Package ‘saens’

November 18, 2024

Type Package

Title Small Area Estimation with Cluster Information for Estimation of Non-Sampled Areas

Version 0.1.2

Description Implementation of small area estimation (Fay-Herriot model) with EBLUP (Empirical Best Linear Unbiased Prediction) Approach for non-sampled area estimation by adding cluster information and assuming that there are similarities among particular areas. See also Rao & Molina (2015, ISBN:978-1-118-73578-7) and Anisa et al. (2013) <doi:10.9790/5728-10121519>.

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`>

Repository CRAN

Date/Publication 2024-11-18 04:40:03 UTC

Contents

AIC.eblupres *Akaike's An Information Criterion.*

Description

Generic function calculating Akaike's "An Information Criterion" for EBLUP model

Usage

```
## S3 method for class 'eblupres'
AIC(object, ...)

## S3 method for class 'eblupres'
BIC(object, ...)
```

Arguments

object	EBLUP model.
...	further arguments passed to or from other methods.

Value

AIC value.

Examples

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

autoplot *Create a complete ggplot appropriate to a particular data type*

Description

autoplot() uses ggplot2 to draw a particular plot for an object of a particular class in a single command. This defines the S3 generic that other classes and packages can extend.

Usage

```
autoplot(object, ...)
```

Arguments

object	an object, whose class will determine the behaviour of autoplot
...	other arguments passed to specific methods

Value

a ggplot object

See Also

autolayer(), ggplot() and fortify()

autoplot.eblupres *Autoplot EBLUP results.*

Description

Autoplot EBLUP results.

Usage

```
## S3 method for class 'eblupres'
autoplot(object, variable = "RSE", ...)
```

Arguments

object	EBLUP model.
variable	variable to plot.
...	further arguments passed to or from other methods.

Value

plot.

Examples

```
library(saens)

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

`coef.eblupres` *Extract Model Coefficients.*

Description

Extract Model Coefficients.

Usage

```
## S3 method for class 'eblupres'
coef(object, ...)
```

Arguments

`object`	EBLUP model.
`...`	further arguments passed to or from other methods.

Value

model coefficients

Examples

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

`eblupfh` *EBLUPs based on a Fay-Herriot Model.*

Description

This function gives the Empirical Best Linear Unbiased Prediction (EBLUP) or Empirical Best (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: loglikehood, AIC, and BIC

References

1. Rao, J. N., & Molina, I. (2015). Small area estimation. John Wiley & Sons.

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 Empirical Best Linear Unbiased Prediction (EBLUP) or Empirical Best (EB) predictor based on a Fay-Herriot model with cluster information for non-sampled areas.

Usage

```
eblupfh_cluster(
  formula,
  data,
  vardir,
  cluster,
  method = "REML",
  mse_method = "jackknife",
  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’

`mse_method` MSE estimating method can be chosen between ’default’ and ’jackknife’

`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: loglikehood, 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 Internasionall 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)
```

eblupfh_ns *Synthetic Estimator.*

Description

Synthetic estimator is one of the simple methods to obtain predicted values of mean specific area parameters, which the direct estimates are unknown. Based on estimated of parameter coefficient models using Empirical Best Unbiased Prediction (EBLUP), the synthetic estimator is obtained by calibrating the estimated parameter coefficient to the auxiliary variables.

Usage

```
eblupfh_ns(
  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 is defined as response ~ auxiliary variables, where the response variable, of numeric type, may contain NA values. When the response variable contains NA, it will be estimated using a synthetic estimator.

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: loglikehood, AIC, and BIC

References

1. Rao, J. N., & Molina, I. (2015). Small area estimation. John Wiley & Sons.

Examples

```
library(saens)

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

`logLik.eblupres` *Extract Log-Likelihood.*

Description

Extract Log-Likelihood.

Usage

```
## S3 method for class 'eblupres'
logLik(object, ...)
```

Arguments

`object`	EBLUP model.
`...`	further arguments passed to or from other methods.

Value

Log-Likehood value

Examples

```
library(saens)

model1 <- eblupfh_cluster(y ~ x1 + x2 + x3, data = mys, vardir = "var", cluster = "clust")
logLik(model1)
```

`mys` *mys: mean years of schooling people with disabilities in Papua Island, Indonesia.*

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

n Number of eligible samples

weight Weight

Source

https://www.bps.go.id

`summary.eblupres` *Summarizing EBLUP Model Fits.*

Description

‘summary‘ method for class "eblupres".

Usage

```
## S3 method for class 'eblupres'
summary(object, ...)
```

Arguments

`object`	EBLUP model.
`...`	further arguments passed to or from other methods.

Value

The function returns 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 (

Examples

```
library(saens)

model1 <- eblupfh_cluster(y ~ x1 + x2 + x3, data = mys, vardir = "var", cluster = "clust")
summary(model1)
```

Index