Package ‘surf’

April 6, 2021

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>


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 ................................................................. 2
LFS79.0910 ................................................................. 2
svyflow ................................................................. 3

Index

LFS79.0809  \hspace{1cm} 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 probabilities probs.

References


LFS79.0910  \hspace{1cm} 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 probabilities probs.
svyflow

References

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
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.

Details

It is important to distinguish "missing" responses from "unapplicable" 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 unapplicable. 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 contains estimates of: the initial response probabilities 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


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

* real
  LFS79.0809, 2
  LFS79.0910, 2

LFS79.0809, 2
LFS79.0910, 2

svyflow, 3