Package 'sregsurvey'

January 9, 2022

Type Package

Title Semiparametric Model-Assisted Estimation in Finite Populations

Version 0.1.2

Maintainer Carlos Alberto Cardozo Delgado <cardozorpackages@gmail.com>

Description It is a framework to fit semiparametric regression estimators for the total parameter of a finite population
when the interest variable is asymmetric distributed. The main references for this package are: Sarndal C.E., Swensson B., and Wretman J. (2003,ISBN: 978-0-387-40620-6, ``Model Assisted Survey Sampling." Springer-Verlag)
and Cardozo C.A and Alonso-Malaver C.E. (2021). ``Semi-parametric model assisted estimation in finite populations." In preparation.

License GPL-3

Encoding UTF-8

RoxygenNote 7.1.2

Suggests survey

Imports gamlss, gamlss.dist, TeachingSampling, methods, dplyr, caret, magrittr

NeedsCompilation no

Author Carlos Alberto Cardozo Delgado [aut, cre, cph],
Carlos E. Alonso-Malaver [aut]

Repository CRAN

Date/Publication 2022-01-09 16:50:02 UTC

R topics documented:

sreg_ber	*Semiparametric Model-Assisted Estimation under a Bernoulli Sampling Design*

Description

`sreg_ber` is used to estimate the total parameter of a finite population generated from a semiparametric generalized gamma population under a Bernoulli sampling design.

Usage

```
sreg_ber(location_formula, scale_formula, data, pi, ...)
```

Arguments

location_formula
:   a symbolic description of the systematic component of the location model to be fitted.

scale_formula
:   a symbolic description of the systematic component of the scale model to be fitted.

data
:   a data frame, list containing the variables in the model.

pi
:   numeric, represents the first order probability. Default value is 0.5.

...
:   further parameters accepted by caret and survey functions.

Value

`sampling_design` is the name of the sampling design used in the estimation process.

`N` is the population size.

`n` is the random sample size used in the estimation process.

`first_order_probabilities` vector of the first order probabilities used in the estimation process.

`sample` is the random sample used in the estimation process.

`estimated_total_y_sreg` is the SREG estimate of the total parameter of the finite population.

Author(s)

Carlos Alberto Cardozo Delgado <cardozorpackages@gmail.com>

References

Cardozo C.A, Alonso C. (2021) Semi-parametric model assisted estimation in finite populations. In preparation.

Cardozo C.A., Paula G., and Vanegas L. (2021). Generalized log-gamma semiparametric models with P-spline smoothing. Submitted.

Sarndal C.E., Swensson B., and Wretman J. (2003). Model Assisted Survey Sampling. Springer-Verlag.

Examples

```
#This example use the data set 'apipop' of the survey package.
library(sregsurvey)
library(survey)
library(magrittr)
library(dplyr)
library(gamlss)
data(api)
attach(apipop)
Apipop <- filter(apipop,full!= 'NA')
Apipop <- filter(Apipop, stype == 'H')
Apipop <- Apipop %>% dplyr::select(api00,grad.sch,full)
sreg_ber(api00 ~  pb(grad.sch), scale_formula = ~ full - 1, data= Apipop, pi=0.25)
# The total population value is
sum(Apipop$api00)
```

---

sreg_pips	*Semiparametric Model-Assisted Estimation under a Proportional to Size Sampling Design*

Description

sreg_pips is used to estimate the total parameter of a finite population generated from a semi-parametric generalized gamma population under a proportional to size without-replacement sampling design.

Usage

```
sreg_pips(location_formula, scale_formula, data, x, n, ...)
```

Arguments

location_formula
: a symbolic description of the systematic component of the location model to be fitted.

scale_formula
: a symbolic description of the systematic component of the scale model to be fitted.

data
: a data frame, list containing the variables in the model.

x
: vector, an auxiliary variable to calculate the inclusion probabilities of each unit.

n
: numeric, sample size.

...
: further parameters accepted by caret and survey functions.

Value

sampling_design is the name of the sampling design used in the estimation process.

N is the population size.

n is the sample size used in the estimation process.

first_order_probabilities vector of the first order probabilities used in the estimation process.

sample is the random sample used in the estimation process.

estimated_total_y_sreg is the SREG estimate of the total parameter of the finite population.

Author(s)

Carlos Alberto Cardozo Delgado <cardozorpackages@gmail.com>

References

Cardozo C.A, Alonso C. (2021) Semi-parametric model assisted estimation in finite populations. In preparation.

Cardozo C.A., Paula G., and Vanegas L. (2021). Generalized log-gamma semiparametric models with P-spline smoothing. Submitted.

Sarndal C.E., Swensson B., and Wretman J. (2003). Model Assisted Survey Sampling. Springer-Verlag.

Examples

```
library(sregsurvey)
library(survey)
library(dplyr)
library(gamlss)
data(api)
attach(apipop)
Apipop <- filter(apipop,full!= 'NA')
Apipop <- filter(Apipop, stype == 'H')
Apipop <- Apipop %>% dplyr::select(api00,grad.sch,full,api99)
n=ceiling(0.25*dim(Apipop)[1])
aux_var <- Apipop %>% dplyr::select(api99)
sreg_pips(api00 ~ pb(grad.sch), scale_formula = ~ full - 1, data= Apipop, x= aux_var, n=n)
# The total population value is
sum(Apipop$api00)
```

| sreg_poisson | *Semiparametric Model-Assisted Estimation under a Poisson Sampling Design* |

Description

sreg_poisson is used to estimate the total parameter of a finite population generated from a semi-parametric generalized gamma population under a Poisson sampling design.

Usage

```
sreg_poisson(location_formula, scale_formula, data, pis, ...)
```

Arguments

location_formula
> a symbolic description of the systematic component of the location model to be fitted.

scale_formula   a symbolic description of the systematic component of the scale model to be fitted.

data   a data frame, list containing the variables in the model.

pis   numeric vector, first order inclusion probabilities. Default value 0.1 for each element.

...   further parameters accepted by caret and survey functions.

Value

sampling_design is the name of the sampling design used in the estimation process.

N is the population size.

n is the random sample size used in the estimation process.

first_order_probabilities vector of the first order probabilities used in the estimation process.

sample is the random sample used in the estimation process.

estimated_total_y_sreg is the SREG estimate of the total parameter of the finite population.

Author(s)

Carlos Alberto Cardozo Delgado <cardozorpackages@gmail.com>

References

Cardozo C.A, Alonso C. (2021) Semi-parametric model assisted estimation in finite populations. In preparation.

Cardozo C.A., Paula G., and Vanegas L. (2021). Generalized log-gamma semiparametric models with P-spline smoothing. Submitted.

Sarndal C.E., Swensson B., and Wretman J. (2003). Model Assisted Survey Sampling. Springer-Verlag.

Examples

```
library(sregsurvey)
library(survey)
library(dplyr)
library(gamlss)
data(api)
attach(apipop)
Apipop <- filter(apipop,full!= 'NA')
```

```
Apipop <- filter(Apipop, stype == 'H')
Apipop <- Apipop %>% dplyr::select(api00,grad.sch,full)
sreg_poisson(api00 ~  pb(grad.sch), scale_formula = ~ full - 1, data= Apipop)
# The total population value is
sum(Apipop$api00)
```

---

sreg_srswr	*Semiparametric Model-Assisted Estimation under a Simple Random Sampling Without Replace Sampling Design*

Description

`sreg_srswr` is used to estimate the total parameter of a finite population generated from a semi-parametric generalized gamma population under a simple random sampling without-replacement sampling design.

Usage

```
sreg_srswr(
  location_formula,
  scale_formula,
  data,
  fraction,
  format = "COMPLETE",
  ...
)
```

Arguments

location_formula

  a symbolic description of the systematic component of the location model to be fitted.

scale_formula a symbolic description of the systematic component of the scale model to be fitted.

data a data frame, list containing the variables in the model.

fraction numeric, represents a fraction of the size of the population. Default value is 0.2.

format character, represents the type of summary of the methodology, 'SIMPLE' or 'COMPLETE'. Default value is 'COMPLETE'.

... further parameters accepted by caret and survey functions.

Value

`sampling_design` is the name of the sampling design used in the estimation process.

`N` is the population size.

`n` is the fixed sample size used in the estimation process.

first_order_probabilities vector of the first order probabilities used in the estimation process.

sample is the random sample used in the estimation process.

estimated_total_y_sreg is the SREG estimate of the total parameter of the finite population.

Author(s)

Carlos Alberto Cardozo Delgado <cardozorpackages@gmail.com>

References

Cardozo C.A, Alonso C. (2021) Semi-parametric model assisted estimation in finite populations. In preparation.

Cardozo C.A., Paula G., and Vanegas L. (2021). Generalized log-gamma semiparametric models with P-spline smoothing. Submitted.

Sarndal C.E., Swensson B., and Wretman J. (2003). Model Assisted Survey Sampling. Springer-Verlag.

Examples

```
library(sregsurvey)
library(survey)
library(dplyr)
library(gamlss)
data(api)
attach(apipop)
Apipop <- filter(apipop,full!= 'NA')
Apipop <- filter(Apipop, stype == 'H')
Apipop <- Apipop %>% dplyr::select(api00,grad.sch,full)
fit <- sreg_srswr(api00 ~ pb(grad.sch), scale_formula = ~ full - 1, data= Apipop, fraction=0.25)
# The total population value is
sum(Apipop$api00)
```

---

sreg_stsi	*Semiparametric Model-Assisted Estimation under a Stratified Sampling with Simple Random Sampling Without Replace in each stratum.*

Description

sreg_stsi is used to estimate the total parameter of a finite population generated from a semi-parametric generalized gamma population under a stratified sampling with simple random sampling without-replacement in each stratum.

Usage

```
sreg_stsi(
  location_formula,
  scale_formula,
  stratum,
  data,
  n,
  ss_sizes,
  allocation_type = "PA",
  aux_x,
  ...
)
```

Arguments

location_formula

> a symbolic description of the systematic component of the location model to be fitted.

scale_formula a symbolic description of the systematic component of the scale model to be fitted.

stratum   vector, represents the strata of each unit in the population

data   a data frame, list containing the variables in the model.

n   integer, represents a fixed sample size.

ss_sizes   vector, represents a vector with the sample size in each stratum.

allocation_type

> character, there is two choices, proportional allocation, 'PA', and x-optimal allocation,'XOA'. By default is a 'PA', Sarndal et. al. (2003).

aux_x   vector, represents an auxiliary variable to help to calculate the sample sizes by the x-optimum allocation method, Sarndal et. al. (2003). This option is validated only when the argument allocation_type is equal to 'XOA'.

...   further parameters accepted by caret and survey functions.

Value

sampling_design is the name of the sampling design used in the estimation process.

N is the population size.

H is the number of strata.

Ns is the population strata sizes.

allocation_type is the method used to calculate sample strata sizes.

global_n is the global sample size used in the estimation process.

first_order_probabilities vector of the first order probabilities used in the estimation process.

sample is the random sample used in the estimation process.

estimated_total_y_sreg is the SREG estimate of the total parameter of the finite population.

Author(s)

Carlos Alberto Cardozo Delgado <cardozorpackages@gmail.com>

References

Cardozo C.A, Alonso C. (2021) Semi-parametric model assisted estimation in finite populations. In preparation.

Cardozo C.A., Paula G., and Vanegas L. (2021). Generalized log-gamma semiparametric models with P-spline smoothing. Submitted.

Sarndal C.E., Swensson B., and Wretman J. (2003). Model Assisted Survey Sampling. Springer-Verlag.

Examples

```
library(sregsurvey)
library(survey)
library(dplyr)
library(magrittr)
library(gamlss)
data(api)
attach(apipop)
Apipop <- filter(apipop,full!= 'NA')
Apipop <- Apipop %>% dplyr::select(api00,grad.sch,full,stype)
dim(Apipop)
sreg_stsi(api00 ~ pb(grad.sch),scale_formula = ~ full - 1, n=400, stratum = 'stype', data = Apipop)
# The total population value is
sum(Apipop$api00)
```

Index