Package ‘saeczi’

April 15, 2024

Type Package
Title Small Area Estimation for Continuous Zero Inflated Data
Version 0.1.3
Maintainer Josh Yamamoto <joshuyamamoto5@gmail.com>
Description Provides functionality to fit a zero-inflated estimator for small area estimation. This estimator is a combines a linear mixed effects regression model and a logistic mixed effects regression model via a two-stage modeling approach. The estimator's mean squared error is estimated via a parametric bootstrap method. Chandra and others (2012, <doi:10.1080/03610918.2011.598991>) introduce and describe this estimator and mean squared error estimator. White and others (2024+, <doi:10.48550/arXiv.2402.03263>) describe the applicability of this estimator to estimation of forest attributes and further assess the estimator's properties.
License MIT + file LICENSE
Encoding UTF-8
LazyData true
Imports dplyr, lme4, purrr, progressr, furrr, future, rlang, Rcpp
RoxygenNote 7.3.1
Suggests testthat (>= 3.0.0)
Config/testthat/edition 3
Depends R (>= 4.1.0)
LinkingTo Rcpp, RcppEigen
NeedsCompilation yes
Author Josh Yamamoto [aut, cre], Dinan Elsyad [aut], Grayson White [aut], Julian Schmitt [aut], Niels Korsgaard [aut], Kelly McConville [aut], Kate Hu [aut]
Repository CRAN
Date/Publication 2024-04-15 15:00:03 UTC
R topics documented:

- pop
- saeczi
- samp

Index

pop  FIA Population Level Auxiliary Data for Oregon

Description

FIA Population Level Auxiliary Data for Oregon

Usage

pop

Format

An object of class data.frame with 10060 rows and 10 columns.

saeczi  Fit a zero-inflation estimator to a dataset

Description

Calculate the domain predictions using the zero-inflation estimator, and outputs those domain-level predictions, in the form of a dataframe. It contains the estimates for each domain, as well as the mean squared error estimates should the user choose. The output of the function is a list, with the first item being said dataframe, and the second being the R squared value of the model.

Usage

saeczi(
samp_dat,  
pop_dat,  
lin_formula,  
log_formula = lin_formula,  
domain_level,  
B = 100L,  
mse_est = FALSE,  
estimand = "means",  
parallel = FALSE
)

Arguments

samp_dat A dataframe with domains, predictor variables, and the response variable of a sample

pop_dat A dataframe with domains and predictor variables of a population

lin_formula model formula for the linear regression model

log_formula model formula for the logistic regression model

domain_level A string of the column name in the dataframes that reflect the domain level

B An integer of the number of reps desired for the bootstrap

mse_est A boolean that specifies if the user

estimand A string specifying whether the estimates should be 'totals' or 'means'.

parallel Compute MSE estimation in parallel

Details

The arguments ‘lin_formula’, and ‘log_formula’ can be unquoted or quoted. The function can handle both forms.

The two datasets (pop_dat and samp_dat) must have the same column names for the domain level, as well as the predictor variables for the function to work.

Value

An object of class ‘zi_mod’ with defined ‘print()’ and ‘summary()’ methods. The object is structured like a list and contains the following elements:

* call: The original function call

* res: A data.frame containing the estimates and mse estimates

* lin_mod: The modeling object used to fit the original linear model

* log_mod: The modeling object used to fit the original logistic model

Examples

data(pop)
data(samp)

lin_formula <- DRYBIO_AG_TPA_live_ADJ ~ tcc16 + elev

result <- saeczi(samp,
pop, 
lin_formula,
log_formula = lin_formula,
domain_level = "COUNTYFIPS",
mse_est = FALSE)
FIA sample data for Oregon

Description

FIA sample data for Oregon

Usage

samp

Format

An object of class tbl_df (inherits from tbl.data.frame) with 1494 rows and 11 columns.
Index

* datasets
  pop, 2
  samp, 4

pop, 2

saeczi, 2
samp, 4