)

#####
## Example 2: Linear models involving discrete auxiliary information
#####

# Draws a simple random sample without replacement
data(Lucy)
data(Marco)
N <- dim(Marco)[1]
n <- 400
sam <- S.SI(N,n)
# The information about the sample units is stored in an object called data
data <- Lucy[sam,]
attach(data)
names(data)
# The auxiliary information
Doma<-Domains(Level)
# Vector of inclusion probabilities for the units in the sample
Pik<-rep(n/N,n)

##### Poststratified common mean model #####

estima<-data.frame(Income, Employees, Taxes)
E.Beta(estima,Doma,Pik,ck=1,b0=FALSE)

##### Poststratified common ratio model #####

estima<-data.frame(Income, Employees)
x<-Doma*Taxes
E.Beta(estima,x,Pik,ck=1,b0=FALSE)
```

E.piPS *Estimation of the Population Total under Probability Proportional to Size Sampling Without Replacement*

Description

Computes the Horvitz-Thompson estimator of the population total according to a π PS sampling design

Usage

E.piPS(y, Pik)

Arguments

y Vector, matrix or data frame containig the recollected information of the variables of interest for every unit in the selected sample

Pik Vector of inclusion probabilities for each unit in the selected sample

Details

Returns the estimation of the population total of every single variable of interest, its estimated variance and its estimated coefficient of variation under a π PPS sampling design. This function uses the results of approximate expressions for the estimated variance of the Horvitz-Thompson estimator

Value

The function returns a data matrix whose columns correspond to the estimated parameters of the variables of interest

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

Matei, A. and Tillé, Y. (2005), Evaluation of Variance Approximations and Estimators in Maximun Entropy Sampling with Unequal Probability and Fixed Sample Design. *Journal of Official Statistics*. Vol 21, 4, 543-570.

Särndal, C-E. and Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*. Springer.

Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*. Editorial Universidad Santo Tomás.

See Also

[S.piPS](#)

Examples

```

# Uses the Marco and Lucy data to draw a sample according to a piPS
# without replacement design
data(Marco)
data(Lucy)
attach(Lucy)
# The inclusion probability of each unit is proportional to the variable Income
# The selected sample of size n=400
n <- 400
res <- S.piPS(n, Income)
sam <- res[,1]
# The information about the units in the sample is stored in an object called data
data <- Lucy[sam,]
attach(data)
names(data)
# Pik.s is the inclusion probability of every single unit in the selected sample
Pik.s <- res[,2]
# The variables of interest are: Income, Employees and Taxes
# This information is stored in a data frame called estima
estima <- data.frame(Income, Employees, Taxes)
E.piPS(estima,Pik.s)
# Same results than HT function
HT(estima, Pik.s)

```

E.PO

Estimation of the Population Total under Poisson Sampling Without Replacement

Description

Computes the Horvitz-Thompson estimator of the population total according to a PO sampling design

Usage

```
E.PO(y, Pik)
```

Arguments

y	Vector, matrix or data frame containig the recollected information of the variables of interest for every unit in the selected sample
Pik	Vector of inclusion probabilities for each unit in the selected sample

Details

Returns the estimation of the population total of every single variable of interest, its estimated variance and its estimated coefficient of variation under a PO sampling design

Value

The function returns a data matrix whose columns correspond to the estimated parameters of the variables of interest

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

Särndal, C-E. and Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*. Springer.
 Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*. Editorial Universidad Santo Tomás.

See Also

[S.P0](#)

Examples

```
# Uses the Marco and Lucy data to draw a Poisson sample
data(Marco)
data(Lucy)
attach(Lucy)
N <- dim(Lucy)[1]
# The population size is 2396. The expected sample size is 400
# The inclusion probability is proportional to the variable Income
n<-400
Pik<-n*Income/sum(Income)
# The selected sample
sam <- S.P0(N,Pik)
# The information about the units in the sample is stored in an object called data
data <- Lucy[sam,]
attach(data)
names(data)
# The inclusion probabilities of each unit in the selected sample
inclusion <- Pik[sam]
# The variables of interest are: Income, Employees and Taxes
# This information is stored in a data frame called estima
estima <- data.frame(Income, Employees, Taxes)
E.P0(estima,inclusion)
```

E.PPS

Estimation of the Population Total under Probability Proportional to Size Sampling With Replacement

Description

Computes the Hansen-Hurwitz estimator of the population total according to a probability proportional to size sampling with replacement design

Usage

E.PPS(y, pk)

Arguments

y Vector, matrix or data frame containing the recollected information of the variables of interest for every unit in the selected sample

pk A vector containing selection probabilities for each unit in the sample

Details

Returns the estimation of the population total of every single variable of interest, its estimated variance and its estimated coefficient of variation estimated under a probability proportional to size sampling with replacement design

Value

The function returns a data matrix whose columns correspond to the estimated parameters of the variables of interest

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

Särndal, C-E. and Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*. Springer.

Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*. Editorial Universidad Santo Tomás.

See Also

[S.PPS](#), [HH](#)

Examples

```
# Uses the Marco and Lucy data to draw a random sample according to a
# PPS with replacement design
data(Marco)
data(Lucy)
attach(Lucy)
# The selection probability of each unit is proportional to the variable Income
res <- S.PPS(400,Income)
# The selected sample
sam <- res[,1]
# The information about the units in the sample is stored in an object called data
data <- Lucy[sam,]
attach(data)
names(data)
# pk.s is the selection probability of each unit in the selected sample
```

```
pk.s <- res[,2]
# The variables of interest are: Income, Employees and Taxes
# This information is stored in a data frame called estima
estima <- data.frame(Income, Employees, Taxes)
E.PPS(estima, pk.s)
```

E.Quantile

Estimation of a Population quantile

Description

Computes the estimation of a population quantile using the principles of the Horvitz-Thompson estimator

Usage

```
E.Quantile(y, Qn, Pik)
```

Arguments

y	Vector, matrix or data frame containig the recollected information of the variables of interest for every unit in the selected sample
Qn	Quantile of interest
Pik	A vetor containing inclusion probabilities for each unit in the sample. If missing, the function will asign the same weights to each unit in the sample

Details

Returns the estimation of the population quantile of every single variable of interest

Value

The function returns a vector whose entries correspond to the estimated quantiles of the variables of interest

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

Särndal, C-E. and Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*. Springer.
Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*. Editorial Universidad Santo Tomás.

See Also

[HT](#)

Examples

```
#####
## Example 1
#####
# Vector U contains the label of a population of size N=5
U <- c("Yves", "Ken", "Erik", "Sharon", "Leslie")
# Vectors y and x give the values of the variables of interest
y<-c(32, 34, 46, 89, 35)
x<-c(52, 60, 75, 100, 50)
z<-cbind(y,x)
# Inclusion probabilities for a design of size n=2
Pik<-c(0.58, 0.34, 0.48, 0.33, 0.27)
# Estimation of the sample median
E.Quantile(y, 0.5)
# Estimation of the sample Q1
E.Quantile(x, 0.25)
# Estimation of the sample Q3
E.Quantile(z, 0.75)
# Estimation of the sample median
E.Quantile(z, 0.5, Pik)

#####
## Example 2
#####
# Uses the Marco and Lucy data to draw a PPS sample with replacement
data(Marco)
data(Lucy)
attach(Lucy)

# The selection probability of each unit is proportional to the variable Income
# The sample size is m=400
res <- S.PPS(400,Income)
# The selected sample
sam <- res[,1]
# The information about the units in the sample is stored in an object called data
data <- Lucy[sam,]
attach(data)
# The vector of selection probabilities of units in the sample
pk.s <- res[,2]
# The vector of inclusion probabilities of units in the sample
Pik.s<-1-(1-pk.s)^400
# The information about the sample units is stored in an object called data
data <- Lucy[sam,]
attach(data)
names(data)
# The variables of interest are: Income, Employees and Taxes
# This information is stored in a data frame called estima
estima <- data.frame(Income, Employees, Taxes)
# Estimation of sample median
E.Quantile(estima,0.5,Pik.s)
```

E. SI *Estimation of the Population Total under Simple Random Sampling Without Replacement*

Description

Computes the Horvitz-Thompson estimator of the population total according to an SI sampling design

Usage

E.SI(N, n, y)

Arguments

N	Population size
n	Sample size
y	Vector, matrix or data frame containig the recollected information of the variables of interest for every unit in the selected sample

Details

Returns the estimation of the population total of every single variable of interest, its estimated variance and its estimated coefficient of variation under an SI sampling design

Value

The function returns a data matrix whose columns correspond to the estimated parameters of the variables of interest

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

Särndal, C-E. and Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*. Springer.
Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*. Editorial Universidad Santo Tomás.

See Also

[S.SI](#)

Examples

```
#####
## Example 1
#####
# Uses the Marco Lucy data to draw a random sample of units according to a SI design
data(Marco)
data(Lucy)

N <- dim(Marco)[1]
n <- 400
sam <- S.SI(N,n)
# The information about the units in the sample is stored in an object called data
data <- Lucy[sam,]
attach(data)
names(data)
# The variables of interest are: Income, Employees and Taxes
# This information is stored in a data frame called estima
estima <- data.frame(Income, Employees, Taxes)
E.SI(N,n,estima)

#####
## Example 2
#####
# Following with Example 1. The variable SPAM is a domain of interest
Doma <- Domains(SPAM)
# This function allows to estimate the parameters of the variables of interest
# for every category in the domain SPAM
estima <- data.frame(Income, Employees, Taxes)
SPAM.no <- estima*Doma[,1]
SPAM.yes <- estima*Doma[,2]
E.SI(N,n,SPAM.no)
E.SI(N,n,SPAM.yes)
```

E.STPPS

Estimation of the Population Total under Stratified Probability Proportional to Size Sampling With Replacement

Description

Computes the Hansen-Hurwitz estimator of the population total according to a probability proportional to size sampling with replacement design

Usage

E.STPPS(y, pk, mh, S)

Arguments

y	Vector, matrix or data frame containig the recollected information of the variables of interest for every unit in the selected sample
pk	A vetor containing selection probabilities for each unit in the sample
mh	Vector of sample size in each stratum
S	Vector identifying the membership to the strata of each unit in selected sample

Details

Returns the estimation of the population total of every single variable of interest, its estimated variance and its estimated coefficient of variation in all of the stratum and finally in the entire population

Value

The function returns an array composed by several matrices representing each variable of interest. The columns of each matrix correspond to the estimated parameters of the variables of interest in each stratum and in the entire population

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

Särndal, C-E. and Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*. Springer.
 Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*. Editorial Universidad Santo Tomás.

See Also

[S.STPPS](#)

Examples

```
# Uses the Marco and Lucy data to draw a stratified random sample
# according to a PPS design in each stratum
data(Marco)
data(Lucy)
attach(Lucy)
# Level is the stratifying variable
summary(Level)
# Defines the sample size at each stratum
m1<-14
m2<-123
m3<-263
mh<-c(m1,m2,m3)
# Draws a stratified sample
res<-S.STPPS(Level, Income, mh)
# The selected sample
```

```

sam<-res[,1]
# The selection probability of each unit in the selected sample
pk <- res[,2]
pk
# The information about the units in the sample is stored in an object called data
data <- Lucy[sam,]
attach(data)
names(data)
# The variables of interest are: Income, Employees and Taxes
# This information is stored in a data frame called estima
estima <- data.frame(Income, Employees, Taxes)
E.STPPS(estima,pk,mh,Level)

```

E.STSI

Estimation of the Population Total under Stratified Simple Random Sampling Without Replacement

Description

Computes the Horvitz-Thompson estimator of the population total according to a STSI sampling design

Usage

```
E.STSI(S, Nh, nh, y)
```

Arguments

S	Vector identifying the membership to the strata of each unit in the population
Nh	Vector of stratum sizes
nh	Vector of sample sizes in each stratum
y	Vector, matrix or data frame containig the recollected information of the variables of interest for every unit in the selected sample

Details

Returns the estimation of the population total of every single variable of interest, its estimated variance and its estimated coefficient of variation in all of the strata and finally in the entire population

Value

The function returns an array composed by several matrices representing each variable of interest. The columns of each matrix correspond to the estimated parameters of the variables of interest in each stratum and in the entire population

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

Särndal, C-E. and Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*. Springer.
 Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*.
 Editorial Universidad Santo Tomás.

See Also

[S.STSI](#)

Examples

```
#####
## Example 1
#####
# Uses the Marco and Lucy data to draw a stratified random sample
# according to a SI design in each stratum
data(Marco)
data(Lucy)
attach(Lucy)
# Level is the stratifying variable
summary(Level)
# Defines the size of each stratum
N1<-summary(Level)[[1]]
N2<-summary(Level)[[2]]
N3<-summary(Level)[[3]]
N1;N2;N3
Nh <- c(N1,N2,N3)
# Defines the sample size at each stratum
n1<-14
n2<-123
n3<-263
nh<-c(n1,n2,n3)
# Draws a stratified sample
sam <- S.STSI(Level, Nh, nh)
# The information about the units in the sample is stored in an object called data
data <- Lucy[sam,]
attach(data)
names(data)
# The variables of interest are: Income, Employees and Taxes
# This information is stored in a data frame called estima
estima <- data.frame(Income, Employees, Taxes)
E.STSI(Level,Nh,nh,estima)

#####
## Example 2
#####
# Following with Example 1. The variable SPAM is a domain of interest
Doma <- Domains(SPAM)
# This function allows to estimate the parameters of the variables of interest
# for every category in the domain SPAM
SPAM.no <- estima*Doma[,1]
SPAM.yes <- estima*Doma[,2]
```


Examples

```
# Uses the marco and Lucy data to draw a Systematic sample
data(Marco)
data(Lucy)

N <- dim(Marco)[1]
# The population is divided in 6 groups of size 399 or 400
# The selected sample
sam <- S.SY(N,6)
# The information about the units in the sample is stored in an object called data
data <- Lucy[sam,]
attach(data)
names(data)
# The variables of interest are: Income, Employees and Taxes
# This information is stored in a data frame called estima
estima <- data.frame(Income, Employees, Taxes)
E.SY(N,6,estima)
```

E.WR	<i>Estimation of the Population Total under Simple Random Sampling With Replacement</i>
------	-----------------------------------------------------------------------------------------

Description

Computes the Hansen-Hurwitz estimator of the population total according to a simple random sampling with replacement design

Usage

```
E.WR(N, m, y)
```

Arguments

N	Population size
m	Sample size
y	Vector, matrix or data frame containig the recollected information of the variables of interest for every unit in the selected sample

Details

Returns the estimation of the population total of every single variable of interest, its estimated variance and its estimated coefficient of variation estimated under an simple random with replacement design

Value

The function returns a data matrix whose columns correspond to the estimated parameters of the variables of interest

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

Särndal, C-E. and Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*. Springer.
Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*. Editorial Universidad Santo Tomás.

See Also

[S.WR](#)

Examples

```
# Uses the Marco and Lucy data to draw a random sample according to a WR design
data(Marco)
data(Lucy)

N <- dim(Marco)[1]
m <- 400
sam <- S.WR(N,m)
# The information about the units in the sample is stored in an object called data
data <- Lucy[sam,]
attach(data)
names(data)
# The variables of interest are: Income, Employees and Taxes
# This information is stored in a data frame called estima
estima <- data.frame(Income, Employees, Taxes)
E.WR(N,m,estima)
```

GREG.SI

The Generalized Regression Estimator under SI sampling design

Description

Computes the generalized regression estimator of the population total for several variables of interest under simple random sampling without replacement

Usage

```
GREG.SI(N, n, y, x, tx, b, b0=FALSE)
```

Arguments

N	The population size
n	The sample size
y	Vector, matrix or data frame containig the recollected information of the variables of interest for every unit in the selected sample
x	Vector, matrix or data frame containig the recollected auxiliary information for every unit in the selected sample
tx	Vector containing the populations totals of the auxiliary information
b	Vector of estimated regression coefficients
b0	By default FALSE. The intercept of the regression model

Value

The function returns a vector of total population estimates for each variable of interest.

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

Särndal, C-E. and Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*. Springer.
 Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*. Editorial Universidad Santo Tomás.

See Also

[E.Beta](#)

Examples

```
#####
## Example 1: Linear models involving continuous auxiliary information
#####

# Draws a simple random sample without replacement
data(Marco)
data(Lucy)

N <- dim(Marco)[1]
n <- 400
sam <- S.SI(N,n)
# The information about the units in the sample is stored in an object called data
data <- Lucy[sam,]
attach(data)
names(data)
# Vector of inclusion probabilities for units in the selected sample
Pik<-rep(n/N,n)
```

```
##### common mean model #####

estima<-data.frame(Income, Employees, Taxes)
x <- rep(1,n)
tx <- c(N)
b <- E.Beta(estima,x,Pik,ck=1,b0=FALSE)
GREG.SI(N,n,estima,x,tx, b, b0=FALSE)

##### common ratio model #####

estima<-data.frame(Income)
x <- data.frame(Employees)
tx <- c(151950)
b <- E.Beta(estima,x,Pik,ck=x,b0=FALSE)
GREG.SI(N,n,estima,x,tx, b, b0=FALSE)

##### Simple regression model without intercept #####

estima<-data.frame(Income, Employees)
x <- data.frame(Taxes)
tx <- c(28654)
b <- E.Beta(estima,x,Pik,ck=1,b0=FALSE)
GREG.SI(N,n,estima,x,tx, b, b0=FALSE)

##### Multiple regression model without intercept #####

estima<-data.frame(Income)
x <- data.frame(Employees, Taxes)
tx <- c(151950, 28654)
b <- E.Beta(estima,x,Pik,ck=1,b0=FALSE)
GREG.SI(N,n,estima,x,tx, b, b0=FALSE)

##### Simple regression model with intercept #####

estima<-data.frame(Income, Employees)
x <- data.frame(Taxes)
tx <- c(N,28654)
b <- E.Beta(estima,x,Pik,ck=1,b0=TRUE)
GREG.SI(N,n,estima,x,tx, b, b0=TRUE)

##### Multiple regression model with intercept #####

estima<-data.frame(Income)
x <- data.frame(Employees, Taxes)
tx <- c(N, 151950, 28654)
b <- E.Beta(estima,x,Pik,ck=1,b0=TRUE)
GREG.SI(N,n,estima,x,tx, b, b0=TRUE)

#####
## Example 2: Linear models involving discrete auxiliary information
#####

# Draws a simple random sample without replacement
```

```

data(Marco)
data(Lucy)

N <- dim(Marco)[1]
n <- 400
sam <- S.SI(N,n)
# The information about the units in the sample is stored in an object called data
data <- Lucy[sam,]
attach(data)
names(data)
# Vector of inclusion probabilities for units in the selected sample
Pik<-rep(n/N,n)
# The auxiliary information is discrete type
Doma<-Domains(Level)

##### Poststratified common mean model #####

estima<-data.frame(Income, Employees, Taxes)
tx <- c(83,737,1576)
b <- E.Beta(estima,Doma,Pik,ck=1,b0=FALSE)
GREG.SI(N,n,estima,Doma,tx, b, b0=FALSE)

##### Poststratified common ratio model #####

estima<-data.frame(Income, Employees)
x<-Doma*Taxes
tx <- c(6251,16293,6110)
b <- E.Beta(estima,x,Pik,ck=1,b0=FALSE)
GREG.SI(N,n,estima,x,tx, b, b0=FALSE)

#####
## Example 3: Domains estimation trough the postestratified estimator
#####

# Draws a simple random sample without replacement
data(Marco)
data(Lucy)

N <- dim(Marco)[1]
n <- 400
sam <- S.SI(N,n)
# The information about the units in the sample is stored in an object called data
data <- Lucy[sam,]
attach(data)
names(data)
# Vector of inclusion probabilities for units in the selected sample
Pik<-rep(n/N,n)
# The auxiliary information is discrete type
Doma<-Domains(Level)

##### Poststratified common mean model for the Income total in each poststratum #####

estima<-Doma*Income

```

```

tx <- c(937, 1459)
b <- E.Beta(estima,Doma,Pik,ck=1,b0=FALSE)
GREG.SI(N,n,estima,Doma,tx, b, b0=FALSE)

##### Poststratified common mean model for the Employees total in each poststratum #####

estima<-Doma*Employees
tx <- c(937, 1459)
b <- E.Beta(estima,Doma,Pik,ck=1,b0=FALSE)
GREG.SI(N,n,estima,Doma,tx, b, b0=FALSE)

##### Poststratified common mean model for the Taxes total in each poststratum #####

estima<-Doma*Taxes
tx <- c(937, 1459)
b <- E.Beta(estima,Doma,Pik,ck=1,b0=FALSE)
GREG.SI(N,n,estima,Doma,tx, b, b0=FALSE)

```

HH

The Hansen-Hurwitz Estimator

Description

Computes the Hansen-Hurwitz Estimator estimator of the population total for several variables of interest

Usage

HH(y, pk)

Arguments

y Vector, matrix or data frame containig the recollected information of the variables of interest for every unit in the selected sample

pk A vetor containing selection probabilities for each unit in the selected sample

Details

The Hansen-Hurwitz estimator is given by

$$\sum_{i=1}^m \frac{y_i}{p_i}$$

where y_i is the value of the variables of interest for the i th unit, and p_i is its corresponding selection probability. This estimator is restricted to with replacement sampling designs.

Value

The function returns a data matrix whose columns correspond to the estimated parameters of the variables of interest

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

Särndal, C-E. and Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*. Springer.
 Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*.
 Editorial Universidad Santo Tomás.

See Also

[HT](#)

Examples

```
#####
## Example 1
#####
# Vector U contains the label of a population of size N=5
U <- c("Yves", "Ken", "Erik", "Sharon", "Leslie")
# Vectors y1 and y2 give the values of the variables of interest
y1<-c(32, 34, 46, 89, 35)
y2<-c(1,1,1,0,0)
y3<-cbind(y1,y2)
# The population size is N=5
N <- length(U)
# The sample size is m=2
m <- 2
# pk is the probability of selection of every single unit
pk <- c(0.35, 0.225, 0.175, 0.125, 0.125)
# Selection of a random sample with replacement
sam <- sample(5,2, replace=TRUE, prob=pk)
# The selected sample is
U[sam]
# The values of the variables of interest for the units in the sample
y1[sam]
y2[sam]
y3[sam,]
# The Hansen-Hurwitz estimator
HH(y1[sam],pk[sam])
HH(y2[sam],pk[sam])
HH(y3[sam,],pk[sam])

#####
## Example 2
```

```
#####
# Uses the Marco and Lucy data to draw a simple random sample with replacement
data(Marco)
data(Lucy)

N <- dim(Marco)[1]
m <- 400
sam <- sample(N,m,replace=TRUE)
# The vector of selection probabilities of units in the sample
pk <- rep(1/N,m)
# The information about the units in the sample is stored in an object called data
data <- Lucy[sam,]
attach(data)
names(data)
# The variables of interest are: Income, Employees and Taxes
# This information is stored in a data frame called estima
estima <- data.frame(Income, Employees, Taxes)
HH(estima, pk)

#####
## Example 3 HH is unbiased for with replacement sampling designs
#####

# Vector U contains the label of a population of size N=5
U <- c("Yves", "Ken", "Erik", "Sharon", "Leslie")
# Vector y1 and y2 are the values of the variables of interest
y<-c(32, 34, 46, 89, 35)
# The population size is N=5
N <- length(U)
# The sample size is m=2
m <- 2
# pk is the probability of selection of every single unit
pk <- c(0.35, 0.225, 0.175, 0.125, 0.125)
# p is the probability of selection of every possible sample
p <- p.WR(N,m,pk)
p
sum(p)
# The sample membership matrix for random size without replacement sampling designs
Ind <- nk(N,m)
Ind
# The support with the values of the elements
Qy <- SupportWR(N,m, ID=y)
Qy
# The support with the values of the elements
Qp <- SupportWR(N,m, ID=pk)
Qp
# The HT estimates for every single sample in the support
HH1 <- HH(Qy[1,], Qp[1,])[1,]
HH2 <- HH(Qy[2,], Qp[2,])[1,]
HH3 <- HH(Qy[3,], Qp[3,])[1,]
HH4 <- HH(Qy[4,], Qp[4,])[1,]
HH5 <- HH(Qy[5,], Qp[5,])[1,]
HH6 <- HH(Qy[6,], Qp[6,])[1,]
```

```

HH7 <- HH(Qy[7,], Qp[7,])[1,]
HH8 <- HH(Qy[8,], Qp[8,])[1,]
HH9 <- HH(Qy[9,], Qp[9,])[1,]
HH10 <- HH(Qy[10,], Qp[10,])[1,]
HH11 <- HH(Qy[11,], Qp[11,])[1,]
HH12 <- HH(Qy[12,], Qp[12,])[1,]
HH13 <- HH(Qy[13,], Qp[13,])[1,]
HH14 <- HH(Qy[14,], Qp[14,])[1,]
HH15 <- HH(Qy[15,], Qp[15,])[1,]
# The HT estimates arranged in a vector
Est <- c(HH1, HH2, HH3, HH4, HH5, HH6, HH7, HH8, HH9, HH10, HH11, HH12, HH13,
HH14, HH15)
Est
# The HT is actually design-unbiased
data.frame(Ind, Est, p)
sum(Est*p)
sum(y)

```

HT *The Horvitz-Thompson Estimator*

Description

Computes the Horvitz-Thompson estimator of the population total for several variables of interest

Usage

```
HT(y, Pik)
```

Arguments

y Vector, matrix or data frame containing the recollected information of the variables of interest for every unit in the selected sample

Pik A vector containing the inclusion probabilities for each unit in the selected sample

Details

The Horvitz-Thompson estimator is given by

$$\sum_{k \in U} \frac{y_k}{\pi_k}$$

where y_k is the value of the variables of interest for the k th unit, and π_k its corresponding inclusion probability. This estimator could be used for without replacement designs as well as for with replacement designs.

Value

The function returns a vector of total population estimates for each variable of interest.

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

Särndal, C-E. and Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*. Springer.
 Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*.
 Editorial Universidad Santo Tomás.

See Also

[HH](#)

Examples

```
#####
## Example 1
#####
# Uses the Marco and Lucy data to draw a simple random sample without replacement
data(Marco)
data(Lucy)

N <- dim(Marco)[1]
n <- 400
sam <- sample(N,n)
# The vector of inclusion probabilities for each unit in the sample
Pik <- rep(n/N,n)
# The information about the units in the sample is stored in an object called data
data <- Lucy[sam,]
attach(data)
names(data)
# The variables of interest are: Income, Employees and Taxes
# This information is stored in a data frame called estima
estima <- data.frame(Income, Employees, Taxes)
HT(estima, Pik)

#####
## Example 2
#####
# Uses the Marco and Lucy data to draw a simple random sample with replacement
data(Marco)
data(Lucy)

N <- dim(Marco)[1]
m <- 400
sam <- sample(N,m,replace=TRUE)
# The vector of selection probabilities of units in the sample
pk <- rep(1/N,m)
# Computation of the inclusion probabilities
Pik <- 1-(1-pk)^m
# The information about the units in the sample is stored in an object called data
data <- Lucy[sam,]
```

```

attach(data)
names(data)
# The variables of interest are: Income, Employees and Taxes
# This information is stored in a data frame called estima
estima <- data.frame(Income, Employees, Taxes)
HT(estima, Pik)

#####
## Example 3
#####
# Without replacement sampling
# Vector U contains the label of a population of size N=5
U <- c("Yves", "Ken", "Erik", "Sharon", "Leslie")
# Vector y1 and y2 are the values of the variables of interest
y1<-c(32, 34, 46, 89, 35)
y2<-c(1,1,1,0,0)
y3<-cbind(y1,y2)
# The population size is N=5
N <- length(U)
# The sample size is n=2
n <- 2
# The sample membership matrix for fixed size without replacement sampling designs
Ind <- Ik(N,n)
# p is the probability of selection of every possible sample
p <- c(0.13, 0.2, 0.15, 0.1, 0.15, 0.04, 0.02, 0.06, 0.07, 0.08)
# Computation of the inclusion probabilities
inclusion <- Pik(p, Ind)
# Selection of a random sample
sam <- sample(5,2)
# The selected sample
U[sam]
# The inclusion probabilities for these two units
inclusion[sam]
# The values of the variables of interest for the units in the sample
y1[sam]
y2[sam]
y3[sam,]
# The Horvitz-Thompson estimator
HT(y1[sam],inclusion[sam])
HT(y2[sam],inclusion[sam])
HT(y3[sam,],inclusion[sam])

#####
## Example 4
#####
# Following Example 3... With replacement sampling
# The population size is N=5
N <- length(U)
# The sample size is m=2
m <- 2
# pk is the probability of selection of every single unit
pk <- c(0.9, 0.025, 0.025, 0.025, 0.025)
# Computation of the inclusion probabilities

```

```

Pik <- 1-(1-pk)^m
# Selection of a random sample with replacement
sam <- sample(5,2, replace=TRUE, prob=pk)
# The selected sample
U[sam]
# The inclusion probabilities for these two units
inclusion[sam]
# The values of the variables of interest for the units in the sample
y1[sam]
y2[sam]
y3[sam,]
# The Horvitz-Thompson estimator
HT(y1[sam],inclusion[sam])
HT(y2[sam],inclusion[sam])
HT(y3[sam,],inclusion[sam])

#####
## Example 5 HT is unbiased for without replacement sampling designs
##           Fixed sample size
#####

# Vector U contains the label of a population of size N=5
U <- c("Yves", "Ken", "Erik", "Sharon", "Leslie")
# Vector y1 and y2 are the values of the variables of interest
y<-c(32, 34, 46, 89, 35)
# The population size is N=5
N <- length(U)
# The sample size is n=2
n <- 2
# The sample membership matrix for fixed size without replacement sampling designs
Ind <- Ik(N,n)
Ind
# p is the probability of selection of every possible sample
p <- c(0.13, 0.2, 0.15, 0.1, 0.15, 0.04, 0.02, 0.06, 0.07, 0.08)
sum(p)
# Computation of the inclusion probabilities
inclusion <- Pik(p, Ind)
inclusion
sum(inclusion)
# The support with the values of the elements
Qy <-Support(N,n,ID=y)
Qy
# The HT estimates for every single sample in the support
HT1<- HT(y[Ind[1,]==1], inclusion[Ind[1,]==1])
HT2<- HT(y[Ind[2,]==1], inclusion[Ind[2,]==1])
HT3<- HT(y[Ind[3,]==1], inclusion[Ind[3,]==1])
HT4<- HT(y[Ind[4,]==1], inclusion[Ind[4,]==1])
HT5<- HT(y[Ind[5,]==1], inclusion[Ind[5,]==1])
HT6<- HT(y[Ind[6,]==1], inclusion[Ind[6,]==1])
HT7<- HT(y[Ind[7,]==1], inclusion[Ind[7,]==1])
HT8<- HT(y[Ind[8,]==1], inclusion[Ind[8,]==1])
HT9<- HT(y[Ind[9,]==1], inclusion[Ind[9,]==1])
HT10<- HT(y[Ind[10,]==1], inclusion[Ind[10,]==1])

```

```

# The HT estimates arranged in a vector
Est <- c(HT1, HT2, HT3, HT4, HT5, HT6, HT7, HT8, HT9, HT10)
Est
# The HT is actually desgn-unbiased
data.frame(Ind, Est, p)
sum(Est*p)
sum(y)

#####
## Example 6 HT is unbiased for without replacement sampling designs
##           Random sample size
#####

# Vector U contains the label of a population of size N=5
U <- c("Yves", "Ken", "Erik", "Sharon", "Leslie")
# Vector y1 and y2 are the values of the variables of interest
y<-c(32, 34, 46, 89, 35)
# The population size is N=5
N <- length(U)
# The sample membership matrix for random size without replacement sampling designs
Ind <- IkRS(N)
Ind
# p is the probability of selection of every possible sample
p <- c(0.59049, 0.06561, 0.06561, 0.06561, 0.06561, 0.06561, 0.00729, 0.00729,
0.00729, 0.00729, 0.00729, 0.00729, 0.00729, 0.00729, 0.00729, 0.00081,
0.00081, 0.00081, 0.00081, 0.00081, 0.00081, 0.00081, 0.00081, 0.00081, 0.00081,
0.00009, 0.00009, 0.00009, 0.00009, 0.00009, 0.00001)
sum(p)
# Computation of the inclusion probabilities
inclusion <- Pik(p, Ind)
inclusion
sum(inclusion)
# The support with the values of the elements
Qy <-SupportRS(N, ID=y)
Qy
# The HT estimates for every single sample in the support
HT1<- HT(y[Ind[1,]==1], inclusion[Ind[1,]==1])
HT2<- HT(y[Ind[2,]==1], inclusion[Ind[2,]==1])
HT3<- HT(y[Ind[3,]==1], inclusion[Ind[3,]==1])
HT4<- HT(y[Ind[4,]==1], inclusion[Ind[4,]==1])
HT5<- HT(y[Ind[5,]==1], inclusion[Ind[5,]==1])
HT6<- HT(y[Ind[6,]==1], inclusion[Ind[6,]==1])
HT7<- HT(y[Ind[7,]==1], inclusion[Ind[7,]==1])
HT8<- HT(y[Ind[8,]==1], inclusion[Ind[8,]==1])
HT9<- HT(y[Ind[9,]==1], inclusion[Ind[9,]==1])
HT10<- HT(y[Ind[10,]==1], inclusion[Ind[10,]==1])
HT11<- HT(y[Ind[11,]==1], inclusion[Ind[11,]==1])
HT12<- HT(y[Ind[12,]==1], inclusion[Ind[12,]==1])
HT13<- HT(y[Ind[13,]==1], inclusion[Ind[13,]==1])
HT14<- HT(y[Ind[14,]==1], inclusion[Ind[14,]==1])
HT15<- HT(y[Ind[15,]==1], inclusion[Ind[15,]==1])
HT16<- HT(y[Ind[16,]==1], inclusion[Ind[16,]==1])
HT17<- HT(y[Ind[17,]==1], inclusion[Ind[17,]==1])

```

```

HT18<- HT(y[Ind[18,]==1], inclusion[Ind[18,]==1])
HT19<- HT(y[Ind[19,]==1], inclusion[Ind[19,]==1])
HT20<- HT(y[Ind[20,]==1], inclusion[Ind[20,]==1])
HT21<- HT(y[Ind[21,]==1], inclusion[Ind[21,]==1])
HT22<- HT(y[Ind[22,]==1], inclusion[Ind[22,]==1])
HT23<- HT(y[Ind[23,]==1], inclusion[Ind[23,]==1])
HT24<- HT(y[Ind[24,]==1], inclusion[Ind[24,]==1])
HT25<- HT(y[Ind[25,]==1], inclusion[Ind[25,]==1])
HT26<- HT(y[Ind[26,]==1], inclusion[Ind[26,]==1])
HT27<- HT(y[Ind[27,]==1], inclusion[Ind[27,]==1])
HT28<- HT(y[Ind[28,]==1], inclusion[Ind[28,]==1])
HT29<- HT(y[Ind[29,]==1], inclusion[Ind[29,]==1])
HT30<- HT(y[Ind[30,]==1], inclusion[Ind[30,]==1])
HT31<- HT(y[Ind[31,]==1], inclusion[Ind[31,]==1])
HT32<- HT(y[Ind[32,]==1], inclusion[Ind[32,]==1])
# The HT estimates arranged in a vector
Est <- c(HT1, HT2, HT3, HT4, HT5, HT6, HT7, HT8, HT9, HT10, HT11, HT12, HT13,
HT14, HT15, HT16, HT17, HT18, HT19, HT20, HT21, HT22, HT23, HT24, HT25, HT26,
HT27, HT28, HT29, HT30, HT31, HT32)
Est
# The HT is actually design-unbiased
data.frame(Ind, Est, p)
sum(Est*p)
sum(y)

#####
## Example 7 HT is unbiased for with replacement sampling designs
#####

# Vector U contains the label of a population of size N=5
U <- c("Yves", "Ken", "Erik", "Sharon", "Leslie")
# Vector y1 and y2 are the values of the variables of interest
y<-c(32, 34, 46, 89, 35)
# The population size is N=5
N <- length(U)
# The sample size is m=2
m <- 2
# pk is the probability of selection of every single unit
pk <- c(0.35, 0.225, 0.175, 0.125, 0.125)
# p is the probability of selection of every possible sample
p <- p.WR(N,m,pk)
p
sum(p)
# The sample membership matrix for random size without replacement sampling designs
Ind <- IkWR(N,m)
Ind
# The support with the values of the elements
Qy <- SupportWR(N,m, ID=y)
Qy
# Computation of the inclusion probabilities
pik <- 1-(1-pk)^m
pik
# The HT estimates for every single sample in the support

```

```
HT1 <- HT(y[Ind[1,]==1], pik[Ind[1,]==1])
HT2 <- HT(y[Ind[2,]==1], pik[Ind[2,]==1])
HT3 <- HT(y[Ind[3,]==1], pik[Ind[3,]==1])
HT4 <- HT(y[Ind[4,]==1], pik[Ind[4,]==1])
HT5 <- HT(y[Ind[5,]==1], pik[Ind[5,]==1])
HT6 <- HT(y[Ind[6,]==1], pik[Ind[6,]==1])
HT7 <- HT(y[Ind[7,]==1], pik[Ind[7,]==1])
HT8 <- HT(y[Ind[8,]==1], pik[Ind[8,]==1])
HT9 <- HT(y[Ind[9,]==1], pik[Ind[9,]==1])
HT10 <- HT(y[Ind[10,]==1], pik[Ind[10,]==1])
HT11 <- HT(y[Ind[11,]==1], pik[Ind[11,]==1])
HT12 <- HT(y[Ind[12,]==1], pik[Ind[12,]==1])
HT13 <- HT(y[Ind[13,]==1], pik[Ind[13,]==1])
HT14 <- HT(y[Ind[14,]==1], pik[Ind[14,]==1])
HT15 <- HT(y[Ind[15,]==1], pik[Ind[15,]==1])
# The HT estimates arranged in a vector
Est <- c(HT1, HT2, HT3, HT4, HT5, HT6, HT7, HT8, HT9, HT10, HT11, HT12, HT13,
HT14, HT15)
Est
# The HT is actually design-unbiased
data.frame(Ind, Est, p)
sum(Est*p)
sum(y)
```

Description

Creates a matrix of values (0, if the unit belongs to a specified sample and 1, otherwise) for every possible sample under fixed sample size designs without replacement

Usage

```
Ik(N, n)
```

Arguments

N	Population size
n	Sample size

Value

The function returns a matrix of $\binom{N}{n}$ rows and N columns. The k th column corresponds to the sample membership indicator, of the k th unit, to a possible sample.

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

Särndal, C-E. and Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*. Springer.
Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*. Editorial Universidad Santo Tomás.

See Also

[Support](#), [Pik](#)

Examples

```
# Vector U contains the label of a population of size N=5
U <- c("Yves", "Ken", "Erik", "Sharon", "Leslie")
N <- length(U)
n <- 2
# The sample membership matrix for fixed size without replacement sampling designs
Ik(N,n)
# The first unit, Yves, belongs to the first four possible samples
```

Description

Creates a matrix of values (0, if the unit belongs to a specified sample and 1, otherwise) for every possible sample under random sample size designs without replacement

Usage

```
IkRS(N)
```

Arguments

N Population size

Value

The function returns a matrix of 2^N rows and N columns. The k th column corresponds to the sample membership indicator, of the k th unit, to a possible sample.

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

Särndal, C-E. and Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*. Springer.
Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*. Editorial Universidad Santo Tomás.

See Also

[SupportRS](#), [Pik](#)

Examples

```
# Vector U contains the label of a population of size N=5
U <- c("Yves", "Ken", "Erik", "Sharon", "Leslie")
N <- length(U)
n <- 3
# The sample membership matrix for fixed size without replacement sampling designs
IkRS(N)
# The first sample is a null one and the last sample is a census
```

IkWR	<i>Sample Membership Indicator for with Replacements sampling designs</i>
------	---------------------------------------------------------------------------

Description

Creates a matrix of values (1, if the unit belongs to a specified sample and 0, otherwise) for every possible sample under fixed sample size designs without replacement

Usage

```
IkWR(N, m)
```

Arguments

N	Population size
m	Sample size

Value

The function returns a matrix of $\binom{N+m-1}{m}$ rows and N columns. The k th column corresponds to the sample membership indicator, of the k th unit, to a possible sample. It returns a value of 1 even if the element is selected more than once in a with replacement sample.

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

Särndal, C-E. and Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*. Springer.
 Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*. Editorial Universidad Santo Tomás.

See Also

[nk](#), [Support](#), [Pik](#)

Examples

```
# Vector U contains the label of a population of size N=5
U <- c("Yves", "Ken", "Erik", "Sharon", "Leslie")
N <- length(U)
m <- 2
# The sample membership matrix for fixed size without replacement sampling designs
IkWR(N,m)
```

Description

Adjustment of a table on the margins

Usage

```
IPFP(Table, Col.knw, Row.knw, tol=0.0001)
```

Arguments

Table	A contingency table
Col.knw	A vector containing the true totals of the columns
Row.knw	A vector containing the true totals of the Rows
tol	The control value, by default equal to 0.0001

Details

Adjust a contingency table on the know margins of the population with the Raking Ratio method

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

Deming, W. & Stephan, F. (1940), On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics*, 11, 427-444.
 Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*. Editorial Universidad Santo Tomás.

Examples

```
#####
## Example 1
#####
# Some example of Ardilly and Tille
Table <- matrix(c(80,90,10,170,80,80,150,210,130),3,3)
rownames(Table) <- c("a1", "a2", "a3")
colnames(Table) <- c("b1", "b2", "b3")
# The table with labels
Table
# The known and true margins
Col.knw <- c(150,300,550)
Row.knw <- c(430,360,210)
# The adjusted table
```

```

IPFP(Table,Col.knw,Row.knw,tol=0.0001)

#####
## Example 2
#####
# Draws a simple random sample
data(Marco)
data(Lucy)

N<-dim(Lucy)[1]
n<-400
sam<-sample(N,n)
data<-Lucy[sam,]
attach(data)
dim(data)
# Two domains of interest
Doma1<-Domains(Level)
Doma2<-Domains(SPAM)
# Cross tabulate of domains
SPAM.no<-Doma2[,1]*Doma1
SPAM.yes<-Doma2[,2]*Doma1
# Estimation
E.SI(N,n,Doma1)
E.SI(N,n,Doma2)
est1 <-E.SI(N,n,SPAM.no)
est2 <-E.SI(N,n,SPAM.yes)
est1;est2
# The contingency table estimated from above
Table <- cbind(est1[1,],est2[1,])
rownames(Table) <- c("Big", "Medium", "Small")
colnames(Table) <- c("SPAM.no", "SPAM.yes")
# The known and true margins
Col.knw <- c(937,1459)
Row.knw<- c(83,737,1576)
# The adjusted table
IPFP(Table,Col.knw,Row.knw,tol=0.0001)

```

Lucy

Some Business Population Database

Description

This data set corresponds to some financial variables of 2396 industrial companies of a city in a particular fiscal year.

Usage

Lucy

Format

ID The identifier of the company. It correspond to an alphanumeric sequence (two letters and three digits)

Ubication The address of the principal office of the company in the city

Level The industrial companies are discriminated according to the Taxes declared. There are small, medium and big companies

Zone The city is divided by geographical zones. A company is classified in a particular zone according to its address

Income The total amount of a company's earnings (or profit) in the previous fiscal year. It is calculated by taking revenues and adjusting for the cost of doing business

Employees The total number of persons working for the company in the previous fiscal year

Taxes The total amount of a company's income Tax

SPAM Indicates if the company uses the Internet and WEBmail options in order to make self-propaganda.

References

Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*. Editorial Universidad Santo Tomás.

See Also

[Marco](#)

Examples

```
data(Lucy)
attach(Lucy)
# The variables of interest are: Income, Employees and Taxes
# This information is stored in a data frame called estima
estima <- data.frame(Income, Employees, Taxes)
# The population totals
colSums(estima)
# Some parameters of interest
table(SPAM,Level)
xtabs(Income ~ Level+SPAM)
# Correlations among characteristics of interest
cor(estima)
# Some useful histograms
hist(Income)
hist(Taxes)
hist(Employees)
# Some useful plots
boxplot(Income ~ Level)
barplot(table(Level))
pie(table(SPAM))
```

Marco

Sampling frame if the Lucy population

Description

This data set corresponds to the sampling frame of Lucy population. It is considered a device to identifying and ubicating all of the 2396 industrial companies of the city.

Usage

Marco

Format

ID The identifier of the company. It correspond to an alphanumeric sequence (two letters and three digits)

Ubication The address of the principal office of the company in the city

Level The industrial companies are discriminated according to the Taxes declared. There are small, medium and big companies

Zone The city is divided by geographical zones. A company is classified in a particular zone according to its address

References

Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*. Editorial Universidad Santo Tomás.

See Also

[Lucy](#)

Examples

```
data(Marco)
summary(Marco$Zone)
```

nk *Sample Selection Indicator for With Replacement Sampling Designs*

Description

Creates a matrix of values (0, if the unit does not belongs to a specified sample, 1, if the unit is selected once in the sample, 2, if the unit is selected twice in the sample, etc.) for every possible sample under fixed sample size designs with replacement

Usage

```
nk(N, m)
```

Arguments

N	Population size
m	Sample size

Value

The function returns a matrix of $\binom{N+m-1}{m}$ rows and N columns. The k th column corresponds to the sample selection indicator, of the k th unit, to a possible sample.

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

Särndal, C-E. and Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*. Springer.
Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*. Editorial Universidad Santo Tomás.

See Also

[SupportWR](#), [Pik](#)

Examples

```
# Vector U contains the label of a population of size N=5
U <- c("Yves", "Ken", "Erik", "Sharon", "Leslie")
N <- length(U)
m <- 2
# The sample membership matrix for fixed size without replacement sampling designs
nk(N,m)
```

OrderWR

Pseudo-Support for Fixed Size With Replacement Sampling Designs

Description

Creates a matrix containing every possible ordered sample under fixed sample size with replacement designs

Usage

```
OrderWR(N,m, ID=FALSE)
```

Arguments

N	Population size
m	Sample size
ID	By default FALSE, a vector of values (numeric or string) identifying each unit in the population

Details

The number of samples in a with replacement support is not equal to the number of ordered samples induced by a with replacement sampling design.

Value

The function returns a matrix of N^m rows and m columns. Each row of this matrix corresponds to a possible ordered sample.

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>. The author acknowledges to Hanwen Zhang <hanwenzhang@usantotomas.edu.co> for valuable suggestions.

References

Tillé, Y. (2006), *Sampling Algorithms*. Springer
Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*. Editorial Universidad Santo Tomás

See Also

[SupportWR](#), [Support](#)

Examples

```

# Vector U contains the label of a population
U <- c("Yves", "Ken", "Erik", "Sharon", "Leslie")
N <- length(U)
# Under this context, there are five (5) possible ordered samples
OrderWR(N,1)
# The same output, but labeled
OrderWR(N,1,ID=U)
# y is the variable of interest
y<-c(32,34,46,89,35)
OrderWR(N,1,ID=y)

# If the sample size is m=2, there are (25) possible ordered samples
OrderWR(N,2)
# The same output, but labeled
OrderWR(N,2,ID=U)
# y is the variable of interest
y<-c(32,34,46,89,35)
OrderWR(N,2,ID=y)

# Note that the number of ordered samples is not equal to the number of
# samples in a well defined with-replacement support
OrderWR(N,2)
SupportWR(N,2)

OrderWR(N,4)
SupportWR(N,4)

```

p.WR

Generalization of every with replacement sampling design

Description

Computes the selection probability (sampling design) of each with replacement sample

Usage

p.WR(N, m, pk)

Arguments

N	Population size
m	Sample size
pk	A vector containing selection probabilities for each unit in the population

Details

Every with replacement sampling design is a particular case of a multinomial distribution.

$$p(\mathbf{S} = \mathbf{s}) = \frac{m!}{n_1!n_2!\cdots n_N!} \prod_{k=1}^N p_k^{n_k}$$

where n_k is the number of times that the k -th unit is selected in a sample.

Value

The function returns a vector of selection probabilities for every with-replacement sample.

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

Särndal, C-E. and Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*. Springer.
 Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*.
 Editorial Universidad Santo Tomás

Examples

```
#####
## Example 1
#####
# With replacement simple random sampling
# Vector U contains the label of a population of size N=5
U <- c("Yves", "Ken", "Erik", "Sharon", "Leslie")
# Vector pk is the selection probability of the units in the finite population
pk <- c(0.2, 0.2, 0.2, 0.2, 0.2)
sum(pk)
N <- length(pk)
m <- 3
# The sampling design
p <- p.WR(N, m, pk)
p
sum(p)

#####
## Example 2
#####
# With replacement PPS random sampling
# Vector U contains the label of a population of size N=5
U <- c("Yves", "Ken", "Erik", "Sharon", "Leslie")
# Vector x is the auxiliary information and y is the variables of interest
x<-c(32, 34, 46, 89, 35)
y<-c(52, 60, 75, 100, 50)
# Vector pk is the selection probability of the units in the finite population
pk <- x/sum(x)
```

```

sum(pk)
N <- length(pk)
m <- 3
# The smapling design
p <- p.WR(N, m, pk)
p
sum(p)

```

Pik *Inclusion Probabilities for Fixed Size Without Replacement Sampling Designs*

Description

Computes the first-order inclusion probability of each unit in the population given a fixed sample size design

Usage

Pik(p, Ind)

Arguments

p A vector containing the selection probabilities of a fixed size without replacement sampling design. The sum of the values of this vector must be one

Ind A sample membership indicator matrix

Details

The inclusion probability of the k th unit is defined as the probability that this unit will be included in a sample, it is denoted by π_k and obtained from a given sampling design as follows:

$$\pi_k = \sum_{s \ni k} p(s)$$

Value

The function returns a vector of inclusion probabilities for each unit in the finite population.

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

Särndal, C-E. and Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*. Springer.

Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*. Editorial Universidad Santo Tomás.

See Also[HT](#)**Examples**

```

# Vector U contains the label of a population of size N=5
U <- c("Yves", "Ken", "Erik", "Sharon", "Leslie")
N <- length(U)
# The sample size is n=2
n <- 2
# The sample membership matrix for fixed size without replacement sampling designs
Ind <- Ik(N,n)
# p is the probability of selection of every sample.
p <- c(0.13, 0.2, 0.15, 0.1, 0.15, 0.04, 0.02, 0.06, 0.07, 0.08)
# Note that the sum of the elements of this vector is one
sum(p)
# Computation of the inclusion probabilities
inclusion <- Pik(p, Ind)
inclusion
# The sum of inclusion probabilities is equal to the sample size n=2
sum(inclusion)

```

PikHol

*Optimal Inclusion Probabilities Under Multi-purpose Sampling***Description**

Computes the population vector of optimal inclusion probabilities under the Holmberg's Approach

Usage

```
PikHol(n,sigma,e)
```

Arguments

n	Vector of optimal sample sizes for each of the characteristics of interest.
sigma	A matrix containing the size measures for each characteristic of interest.
e	Maximum allowed error under the ANOREL approach.

Details

Assuming that all of the characteristics of interest are equally important, the Holmberg's sampling design yields the following inclusion probabilities

$$\pi_{(opt)k} = \frac{n^* \sqrt{a_{qk}}}{\sum_{k \in U} \sqrt{a_{qk}}}$$

where

$$n^* \geq \frac{(\sum_{k \in U} \sqrt{a_{qk}})^2}{(1+c)Q + \sum_{k \in U} a_{qk}}$$

and

$$a_{qk} = \sum_{q=1}^Q \frac{\sigma_{qk}^2}{\sum_{k \in U} \left(\frac{1}{\pi_{qk}} - 1 \right) \sigma_{qk}^2}$$

Note that σ_{qk}^2 is a size measure associated with the k-th element in the q-th characteristic of interest.

Value

The function returns a vector of inclusion probabilities.

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

- Holmberg, A. (2002), On the Choice of Sampling Design under GREG Estimation in Multiparameter Surveys. *RD Department, Statistics Sweden*.
- Särndal, C-E. and Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*. Springer.
- Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*. Editorial Universidad Santo Tomás

Examples

```
# Uses the Marco and Lucy data to draw an optimal sample
# in a multipurpose survey context
data(Lucy)
attach(Lucy)
# Different sample sizes for two characteristics of interest: Employees and Taxes
N <- dim(Lucy)[1]
n <- c(350,400)
# The size measure is the same for both characteristics of interest,
# but the relationship in between is different
sigy1 <- sqrt(Income^(1))
sigy2 <- sqrt(Income^(2))
# The matrix containign the size measures for each characteristics of interest
sigma<-cbind(sigy1,sigy2)
# The vector of optimal inclusion probabilities under the Holmberg's approach
Piks<-PikHol(n,sigma,0.03)
# The optimal sample size is given by the sum of piks
sum(Piks)
# Performing the S.piPS function in order to select the optimal sample of size n=400
res<-S.piPS(375,Piks)
sam <- res[,1]
# The information about the units in the sample is stored in an object called data
data <- Lucy[sam,]
attach(data)
names(data)
```

```
# Pik.s is the vector of inclusion probability of every single unit
# in the selected sample
Pik.s <- res[,2]
# The variables of interest are: Income, Employees and Taxes
# This information is stored in a data frame called estima
estima <- data.frame(Income, Employees, Taxes)
E.piPS(estima,Pik.s)
```

Pikl *Second Order Inclusion Probabilities for Fixed Size Without Replacement Sampling Designs*

Description

Computes the second-order inclusion probabilities of each pair of units in the population given a fixed sample size design

Usage

```
Pikl(N, n, p)
```

Arguments

N	Population size
n	Sample size
p	A vector containing the selection probabilities of a fixed size without replacement sampling design. The sum of the values of this vector must be one

Details

The second-order inclusion probability of the kl th units is defined as the probability that unit k and unit l will be both included in a sample; it is denoted by π_{kl} and obtained from a given sampling design as follows:

$$\pi_{kl} = \sum_{s \ni k, l} p(s)$$

Value

The function returns a symmetric matrix of size $N \times N$ containing the second-order inclusion probabilities for each pair of units in the finite population.

Author(s)

Helbert Novoa <helbertnova@gmail.com> with contributions from Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co> and Hanwen Zhang <hanwenzhang@usantotomas.edu.co>

References

Särndal, C-E. and Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*. Springer.
 Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*.
 Editorial Universidad Santo Tomás.

See Also

[VarHT](#), [Deltak1](#), [Pik](#)

Examples

```
# Vector U contains the label of a population of size N=5
U <- c("Yves", "Ken", "Erik", "Sharon", "Leslie")
N <- length(U)
# The sample size is n=2
n <- 2
# p is the probability of selection of every sample.
p <- c(0.13, 0.2, 0.15, 0.1, 0.15, 0.04, 0.02, 0.06, 0.07, 0.08)
# Note that the sum of the elements of this vector is one
sum(p)
# Computation of the second-order inclusion probabilities
Pikl(N, n, p)
```

 PikPPS

Inclusion Probabilities in Proportional to Size Sampling Designs

Description

For a given sample size, this function returns a vector of first order inclusion probabilities for a sampling design proportional to an auxiliary variable

Usage

```
PikPPS(n, x)
```

Arguments

n	Integer indicating the sample size
x	Vector of auxiliary information for each unit in the population

Details

For a given vector of auxiliary information with value x_k for the k -th unit and population total t_x , the following expression

$$\pi_k = n \times \frac{x_k}{t_x}$$

is not always less than unity. A sequential algorithm must be used in order to ensure that for every unit in the population the inclusion probability gives less or equal to unity.

Value

The function returns a vector of inclusion probabilities of size N . Every element of this vector is a value between zero and one.

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

Särndal, C-E. and Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*. Springer.
 Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*.
 Editorial Universidad Santo Tomás.

See Also

[PikHol](#), [E.piPS](#), [S.piPS](#)

Examples

```
#####
## Example 1
#####
x <- c(30,41,50,170,43,200)
n <- 3
# Two elements yields values bigger than one
n*x/sum(x)
# With this functions, all of the values are between zero and one
PikPPS(n,x)
# The sum is equal to the sample size
sum(PikPPS(n,x))

#####
## Example 2
#####
# Vector U contains the label of a population of size N=5
U <- c("Yves", "Ken", "Erik", "Sharon", "Leslie")
# The auxiliary information
x <- c(52, 60, 75, 100, 50)
# Gives the inclusion probabilities for the population accordin to a
# proportional to size design without replacement of size n=4
pik <- PikPPS(4,x)
pik
# The selected sample is
sum(pik)

#####
## Example 3
#####
# Uses the Marco and Lucy data to compute teh vector of inclusion probabilities
# accordind to a piPS without replacement design
```

```
data(Marco)
data(Lucy)
attach(Lucy)
# The sample size
n=600
# The selection probability of each unit is proportional to the variable Income
pik <- PikPPS(n,Income)
# The inclusion probabilities of the units in the sample
pik
# The sum of the values in pik is equal to the sample size
sum(pik)
# According to the design some elements must be selected
# They are called forced inclusion units
which(pik==1)
```

S.BE

Bernoulli Sampling Without Replacement

Description

Draws a Bernoulli sample without replacement of expected size n from a population of size N

Usage

```
S.BE(N, prob)
```

Arguments

N	Population size
prob	Inclusion probability for each unit in the population

Details

The selected sample is drawn according to a sequential procedure algorithm based on a uniform distribution. The Bernoulli sampling design is not a fixed sample size one.

Value

The function returns a vector of size N . Each element of this vector indicates if the unit was selected. Then, if the value of this vector for unit k is zero, the unit k was not selected in the sample; otherwise, the unit was selected in the sample.

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

- Särndal, C-E. and Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*. Springer.
- Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*. Editorial Universidad Santo Tomás.
- Tillé, Y. (2006), *Sampling Algorithms*. Springer.

See Also

[E.BE](#)

Examples

```
#####
## Example 1
#####
# Vector U contains the label of a population of size N=5
U <- c("Yves", "Ken", "Erik", "Sharon", "Leslie")
# Draws a Bernoulli sample without replacement of expected size n=3
# The inclusion probability is 0.6 for each unit in the population
sam <- S.BE(5,0.6)
sam
# The selected sample is
U[sam]

#####
## Example 2
#####
# Uses the Marco and Lucy data to draw a Bernoulli sample
data(Marco)
data(Lucy)
attach(Lucy)
N <- dim(Marco)[1]
# The population size is 2396. If the expected sample size is 400,
# then, the inclusion probability must be 400/2396=0.1669
sam <- S.BE(N,0.1669)
# The information about the units in the sample is stored in an object called data
data <- Lucy[sam,]
data
dim(data)
```

Description

Draws a probability proportional to size sample without replacement of size n from a population of size N .

Usage

```
S.piPS(n, x, e)
```

Arguments

x	Vector of auxiliary information for each unit in the population
n	Sample size
e	By default, a vector of size N of independent random numbers drawn from the <i>Uniform</i> (0, 1)

Details

The selected sample is drawn according to the Sunter method (sequential-list procedure)

Value

The function returns a matrix of m rows and two columns. Each element of the first column indicates the unit that was selected. Each element of the second column indicates the selection probability of this unit

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

Särndal, C-E. and Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*. Springer.
 Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*. Editorial Universidad Santo Tomás.

See Also

[E.piPS](#)

Examples

```
#####
## Example 1
#####
# Vector U contains the label of a population of size N=5
U <- c("Yves", "Ken", "Erik", "Sharon", "Leslie")
# The auxiliary information
x <- c(52, 60, 75, 100, 50)
# Draws a piPS sample without replacement of size n=3
res <- S.piPS(3,x)
res
sam <- res[,1]
sam
# The selected sample is
U[sam]
```

```
#####
## Example 2
#####
# Uses the Marco and Lucy data to draw a random sample of units accordind to a
# piPS without replacement design
data(Marco)
data(Lucy)
attach(Lucy)
# The selection probability of each unit is proportional to the variable Income
res <- S.piPS(400,Income)
# The selected sample
sam <- res[,1]
# The inclusion probabilities of the units in the sample
Pik.s <- res[,2]
# The information about the units in the sample is stored in an object called data
data <- Lucy[sam,]
data
dim(data)
```

S.PO

Poisson Sampling

Description

Draws a Poisson sample of expected size n from a population of size N

Usage

S.PO(N , Pik)

Arguments

N	Population size
Pik	Vector of inclusion probabilities for each unit in the population

Details

The selected sample is drawn according to a sequential procedure algorithm based on a uniform distribution. The Poisson sampling design is not a fixed sample size one.

Value

The function returns a vector of size N . Each element of this vector indicates if the unit was selected. Then, if the value of this vector for unit k is zero, the unit k was not selected in the sample; otherwise, the unit was selected in the sample.

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

- Särndal, C-E. and Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*. Springer.
- Gutiérrez, H.A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*. Editorial Universidad Santo Tomás.
- Till'è, Y. (2006), *Sampling Algorithms*. Springer.

See Also

[E.P0](#)

Examples

```
#####
## Example 1
#####
# Vector U contains the label of a population of size N=5
U <- c("Yves", "Ken", "Erik", "Sharon", "Leslie")
# Draws a Bernoulli sample without replacement of expected size n=3
# "Erik" is drawn in every possible sample because its inclusion probability is one
Pik <- c(0.5, 0.2, 1, 0.9, 0.5)
sam <- S.P0(5,Pik)
sam
# The selected sample is
U[sam]

#####
## Example 2
#####
# Uses the Marco and Lucy data to draw a Poisson sample
data(Marco)
data(Lucy)
attach(Lucy)
N <- dim(Lucy)[1]
# The population size is 2396. The expected sample size is 400,
# The inclusion probability is proportional to the variable Income
n<-400
Pik<-n*Income/sum(Income)
# None element of Pik bigger than one
which(Pik>1)
# The selected sample
sam <- S.P0(N,Pik)
# The information about the units in the sample is stored in an object called data
data <- Lucy[sam,]
data
dim(data)
```

Description

Draws a probability proportional to size sample with replacement of size m from a population of size N

Usage

```
S.PPS(m, x)
```

Arguments

<code>m</code>	Sample size
<code>x</code>	Vector of auxiliary information for each unit in the population

Details

The selected sample is drawn according to the cumulative total method (sequential-list procedure)

Value

The function returns a matrix of m rows and two columns. Each element of the first column indicates the unit that was selected. Each element of the second column indicates the selection probability of this unit

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

Särndal, C-E. and Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*. Springer.
Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*. Editorial Universidad Santo Tomás.

See Also

[E.PPS](#)

Examples

```
#####
## Example 1
#####
# Vector U contains the label of a population of size N=5
U <- c("Yves", "Ken", "Erik", "Sharon", "Leslie")
# The auxiliary information
x <- c(52, 60, 75, 100, 50)
# Draws a PPS sample with replacement of size m=3
res <- S.PPS(3,x)
sam <- res[,1]
# The selected sample is
```

```

U[sam]

#####
## Example 2
#####
# Uses the Marco and Lucy data to draw a random sample according to a
# PPS with replacement design
data(Lucy)
attach(Lucy)
# The selection probability of each unit is proportional to the variable Income
res<-S.PPS(400,Income)
# The selected sample
sam <- res[,1]
# The information about the units in the sample is stored in an object called data
data <- Lucy[sam,]
data
dim(data)

```

S.SI

Simple Random Sampling Without Replacement

Description

Draws a simple random sample without replacement of size n from a population of size N

Usage

```
S.SI(N, n, e=runif(N))
```

Arguments

N	Population size
n	Sample size
e	By default, a vector of size N of independent random numbers drawn from the <i>Uniform</i> (0, 1)

Details

The selected sample is drawn according to a selection-rejection (list-sequential) algorithm

Value

The function returns a vector of size N . Each element of this vector indicates if the unit was selected. Then, if the value of this vector for unit k is zero, the unit k was not selected in the sample; otherwise, the unit was selected in the sample.

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

- Särndal, C-E. and Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*. Springer.
- Fan, C.T., Muller, M.E., Rezucha, I. (1962), Development of sampling plans by using sequential (item by item) selection techniques and digital computer, *Journal of the American Statistical Association*, 57, 387-402.
- Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*. Editorial Universidad Santo Tomás.

See Also

[E.SI](#)

Examples

```
#####
## Example 1
#####
# Vector U contains the label of a population of size N=5
U <- c("Yves", "Ken", "Erik", "Sharon", "Leslie")
# Fixes the random numbers in order to select a sample
# Ideal for teaching purposes in the blackboard
e <- c(0.4938, 0.7044, 0.4585, 0.6747, 0.0640)
# Draws a simple random sample without replacement of size n=3
sam <- S.SI(5,3,e)
sam
# The selected sample is
U[sam]

#####
## Example 2
#####
# Uses the Marco and Lucy data to draw a random sample according to a SI design
data(Marco)
data(Lucy)

N <- dim(Lucy)[1]
n <- 400
sam<-S.SI(N,n)
# The information about the units in the sample is stored in an object called data
data <- Lucy[sam,]
data
dim(data)
```

Description

Draws a probability proportional to size simple random sample with replacement of size m_h in stratum h of size N_h

Usage

```
S.STPPS(S, x, mh)
```

Arguments

S	Vector identifying the membership to the strata of each unit in the population
x	Vector of auxiliary information for each unit in the population
mh	Vector of sample size in each stratum

Details

The selected sample is drawn according to the cumulative total method (sequential-list procedure) in each stratum

Value

The function returns a matrix of $m = m_1 + \dots + m_h$ rows and two columns. Each element of the first column indicates the unit that was selected. Each element of the second column indicates the selection probability of this unit

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

Särndal, C-E. and Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*. Springer.
 Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*. Editorial Universidad Santo Tomás.

See Also

[E.STPPS](#)

Examples

```
#####
## Example 1
#####
# Vector U contains the label of a population of size N=5
U <- c("Yves", "Ken", "Erik", "Sharon", "Leslie")
# The auxiliary information
x <- c(52, 60, 75, 100, 50)
# Vector Strata contains an indicator variable of stratum membership
Strata <- c("A", "A", "A", "B", "B")
# Then sample size in each stratum
mh <- c(2,2)
# Draws a stratified PPS sample with replacement of size n=4
res <- S.STPPS(Strata, x, mh)
# The selected sample
```

```

sam <- res[,1]
U[sam]
# The selection probability of each unit selected to be in the sample
pk <- res[,2]
pk

#####
## Example 2
#####
# Uses the Marco and Lucy data to draw a stratified random sample
# according to a PPS design in each stratum
data(Marco)
data(Lucy)
attach(Lucy)
# Level is the stratifying variable
summary(Level)
# Defines the sample size at each stratum
m1<-14
m2<-123
m3<-263
mh<-c(m1,m2,m3)
# Draws a stratified sample
res<-S.STPPS(Level, Income, mh)
# The selected sample
sam<-res[,1]
# The information about the units in the sample is stored in an object called data
data <- Lucy[sam,]
data
dim(data)
# The selection probability of each unit selected in the sample
pk <- res[,2]
pk

```

S.STSI

Stratified sampling applying SI design in all strata

Description

Draws a simple random sample without replacement of size n_h in stratum h of size N_h

Usage

S.STSI(S, Nh, nh)

Arguments

S	Vector identifying the membership to the strata of each unit in the population
Nh	Vector of stratum sizes
nh	Vector of sample size in each stratum

Details

The selected sample is drawn according to a selection-rejection (list-sequential) algorithm in each stratum

Value

The function returns a vector of size $n = n_1 + \dots + n_H$. Each element of this vector indicates the unit that was selected.

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

Särndal, C-E. and Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*. Springer.
Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*. Editorial Universidad Santo Tomás.

See Also

[E.STSI](#)

Examples

```
#####
## Example 1
#####
# Vector U contains the label of a population of size N=5
U <- c("Yves", "Ken", "Erik", "Sharon", "Leslie")
# Vector Strata contains an indicator variable of stratum membership
Strata <- c("A", "A", "A", "B", "B")
Strata
# The stratum sizes
Nh <- c(3,2)
# Then sample size in each stratum
nh <- c(2,1)
# Draws a stratified simple random sample without replacement of size n=3
sam <- S.STSI(Strata, Nh, nh)
sam
# The selected sample is
U[sam]

#####
## Example 2
#####
# Uses the Marco and Lucy data to draw a stratified random sample
# accordind to a SI design in each stratum
data(Marco)
data(Lucy)
attach(Marco)
```

```

# Level is the stratifying variable
summary(Level)
# Defines the size of each stratum
N1<-summary(Level)[[1]]
N2<-summary(Level)[[2]]
N3<-summary(Level)[[3]]
N1;N2;N3
Nh <- c(N1,N2,N3)
# Defines the sample size at each stratum
n1<-14
n2<-123
n3<-263
nh<-c(n1,n2,n3)
# Draws a stratified sample
sam <- S.STSI(Level, Nh, nh)
# The information about the units in the sample is stored in an object called data
data <- Lucy[sam,]
data
dim(data)

```

S.SY

Systematic Sampling

Description

Draws a Systematic sample of size n from a population of size N

Usage

S.SY(N, a)

Arguments

N	Population size
a	Number of groups dividing the population

Details

The selected sample is drawn according to a random start.

Value

The function returns a vector of size n . Each element of this vector indicates the unit that was selected.

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>. The author acknowledges to Kristýna Stodolová <Kristyna.Stodolova@seznam.cz> for valuable suggestions.

References

- Madow, L.H. and Madow, W.G. (1944), On the theory of systematic sampling. *Annals of Mathematical Statistics*. 15, 1-24.
- Särndal, C-E. and Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*. Springer.
- Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*. Editorial Universidad Santo Tomás.

See Also

[E.SY](#)

Examples

```
#####
## Example 1
#####
# Vector U contains the label of a population of size N=5
U <- c("Yves", "Ken", "Erik", "Sharon", "Leslie")
# The population of size N=5 is divided in a=2 groups
# Draws a Systematic sample.
sam <- S.SY(5,2)
sam
# The selected sample is
U[sam]
# There are only two possible samples

#####
## Example 2
#####
# Uses the Marco and Lucy data to draw a Systematic sample
data(Marco)

N <- dim(Marco)[1]
# The population is divided in 6 groups of size 399 or 400
# The selected sample
sam <- S.SY(N,6)
# The information about the units in the sample is stored in an object called data
data <- Marco[sam,]
data
dim(data)
```

Description

Draws a simple random sample with replacement of size m from a population of size N

Usage

S.WR(N, m)

Arguments

N	Population size
m	Sample size

Details

The selected sample is drawn according to a sequential procedure algorithm based on a binomial distribution

Value

The function returns a vector of size m . Each element of this vector indicates the unit that was selected.

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

Tillé, Y. (2006), *Sampling Algorithms*. Springer.
 Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*. Editorial Universidad Santo Tomás.

See Also

[E.WR](#)

Examples

```
#####
## Example 1
#####
# Vector U contains the label of a population of size N=5
U <- c("Yves", "Ken", "Erik", "Sharon", "Leslie")
# Draws a simple random sample with replacement of size m=3
sam <- S.WR(5,3)
sam
# The selected sample
U[sam]

#####
## Example 2
#####
# Uses the Marco and Lucy data to draw a random sample of units according to a
# simple random sampling with replacement design
data(Marco)
```

```
data(Lucy)

N <- dim(Marco)[1]
m <- 400
sam<-S.WR(N,m)
# The information about the units in the sample is stored in an object called data
data <- Lucy[sam,]
data
dim(data)
```

Support

Sampling Support for Fixed Size Without Replacement Sampling Designs

Description

Creates a matrix containing every possible sample under fixed sample size designs

Usage

```
Support(N, n, ID=FALSE)
```

Arguments

N	Population size
n	Sample size
ID	By default FALSE, a vector of values (numeric or string) identifying each unit in the population

Details

A support is defined as the set of samples such that for any sample in the support, all the permutations of the coordinates of the sample are also in the support

Value

The function returns a matrix of $\text{binom}Nn$ rows and n columns. Each row of this matrix corresponds to a possible sample

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

Tillé, Y. (2006), *Sampling Algorithms*. Springer
Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*. Editorial Universidad Santo Tomás

See Also[Ik](#)**Examples**

```

# Vector U contains the label of a population
U <- c("Yves", "Ken", "Erik", "Sharon", "Leslie")
N <- length(U)
n <- 2
# The support for fixed size without replacement sampling designs
# Under this context, there are ten (10) possibles samples
Support(N,n)
# The same support, but labeled
Support(N,n,ID=U)
# y is the variable of interest
y<-c(32,34,46,89,35)
# The following output is very useful when checking
# the design-unbiasedness of an estimator
Support(N,n,ID=y)

```

SupportRS

*Sampling Support for Random Size Without Replacement Sampling Designs***Description**

Creates a matrix containing every possible sample under random sample size designs

Usage

```
SupportRS(N, ID=FALSE)
```

Arguments

N	Population size
ID	By default FALSE, a vector of values (numeric or string) identifying each unit in the population

Details

A support is defined as the set of samples such that for any sample in the support, all the permutations of the coordinates of the sample are also in the support

Value

The function returns a matrix of 2^N rows and N columns. Each row of this matrix corresponds to a possible sample

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

Tillé, Y. (2006), *Sampling Algorithms*. Springer
 Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*.
 Editorial Universidad Santo Tomás

See Also

[IkRS](#)

Examples

```
# Vector U contains the label of a population
U <- c("Yves", "Ken", "Erik", "Sharon", "Leslie")
N <- length(U)
# The support for fixed size without replacement sampling designs
# Under this context, there are ten (10) possibles samples
SupportRS(N)
# The same support, but labeled
SupportRS(N, ID=U)
# y is the variable of interest
y<-c(32,34,46,89,35)
# The following output is very useful when checking
# the design-unbiasedness of an estimator
SupportRS(N, ID=y)
```

SupportWR

Sampling Support for Fixed Size With Replacement Sampling Designs

Description

Creates a matrix containing every possible sample under fixed sample size with replacement designs

Usage

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

Arguments

N	Population size
m	Sample size
ID	By default FALSE, a vector of values (numeric or string) identifying each unit in the population

Details

A support is defined as the set of samples such that, for any sample in the support, all the permutations of the coordinates of the sample are also in the support

Value

The function returns a matrix of $\binom{N+m-1}{m}$ rows and m columns. Each row of this matrix corresponds to a possible sample

Author(s)

Jorge Eduardo Ortiz Pinilla <jorgeortiz@usantotomas.edu.co> with contributions from Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

Ortiz, J. E. (2009), *Simulación y métodos estadísticos*. Editorial Universidad Santo Tomás.
 Tillé, Y. (2006), *Sampling Algorithms*. Springer.
 Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*. Editorial Universidad Santo Tomás.

See Also

[Support](#)

Examples

```
# Vector U contains the label of a population
U <- c("Yves", "Ken", "Erik", "Sharon", "Leslie")
N <- length(U)
m <- 2
# The support for fixed size without replacement sampling designs
# Under this context, there are ten (10) possibles samples
SupportWR(N, m)
# The same support, but labeled
SupportWR(N, m, ID=U)
# y is the variable of interest
y<-c(32,34,46,89,35)
# The following output is very useful when checking
# the design-unbiasedness of an estimator
SupportWR(N, m, ID=y)
```

 T.SIC *Computation of Population Totals for Clusters*

Description

Computes the population total of the characteristics of interest in clusters

Usage

T.SIC(y, Cluster)

Arguments

y	Vector, matrix or data frame containig the recollected information of the variables of interest for every unit in the selected sample
Cluster	Vector identifying the membership to the cluster of each unit in the selected sample of clusters

Value

The function returns a matrix of clusters totals. The columns of each matrix correspond to the totals of the variables of interest in each cluster

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

Särndal, C-E. and Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*. Springer.
 Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimacion de parámetros*. Editorial Universidad Santo Tomás.

See Also

[S.SI](#), [E.SI](#)

Examples

```
#####
## Example 1
#####
# Vector U contains the label of a population of size N=5
U <- c("Yves", "Ken", "Erik", "Sharon", "Leslie")
# Vector y1 and y2 are the values of the variables of interest
y1<-c(32, 34, 46, 89, 35)
y2<-c(1,1,1,0,0)
y3<-cbind(y1,y2)
```

```

# Vector Cluster contains a indicator variable of cluster membership
Cluster <- c("C1", "C2", "C1", "C2", "C1")
Cluster
# Draws a stratified simple random sample without replacement of size n=3
T.SIC(y1,Cluster)
T.SIC(y2,Cluster)
T.SIC(y3,Cluster)

#####
## Example 2 Sampling and estimation in Cluster smapling
#####
# Uses the Marco and Lucy data to draw a clusters sample according to a SI design
# Zone is the clustering variable
data(Marco)
data(Lucy)
attach(Marco)
summary(Zone)
# The population of clusters
UI<-c("A","B","C","D","E")
NI=length(UI)
# The sample size
nI=2
# Draws a simple random sample of two clusters
samI<-S.SI(NI,nI)
dataI<-UI[samI]
dataI
# The information about each unit in the cluster is saved in Lucy1 and Lucy2
data(Lucy)
Lucy1<-Lucy[which(Zone==dataI[1]),]
Lucy2<-Lucy[which(Zone==dataI[2]),]
LucyI<-rbind(Lucy1,Lucy2)
attach(LucyI)
# The clustering variable is Zone
Cluster <- as.factor(as.integer(Zone))
# The variables of interest are: Income, Employees and Taxes
# This information is stored in a data frame called estima
estima <- data.frame(Income, Employees, Taxes)
y<-T.SIC(estima,Cluster)
# Estimation of the Population total
E.SI(NI,nI,y)

```

VarHT

Variance of the Horvitz-Thompson Estimator

Description

Computes the theoretical variance of the Horvitz-Thompson estimator given a without replacement fixed sample size design

Usage

```
VarHT(y, N, n, p)
```

Arguments

y	Vector containig the recollected information of the characteristic of interest for every unit in the population
N	Population size
n	Sample size
p	A vector containing the selection probabilities of a fixed size without replacement sampling design. The sum of the values of this vector must be one

Details

The variance of the Horvitz-Thompson estimator, under a given sampling design p , is given by

$$Var_p(\hat{t}_{y,\pi}) = \sum_{k \in U} \sum_{l \in U} \Delta_{kl} \frac{y_k}{\pi_k} \frac{y_l}{\pi_l}$$

Value

The function returns the value of the theoretical variances of the Horviz-Thompson estimator.

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

Särndal, C-E. and Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*. Springer.
 Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*. Editorial Universidad Santo Tomás.

See Also

[HT](#), [Deltakl](#), [Pikl](#), [Pik](#)

Examples

```
# Without replacement sampling
# Vector U contains the label of a population of size N=5
U <- c("Yves", "Ken", "Erik", "Sharon", "Leslie")
# Vector y1 and y2 are the values of the variables of interest
y1<-c(32, 34, 46, 89, 35)
y2<-c(1,1,1,0,0)
# The population size is N=5
N <- length(U)
# The sample size is n=2
n <- 2
# p is the probability of selection of every possible sample
```

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

# Calculates the theoretical variance of the HT estimator
VarHT(y1, N, n, p)
VarHT(y2, N, n, p)
```

Wk

*The Calibration Weights***Description**

Computes the calibration weights for the estimation of the population total of several variables of interest

Usage

```
Wk(x, tx, Pik, ck, b0)
```

Arguments

x	Vector, matrix or data frame containing the recollected auxiliary information for every unit in the selected sample
tx	Vector containing the populations totals of the auxiliary information
Pik	A vector containing inclusion probabilities for each unit in the sample
ck	A vector of weights induced by the structure of variance of the supposed model
b0	By default FALSE. The intercept of the regression model

Details

The calibration weights satisfy the following expression

$$\sum_{k \in S} w_k x_k = \sum_{k \in U} x_k$$

Value

The function returns a vector of calibrated weights.

Author(s)

Hugo Andrés Gutiérrez Rojas <hugogutierrez@usantotomas.edu.co>

References

Särndal, C-E. and Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*. Springer.
 Gutiérrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*.
 Editorial Universidad Santo Tomás

Examples

```
#####
## Example 1
#####
# Without replacement sampling
# Vector U contains the label of a population of size N=5
U <- c("Yves", "Ken", "Erik", "Sharon", "Leslie")
# Vector x is the auxiliary information and y is the variables of interest
x<-c(32, 34, 46, 89, 35)
y<-c(52, 60, 75, 100, 50)
# pik is some vector of inclusion probabilities in the sample
# In this case the sample size is equal to the population size
pik<-rep(1,5)
w1<-Wk(x,tx=236,pik,ck=1,b0=FALSE)
sum(x*w1)
# Draws a sample size without replacement
sam<-sample(5,4)
pik<-rep(4/5,5)
# The auxiliary information an variable of interest in the selected smaple
x.s<-x[sam]
y.s<-y[sam]
# The vector of inclusion probabilities in the selected smaple
pik.s<-pik[sam]
# Calibration weights under some specifics model
w2<-Wk(x.s,tx=236,pik.s,ck=1,b0=FALSE)
sum(x.s*w2)
w3<-Wk(x.s,tx=c(5,236),pik.s,ck=1,b0=TRUE)
sum(x.s*w3)
w4<-Wk(x.s,tx=c(5,236),pik.s,ck=x.s,b0=TRUE)
sum(x.s*w4)
w5<-Wk(x.s,tx=236,pik.s,ck=x.s,b0=FALSE)
sum(x.s*w5)

#####
## Example 2: Linear models involving continuous auxiliary information
#####

# Draws a simple random sample without replacement
data(Marco)
data(Lucy)

N <- dim(Marco)[1]
n <- 400
sam <- S.SI(N,n)
# The information about the units in the sample is stored in an object called data
data <- Lucy[sam,]
attach(data)
names(data)
# Vector of inclusion probabilities for units in the selected sample
Pik<-rep(n/N,n)

##### common ratio model #####
```

```

estima<-data.frame(Income)
x <- Employees
tx <- c(151950)
w <- Wk(x,tx,Pik,ck=1,b0=FALSE)
sum(x*w)
# The calibration estimation
colSums(estima*w)

##### Simple regression model without intercept #####

estima<-data.frame(Income, Employees)
x <- Taxes
tx <- c(28654)
w<-Wk(x,tx,Pik,ck=x,b0=FALSE)
sum(x*w)
# The calibration estimation
colSums(estima*w)

##### Multiple regression model without intercept #####

estima<-data.frame(Income)
x <- cbind(Employees, Taxes)
tx <- c(151950, 28654)
w <- Wk(x,tx,Pik,ck=1,b0=FALSE)
sum(x[,1]*w)
sum(x[,2]*w)
# The calibration estimation
colSums(estima*w)

##### Simple regression model with intercept #####

estima<-data.frame(Income, Employees)
x <- Taxes
tx <- c(N,28654)
w <- Wk(x,tx,Pik,ck=1,b0=TRUE)
sum(1*w)
sum(x*w)
# The calibration estimation
colSums(estima*w)

##### Multiple regression model with intercept #####

estima<-data.frame(Income)
x <- cbind(Employees, Taxes)
tx <- c(N, 151950, 28654)
w <- Wk(x,tx,Pik,ck=1,b0=TRUE)
sum(1*w)
sum(x[,1]*w)
sum(x[,2]*w)
# The calibration estimation
colSums(estima*w)

```

```
#####
## Example 3: Linear models involving discrete auxiliary information
#####

# Draws a simple random sample without replacement
data(Marco)
data(Lucy)

N <- dim(Marco)[1]
n <- 400
sam <- S.SI(N,n)
# The information about the units in the sample is stored in an object called data
data <- Lucy[sam,]
attach(data)
names(data)
# Vector of inclusion probabilities for units in the selected sample
Pik<-rep(n/N,n)
# The auxiliary information is discrete type
Doma<-Domains(Level)

##### Poststratified common mean model #####

estima<-data.frame(Income, Employees, Taxes)
tx <- c(83,737,1576)
w <- Wk(Doma,tx,Pik,ck=1,b0=FALSE)
sum(Doma[,1]*w)
sum(Doma[,2]*w)
sum(Doma[,3]*w)
# The calibration estimation
colSums(estima*w)

##### Poststratified common ratio model #####

estima<-data.frame(Income, Employees)
x<-Doma*Taxes
tx <- c(6251,16293,6110)
w <- Wk(x,tx,Pik,ck=1,b0=FALSE)
sum(x[,1]*w)
sum(x[,2]*w)
sum(x[,3]*w)
# The calibration estimation
colSums(estima*w)
```

Index

*Topic **datasets**

Lucy, [44](#)
Marco, [46](#)

*Topic **survey**

Deltak1, [2](#)
Domains, [4](#)
E.2SI, [5](#)
E.BE, [9](#)
E.Beta, [10](#)
E.piPS, [13](#)
E.PO, [14](#)
E.PPS, [15](#)
E.Quantile, [17](#)
E.SI, [19](#)
E.STPPS, [20](#)
E.STSI, [22](#)
E.SY, [24](#)
E.WR, [25](#)
GREG.SI, [26](#)
HH, [30](#)
HT, [33](#)
Ik, [39](#)
IkRS, [41](#)
IkWR, [42](#)
IPFP, [43](#)
nk, [47](#)
OrderWR, [48](#)
p.WR, [49](#)
Pik, [51](#)
PikHol, [52](#)
Pik1, [54](#)
PikPPS, [55](#)
S.BE, [57](#)
S.piPS, [58](#)
S.PO, [60](#)
S.PPS, [61](#)
S.SI, [63](#)
S.STPPS, [64](#)
S.STSI, [66](#)

S.SY, [68](#)
S.WR, [69](#)
Support, [71](#)
SupportRS, [72](#)
SupportWR, [73](#)
T.SIC, [75](#)
VarHT, [76](#)
Wk, [78](#)

Deltak1, [2](#), [55](#), [77](#)
Domains, [4](#)

E.2SI, [5](#)
E.BE, [8](#), [58](#)
E.Beta, [10](#), [27](#)
E.piPS, [13](#), [56](#), [59](#)
E.PO, [14](#), [61](#)
E.PPS, [15](#), [62](#)
E.Quantile, [17](#)
E.SI, [4](#), [19](#), [64](#), [75](#)
E.STPPS, [20](#), [65](#)
E.STSI, [22](#), [67](#)
E.SY, [24](#), [69](#)
E.WR, [25](#), [70](#)

GREG.SI, [11](#), [26](#)

HH, [16](#), [30](#), [34](#)
HT, [17](#), [31](#), [33](#), [52](#), [77](#)

Ik, [39](#), [72](#)
IkRS, [41](#), [73](#)
IkWR, [42](#)
IPFP, [43](#)

Lucy, [44](#), [46](#)

Marco, [45](#), [46](#)

nk, [42](#), [47](#)

OrderWR, [48](#)

p. WR, 49
Pik, 3, 40–42, 47, 51, 55, 77
PikHol, 52, 56
PikI, 3, 54, 77
PikPPS, 55

S. BE, 9, 57
S. piPS, 13, 56, 58
S. PO, 15, 60
S. PPS, 16, 61
S. SI, 6, 19, 63, 75
S. STPPS, 21, 64
S. STSI, 23, 66
S. SY, 24, 68
S. WR, 26, 69
Support, 40, 42, 48, 71, 74
SupportRS, 41, 72
SupportWR, 47, 48, 73

T. SIC, 75

VarHT, 3, 55, 76

Wk, 78