

Package ‘saens’

February 14, 2024

Type Package

Title 'EBLUP Fay-Herriot' Model for Estimation of Non-Sampled Areas
with Cluster Information

Version 0.1.0

Description Small area estimation ('Fay-Herriot' model) for non-
sampled area estimation with cluster information.

License MIT + file LICENSE

URL <https://github.com/Alfrzlp/sae-ns>

BugReports <https://github.com/Alfrzlp/sae-ns/issues>

Encoding UTF-8

LazyData true

Depends R (>= 4.00)

RoxygenNote 7.2.0

Imports cli, dplyr, ggplot2, methods, rlang, stats, tidyr

NeedsCompilation no

Author Ridson Al Farizal P [aut, cre, cph]

(<https://orcid.org/0000-0003-0617-0214>),

Azka Ubaidillah [aut] (<https://orcid.org/0000-0002-3597-0459>)

Maintainer Ridson Al Farizal P <alfrzlp@gmail.com>

R topics documented:

eblupfh	2
eblupfh_cluster	4
milk	6
mys	7
saens	7
Index	8

eblupfh

EBLUPs based on a Fay-Herriot Model.

Description

This function gives the 'EBLUP' (or EB predictor under normality) based on a 'Fay-Herriot' model.

Usage

```
eblupfh(
  formula,
  data,
  vardir,
  method = "REML",
  maxiter = 100,
  precision = 1e-04,
  scale = FALSE,
  print_result = TRUE
)
```

Arguments

formula	an object of class formula that contains a description of the model to be fitted. The variables included in the formula must be contained in the data.
data	a data frame or a data frame extension (e.g. a tibble).
vardir	vector or column names from data that contain variance sampling from the direct estimator for each area.
method	Fitting method can be chosen between 'ML' and 'REML'.
maxiter	maximum number of iterations allowed in the Fisher-scoring algorithm. Default is 100 iterations.
precision	convergence tolerance limit for the Fisher-scoring algorithm. Default value is 0.0001.
scale	scaling auxiliary variable or not, default value is FALSE.
print_result	print coefficient or not, default value is TRUE.

Details

The model has a form that is response ~ auxiliary variables. where numeric type response variables can contain NA. When the response variable contains NA it will be estimated with cluster information.

Value

The function returns a list with the following objects (df_res and fit): df_res a data frame that contains the following columns:

- y variable response

- eblup estimated results for each area
- random_effect random effect for each area
- vardir variance sampling from the direct estimator for each area
- mse Mean Square Error
- rse Relative Standart Error (%)

fit a list containing the following objects:

- estcoef a data frame with the estimated model coefficients in the first column (beta), their asymptotic standard errors in the second column (std.error), the t-statistics in the third column (tvalue) and the p-values of the significance of each coefficient in last column (pvalue)
- model_formula model formula applied
- method type of fitting method applied (ML or REML)
- random_effect_var estimated random effect variance
- convergence logical value that indicates the Fisher-scoring algorithm has converged or not
- n_iter number of iterations performed by the Fisher-scoring algorithm.
- goodness vector containing several goodness-of-fit measures: loglikelihood, AIC, and BIC

References

1. Rao, J. N., & Molina, I. (2015). Small area estimation. John Wiley & Sons.
2. Anisa, R., Kurnia, A., & Indahwati, I. (2013). Cluster information of non-sampled area in small area estimation. E-Prosiding Internasional Departemen Statistika FMIPA Universitas Padjadjaran, 1(1), 69-76.

Examples

```
library(saens)

m1 <- eblupfh(y ~ x1 + x2 + x3, data = na.omit(mys), vardir = "var")
m1 <- eblupfh(y ~ x1 + x2 + x3, data = na.omit(mys), vardir = ~var)
```

eblupfh_cluster

EBLUPs based on a Fay-Herriot Model with Cluster Information.

Description

This function gives the 'EBLUP' (or EB predictor under normality) based on a 'Fay-Herriot' model with cluster information for non-sampled areas.

Usage

```
eblupfh_cluster(
  formula,
  data,
  vardir,
  cluster,
  method = "REML",
  maxiter = 100,
  precision = 1e-04,
  scale = FALSE,
  print_result = TRUE
)
```

Arguments

formula	an object of class formula that contains a description of the model to be fitted. The variables included in the formula must be contained in the data.
data	a data frame or a data frame extension (e.g. a tibble).
vardir	vector or column names from data that contain variance sampling from the direct estimator for each area.
cluster	vector or column name from data that contain cluster information.
method	Fitting method can be chosen between 'ML' and 'REML'
maxiter	maximum number of iterations allowed in the Fisher-scoring algorithm. Default is 100 iterations.
precision	convergence tolerance limit for the Fisher-scoring algorithm. Default value is 0.0001.
scale	scaling auxiliary variable or not, default value is FALSE.
print_result	print coefficient or not, default value is TRUE.

Details

The model has a form that is response ~ auxiliary variables. where numeric type response variables can contain NA. When the response variable contains NA it will be estimated with cluster information.

Value

The function returns a list with the following objects `df_res` and `fit`: `df_res` a data frame that contains the following columns:

- `y` variable response
- `eblup` estimated results for each area
- `random_effect` random effect for each area
- `vardir` variance sampling from the direct estimator for each area
- `mse` Mean Square Error
- `cluster` cluster information for each area
- `rse` Relative Standart Error (%)

`fit` a list containing the following objects:

- `estcoef` a data frame with the estimated model coefficients in the first column (`beta`), their asymptotic standard errors in the second column (`std.error`), the t-statistics in the third column (`tvalue`) and the p-values of the significance of each coefficient in last column (`pvalue`)
- `model_formula` model formula applied
- `method` type of fitting method applied (ML or REML)
- `random_effect_var` estimated random effect variance
- `convergence` logical value that indicates the Fisher-scoring algorithm has converged or not
- `n_iter` number of iterations performed by the Fisher-scoring algorithm.
- `goodness` vector containing several goodness-of-fit measures: loglikelihood, AIC, and BIC

References

1. Rao, J. N., & Molina, I. (2015). Small area estimation. John Wiley & Sons.
2. Anisa, R., Kurnia, A., & Indahwati, I. (2013). Cluster information of non-sampled area in small area estimation. E-Prosiding Internasional Departemen Statistika FMIPA Universitas Padjadjaran, 1(1), 69-76.

Examples

```
library(saens)

m1 <- eblupfh_cluster(y ~ x1 + x2 + x3, data = mys, vardir = "var", cluster = "clust")
m1 <- eblupfh_cluster(y ~ x1 + x2 + x3, data = mys, vardir = ~var, cluster = ~clust)
```

milk

milk: Data on fresh milk expenditure.

Description

Data on fresh milk expenditure, used by Arora and Lahiri (1997) and by You and Chapman (2006).

Usage

```
milk
```

Format

A data frame with 43 observations on the following 6 variables.

SmallArea areas of inferential interest.

ni sample sizes of small areas.

yi average expenditure on fresh milk for the year 1989 (direct estimates for the small areas).

SD estimated standard deviations of yi.

var variance sampling from the direct estimator (yi) for each area

CV estimated coefficients of variation of yi.

MajorArea major areas created by You and Chapman (2006). These areas have similar direct estimates and produce a large CV reduction when using a FH model.

References

1. Arora, V. and Lahiri, P. (1997). On the superiority of the Bayesian method over the BLUP in small area estimation problems. *Statistica Sinica* 7, 1053-1063.
2. You, Y. and Chapman, B. (2006). Small area estimation using area level models and estimated sampling variances. *Survey Methodology* 32, 97-103.

mys	<i>mys: mean years of schooling people with disabilities in Papua Island, Indonesia.</i>
-----	--

Description

A dataset containing the mean years of schooling people with disabilities in Papua Island, Indonesia in 2021.

Usage

```
mys
```

Format

A data frame with 42 rows and 7 variables with 10 domains are non-sampled areas.

area regency municipality

y mean years of schooling people with disabilities

var variance sampling from the direct estimator for each area

rse relative standard error (%)

x1 Number of Elementary Schools

x2 Number of Junior High Schools

x3 Number of Senior High Schools

clust Cluster

Source

<https://www.bps.go.id>

saens	<i>saens : Small Area Estimation (sae) for Non-sampled Areas.</i>
-------	---

Description

Small area estimation for non sampled areas using cluster information.

Author(s)

Ridson Al Farizal P

References

1. Rao, J. N., & Molina, I. (2015). Small area estimation. John Wiley & Sons.
2. Anisa, R., Kurnia, A., & Indahwati, I. (2013). Cluster information of non-sampled area in small area estimation. E-Prosidings Internasional Departemen Statistika FMIPA Universitas Padjadjaran, 1(1), 69-76.

Index

* **datasets**

milk, [6](#)

mys, [7](#)

eblupfh, [2](#)

eblupfh_cluster, [4](#)

milk, [6](#)

mys, [7](#)

saens, [7](#)