Package 'saezero'

May 12, 2020

Type Package

Title Small Area Estimation under a Zero Inflated Lognormal Model with Correlated Random Area Effects

Version 0.1.0

Date 2020-03-22

Author Xiaodan Lyu <annielyu8@gmail.com>

Maintainer Xiaodan Lyu <annielyu8@gmail.com>

URL https://github.com/XiaodanLyu/saezero

Description Functions for small area estimation under a zero inflated lognormal model with correlated random area effects.
The functions in this package are designed and named after the package sae (Molina and Marhuenda, 2015).
Special thanks to Dr. Emily Berg <emilyb@iastate.edu> whose R code serves as a great guide to develop part of the functions in this package.

License GPL-2

Encoding UTF-8

LazyData true

Imports lme4, statmod, stats, sae

RoxygenNote 7.0.2

Suggests knitr, rmarkdown

VignetteBuilder knitr

R topics documented:

---

as.2pdata	*Converts a data frame to a list made for fitting LBH model*

Description

The default method transforms the data into a format that fits the framework of the unit level model of Lyu, Berg and Hofmann.

Usage

```
as.2pdata(
  f_pos,
  f_zero = f_pos,
  f_area,
  data,
  transform = "BoxCox",
  lambda = 0
)
```

Arguments

f_pos	an object of class [formula](): a symbolic description of the fixed effect model to be fitted to the positive part.
f_zero	an object of class [formula](): a symbolic description of the fixed effect model to be fitted to the binary part. Default value is to using the same formula as the positive part (f_pos).
f_area	an object of class [formula](): a symbolic description of the area code to be fitted to both the positive part and the negative part.
data	sample data containing the variables named in f_pos, f_zero and f_area. Any sampling units containing missing entries in the model data frame are removed.
transform	type of transformation for the positive responses to be chosen between the "Box-Cox" and "power" families. Default value is "Boxcox".
lambda	value for the parameter of the family of transformations specified in transform. Default value is 0, which gives the log transformation for the two possible families.

Details

The response variable in the formula f_zero is ignored. I(y>0) will be used for the binary part where y is the response variable in the formula f_pos.

Although this function allows a general transformation family for the positive part, the Empirical Bayes predictor of area means developed in the paper of Lyu, Berg and Hofmann ([ebLBH]()) assumes a log transformation is used. A general transformation family is provided in this function so as to facilitate a profile likelihood analysis together with the function [mleLBH]() to check the assumption of a log transformation for the positive part, i.e., whether lambda is significantly different from 0. See here for an example of profiling lambda with respect to the BoxCox transformation family.

Value

An object of the class "2pdata" which is a list with the following components:

- `lys`: transformed response vector for the positive part
- `Xs1`: model matrix for the positive part
- `deltas`: binary response vector for the binary part
- `Xs0`: model matrix for the binary part
- `area`: vector with the area code

See Also

[bxcx](#)

Examples

```
erosion_2p <- as.2pdata(f_pos = RUSLE2~logR+logK+logS,
                        f_zero = ~logR+logS+crop2+crop3,
                        f_area = ~cty, data = erosion)
```

---

ebLBH   *EB estimator of area means and associated MSE estimator*

Description

Obtains numerical approximations of EB estimators of area means under the unit level model of Lyu, Berg and Hofmann when the values of auxiliary variables for population units and the model parameter estimates are available.

Usage

```
ebLBH(Xaux, f_q = ~1, data_2p, fit, fullpop = FALSE)
```

Arguments

Xaux	matrix or data frame containing covariates, the area code and the variables named in f_q for population units.
f_q	an object of class [formula](#): a symbolic description of the number of population units with the same covariates. Default value is ~1.
data_2p	a two-part data object returned by [as.2pdata](#).
fit	a list of model parameter estimates containing at least fixed effects coefficients and variance components (named as the return value of [mleLBH](#)).
fullpop	a boolean variable indicating whether Xaux contains covariates information for the full population (TRUE) or just the out-of-sample units (FALSE). Default value is FALSE. The details of this indicator are given under Details.

Details

When Xaux contains only the covariates of the out-of-sample units (`fullpop = FALSE`), observed response is used for the sampled units when calculating the EB estimator. When Xaux contains the covariates of the full population (`fullpop = TRUE`), unit-level EB prediction is used for the sampled units. This is reasonable when the sampling fraction is extremely small in each area (e.g., Battese, Harter and Fuller (1988)).

Value

A data frame with the number of rows equal to the number of unique areas in Xaux:

- `area`: area codes
- `eb`: EB estimator of area means
- `mse`: the One-step MSE estimator

See Also

as.2pdata, mleLBH

Examples

```
erosion_2p <- as.2pdata(f_pos = RUSLE2~logR+logK+logS,
                        f_zero = ~logR+logS+crop2+crop3,
                        f_area = ~cty, data = erosion)
fit <- mleLBH(erosion_2p)
predictions <- ebLBH(Xaux, f_q = ~cnt, erosion_2p, fit, fullpop = TRUE)
```

---

erosion	*Simulated erosion data*

Description

Simulated data that mimics (the main features of) the real soil erosion data and other related variables for 64 South Dakota counties.

Usage

```
erosion
```

Format

A data frame with 646 rows and 10 variables.

- ctylab: county name
- cty: county code
- mukey: soil map unit key
- crop: crop category from the CDL data
- logR: `log` USLE rainfall factor
- logK: `log` USLE erosion erodibility factor
- logS: `log` USLE soil slope factor
- crop2: indicator of soybean
- crop3: indicator of spring wheat
- RUSLE2: soil erosion loss in tons per year

See Also

[Xaux](#)

---

map_sd   *South Dakota map data*

Description

South Dakota county map coordinates in longitude and latitude with FIPS.

Usage

```
map_sd
```

Format

An object of class `data.frame` with 1414 rows and 8 columns.

---

mleLBH   *MLE model parameter estimator*

Description

Fits by ML method the unit level model of Lyu, Berg and Hofmann. The specified link function is used in the binary part.

Usage

```
mleLBH(data_2p, link = "logit")
```

Arguments

data_2p   a two-part data object returned by [as.2pdata](#).

link   a specification for the link function used to model the binary part. The accepted link functions are `logit`, `probit`, `cauchit`, `log` and `cloglog`. Default value is "logit".

Value

The function returns a list with the following objects:

- `fixed`: list with the estimated values of the fixed regression coefficient in the positive part (p1) and in the binary part (p0).
- `random`: data frame with the predicted random effects in the positive part (p1) and the binary part (p0).
- `errorvar`: estimated model error variance in the positive part.
- `refvar1`: estimated random effects variance in the positive part.
- `refvar0`: estimated random effects variance in the binary part.

- `refcor`: estimated correlation coefficient of the random effects between the two parts.
- `loglik`: log-likelihood accomodating a general transformation family and `lambda` in the positive part as specified in `as.2pdata`.
- `residuals`: the marginal (`mar`) and conditional (`con`) residuals from the model fit in the positive part (`p1`, cases with nonpositive response are NA.).
- `fit0`: model parameter estimator under independence assumption.

See Also

`as.2pdata`

Examples

```
erosion_2p <- as.2pdata(f_pos = RUSLE2~logR+logK+logS,
                        f_zero = ~logR+logS+crop2+crop3,
                        f_area = ~cty, data = erosion)
fit <- mleLBH(erosion_2p)
```

---

simLBH   *Simulate unit responses*

Description

Simulate responses given the model parameter and covariates under the unit level model of Lyu, Berg and Hofmann.

Usage

```
simLBH(fit, Xs, f_pos, f_zero = f_pos, f_area, link = "logit")
```

Arguments

`fit`	a list of model parameter estimates containing at least fixed effects coefficients and variance components (named as the return value of `mleLBH`).
`Xs`	covariates matrix or data frame containing the variables named in `f_pos`, `f_zero` and `f_area`.
`f_pos`	an object of class `formula`: a symbolic description of the fixed effect model to be fitted to the positive part.
`f_zero`	an object of class `formula`: a symbolic description of the fixed effect model to be fitted to the binary part. Default value is to using the same formula as the positive part (`f_pos`).
`f_area`	an object of class `formula`: a symbolic description of the area code to be fitted to both the positive part and the negative part.
`link`	a specification for the link function used to model the binary part. The accepted link functions are `logit`, `probit`, `cauchit`, `log` and `cloglog`. Default value is "logit".

Value

response vector

Examples

```
erosion_2p <- as.2pdata(f_pos = RUSLE2~logR+logK+logS,
                        f_zero = ~logR+logS+crop2+crop3,
                        f_area = ~cty, data = erosion)
fit <- mleLBH(erosion_2p)
responses <- simLBH(fit, Xaux, f_pos = ~logR+logK+logS,
                    f_zero = ~logR+logS+crop2+crop3, f_area = ~cty)
```

Xaux   *Values of auxiliary variables for 64 domains*

Description

Values of auxiliary variables for full population units within 64 domains of data set erosion.

Usage

```
Xaux
```

Format

A data frame with 16,580 rows and 10 variables.

- ctylab: county name
- cty: county code
- mukey: soil map unit key
- crop: crop category from the CDL data
- logR: log USLE rainfall factor
- logK: log USLE erosion erodibility factor
- logS: log USLE soil slope factor
- crop2: indicator of soybean
- crop3: indicator of spring wheat
- cnt: number of crop pixels in soil map unit segment within county, from the CDL data

Source

viscover: a live Shiny tool featured in the RStudio Shiny gallery, which demonstrates how we integrate the USDA-NRCS Soil data and the USDA-NASS Cropland data in order to produce this dataset of auxiliary information for the cropland population in South Dakota.

Index