Package 'surf'

October 14, 2022

Type Package

Title Survey-Based Gross Flows Estimation

Version 1.0.0

Author Guilherme Jacob

Date 2021-04-02

Maintainer Guilherme Jacob <guilhermejacob91@gmail.com>

Description Estimation of gross flows under non-response and complex sampling designs, using Gutiérrez, Nascimento Silva and Trujillo (2014) <https://www150.statcan.gc.ca/n1/pub/12-001-x/2014002/article/14113-eng.pdf> complex sampling extension of the non-response model developed by Stasny (1987) <https://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/some-markov-chain-models-for-nonresponse-in-estimating-gross-labor-force-flows.pdf>.
It estimates the gross flows process under non-response by modelling the observable cross-tabulation counts as a two-stage Markov Chain process, combining (1) the unobservable Markov Chain describing the transition of states; and (2) the non-response process, given by the initial response probabilities and the response/non-response transition probabilities.

BugReports https://github.com/guilhermejacob/surf

Depends R (>= 3.5.0), stats, Matrix

Imports numDeriv, survey, MASS, abind

License GPL-3

LazyData true

Suggests testthat, sampling, rmarkdown, knitr

RoxygenNote 7.1.1

Encoding UTF-8

NeedsCompilation no

Repository CRAN

Date/Publication 2021-04-06 08:10:03 UTC

R topics documented:

LFS79.0809	*Artificial 1979 Aug-Sep Canadian LFS dataset*

Description

Artificial dataset based on Stasny (1987) LFS counts for the 1979 August-September flows.

Usage

```
data("LFS79.0809")
```

Format

A data frame with three columns: y1 and y2 describing individual job status (E = "Employed"; U; "Unemployed"; N = "Not in the labor force") in each month, and a vector of selection probailities probs.

References

STASNY, E. A. Some Markov-chain models for nonresponse in estimating gross labor force flows. *Journal of Official Statistics*, v. 3, n. 4, p. 359, 1987.

LFS79.0910	*Artificial 1979 Sep-Oct LFS dataset*

Description

Artificial dataset based on Stasny (1987) LFS counts for the 1979 September-October flows.

Usage

```
data("LFS79.0910")
```

Format

A data frame with three columns: y1 and y2 describing individual job status (E = "Employed"; U; "Unemployed"; N = "Not in the labor force") in each month, and a vector of selection probailities probs.

References

STASNY, E. A. Some Markov-chain models for nonresponse in estimating gross labor force flows. *Journal of Official Statistics*, v. 3, n. 4, p. 359, 1987.

---

svyflow	*Gross flow estimation between categories*

Description

Compute gross flows for data from complex surveys with repeated samples.

Usage

```
## S3 method for class 'survey.design2'
svyflow(
  x,
  design,
  model = c("A", "B", "C", "D"),
  tol = 1e-04,
  maxit = 5000,
  verbose = FALSE,
  as.zero.flows = FALSE,
  influence = FALSE,
  ...
)

## S3 method for class 'svyrep.design'
svyflow(
  x,
  design,
  model = c("A", "B", "C", "D"),
  tol = 1e-04,
  maxit = 5000,
  verbose = FALSE,
  as.zero.flows = FALSE,
  influence = FALSE,
  ...
)
```

Arguments

x	a one-sided formula indicating a *factor* variable.
design	survey design object
model	Stasny (1987) model for the non-response process. Possibilities: `"A"`, `"B"`, `"C"`, `"D"`. Defaults to `model = "A"`.
tol	Tolerance for iterative proportional fitting. Defaults to `1e-4`.

maxit	Maximum number of iterations for iterative proportional fitting. Defaults to `maxit = 5000`.
verbose	Print proportional fitting iterations. Defaults to `verbose = FALSE`.
as.zero.flows	Should zeroes in the observed gross flows should be considered as zeroes in the population transition probability matrix? Defaults to `as.zero.flows = FALSE`.
influence	Should influence functions estimates be stored? Defaults to `influence = FALSE`.
...	future expansion.

Details

It is important to distinguish "missing" responses from "unaplicable" responses. This is feasible by subsetting the design for only applicable responses (with actual missing responses, if that is the case). For instance, suppose that we have two variables encoded as employed/unemployed, with NAs if the response is missing or is unaplicable. An NA might be a person that did not respond *or* a person who was under the working-age at the time of the survey. It is important to distinguish across those, as only one of those cases is an actual non-response. You could do that by looking for people who were in working age in any round, for instance. This can be done by using `subset`, as you should for a `survey design` object.

Value

Objects of class `flowstat`, a list of `svystat` and `svymstat` (a matrix version of `svystat`) objects. The `flowstat` object contais estimates of: the initial response probababilities `psi`, the response/response transition probabilities `rho`, the non-response/non-response transition probabilities `tau`, the (non-response corrected) initial and final distributions across categories `eta` and `gamma`, the (non-response corrected) transition probability matrix `pij`, the (non-response corrected) gross flows matrix `muij`, and the vector of net flows `delta`. These objects have methods for coef, vcov, SE, and cv.

A Rao-Scott Corrected Chi^2 test is also calculated.

Author(s)

Guilherme Jacob

References

STASNY, E. A. Some Markov-chain models for nonresponse in estimating gross labor force flows. *Journal of Official Statistics*, v. 3, n. 4, p. 359, 1987.

GUTIERREZ, H. A.; TRUJILLO, L.; SILVA, P. L. N. The estimation of gross flows in complex surveys with random nonresponse. *Survey Methodology*, v. 40, n. 2, p. 285–321, dec. 2014. URL https://www150.statcan.gc.ca/n1/en/catalogue/12-001-X201400214113.

LUMLEY, T. *Complex Surveys:* A guide to analysis using R. Hoboken: John Wiley & Sons, 2010. (Wiley Series in Survey Methodology). ISBN 978-0-470-28430-8.

Examples

```
# load library
library( survey )
library( surf )

# load data
data( "LFS79.0809" )

# create surf design object
lfs.des <- svydesign( ids = ~0 , probs = ~ prob , data = LFS79.0809 , nest = TRUE )

# flow estimates
estflows <- svyflow( ~y1+y2 , design = lfs.des )
coef( estflows$muij )
SE( estflows$muij )
```

Index