Package: stratifiedyh (via r-universe)

May 12, 2026

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 https://duanyh16.r-universe.dev

Date/Publication 2024-09-19 02:58:09 UTC

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

RemoteRef HEAD

RemoteSha 812d4dc53c385d9cc7e2468d5fd4fa469b5600bb

Contents

cluster_labels 2
custom_transform 2
stratified_custom_labels 3
stratified_labels 4
systematic_labels 4

Index **6**

`cluster_labels` *Cluster Sampling and Labeling*

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` *Apply Custom Transformation to Data Column*

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.

Value

A data frame with the transformed column.

Examples

```
result <- custom_transform(iris, selected_column = "Sepal.Length", transformation_type = "scale")
```

---

stratified_custom_labels

Stratify and Assign Custom Labels to Data

Description

This function stratifies data based on a specified grouping column and assigns custom labels according to a given percentage.

Usage

```
stratified_custom_labels(df, group_col, label_percentage, label1, label2)
```

Arguments

df	A data frame to be stratified.
group_col	A character string specifying the column name to group by.
label_percentage	A numeric value between 0 and 100 indicating the percentage of the first label to assign within each group.
label1	A character string representing the first label.
label2	A character string representing the second label.

Value

A data frame with an additional column "Custom_Labels" containing the stratified custom labels.

Examples

```
result <- stratified_custom_labels(iris, group_col = "Species",
                                   label_percentage = 50,
                                   label1 = "High", label2 = "Low")
```

stratified_labels *Stratify and Assign Labels to Data*

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 *Systematic Sampling and Labeling*

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