

Package ‘mase’

January 17, 2024

Type Package

Title Model-Assisted Survey Estimators

Version 0.1.5.2

Date 2024-02-17

Encoding UTF-8

Maintainer Kelly McConville <kmconville@fas.harvard.edu>

Description A set of model-assisted survey estimators and corresponding variance estimators for single stage, unequal probability, without replacement sampling designs. All of the estimators can be written as a generalized regression estimator with the Horvitz-Thompson, ratio, post-stratified, and regression estimators summarized by Sarndal et al. (1992, ISBN:978-0-387-40620-6). Two of the estimators employ a statistical learning model as the assisting model: the elastic net regression estimator, which is an extension of the lasso regression estimator given by McConville et al. (2017) <doi:10.1093/jssam/smw041>, and the regression tree estimator described in McConville and Toth (2017) <arXiv:1712.05708>. The variance estimators which approximate the joint inclusion probabilities can be found in Berger and Tille (2009) <doi:10.1016/S0169-7161(08)00002-3> and the bootstrap variance estimator is presented in Mashreghi et al. (2016) <doi:10.1214/16-SS113>.

License GPL-2

Imports glmnet, survey, dplyr, tidyr, rpms, boot, stats, Rdpack, ellipsis, Rcpp

Suggests roxygen2, testthat (>= 3.0.0), knitr, rmarkdown

Depends R (>= 4.1.0)

Collate 'gregt.R' 'varMase.R' 'GREG.R' 'IdahoPop.R' 'IdahoSamp.R' 'RcppExports.R' 'gregElasticNett.R' 'gregElasticNet.R' 'gregTree.R' 'gregtreet.R' 'htt.R' 'horvitzThompson.R' 'logisticGregElasticNett.R' 'logisticGregt.R' 'mase-package.R' 'modifiedGreg.R' 'modifiedGregt.R' 'modifiedLogisticGregt.R' 'postStratt.R' 'postStrat.R' 'ratio.R' 'ratioEstimator.R' 'ratioEstimator.R'

RoxygenNote 7.2.3

RdMacros Rdpack

Config/testthat/edition 3

LazyData true

LinkingTo Rcpp, RcppEigen

NeedsCompilation yes

Author Kelly McConville [cre, aut, cph],

Josh Yamamoto [aut],

Becky Tang [aut],

George Zhu [aut],

Sida Li [ctb],

Shirley Chueng [ctb],

Daniell Toth [ctb]

Repository CRAN

Date/Publication 2024-01-17 16:50:02 UTC

R topics documented:

greg	2
gregElasticNet	5
gregTree	7
horvitzThompson	9
IdahoPop	11
IdahoSamp	11
modifiedGreg	12
postStrat	14
ratio	16
ratioEstimator	18

Index	21
--------------	-----------

greg	<i>Compute a generalized regression estimator</i>
------	---

Description

Calculates a generalized regression estimator for a finite population mean/proportion or total based on sample data collected from a complex sampling design and auxiliary population data.

Usage

```
greg(
  y,
  xsample,
  xpop,
  pi = NULL,
```

```

model = "linear",
pi2 = NULL,
var_est = FALSE,
var_method = "LinHB",
datatype = "raw",
N = NULL,
modelselect = FALSE,
lambda = "lambda.min",
B = 1000,
fpc = TRUE,
messages = TRUE
)

```

Arguments

y	A numeric vector of the sampled response variable.
xsample	A data frame of the auxiliary data in the sample.
xpop	A data frame of population level auxiliary information. It must contain the same names as xsample. If datatype = "raw", must contain unit level data. If datatype = "totals" or "means", then contains one row of aggregated, population totals or means for the auxiliary data. Default is "raw".
pi	A numeric vector of inclusion probabilities for each sampled unit in y. If NULL, then simple random sampling without replacement is assumed.
model	A string that specifies the regression model to utilize. Options are "linear" or "logistic".
pi2	A square matrix of the joint inclusion probabilities. Needed for the "LinHT" variance estimator.
var_est	A logical indicating whether or not to compute a variance estimator. Default is FALSE.
var_method	The method to use when computing the variance estimator. Options are a Taylor linearized technique: "LinHB" = Hajek-Berger estimator, "LinHH" = Hansen-Hurwitz estimator, "LinHTSRS" = Horvitz-Thompson estimator under simple random sampling without replacement, and "LinHT" = Horvitz-Thompson estimator or a resampling technique: "bootstrapSRS" = bootstrap variance estimator under simple random sampling without replacement. The default is "LinHB".
datatype	A string that specifies the form of population auxiliary data. The possible values are "raw", "totals" or "means" for whether the user is providing population data at the unit level, aggregated to totals, or aggregated to means. Default is "raw".
N	A numeric value of the population size. If NULL, it is estimated with the sum of the inverse of the pis.
modelselect	A logical for whether or not to run lasso regression first and then fit the model using only the predictors with non-zero lasso coefficients. Default is FALSE.
lambda	A string specifying how to tune the lasso hyper-parameter. Only used if modelselect = TRUE and defaults to "lambda.min". The possible values are "lambda.min", which is the lambda value associated with the minimum cross validation error

	or "lambda.lse", which is the lambda value associated with a cross validation error that is one standard error away from the minimum, resulting in a smaller model.
B	The number of bootstrap samples if computing the bootstrap variance estimator. Default is 1000.
fpc	Default to TRUE, logical for whether or not the variance calculation should include a finite population correction when calculating the "LinHTSRS" or the "SRSbootstrap" variance estimator.
messages	A logical indicating whether to output the messages internal to mase. Default is TRUE.

Value

A list of output containing:

- * pop_total: Estimate of population total.
- * pop_mean: Estimate of the population mean (or proportion).
- * weights: Survey weights produced by GREG (linear model only).
- * pop_total_var: Estimated variance of population total estimate.
- * pop_mean_var: Estimated variance of population mean estimate.

References

Cassel C~M, Sarndal C~E, Wretman J~H (1976). "Some results on generalized difference estimation and generalized regression estimation for finite populations." *Biometrika*, **63**, 615–620.

Sarndal C~E, Swensson B, Wretman J (1992). *Model Assisted Survey Sampling*. Springer-Verlag, New York.

Examples

```
library(dplyr)
data(IdahoPop)
data(IdahoSamp)

xsample <- filter(IdahoSamp, COUNTYFIPS == "16055")
xpop <- filter(IdahoPop, COUNTYFIPS == "16055")

greg(y = xsample$BA_TPA_ADJ,
     N = xpop$npixels,
     xsample = xsample[c("tcc", "elev")],
     xpop = xpop[c("tcc", "elev")],
     var_est = TRUE,
     var_method = "LinHB",
     datatype = "means")
```

gregElasticNet

*Compute an elastic net regression estimator***Description**

Calculates a lasso, ridge or elastic net generalized regression estimator for a finite population mean/proportion or total based on sample data collected from a complex sampling design and auxiliary population data.

Usage

```
gregElasticNet(
  y,
  xsample,
  xpop,
  pi = NULL,
  alpha = 1,
  model = "linear",
  pi2 = NULL,
  var_est = FALSE,
  var_method = "LinHB",
  datatype = "raw",
  N = NULL,
  lambda = "lambda.min",
  B = 1000,
  cvfolds = 10,
  weights_method = "ridge",
  eta = 1e-04,
  fpc = TRUE,
  messages = TRUE
)
```

Arguments

y	A numeric vector of the sampled response variable.
xsample	A data frame of the auxiliary data in the sample.
xpop	A data frame of population level auxiliary information. It must contain the same names as xsample. If datatype = "raw", must contain unit level data. If datatype = "totals" or "means", then contains one row of aggregated, population totals or means for the auxiliary data. Default is "raw".
pi	A numeric vector of inclusion probabilities for each sampled unit in y. If NULL, then simple random sampling without replacement is assumed.
alpha	A numeric value between 0 and 1 which signifies the mixing parameter for the lasso and ridge penalties in the elastic net. When alpha = 1, only a lasso penalty is used. When alpha = 0, only a ridge penalty is used. Default is alpha = 1.

model	A string that specifies the regression model to utilize. Options are "linear" or "logistic".
pi2	A square matrix of the joint inclusion probabilities. Needed for the "LinHT" variance estimator.
var_est	A logical indicating whether or not to compute a variance estimator. Default is FALSE.
var_method	The method to use when computing the variance estimator. Options are a Taylor linearized technique: "LinHB" = Hajek-Berger estimator, "LinHH" = Hansen-Hurwitz estimator, "LinHTSRS" = Horvitz-Thompson estimator under simple random sampling without replacement, and "LinHT" = Horvitz-Thompson estimator under simple random sampling without replacement. The default is "LinHB".
datatype	A string that specifies the form of population auxiliary data. The possible values are "raw", "totals" or "means" for whether the user is providing population data at the unit level, aggregated to totals, or aggregated to means. Default is "raw".
N	A numeric value of the population size. If NULL, it is estimated with the sum of the inverse of the pis.
lambda	A string specifying how to tune the lambda hyper-parameter. Only used if modelselect = TRUE and defaults to "lambda.min". The possible values are "lambda.min", which is the lambda value associated with the minimum cross validation error or "lambda.1se", which is the lambda value associated with a cross validation error that is one standard error away from the minimum, resulting in a smaller model.
B	The number of bootstrap samples if computing the bootstrap variance estimator. Default is 1000.
cvfolds	The number of folds for the cross-validation to tune lambda.
weights_method	A string specifying which method to use to calculate survey weights. Currently, "ridge" is the only option. The "ridge" method uses a ridge regression approximation to calculate weights (see McConville et al (2017), section 3.2 for details). Support for "calibration" to come soon, which employs the model calibration method of Wu and Sitter (2001).
eta	A small positive number. Defaults to 0.0001. See McConville et al (2017), section 3.2 for details.
fpc	Default to TRUE, logical for whether or not the variance calculation should include a finite population correction when calculating the "LinHTSRS" or the "SRSbootstrap" variance estimator.
messages	A logical indicating whether to output the messages internal to mase. Default is TRUE.

Value

A list of output containing:

* pop_total: Estimate of population total.

* coefficients: Elastic net coefficient estimates.

- * pop_mean: Estimate of the population mean (or proportion).
- * pop_total_var: Estimated variance of population total estimate.
- * pop_mean_var: Estimated variance of population mean estimate.

References

McConville K~S, Breidt F~J, Lee T~C~M, Moisen G~G (2017). "Model-Assisted Survey Regression Estimation with the Lasso." *Journal of Survey Statistics and Methodology*, **5**, 131-158.

Examples

```
library(dplyr)
data(IdahoPop)
data(IdahoSamp)
xsample <- filter(IdahoSamp, COUNTYFIPS == "16055")
xpop <- filter(IdahoPop, COUNTYFIPS == "16055")
gregElasticNet(y = xsample$BA_TPA_ADJ,
               N = xpop$npixels,
               xsample = xsample[c("tcc", "elev", "ppt", "tmean")],
               xpop = xpop[c("tcc", "elev", "ppt", "tmean")],
               var_est = TRUE,
               var_method = "LinHB",
               datatype = "means",
               alpha = 0.5)
```

gregTree

Compute a regression tree estimator

Description

Calculates a regression tree estimator for a finite population mean/proportion or total based on sample data collected from a complex sampling design and auxiliary population data.

Usage

```
gregTree(
  y,
  xsample,
  xpop,
  pi = NULL,
  pi2 = NULL,
  var_est = FALSE,
  var_method = "LinHB",
  B = 1000,
  pval = 0.05,
  perm_reps = 500,
  bin_size = NULL,
```

```

    fpc = TRUE,
    messages = TRUE
  )

```

Arguments

<code>y</code>	A numeric vector of the sampled response variable.
<code>xsample</code>	A data frame of the auxiliary data in the sample.
<code>xpop</code>	A data frame of population level auxiliary information. It must contain the same names as <code>xsample</code> . If <code>datatype = "raw"</code> , must contain unit level data. If <code>datatype = "totals"</code> or <code>"means"</code> , then contains one row of aggregated, population totals or means for the auxiliary data. Default is <code>"raw"</code> .
<code>pi</code>	A numeric vector of inclusion probabilities for each sampled unit in <code>y</code> . If <code>NULL</code> , then simple random sampling without replacement is assumed.
<code>pi2</code>	A square matrix of the joint inclusion probabilities. Needed for the <code>"LinHT"</code> variance estimator.
<code>var_est</code>	A logical indicating whether or not to compute a variance estimator. Default is <code>FALSE</code> .
<code>var_method</code>	The method to use when computing the variance estimator. Options are a Taylor linearized technique: <code>"LinHB"</code> = Hajek-Berger estimator, <code>"LinHH"</code> = Hansen-Hurwitz estimator, <code>"LinHTSRS"</code> = Horvitz-Thompson estimator under simple random sampling without replacement, and <code>"LinHT"</code> = Horvitz-Thompson estimator or a resampling technique: <code>"bootstrapSRS"</code> = bootstrap variance estimator under simple random sampling without replacement. The default is <code>"LinHB"</code> .
<code>B</code>	The number of bootstrap samples if computing the bootstrap variance estimator. Default is 1000.
<code>pval</code>	Designated p-value level to reject null hypothesis in permutation test used to fit the regression tree. Default value is 0.05.
<code>perm_reps</code>	An integer specifying the number of permutations for each permutation test run to fit the regression tree. Default value is 500.
<code>bin_size</code>	A integer specifying the minimum number of observations in each node.
<code>fpc</code>	Default to <code>TRUE</code> , logical for whether or not the variance calculation should include a finite population correction when calculating the <code>"LinHTSRS"</code> or the <code>"SRSbootstrap"</code> variance estimator.
<code>messages</code>	A logical indicating whether to output the messages internal to <code>mase</code> . Default is <code>TRUE</code> .

Value

A list of output containing:

- * `pop_total`: Estimate of population total.
- * `pop_mean`: Estimate of the population mean (or proportion).
- * `weights`: Survey weights produced by `gregTree`.
- * `pop_total_var`: Estimated variance of population total estimate.
- * `pop_mean_var`: Estimated variance of population mean estimate.

References

McConville K-S, Toth D (2018). “Automated selection of post-strata using a model-assisted regression tree estimator.” *Scandinavian Journal of Statistics*.

Examples

```
library(dplyr)
data(IdahoPop)
data(IdahoSamp)

xsample <- filter(IdahoSamp, COUNTYFIPS == "16055")
xpop <- filter(IdahoSamp, COUNTYFIPS == "16055")

gregTree(y = xsample$BA_TPA_ADJ,
         xsample = xsample[c("tcc", "elev")],
         xpop = xpop[c("tcc", "elev")],
         var_est = TRUE)
```

horvitzThompson

Compute the Horvitz-Thompson Estimator

Description

Calculate the Horvitz-Thompson Estimator for a finite population mean/proportion or total based on sample data collected from a complex sampling design.

Usage

```
horvitzThompson(
  y,
  pi = NULL,
  N = NULL,
  pi2 = NULL,
  var_est = FALSE,
  var_method = "LinHB",
  B = 1000,
  fpc = TRUE,
  messages = TRUE
)
```

Arguments

y	A numeric vector of the sampled response variable.
pi	A numeric vector of inclusion probabilities for each sampled unit in y. If NULL, then simple random sampling without replacement is assumed.
N	A numeric value of the population size. If NULL, it is estimated with the sum of the inverse of the pis.

pi2	A square matrix of the joint inclusion probabilities. Needed for the "LinHT" variance estimator.
var_est	A logical indicating whether or not to compute a variance estimator. Default is FALSE.
var_method	The method to use when computing the variance estimator. Options are a Taylor linearized technique: "LinHB"= Hajek-Berger estimator, "LinHH" = Hansen-Hurwitz estimator, "LinHTSRS" = Horvitz-Thompson estimator under simple random sampling without replacement, and "LinHT" = Horvitz-Thompson estimator or a resampling technique: "bootstrapSRS" = bootstrap variance estimator under simple random sampling without replacement. The default is "LinHB".
B	The number of bootstrap samples if computing the bootstrap variance estimator. Default is 1000.
fpc	Default to TRUE, logical for whether or not the variance calculation should include a finite population correction when calculating the "LinHTSRS" or the "SRSbootstrap" variance estimator.
messages	A logical indicating whether to output the messages internal to mase. Default is TRUE.

Value

List of output containing:

- * pop_total: Estimate of population total.
- * pop_mean: Estimate of population mean.
- * pop_total_var: Estimated variance of population total estimate.
- * pop_mean_var: Estimated variance of population mean estimate.

References

Horvitz DG, Thompson DJ (1952). "A generalization of sampling without replacement from a finite universe." *Journal of the American Statistical Association*, **47**, 663-685.

Examples

```
library(dplyr)
data(IdahoSamp)
data(IdahoPop)
xsample <- filter(IdahoSamp, COUNTYFIPS == "16055")
xpop <- filter(IdahoPop, COUNTYFIPS == "16055")

horvitzThompson(y = xsample$BA_TPA_ADJ,
                N = xpop$npixels,
                var_est = TRUE,
                var_method = "LinHTSRS")
```

IdahoPop

FIA Population Level Auxiliary Data for Idaho

Description

FIA Population Level Auxiliary Data for Idaho

Usage

IdahoPop

Format

A data frame with 44 rows and 6 columns:

COUNTYFIPS Unique identifier of the county to which a plot belongs

tcc Tree Canopy Cover

elev Elevation

ppt Precipitation

tmean Mean Temperature

tnt.1 Proportion of pixels in a county that are classified as Non-Tree

tnt.2 Proportion of pixels in a county that are classified as Tree

npixels Number of pixels in a county

Source

<<https://www.fia.fs.usda.gov/library/database-documentation/index.php#FIADB>>

IdahoSamp

FIA Sample Plot-Level Data for Idaho

Description

FIA Sample Plot-Level Data for Idaho

Usage

IdahoSamp

Format

A data frame with 3,753 rows and 5 columns:

COUNTYFIPS Unique identifier of the county to which a plot belongs

tcc Tree Canopy Cover

elev Elevation

ppt Precipitation

tmean Mean Temperature

tnt Tree/Non-Tree Indicator: 1 being Non-tree, 2 being tree

BA_TPA_ADJ Basal Area in the units of trees-per-acre

Source

<<https://www.fia.fs.usda.gov/library/database-documentation/index.php#FIADB>>

modifiedGreg

Compute a modified generalized regression estimator

Description

Calculates a modified generalized regression estimator for a finite population mean/proportion or total based on sample data collected from a complex sampling design and auxiliary population data.

Usage

```
modifiedGreg(
  y,
  xsample,
  xpop,
  domains,
  pi = NULL,
  pi2 = NULL,
  datatype = "raw",
  model = "linear",
  var_est = F,
  var_method = "LinHB",
  modelselect = FALSE,
  lambda = "lambda.min",
  domain_col_name = NULL,
  estimation_domains = NULL,
  N = NULL,
  B = 1000,
  fpc = TRUE,
  messages = TRUE
)
```

Arguments

y	A vector of the response values from the sample
xsample	A data frame of the auxiliary data in the sample.
xpop	A data frame of population level auxiliary information. It must contain all of the names from xsample. If datatype = "raw", must contain unit level data. If datatype = "totals" or "means", then contains one row of aggregated, population totals or means for the auxiliary data and must include a column labeled N with the population sizes for each domain. Default is "raw".
domains	A vector of the specific domain that each row of xsample belongs to.
pi	First order inclusion probabilities.
pi2	Second order inclusion probabilities.
datatype	A string that specifies the form of population auxiliary data. The possible values are "raw", "totals" or "means" for whether the user is providing population data at the unit level, aggregated to totals, or aggregated to means. Default is "raw".
model	A string that specifies the regression model to utilize. Options are "linear" or "logistic".
var_est	A logical value that specifies whether variance estimation should be performed.
var_method	A string that specifies the variance method to utilize.
modelselect	A logical for whether or not to run lasso regression first and then fit the model using only the predictors with non-zero lasso coefficients. Default is FALSE.
lambda	A string specifying how to tune the lasso hyper-parameter. Only used if modelselect = TRUE and defaults to "lambda.min". The possible values are "lambda.min", which is the lambda value associated with the minimum cross validation error or "lambda.1se", which is the lambda value associated with a cross validation error that is one standard error away from the minimum, resulting in a smaller model.
domain_col_name	A string that specifies the name of the column that contains the domain values in xpop.
estimation_domains	A vector of domain values over which to produce estimates. If NULL, estimation will be performed over all of the domains included in xpop.
N	The total population size.
B	The number of bootstrap iterations to perform when var_method = "bootstrap-SRS"
fpc	Default to TRUE, logical for whether or not the variance calculation should include a finite population correction when calculating the "LinHTSRS" or the "SRSbootstrap" variance estimator.
messages	A logical indicating whether to output the messages internal to mase. Default is TRUE.

References

RAO J, MOLINA I (2015). *Small Area Estimation*. Wiley, New Jersey.

Examples

```
library(dplyr)
data(IdahoPop)
data(IdahoSamp)

modifiedGreg(y = IdahoSamp$BA_TPA_ADJ,
             xsample = IdahoSamp[c("tcc", "elev")],
             xpop = IdahoPop[c("COUNTYFIPS", "tcc", "elev", "npixels")] |> rename(N = npixels),
             domains = IdahoSamp$COUNTYFIPS,
             datatype = "means",
             N = sum(IdahoPop$npixels),
             var_est = TRUE)
```

postStrat

Compute a post-stratified estimator

Description

Calculates a post-stratified estimator for a finite population mean/proportion or total based on sample data collected from a complex sampling design and a single, categorical auxiliary population variable.

Usage

```
postStrat(
  y,
  xsample,
  xpop,
  pi = NULL,
  N = NULL,
  var_est = FALSE,
  var_method = "LinHB",
  pi2 = NULL,
  datatype = "raw",
  B = 1000,
  fpc = TRUE,
  messages = TRUE
)
```

Arguments

y	A numeric vector of the sampled response variable.
xsample	A vector containing the post-stratum for each sampled unit.
xpop	A vector or data frame, depending on datatype. If datatype = "raw", then a vector containing the post-stratum for each population unit. If datatype = "totals" or "means", then a data frame, where the first column lists the possible post-strata and the second column contains the population total or proportion in each post-stratum.

pi	A numeric vector of inclusion probabilities for each sampled unit in y. If NULL, then simple random sampling without replacement is assumed.
N	A numeric value of the population size. If NULL, it is estimated with the sum of the inverse of the pis.
var_est	Default to FALSE, logical for whether or not to compute estimate of variance
var_method	The method to use when computing the variance estimator. Options are a Taylor linearized technique: "LinHB"= Hajek-Berger estimator, "LinHH" = Hansen-Hurwitz estimator, "LinHTSRS" = Horvitz-Thompson estimator under simple random sampling without replacement, and "LinHT" = Horvitz-Thompson estimator or a resampling technique: "bootstrapSRS" = bootstrap variance estimator under simple random sampling without replacement, "SRSunconditional" = simple random sampling variance estimator which accounts for random strata.
pi2	A square matrix of the joint inclusion probabilities. Needed for the "LinHT" variance estimator.
datatype	Default to "raw", takes values "raw", "totals" or "means" for whether the user is providing the raw population stratum memberships, the population totals of each stratum, or the population proportions of each stratum.
B	The number of bootstrap samples if computing the bootstrap variance estimator. Default is 1000.
fpc	Default to TRUE, logical for whether or not the variance calculation should include a finite population correction when calculating the "LinHTSRS", "SRSunconditional", or the "SRSbootstrap" variance estimator.
messages	A logical indicating whether to output the messages internal to mase. Default is TRUE.

Value

A list of output containing:

- * pop_total: Estimate of population total.
- * pop_mean: Estimate of the population mean (or proportion).
- * pop_total_var: Estimated variance of population total estimate.
- * pop_mean_var: Estimated variance of population mean estimate.
- * strat_ests: Table of total and mean estimates for each strata.
- * weights: Survey weights produced by PS.

References

- Cochran W~G (1977). *Sampling Techniques*, 3rd edition. John Wiley & Sons, New York.
- Sarndal C~E, Swensson B, Wretman J (1992). *Model Assisted Survey Sampling*. Springer-Verlag, New York.

Examples

```

library(tidyr)
library(dplyr)

data(IdahoPop)
data(IdahoSamp)

xsample <- filter(IdahoSamp, COUNTYFIPS == "16055")
xpop <- filter(IdahoPop, COUNTYFIPS == "16055")

pop <- xpop[c("tnt.1", "tnt.2")] |>
  pivot_longer(everything(), names_to = "tnt", values_to = "prop") |>
  mutate(tnt = as.numeric(gsub("\\D", "", tnt)))

postStrat(y = xsample$BA_TPA_ADJ,
          N = xpop$npixels,
          xsample = xsample$tnt,
          xpop = pop,
          datatype = "means",
          var_est = TRUE,
          var_method = "SRSunconditional")

```

ratio

*Compute a ratio of two estimators***Description**

Compute a ratio of two estimators

Usage

```

ratio(
  y_num,
  y_den,
  xsample,
  xpop,
  pi = NULL,
  pi2 = NULL,
  N = NULL,
  estimator = NULL,
  var_est = F,
  var_method = "LinHTSRS",
  datatype = "raw",
  fpc = TRUE,
  messages = TRUE,
  ...
)

```


Arguments

y_num	A vector containing the response value for each sampled unit in the numerator
y_den	A vector containing the response value for each sampled unit in the denominator
xsample	A data frame of the auxiliary data in the sample.
xpop	A data frame of population level auxiliary information. It must contain the same names as xsample. If datatype = "raw", must contain unit level data. If datatype = "totals" or "means", then contains one row of aggregated, population totals or means for the auxiliary data. Default is "raw".
pi	A numeric vector of inclusion probabilities for each sampled unit in y. If NULL, then simple random sampling without replacement is assumed.
pi2	A square matrix of the joint inclusion probabilities. Needed for the "LinHT" variance estimator.
N	A numeric value of the population size. If NULL, it is estimated with the sum of the inverse of the pis.
estimator	A string containing the name of the estimators of which you are taking a ratio of. The names follow the same format as the functions independently do in mase. Options are "horvitzThompson", "postStrat", and "greg".
var_est	A logical indicating whether or not to compute a variance estimator. Default is FALSE.
var_method	The method to use when computing the variance estimator. Options are a Taylor linearized technique: "LinHB" = Hajek-Berger estimator, "LinHH" = Hansen-Hurwitz estimator, "LinHTSRS" = Horvitz-Thompson estimator under simple random sampling without replacement, and "LinHT" = Horvitz-Thompson estimator or a resampling technique: "bootstrapSRS" = bootstrap variance estimator under simple random sampling without replacement. The default is "LinHB".
datatype	Default to "raw", takes values "raw", "totals" or "means" for whether the user is providing the raw population stratum memberships, the population totals of each stratum, or the population proportions of each stratum.
fpc	Default to TRUE, logical for whether or not the variance calculation should include a finite population correction when calculating the "LinHTSRS" or the "SRSbootstrap" variance estimator.
messages	A logical indicating whether to output the messages internal to mase. Default is TRUE.
...	Any additional arguments that can be passed to mase::horvitzThompson, mase::greg, and mase::postStrat

Value

A list of output containing:

- * ratio_est: Estimate of the ratio of the population totals/means of the two estimators.
- * ratio_var_est: Estimate of the variance of the ratio of two estimators.

References

Cochran W-G (1977). *Sampling Techniques*, 3rd edition. John Wiley & Sons, New York. Sarndal C-E, Swensson B, Wretman J (1992). *Model Assisted Survey Sampling*. Springer-Verlag, New York.

Examples

```
library(survey)
data(api)
ratio(y_num = apisrs$api.stu, y_den = apisrs$enroll, xsample = apisrs$stype,
xpop = apipop$stype, pi = apisrs$pw^(-1), estimator = "postStrat",
var_est = TRUE, var_method = "LinHB", datatype = "raw")
```

ratioEstimator	<i>Compute a ratio estimator</i>
----------------	----------------------------------

Description

Calculates a ratio estimator for a finite population mean/proportion or total based on sample data collected from a complex sampling design and auxiliary population data.

Usage

```
ratioEstimator(
  y,
  xsample,
  xpop,
  datatype = "raw",
  pi = NULL,
  N = NULL,
  pi2 = NULL,
  var_est = FALSE,
  var_method = "LinHB",
  B = 1000,
  fpc = TRUE,
  messages = TRUE
)
```

Arguments

y	A numeric vector of the sampled response variable.
xsample	A numeric vector of the sampled auxiliary variable.
xpop	A numeric vector of population level auxiliary information. Must come in the form of raw data, population total or population mean.

datatype	A string that specifies the form of population auxiliary data. The possible values are "raw", "total" or "mean". If datatype = "raw", then xpop must contain a numeric vector of the auxiliary variable for each unit in the population. If datatype = "total" or "mean", then contains either the population total or population mean for the auxiliary variable.
pi	A numeric vector of inclusion probabilities for each sampled unit in y. If NULL, then simple random sampling without replacement is assumed.
N	A numeric value of the population size. If NULL, it is estimated with the sum of the inverse of the pis.
pi2	A square matrix of the joint inclusion probabilities. Needed for the "LinHT" variance estimator.
var_est	A logical indicating whether or not to compute a variance estimator. Default is FALSE.
var_method	The method to use when computing the variance estimator. Options are a Taylor linearized technique: "LinHB"= Hajek-Berger estimator, "LinHH" = Hansen-Hurwitz estimator, "LinHTSRS" = Horvitz-Thompson estimator under simple random sampling without replacement, and "LinHT" = Horvitz-Thompson estimator or a resampling technique: "bootstrapSRS" = bootstrap variance estimator under simple random sampling without replacement. The default is "LinHB".
B	The number of bootstrap samples if computing the bootstrap variance estimator. Default is 1000.
fpc	Default to TRUE, logical for whether or not the variance calculation should include a finite population correction when calculating the "LinHTSRS" or the "SRSbootstrap" variance estimator.
messages	A logical indicating whether to output the messages internal to mase. Default is TRUE.

Value

List of output containing:

- * pop_total: Estimate of population total.
- * pop_mean: Estimate of population mean.
- * pop_total_var: Estimated variance of population total estimate.
- * pop_mean_var: Estimated variance of population mean estimate.

References

Cochran W~G (1977). *Sampling Techniques*, 3rd edition. John Wiley & Sons, New York. Sarndal C~E, Swensson B, Wretman J (1992). *Model Assisted Survey Sampling*. Springer-Verlag, New York.

Examples

```
library(dplyr)
data(IdahoPop)
data(IdahoSamp)
```

```
xsample <- filter(IdahoSamp, COUNTYFIPS == "16055")
xpop <- filter(IdahoPop, COUNTYFIPS == "16055")

ratioEstimator(y = xsample$BA_TPA_ADJ,
               xsample = xsample$tcc,
               xpop = xpop$tcc,
               datatype = "means",
               N = xpop$npixels)
```

Index

* datasets

IdahoPop, [11](#)

IdahoSamp, [11](#)

greg, [2](#)

gregElasticNet, [5](#)

gregTree, [7](#)

horvitzThompson, [9](#)

IdahoPop, [11](#)

IdahoSamp, [11](#)

modifiedGreg, [12](#)

postStrat, [14](#)

ratio, [16](#)

ratioEstimator, [18](#)