Package ‘BayesSampling’

May 1, 2021

Type Package

Title Bayes Linear Estimators for Finite Population

Version 1.1.0

Date 2021-04-24

Maintainer Pedro Soares Figueiredo <pedrosfig@hotmail.com>

Description Allows the user to apply the Bayes Linear approach to finite population with the Simple Random Sampling - BLE_SRS() - and
the Stratified Simple Random Sampling design - BLE_SSRS() -
(both without replacement), to the Ratio estimator (using auxiliary
information) - BLE_Ratio() - and to categorical data - BLE_Categorical().
The Bayes linear estimation approach is applied to a general linear regression model for finite population prediction in BLE_Reg()
and it is also possible to achieve the design based estimators using vague prior distributions.
Based on Gonçalves, K.C.M, Moura, F.A.S and Migon, H.S.(2014) <https://www150.statcan.gc.ca/n1/en/catalogue/12-001-X201400111886>.

URL https://www150.statcan.gc.ca/n1/en/catalogue/12-001-X201400111886,
https://github.com/pedrosfig/BayesSampling

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

Depends R (>= 3.5)

Imports MASS, Matrix, stats, matrixcalc

Suggests knitr, rmarkdown, TeachingSampling

VignetteBuilder knitr

Language en-US

NeedsCompilation no

Author Pedro Soares Figueiredo [aut, cre]
(<https://orcid.org/0000-0003-2279-2881>),
Kelly C. M. Gonçalves [aut, ths]
(<https://orcid.org/0000-0002-4524-547X>)

Repository CRAN

Date/Publication 2021-05-01 19:00:02 UTC

R topics documented:

`BigCity` *Full Person-level Population Database*

Description

This data set corresponds to some socioeconomic variables from 150266 people of a city in a particular year.

Usage

```
data(BigCity)
```

Format

A data.frame with 150266 rows and 12 variables:

HHID The identifier of the household. It corresponds to an alphanumeric sequence (four letters and five digits).

PersonID The identifier of the person within the household. NOTE it is not a unique identifier of a person for the whole population. It corresponds to an alphanumeric sequence (five letters and two digits).

Stratum Households are located in geographic strata. There are 119 strata across the city.

PSU Households are clustered in cartographic segments defined as primary sampling units (PSU). There are 1664 PSU and they are nested within strata.

Zone Segments clustered within strata can be located within urban or rural areas along the city.

Sex Sex of the person.

Income Per capita monthly income.

Expenditure Per capita monthly expenditure.

Employment A person's employment status.

Poverty This variable indicates whether the person is poor or not. It depends on income.

Source

https://CRAN.R-project.org/package=TeachingSampling

References

Package 'TeachingSampling'; see BigCity

BLE_Categorical *Bayes Linear Method for Categorical Data*

Description

Creates the Bayes Linear Estimator for Categorical Data

Usage

```
BLE_Categorical(ys, n, N, m = NULL, rho = NULL)
```

Arguments

ys	k-vector of sample proportion for each category.
n	sample size.
N	total size of the population.
m	k-vector with the prior proportion of each strata. If NULL, sample proportion for each strata will be used (non-informative prior).
rho	matrix with the prior correlation coefficients between two different units within categories. It must be a symmetric square matrix of dimension k (or k-1). If NULL, non-informative prior will be used.

Value

A list containing the following components:

- `est.prop` - BLE for the sample proportion of each category
- `Vest.prop` - Variance associated with the above
- `Vs.Matrix` - Vs matrix, as defined by the BLE method (should be a positive-definite matrix)
- `R.Matrix` - R matrix, as defined by the BLE method (should be a positive-definite matrix)

Source

https://www150.statcan.gc.ca/n1/en/catalogue/12-001-X201400111886

References

Gonçalves, K.C.M, Moura, F.A.S and Migon, H.S.(2014). Bayes Linear Estimation for Finite Population with emphasis on categorical data. Survey Methodology, 40, 15-28.

Examples

```
# 2 categories
ys <- c(0.2614, 0.7386)
n <- 153
N <- 15288
m <- c(0.7, 0.3)
rho <- matrix(0.1, 1)

Estimator <- BLE_Categorical(ys,n,N,m,rho)
Estimator

ys <- c(0.2614, 0.7386)
n <- 153
N <- 15288
m <- c(0.7, 0.3)
rho <- matrix(0.5, 1)

Estimator <- BLE_Categorical(ys,n,N,m,rho)
Estimator


# 3 categories
ys <- c(0.2, 0.5, 0.3)
n <- 100
N <- 10000
m <- c(0.4, 0.1, 0.5)
mat <- c(0.4, 0.1, 0.1, 0.1, 0.2, 0.1, 0.1, 0.1, 0.6)
rho <- matrix(mat, 3, 3)
```

BLE_Ratio *Ratio BLE*

Description

Creates the Bayes Linear Estimator for the Ratio "estimator"

Usage

```
BLE_Ratio(ys, xs, x_nots, m = NULL, v = NULL, sigma = NULL, n = NULL)
```

Arguments

ys	vector of sample observations or sample mean (`sigma` and `n` parameters will be required in this case).
xs	vector with values for the auxiliary variable of the elements in the sample or sample mean.
x_nots	vector with values for the auxiliary variable of the elements not in the sample.
m	prior mean for the ratio between Y and X. If `NULL`, `mean(ys)/mean(xs)` will be used (non-informative prior).
v	prior variance of the ratio between Y and X (bigger than `sigma^2`). If `NULL`, it will tend to infinity (non-informative prior).
sigma	prior estimate of variability (standard deviation) of the ratio within the population. If `NULL`, sample variance of the ratio will be used.
n	sample size. Necessary only if `ys` and `xs` represent sample means (will not be used otherwise).

Value

A list containing the following components:

- `est.beta` - BLE of Beta
- `Vest.beta` - Variance associated with the above
- `est.mean` - BLE for each individual not in the sample
- `Vest.mean` - Covariance matrix associated with the above
- `est.tot` - BLE for the total
- `Vest.tot` - Variance associated with the above

Source

https://www150.statcan.gc.ca/n1/en/catalogue/12-001-X201400111886

References

Gonçalves, K.C.M, Moura, F.A.S and Migon, H.S.(2014). Bayes Linear Estimation for Finite Population with emphasis on categorical data. Survey Methodology, 40, 15-28.

Examples

```
ys <- c(10,8,6)
xs <- c(5,4,3.1)
x_nots <- c(1,20,13,15,-5)
m <- 2.5
v <- 10
sigma <- 2

Estimator <- BLE_Ratio(ys, xs, x_nots, m, v, sigma)
Estimator
```

```
# Same example but informing sample means and sample size instead of sample observations
ys <- mean(c(10,8,6))
xs <- mean(c(5,4,3.1))
n <- 3
x_nots <- c(1,20,13,15,-5)
m <- 2.5
v <- 10
sigma <- 2

Estimator <- BLE_Ratio(ys, xs, x_nots, m, v, sigma, n)
Estimator
```

BLE_Reg *General BLE case*

Description

Calculates the Bayes Linear Estimator for Regression models (general case)

Usage

```
BLE_Reg(ys, xs, a, R, Vs, x_nots, V_nots)
```

Arguments

ys	response variable of the sample
xs	explicative variable of the sample
a	vector of means from Beta
R	covariance matrix of Beta
Vs	covariance of sample errors
x_nots	values of X for the individuals not in the sample
V_nots	covariance matrix of the individuals not in the sample

Value

A list containing the following components:

- `est.beta` - BLE of Beta
- `Vest.beta` - Variance associated with the above
- `est.mean` - BLE of each individual not in the sample
- `Vest.mean` - Covariance matrix associated with the above
- `est.tot` - BLE for the total
- `Vest.tot` - Variance associated with the above

Source

https://www150.statcan.gc.ca/n1/en/catalogue/12-001-X201400111886

References

Gonçalves, K.C.M, Moura, F.A.S and Migon, H.S.(2014). Bayes Linear Estimation for Finite Population with emphasis on categorical data. Survey Methodology, 40, 15-28.

Examples

```
xs <- matrix(c(1,1,1,1,2,3,5,0),nrow=4,ncol=2)
ys <- c(12,17,28,2)
x_nots <- matrix(c(1,1,1,0,1,4),nrow=3,ncol=2)
a <- c(1.5,6)
R <- matrix(c(10,2,2,10),nrow=2,ncol=2)
Vs <- diag(c(1,1,1,1))
V_nots <- diag(c(1,1,1))

Estimator <- BLE_Reg(ys, xs, a, R, Vs, x_nots, V_nots)
Estimator
```

BLE_SRS *Simple Random Sample BLE*

Description

Creates the Bayes Linear Estimator for the Simple Random Sampling design (without replacement)

Usage

```
BLE_SRS(ys, N, m = NULL, v = NULL, sigma = NULL, n = NULL)
```

Arguments

ys	vector of sample observations or sample mean (sigma and n parameters will be required in this case).
N	total size of the population.
m	prior mean. If NULL, sample mean will be used (non-informative prior).
v	prior variance of an element from the population (bigger than sigma^2). If NULL, it will tend to infinity (non-informative prior).
sigma	prior estimate of variability (standard deviation) within the population. If NULL, sample variance will be used.
n	sample size. Necessary only if ys represent sample mean (will not be used otherwise).

Value

A list containing the following components:

- `est.beta` - BLE of Beta (BLE for every individual)
- `Vest.beta` - Variance associated with the above
- `est.mean` - BLE for each individual not in the sample
- `Vest.mean` - Covariance matrix associated with the above
- `est.tot` - BLE for the total
- `Vest.tot` - Variance associated with the above

Source

https://www150.statcan.gc.ca/n1/en/catalogue/12-001-X201400111886

References

Gonçalves, K.C.M, Moura, F.A.S and Migon, H.S.(2014). Bayes Linear Estimation for Finite Population with emphasis on categorical data. Survey Methodology, 40, 15-28.

Examples

```
ys <- c(5,6,8)
N <- 5
m <- 6
v <- 5
sigma <- 1

Estimator <- BLE_SRS(ys, N, m, v, sigma)
Estimator


# Same example but informing sample mean and sample size instead of sample observations
ys <- mean(c(5,6,8))
N <- 5
n <- 3
m <- 6
v <- 5
sigma <- 1

Estimator <- BLE_SRS(ys, N, m, v, sigma, n)
Estimator
```

BLE_SSRS *Stratified Simple Random Sample BLE*

Description

Creates the Bayes Linear Estimator for the Stratified Simple Random Sampling design (without replacement)

Usage

```
BLE_SSRS(ys, h, N, m = NULL, v = NULL, sigma = NULL)
```

Arguments

`ys`	vector of sample observations or sample mean for each strata (`sigma` parameter will be required in this case).
`h`	vector with number of observations in each strata.
`N`	vector with the total size of each strata.
`m`	vector with the prior mean of each strata. If `NULL`, sample mean for each strata will be used (non-informative prior).
`v`	vector with the prior variance of an element from each strata (bigger than `sigma^2` for each strata). If `NULL`, it will tend to infinity (non-informative prior).
`sigma`	vector with the prior estimate of variability (standard deviation) within each strata of the population. If `NULL`, sample variance of each strata will be used.

Value

A list containing the following components:

- `est.beta` - BLE of Beta (BLE for the individuals in each strata)
- `Vest.beta` - Variance associated with the above
- `est.mean` - BLE for each individual not in the sample
- `Vest.mean` - Covariance matrix associated with the above
- `est.tot` - BLE for the total
- `Vest.tot` - Variance associated with the above

Source

https://www150.statcan.gc.ca/n1/en/catalogue/12-001-X201400111886

References

Gonçalves, K.C.M, Moura, F.A.S and Migon, H.S.(2014). Bayes Linear Estimation for Finite Population with emphasis on categorical data. Survey Methodology, 40, 15-28.

Examples

```
ys <- c(2,-1,1.5, 6,10, 8,8)
h <- c(3,2,2)
N <- c(5,5,3)
m <- c(0,9,8)
v <- c(3,8,1)
sigma <- c(1,2,0.5)

Estimator <- BLE_SSRS(ys, h, N, m, v, sigma)
Estimator


# Same example but informing sample means instead of sample observations
y1 <- mean(c(2,-1,1.5))
y2 <- mean(c(6,10))
y3 <- mean(c(8,8))
ys <- c(y1, y2, y3)
h <- c(3,2,2)
N <- c(5,5,3)
m <- c(0,9,8)
v <- c(3,8,1)
sigma <- c(1,2,0.5)

Estimator <- BLE_SSRS(ys, h, N, m, v, sigma)
Estimator
```

C *calculates the C factor*

Description

calculates the C factor

Usage

```
C(ys, xs, R, Vs)
```

Arguments

ys	response variable of the sample
xs	explicative variable of the sample
R	covariance matrix of Beta
Vs	covariance of sample errors

`create1` *creates vector of 1's to be used in the estimators*

Description

creates vector of 1's to be used in the estimators

Usage

```
create1(y)
```

Arguments

`y`	sample matrix

Value

vector of 1's with size equal to the number of observations in the sample

---

`E_beta` *calculates the BLE for Beta*

Description

calculates the BLE for Beta

Usage

```
E_beta(ys, xs, a, R, Vs)
```

Arguments

`ys`	response variable of the sample
`xs`	explicative variable of the sample
`a`	vector of means from Beta
`R`	covariance matrix of Beta
`Vs`	covariance of sample errors

E_theta_Reg *calculates the BLE for the individuals not in the sample*

Description

calculates the BLE for the individuals not in the sample

Usage

```
E_theta_Reg(ys, xs, a, R, Vs, x_nots)
```

Arguments

ys	response variable of the sample
xs	explicative variable of the sample
a	vector of means from Beta
R	covariance matrix of Beta
Vs	covariance of sample errors
x_nots	values of X for the individuals not in the sample

T_Reg *calculates BLE for the total T*

Description

calculates BLE for the total T

Usage

```
T_Reg(ys, xs, a, R, Vs, x_nots)
```

Arguments

ys	response variable of the sample
xs	explicative variable of the sample
a	vector of means from Beta
R	covariance matrix of Beta
Vs	covariance of sample errors
x_nots	values of X for the individuals not in the sample

VT_Reg *calculates risk matrix associated with the BLE for for the total T*

Description

calculates risk matrix associated with the BLE for for the total T

Usage

```
VT_Reg(ys, xs, a, R, Vs, x_nots, V_nots)
```

Arguments

ys	response variable of the sample
xs	explicative variable of the sample
a	vector of means from Beta
R	covariance matrix of Beta
Vs	covariance of sample errors
x_nots	values of X for the individuals not in the sample
V_nots	covariance matrix of the individuals not in the sample

V_beta *calculates the risk matrix associated with the BLE for Beta*

Description

calculates the risk matrix associated with the BLE for Beta

Usage

```
V_beta(ys, xs, R, Vs)
```

Arguments

ys	response variable of the sample
xs	explicative variable of the sample
R	covariance matrix of Beta
Vs	covariance of sample errors

V_theta_Reg *calculates the risk matrix associated with the BLE for the individuals not in the sample*

Description

calculates the risk matrix associated with the BLE for the individuals not in the sample

Usage

```
V_theta_Reg(ys, xs, R, Vs, x_nots, V_nots)
```

Arguments

`ys`	response variable of the sample
`xs`	explicative variable of the sample
`R`	covariance matrix of Beta
`Vs`	covariance of sample errors
`x_nots`	values of X for the individuals not in the sample
`V_nots`	covariance matrix of the individuals not in the sample

Index