Package: AutoStrataK (via r-universe)

June 30, 2026

Type Package

Title Automatic Optimal Stratification for Survey Sampling

Version 0.1.4

Description Provides tools for the automatic stratification of survey populations using clustering and optimization techniques. The package assists researchers and survey practitioners in constructing homogeneous strata to improve the efficiency and precision of survey estimates. Functions are provided for generating strata, evaluating stratification quality, summarizing stratified populations, and visualizing stratification results. These tools support the design and implementation of efficient survey sampling strategies. The package utilizes standard statistical methods from survey sampling, clustering, and optimization for automatic stratification. Methods are described in Cochran (1977, ISBN:9780471162405) and Lohr (2021, ISBN:9780367354556).

License GPL-3

Encoding UTF-8

Imports dplyr, ggplot2, rlang, stats

Depends R (>= 4.1.0)

Suggests knitr, rmarkdown

VignetteBuilder knitr

Config/roxygen2/version 8.0.0

NeedsCompilation no

Author Khalid Islam [aut, cre]

Maintainer Khalid Islam <drkhalidulislam@gmail.com>

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

Date/Publication 2026-06-30 21:09:40 UTC

RemoteUrl https://github.com/cran/AutoStrataK

RemoteRef HEAD

RemoteSha 6990f2c4b4f417c8cb08c1bc859d4af83006d6a2

Contents

autostrata 2
clustering_strata 3
compare_sampling 3
evaluate_strata 4
optimal_allocation 5
plot.autostrata 5
summary.autostrata 6

Index **7**

`autostrata` *Automatic Stratification*

Description

Automatically partitions a population into homogeneous strata using clustering-based methods. The resulting strata can be used in survey sampling designs to improve the precision and efficiency of estimates.

Usage

```
autostrata(data, target, n_strata = 4)
```

Arguments

`data`	A data frame containing the study variables.
`target`	A target variable used for stratification.
`n_strata`	An integer specifying the desired number of strata.

Value

A data frame containing the original data along with the assigned stratum membership. The returned object has class `"autostrata"`.

Examples

```
data <- data.frame(
  x = c(10, 15, 20, 25, 30, 35),
  y = c(5, 8, 12, 16, 20, 24)
)

result <- autostrata(
  data = data,
  target = y,
  n_strata = 2
)
```

```
head(result)
```

`clustering_strata` *Create Optimal Strata*

Description

Create Optimal Strata

Usage

```
clustering_strata(data, target, n_strata)
```

Arguments

`data`	Input data frame
`target`	Target variable name
`n_strata`	Number of strata

Value

Data frame with stratum assignments

`compare_sampling` *Compare Efficiency*

Description

Compares variance under simple random sampling (SRS) with within-stratum variance after stratification.

Usage

```
compare_sampling(data, target)
```

Arguments

`data`	A stratified data frame containing a variable named `stratum`.
`target`	Character string giving the target variable name.

Value

A numeric value representing relative efficiency. Values greater than 1 indicate improved efficiency due to stratification.

`evaluate_strata` *Evaluate Strata*

Description

Computes measures of stratification quality, including the overall variance of the target variable, the total within-stratum variance, and a homogeneity index.

Usage

```
evaluate_strata(data, target)
```

Arguments

`data`	A stratified data frame containing a column named `stratum`.
`target`	Character string specifying the target variable used for stratification.

Value

A list containing:

overall_variance Variance of the target variable in the full population.

within_variance Sum of within-stratum variances.

homogeneity Homogeneity index, where larger values indicate more homogeneous strata.

Examples

```
data <- data.frame(
  income = c(15, 18, 20, 25, 30, 35, 40, 50),
  stratum = c(1, 1, 2, 2, 3, 3, 4, 4)
)

evaluate_strata(
  data = data,
  target = "income"
)
```

optimal_allocation *Neyman Allocation*

Description

Neyman Allocation

Usage

```
optimal_allocation(data, target, total_sample)
```

Arguments

data	Stratified data frame
target	Target variable name
total_sample	Total sample size

Value

Allocation table

plot.autostrata *Plot Strata*

Description

Plot Strata

Usage

```
## S3 method for class 'autostrata'
plot(x, ...)
```

Arguments

x	Object of class autostrata
...	Additional arguments

Value

A ggplot object

`summary.autostrata` *Summary Method*

Description

Summary Method

Usage

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

Arguments

`object`	Object of class autostrata
`...`	Additional arguments

Value

Printed summary

Index

autostrata, 2

clustering_strata, 3
compare_sampling, 3

evaluate_strata, 4

optimal_allocation, 5

plot.autostrata, 5

summary.autostrata, 6