

Package ‘cmahalanobis’

September 24, 2024

Type Package

Title Calculate Distance Measures for a Given List of Data Frames with Factors

Version 0.4.2

Maintainer Flavio Gioia <flaviogioia.fg@gmail.com>

Description

It provides functions that calculate Mahalanobis distance, Euclidean distance, Manhattan distance and Chebyshev distance between each pair of species in a list of data frames. These metrics are fundamental in various fields, such as cluster analysis, classification, and other applications of machine learning and data mining, where assessing similarity or dissimilarity between data is crucial. The package is designed to be flexible and easily integrated into data analysis workflows, providing reliable tools for evaluating distances in multidimensional contexts.

License GPL-3

Encoding UTF-8

RoxygenNote 7.2.3

Imports stats, ggplot2, reshape2, mice

Suggests rmarkdown, testthat (>= 3.0.0)

NeedsCompilation no

Config/testthat/edition 3

Author Flavio Gioia [aut, cre]

Repository CRAN

Date/Publication 2024-09-23 22:00:22 UTC

Contents

cchebyshev	2
ceulide	3
cmahalanobis	4
cmanhattan	5
generate_report_cchebyshev	6
generate_report_ceulide	7

generate_report_cmahalanobis	8
generate_report_cmanhattan	9
pvaluesccheb	10
pvaluesceucl	11
pvaluescmaha	12
pvaluescmanh	13

Index 14

cchebyshev	<i>Calculate the p_values matrix for each species, using Chebyshev distance as a base.</i>
------------	--

Description

This function takes a dataset, a factor, a p_value method, number of bootstraps and permutation when necessary, and returns a p_values matrix between each pair of species and a plot if the user select TRUE using the Chebyshev distance for the distances calculation.

Usage

```
cchebyshev(
  dataset,
  formula,
  plot = TRUE,
  plot_title = "Chebyshev Distance Between Groups"
)
```

Arguments

dataset	A dataframe.
formula	A factor which you want to calculate Chebyshev distance.
plot	If TRUE, displays a plot of distances.
plot_title	The title of plot.

Value

A matrix containing distances and, optionally, the plot.

Examples

```
# Example with iris dataset

cchebyshev(iris, ~Species, plot = TRUE, plot_title = "Chebyshev Distance Between Groups")

# Example with mtcars dataset

cchebyshev(mtcars, ~am, plot = TRUE, plot_title = "Chebyshev Distance Between Groups")
```

ceulide	<i>Calculate the Euclidean distance of a factor in a dataframe.</i>
---------	---

Description

This function takes a dataframe and a factor in input, and returns a matrix with the Euclidean distances about it.

Usage

```
ceulide(  
  dataset,  
  formula,  
  plot = TRUE,  
  plot_title = "Euclidean Distance Between Groups"  
)
```

Arguments

dataset	A dataframe.
formula	The factor which you want to calculate the Euclidean distances matrix.
plot	If TRUE, shows a plot of the Euclidean distances matrix.
plot_title	The title of the plot.

Details

Calculate Euclidean distance

Value

The matrix containing distances.

Examples

```
# Example with iris dataset  
  
ceulide(iris, ~Species, plot = TRUE, plot_title = "Euclidean Distance Between Groups")  
  
# Example with mtcars dataset  
  
ceulide(mtcars, ~am, plot = TRUE, plot_title = "Euclidean Distance Between Groups")
```

`cmahalanobis`*Calculate the Mahalanobis distance for each species.*

Description

This function takes a dataframe and a factor in input, and returns a matrix with the Mahalanobis distances about it.

Usage

```
cmahalanobis(  
  dataset,  
  formula,  
  plot = TRUE,  
  plot_title = "Mahalanobis Distance Between Groups"  
)
```

Arguments

<code>dataset</code>	A dataframe.
<code>formula</code>	A factor which you want to calculate the Mahalanobis distances matrix.
<code>plot</code>	Logical, if TRUE, a plot of Mahalanobis distances matrix is displayed.
<code>plot_title</code>	The title to be used for the plot if plot is TRUE.

Value

A matrix containing Mahalanobis distances between each pair of groups and the plot.

Examples

```
# Example with the iris dataset  
  
data(iris)  
  
# Calculate the Mahalanobis distance with the cmahalanobis function  
cmahalanobis(iris, ~Species, plot = TRUE, plot_title = "Mahalanobis Distance Between Groups")  
  
# Example with the mtcars dataset  
data(mtcars)  
  
# Calculate the Mahalanobis distance with the cmahalanobis function  
cmahalanobis(mtcars, ~am, plot = TRUE, plot_title = "Mahalanobis Distance Between Groups")
```

`cmanhattan`*Calculate a Manhattan distance of a factor in a dataframe.*

Description

This function takes a dataframe and a factor in input, and returns a matrix with the Manhattan distances about it.

Usage

```
cmanhattan(  
  dataset,  
  formula,  
  plot = TRUE,  
  plot_title = "Manhattan Distance Between Groups"  
)
```

Arguments

<code>dataset</code>	A dataframe.
<code>formula</code>	A factor which you want to calculate Manhattan distance.
<code>plot</code>	If TRUE, show a plot of distances.
<code>plot_title</code>	The title of plot.

Details

Calculate Manhattan distance

Value

A matrix containing distances.

Examples

```
# Example with iris dataset  
  
cmanhattan(iris, ~Species, plot = TRUE, plot_title = "Manhattan Distance Between Groups")  
  
# Example with mtcars dataset  
  
cmanhattan(mtcars, ~am, plot = TRUE, plot_title = "Manhattan Distance Between Groups")
```

`generate_report_cchebyshev`

Generate a Microsoft Word document about the Chebyshev distance matrix and the p-values matrix with corresponding plots.

Description

This function takes a dataframe, a factor and returns a Microsoft Word document about the Chebyshev distance matrix and the p-values matrix with corresponding plots.

Usage

```
generate_report_cchebyshev(  
  dataset,  
  formula,  
  pvalue.method = "chisq",  
  num.permutations = 10,  
  num.bootstraps = 10  
)
```

Arguments

<code>dataset</code>	A dataframe.
<code>formula</code>	A factor which you want to calculate the Chebyshev distance matrix and the p_values matrix.
<code>pvalue.method</code>	A p_value method used to calculate the matrix, the default value is "chisq". Other methods are "permutation" and "bootstrap".
<code>num.permutations</code>	Number of permutation to specify if you select "permutation" in "pvalue.method". The default value is 100.
<code>num.bootstraps</code>	Number of bootstrap to specify if you select "bootstrap" in "p_value method". The default value is 10.

Value

A Microsoft Word document about the Chebyshev distance matrix and the p_values matrix.

Examples

```
# Generate a report about "Species" factor in iris dataset  
generate_report_cchebyshev(iris, ~Species)  
  
# Generate a report about "am" factor in mtcars dataset  
generate_report_cchebyshev(mtcars, ~am)
```

`generate_report_ceuclide`

Generate a Microsoft Word document about the Euclidean distance matrix and the p-values matrix with relative plots.

Description

This function takes a dataframe, a factor and returns a Microsoft Word document about the Euclidean distance matrix and the p-values matrix with relative plots.

Usage

```
generate_report_ceuclide(  
  dataset,  
  formula,  
  pvalue.method = "chisq",  
  num.permutations = 10,  
  num.bootstraps = 10  
)
```

Arguments

<code>dataset</code>	A dataframe.
<code>formula</code>	A factor which you want to calculate the Euclidean distance matrix and the p_values matrix.
<code>pvalue.method</code>	A p_value method used to calculate the matrix, the default value is "chisq". Other methods are "permutation" and "bootstrap".
<code>num.permutations</code>	Number of permutation to specify if you select "permutation" in "pvalue.method". The default value is 100.
<code>num.bootstraps</code>	Number of bootstrap to specify if you select "bootstrap" in "p_value method". The default value is 10.

Value

A Microsoft Word document about the Euclidean distance matrix and the p_values matrix.

Examples

```
# Generate a report about "Species" factor in iris dataset  
generate_report_ceuclide(iris, ~Species)  
  
# Generate a report about "am" factor in mtcars dataset  
generate_report_ceuclide(mtcars, ~am)
```

`generate_report_cmahalanobis`

Generate a Microsoft Word document about Mahalanobis distance matrix and p-values matrix with corresponding plots.

Description

This function takes a dataframe, a factor and returns a Microsoft Word document about Mahalanobis distance matrix and p-values matrix with corresponding plots.

Usage

```
generate_report_cmahalanobis(  
  dataset,  
  formula,  
  pvalue.method = "chisq",  
  num.permutations = 10,  
  num.bootstraps = 10  
)
```

Arguments

<code>dataset</code>	A dataframe.
<code>formula</code>	A factor which you want to calculate Mahalanobis distances matrix and p_values matrix.
<code>pvalue.method</code>	A method with which you want to calculate pvalue matrix. The default method is "chisq". Other methods are "permutation" and "bootstrap".
<code>num.permutations</code>	A number of permutations to define if you choose "permutation".
<code>num.bootstraps</code>	A number of bootstrap to define if you choose "bootstrap".

Value

A Microsoft Word document about Mahalanobis distances matrix and p_values matrix.

Examples

```
# Generate a report about "Species" factor in iris dataset  
generate_report_cmahalanobis(iris, ~Species)  
  
# Generate a report about "am" factor in mtcars dataset  
generate_report_cmahalanobis(mtcars, ~am)
```

`generate_report_cmanhattan`

Generate a Microsoft Word document about the Manhattan distance and the p-values matrices with corresponding plots.

Description

This function takes a dataframe, a factor and returns a Microsoft Word document about the Manhattan distance matrix and the p-values matrix with corresponding plots.

Usage

```
generate_report_cmanhattan(  
  dataset,  
  formula,  
  pvalue.method = "chisq",  
  num.permutations = 10,  
  num.bootstraps = 10  
)
```

Arguments

<code>dataset</code>	A dataframe.
<code>formula</code>	A factor which you want to calculate the Manhattan distance matrix and the p_values matrix.
<code>pvalue.method</code>	A p_value method used to calculate the matrix, the default value is "chisq". Other methods are "permutation" and "bootstrap".
<code>num.permutations</code>	Number of permutation to specify if you select "permutation" in "pvalue.method". The default value is 100.
<code>num.bootstraps</code>	Number of bootstrap to specify if you select "bootstrap" in "p_value method". The default value is 10.

Value

A Microsoft Word document about the Manhattan distance matrix and the p_values matrix.

Examples

```
# Generate a report about "Species" factor in iris dataset  
generate_report_cmanhattan(iris, ~Species)  
  
# Generate a report about "am" factor in mtcars dataset  
generate_report_cmanhattan(mtcars, ~am)
```

pvaluesccheb	<i>Calculate the p_values matrix for each species, using Chebyshev distance as a base.</i>
--------------	--

Description

This function takes a dataset, a factor, a p_value method, number of bootstraps and permutation when necessary, and returns a p_values matrix between each pair of species and a plot if the user select TRUE using Chebyshev distance for the distances calculation.

Usage

```
pvaluesccheb(  
  dataset,  
  formula,  
  pvalue.method = "chisq",  
  num.permutations = 100,  
  num.bootstraps = 10,  
  plot = TRUE  
)
```

Arguments

dataset	A dataframe.
formula	A factor which you want to calculate Chebyshev distance.
pvalue.method	A p_value method used to calculate the matrix, the default value is "chisq". Other methods are "permutation" and "bootstrap".
num.permutations	Number of permutation to specify if you select "permutation" in "pvalue.method". The default value is 100.
num.bootstraps	Number of bootstrap to specify if you select "bootstrap" in "p_value method". The default value is 10.
plot	if TRUE, plot the p_values heatmap. The default value is TRUE.

Value

A list containing a matrix of p_values and, optionally, the plot.

Examples

```
# Calculate p_values of "Species" variable in iris dataset  
pvaluesccheb(iris,~Species, pvalue.method = "chisq", num.permutations = 100, num.bootstraps = 10)  
# Calculate p_values of "am" variable in mtcars dataset  
pvaluesccheb(mtcars,~am, pvalue.method = "chisq", num.permutations = 100, num.bootstraps = 10)
```

pvaluesceocl	<i>Calculate the p_values matrix for each species, using the Euclidean distance as a base.</i>
--------------	--

Description

This function takes a dataset, a factor, a p_value method, number of bootstraps and permutation when necessary, and returns a p_values matrix between each pair of species and a plot if the user select TRUE using Euclidean distance for the distances calculation.

Usage

```
pvaluesceocl(
  dataset,
  formula,
  pvalue.method = "chisq",
  num.permutations = 100,
  num.bootstraps = 10,
  plot = TRUE
)
```

Arguments

dataset	A dataframe.
formula	A factor which you want to calculate the Euclidean distances.
pvalue.method	A p_value method used to calculate the matrix, the default value is "chisq". Other methods are "permutation" and "bootstrap".
num.permutations	Number of permutation to specify if you select "permutation" in "pvalue.method". The default value is 100.
num.bootstraps	Number of bootstrap to specify if you select "bootstrap" in "p_value method". The default value is 10.
plot	if TRUE, plot the p_values heatmap. The default value is TRUE.

Value

A list containing the p_values matrix and, optionally, the plot. #' @examples # Calculate p_values of "Species" variable in iris dataset pvaluesceocl(iris,~Species, pvalue.method = "chisq", num.permutations = 100, num.bootstraps = 10) # Calculate p_values of "am" variable in mtcars dataset pvaluesceocl(mtcars,~am, pvalue.method = "chisq", num.permutations = 100, num.bootstraps = 10)

pvaluescmaha	<i>Calculate p_values matrix for each species, using Mahalanobis distance as a base.</i>
--------------	--

Description

This function takes a dataset, a factor, a p_value method, number of bootstraps and permutation when necessary, and returns a p_values matrix between each pair of the species and a plot if the user select TRUE using Mahalanobis distance for distances calculation.

Usage

```
pvaluescmaha(
  dataset,
  formula,
  pvalue.method = "chisq",
  num.permutations = 100,
  num.bootstraps = 10,
  plot = TRUE
)
```

Arguments

dataset	A dataframe.
formula	A factor which you want to calculate the Mahalanobis distances matrix.
pvalue.method	A p_value method used to calculate the matrix, the default value is "chisq". Other methods are "permutation" and "bootstrap".
num.permutations	Number of permutation to specify if you select "permutation" in "pvalue.method". The default value is 100.
num.bootstraps	Number of bootstrap to specify if you select "bootstrap" in "p_value method". The default value is 10.
plot	if TRUE, plot a p_values heatmap. The default value is TRUE.

Value

A list containing the p-values matrix and, optionally, the plot.

Examples

```
# Calculate p_values of "Species" variable in iris dataset
pvaluescmaha(iris, ~Species, pvalue.method = "chisq", num.permutations = 100, num.bootstraps = 10)
# Calculate p_values of "am" variable in mtcars dataset
pvaluescmaha(mtcars, ~am, pvalue.method = "chisq", num.permutations = 100, num.bootstraps = 10)
```

pvaluescmanh	<i>Calculate the p_values matrix for each species, using Manhattan distance as a base.</i>
--------------	--

Description

This function takes a dataset, a factor, a p_value method, number of bootstraps and permutation when necessary, and returns a p_values matrix between each pair of species and a plot if the user select TRUE using Manhattan distance for the distances calculation.

Usage

```
pvaluescmanh(  
  dataset,  
  formula,  
  pvalue.method = "chisq",  
  num.permutations = 100,  
  num.bootstraps = 10,  
  plot = TRUE  
)
```

Arguments

dataset	A dataframe
formula	A factor which you want to calculate Manhattan distances.
pvalue.method	A p_value method used to calculate the matrix, the default value is "chisq". Other methods are "permutation" and "bootstrap".
num.permutations	Number of permutation to specify if you select "permutation" in "pvalue.method". The default value is 100.
num.bootstraps	Number of bootstrap to specify if you select "bootstrap" in "p_value method". The default value is 10.
plot	if TRUE, plot the p_values heatmap. The default value is TRUE.

Value

A matrix containing a matrix of p_values and, optionally, the plot.

Examples

```
# Calculate p_values of "Species" variable in iris dataset  
pvaluescmanh(iris,~Species, pvalue.method = "chisq", num.permutations = 100, num.bootstraps = 10)  
# Calculate p_values of "am" variable in mtcars dataset  
pvaluescmanh(mtcars,~am, pvalue.method = "chisq", num.permutations = 100, num.bootstraps = 10)
```

Index

`cchebyshev`, 2

`ceuclide`, 3

`cmahalanobis`, 4

`cmanhattan`, 5

`generate_report_cchebyshev`, 6

`generate_report_ceuclide`, 7

`generate_report_cmahalanobis`, 8

`generate_report_cmanhattan`, 9

`pvaluesccheb`, 10

`pvaluesceucl`, 11

`pvaluescmaha`, 12

`pvaluescmanh`, 13