Package: kbaabb (via r-universe)

October 22, 2026

Type Package

Title Generates an Artificial Population Based on the KBAABB Methodology

Version 0.0.1

Description Provides tools for generating an artificial population based on a kNN-based approximation to the approximate Bayesian bootstrap (KBAABB). This methodology is described in White et al. (2025) <doi:10.1093/forestry/cpaf071>.

License CC0

Encoding UTF-8

Imports FNN

Depends R (>= 2.10)

LazyData true

Config/roxygen2/version 8.0.0

Repository https://graysonwhite.r-universe.dev

Date/Publication 2026-10-21 22:31:36 UTC

RemoteUrl https://github.com/graysonwhite/kbaabb

RemoteRef HEAD

RemoteSha 614b24cd61483828c3e7bed60db4bad6eecf81ea

Contents

kbaabb 2
SJC_boundary 4
SJC_population 4
SJC_sample 5

Index **6**

kbaabb *kbaabb*

Description

kbaabb() generates an imputed population dataset based on methodology introduced in White et al. (2025).

Usage

```
kbaabb(
  survey_data,
  population_data,
  formula,
  k = 10,
  strata = NULL,
  center_scale = TRUE,
  seed = NULL,
  keep_survey_id = TRUE,
  id_name = "survey_row_id",
  ...
)
```

Arguments

survey_data	A dataframe containing the survey data to be used for the imputation (i.e. "the sample"). This dataframe must contain all variables listed in formula and strata.
population_data	A dataframe containing the population data for imputation to occur on (i.e. "the population"). This dataframe must contain all auxiliary variables listed in formula and strata.
formula	The formula specified for imputation, taking the form y1 + y2 + . . . + yn ~ x1 + x2 + . . . + xm where {x1, x2, . . . , xm} is the set of auxiliary variables used for imputation and {y1, y2, . . . , yn} are the responses.
k	Integer. The number of neighbors used in the k nearest neighbors imputation
strata	Character. The name of a variable to be used for stratification. If NULL (default), imputation is performed without stratification. Otherwise, stratification occurs based on the variable specified in strata. It is advised to provide strata.
center_scale	Logical. If TRUE (default), auxiliary variables are centered and scaled (mean = 0, variance = 1) based on the population data. Otherwise, the original sample and population dataframes supplied by the user are used in an unmodified form.
seed	A seed to be set for reproducibility.
keep_survey_id	Logical, should the rownames of the survey data which correspond to the imputed values be kept? Allows for a lookup of which rows were used for which imputed observations.

id_name	The name for the column containing the rownames in keep_survey_id.
...	Currently ignored. For extendability.

Value

A kbaabb object including:

- imputed_population_data: The population data
- survey_data: Original sample data
- k: Value for k
- stratified: Indicator of if stratifying occured, as well as
- strata: The stratifying variable
- center_scale: If centering and scaling occured
- formula: The formula used for population imputation

Examples

```
# KBAABB imputation for k = 5, stratifying by `tnt`:
pop = kbaabb(survey_data = SJC_sample,
       population_data = SJC_population,
       formula = biomass ~ tcc + elev,
       k = 5,
       strata = "tnt",
       center_scale = TRUE,
       seed = 37)
head(pop$imputed_population_data)

# and without stratification
pop = kbaabb(survey_data = SJC_sample,
       population_data = SJC_population,
       formula = biomass ~ tcc + elev,
       k = 5,
       seed = 37)
head(pop$imputed_population_data)

# or with multiple imputations
pop = kbaabb(survey_data = SJC_sample,
       population_data = SJC_population,
       formula = biomass + measurement_year ~ tcc + elev,
       k = 5,
       strata = "tnt",
       seed = 37)
head(pop$imputed_population_data)
```

SJC_boundary — *Polygon boundary dataset for San Juan County, Washington, USA*

Description

An ‘sf‘ object containing the polygon boundaries of San Juan County, Washington, USA.

Usage

```
SJC_boundary
```

Format

‘SJC_boundary‘: A data frame and sf object with 124 rows and 2 columns:

island Name of the island that the polygon represents.

geometry ‘sf‘ geometry.

SJC_population — *Population dataset for San Juan County, Washington, USA*

Description

Population level data for each 90m pixel in San Juan County, Washington, USA. For more information on the auxiliary data, see Wieczorek et al. (2023+).

Usage

```
SJC_population
```

Format

‘SJC_population‘: A data frame and sf object with 55,570 rows and 5 columns:

island Name of the island that the pixel occurs on.

tcc Auxiliary data. Tree canopy cover value at given pixel.

elev Auxiliary data. Elevation value at given pixel.

tnt Auxiliary data. Tree (2) and non-tree (1) layer.

geometry ‘sf‘ geometry.

`SJC_sample` *Sample dataset for San Juan County, Washington, USA*

Description

Survey data collected by the USDA Forest Service, Forest Inventory & Analysis Program (FIA). This dataset contains plot-level data for biomass as collected and reported by the FIA. Further, auxiliary data has been matched with this plot level data to match the population dataset.

Usage

```
SJC_sample
```

Format

‘SJC_sample‘: A data frame and sf object with 25 rows and 8 columns:

plt_cn Name of the island that the pixel occurs on.

biomass Average biomass measurement calculated from ground-collected FIA data. Units: short tons/acre.

tcc Auxiliary data. Tree canopy cover value at given FIA plot.

elev Auxiliary data. Elevation value at given FIA plot

wc3cl Auxiliary data. Tree (2) and non-tree (1) value at a given FIA plot.

measurement_year Year in which the FIA plot was measured.

island Name of the island that the FIA plot is on.

geometry ‘sf‘ geometry.

Index

∗ **datasets**
SJC_boundary, 4
SJC_population, 4
SJC_sample, 5

kbaabb, 2

SJC_boundary, 4
SJC_population, 4
SJC_sample, 5