

Package ‘stratifiedyh’

July 23, 2025

Type Package

Title Stratified Sampling and Labeling of Data in R

Version 0.1.0

Description

Provides functions for stratified sampling and assigning custom labels to data, ensuring randomness within groups. The package supports various sampling methods such as stratified, cluster, and systematic sampling. It allows users to apply transformations and customize the sampling process. This package can be useful for statistical analysis and data preparation tasks.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.3.2

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

NeedsCompilation no

Author Duan Yuanheng [aut, cre]

Maintainer Duan Yuanheng <yhyuanheng@gmail.com>

Repository CRAN

Date/Publication 2024-09-18 12:30:02 UTC

Contents

cluster_labels	2
custom_transform	2
stratified_custom_labels	3
stratified_labels	4
systematic_labels	4
Index	6

cluster_labels	<i>Cluster Sampling and Labeling</i>
----------------	--------------------------------------

Description

This function performs cluster sampling on the dataframe and assigns "Yes" or "No" labels to rows based on selected clusters.

Usage

```
cluster_labels(df, group_col, yes_percentage)
```

Arguments

df	A data frame containing the data.
group_col	A character string specifying the column to use for clustering.
yes_percentage	A numeric value between 0 and 100 indicating the percentage of clusters to label as "Yes".

Value

A data frame with an additional column "Clustered_Yes_No" containing the cluster-sampled "Yes"/"No" labels.

Examples

```
result <- cluster_labels(iris, group_col = "Species", yes_percentage = 50)
```

custom_transform	<i>Apply Custom Transformation to Data Column</i>
------------------	---

Description

This function allows the user to apply a custom transformation (scaling, normalization, log transform, or custom function) to a specified numeric column.

Usage

```
custom_transform(df, selected_column, transformation_type)
```

Arguments

df	A data frame containing the data.
selected_column	A character string specifying the column to be transformed.
transformation_type	A character string representing the transformation type: "scale", "normalize", "log", or a custom R function.

stratified_labels	<i>Stratify and Assign Labels to Data</i>
-------------------	---

Description

This function stratifies data based on a specified grouping column and assigns "Yes" or "No" labels according to a given percentage.

Usage

```
stratified_labels(df, group_col, yes_percentage)
```

Arguments

df	A data frame to be stratified.
group_col	A character string specifying the column name to group by.
yes_percentage	A numeric value between 0 and 100 indicating the percentage of "Yes" labels to assign within each group.

Value

A data frame with an additional column "Sampled_Yes_No" containing the stratified "Yes"/"No" labels.

Examples

```
# Example with the iris dataset
result <- stratified_labels(iris, group_col = "Species", yes_percentage = 50)
```

systematic_labels	<i>Systematic Sampling and Labeling</i>
-------------------	---

Description

This function performs systematic sampling on the dataframe and assigns "Yes" or "No" labels to rows based on the specified interval.

Usage

```
systematic_labels(df, group_col, sampling_interval)
```

Arguments

df	A data frame containing the data.
group_col	A character string specifying the column to use for grouping.
sampling_interval	A numeric value representing the interval for systematic sampling.

Value

A data frame with an additional column "Systematic_Yes_No" containing the systematically sampled "Yes"/"No" labels.

Examples

```
result <- systematic_labels(iris, group_col = "Species", sampling_interval = 2)
```

Index

cluster_labels, [2](#)
custom_transform, [2](#)

stratified_custom_labels, [3](#)
stratified_labels, [4](#)
systematic_labels, [4](#)