

Package: cmahalanobis (via r-universe)

October 9, 2026

Type Package

Title Calculate Distance Measures for DataFrames

Version 1.1.0

Date 2026-09-23

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

Description It provides functions that calculate Mahalanobis distance, Euclidean distance, Manhattan distance, Chebyshev distance, Hamming distance, Canberra distance, Minkowski dissimilarity (distance defined for $p \geq 1$), Cosine dissimilarity, Bhattacharyya dissimilarity, Jaccard distance, Hellinger distance, Bray-Curtis dissimilarity, Sorensen-Dice dissimilarity, Pearson, Kendall and Spearman dissimilarities and correlations (these last three on GitHub) between each pair of species in a list of data frames. These statistics 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

Imports stats, graphics, matrixStats

Suggests rmarkdown, testthat ($\geq 3.0.0$)

URL <https://github.com/Hflavio12/cmahalanobis>

NeedsCompilation no

Author Flavio Gioia [aut, cre] (ORCID:
<<https://orcid.org/0009-0000-0326-3840>>)

Config/roxygen2/version 8.0.0

Repository <https://hflavio12.r-universe.dev>

Date/Publication 2026-10-09 12:56:05 UTC

RemoteUrl <https://github.com/hflavio12/cmahanobis>

RemoteRef HEAD

RemoteSha bb4c8bda12f21b96baa658ad234f3b47653f983b

Contents

cbhattacharyya	3
cbraycurtis	4
ccanberra	6
cchebyshev	8
ccosine	10
ceuclyde	12
chamming	13
chellinger	15
cjaccard	17
ckendall	19
cmahanobis	20
cmanhattan	22
cminkowski	24
cpearson	26
csorensendice	28
cspearman	29
generate_report_cbhattacharyya	31
generate_report_cbraycurtis	33
generate_report_ccanberra	34
generate_report_cchebyshev	36
generate_report_ccosine	37
generate_report_ceuclyde	39
generate_report_chamming	41
generate_report_chellinger	42
generate_report_cjaccard	44
generate_report_cmahanobis	45
generate_report_cmanhattan	47
generate_report_cminkowski	48
generate_report_csorensendice	50
pvaluescbatt	52
pvaluescbrcu	53
pvaluesccanb	55
pvaluesccheb	56
pvaluesccosi	58
pvaluesceuclyde	59
pvalueschamm	61
pvalueschell	62
pvaluescjacc	64
pvaluescmaha	65
pvaluescmanh	67

pvaluescmink	68
pvaluescore	70

Index	72
--------------	-----------

cbhattacharyya	<i>Calculate the Bhattacharyya dissimilarities for each pair of factors or for the index.</i>
----------------	---

Description

This function takes a dataframe and a variable or variables (two or more) in input, and returns a matrix or matrices (two or more) with the Bhattacharyya dissimilarities about the factors inside them. You can also select "index" to calculate the Bhattacharyya dissimilarities between each row.

Usage

```
cbhattacharyya(  
  dataset,  
  formula,  
  plot = TRUE,  
  min_group_size = 3,  
  method = "average",  
  max_index_sample = NULL,  
  automatic_encoding = FALSE,  
  na_removal = FALSE,  
  grouping_stat = "mean"  
)
```

Arguments

dataset	A dataframe.
formula	The index of the dataframe, otherwise a variable or variables (two or more) with factors which you want to calculate the Bhattacharyya dissimilarities matrix or matrices (two or more).
plot	Logical, if TRUE, dendrograms with various agglomeration metrics for factors (two or more) are displayed. With "index" in formula, a dendrogram considering the observation is displayed.
min_group_size	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded. For "index", this value is always 1.
method	The agglomeration method for calculating dissimilarities between observations. Available methods are "ward.D", "ward.D2", "single", "complete", "average", "mcquitty", "median" or "centroid".
max_index_sample	A number of random samples from a dataset for which you want to calculate the dissimilarity for index mode, useful with very large dataset.

`automatic_encoding` Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).

`na_removal` Logical, if TRUE, missing value removal on rows is performed.

`grouping_stat` When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations).

Value

According to the option chosen in formula, with "index" the Bhattacharyya dissimilarities matrix will be printed; instead, by specifying variables, the Bhattacharyya dissimilarities matrix or matrices (two or more) between each pair of groups and, optionally, the plot or plots (two or more) will be printed.

Note

If "index" is selected with variables, only dissimilarities between rows are calculated. Therefore, this snippet: `"cbhattacharyya(mtcars, ~am + carb + index)"` will print dissimilarities only considering "index". Optionally, rows with NA values can be omitted and negative values are transformed in 0.

Examples

```
# Example with the CO2 dataset

table(CO2$Plant)

cbhattacharyya(CO2, ~Plant,
  plot = TRUE,
  grouping_stat = 'median',
  automatic_encoding = TRUE)

# Example with the airquality dataset

summary(airquality)

cbhattacharyya(airquality, ~index,
  plot = TRUE,
  na_removal = TRUE,
  max_index_sample = 40)
```

cbraycurtis

Calculate the Bray-Curtis dissimilarities for each pair of factors or for the index.

Description

This function takes a dataframe and a variable or variables (two or more) in input, and returns a matrix or matrices (two or more) with the Bray-Curtis dissimilarities about the factors inside them. You can also select "index" to calculate the Bray-Curtis dissimilarities between each row.

Usage

```
cbraycurtis(  
  dataset,  
  formula,  
  plot = TRUE,  
  min_group_size = 3,  
  method = "average",  
  max_index_sample = NULL,  
  automatic_encoding = FALSE,  
  na_removal = FALSE,  
  grouping_stat = "mean"  
)
```

Arguments

dataset	A dataframe.
formula	The index of the dataframe, otherwise a variable or variables (two or more) with factors which you want to calculate the Bray-Curtis dissimilarities matrix or matrices (two or more).
plot	Logical, if TRUE, dendrograms with various agglomeration metrics for factors (two or more) are displayed. With "index" in formula, a dendrogram considering the observation is displayed.
min_group_size	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded. For "index", this value is always 1.
method	The agglomeration method for calculating dissimilarities between observations. Available methods are "ward.D", "ward.D2", "single", "complete", "average", "mcquitty", "median" or "centroid".
max_index_sample	A number of random samples from a dataset for which you want to calculate the dissimilarity for index mode, useful with very large dataset.
automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
na_removal	Logical, if TRUE, missing value removal on rows is performed.
grouping_stat	When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations).

Value

According to the option chosen in formula, with "index" the Bray-Curtis dissimilarities matrix will be printed; instead, by specifying variables, the Bray-Curtis dissimilarities matrix or matrices (two or more) between each pair of factors and, optionally, the plot or plots (two or more) will be printed.

Note

If "index" is selected with variables, only dissimilarities between rows are calculated. Therefore, this snippet: "cbraycurtis(mtcars, ~am + carb + index)" will print dissimilarities only considering "index". Optionally, rows with NA values are omitted.

Examples

```
# Example with the CO2 dataset

table(CO2$Plant)

cbraycurtis(CO2, ~Plant,
            plot = TRUE,
            grouping_stat = 'mean',
            na_removal = TRUE,
            automatic_encoding = TRUE)

# Example with the airquality dataset

summary(airquality)

cbraycurtis(airquality, ~index,
            plot = TRUE,
            na_removal = TRUE,
            max_index_sample = 40)
```

ccanberra

Calculate the Canberra distances for each pair of factors or for the index.

Description

This function takes a dataframe and a variable or variables (two or more) in input, and returns a matrix or matrices (two or more) with the Canberra distances about the factors inside them. You can also select "index" to calculate the Canberra distances between each row.

Usage

```
ccanberra(
  dataset,
  formula,
```

```

    plot = TRUE,
    min_group_size = 3,
    method = "average",
    max_index_sample = NULL,
    automatic_encoding = FALSE,
    na_removal = FALSE,
    grouping_stat = "mean"
  )

```

Arguments

dataset	A dataframe.
formula	The index of the dataframe, otherwise a variable or variables (two or more) with factors which you want to calculate the Canberra distances matrix or matrices (two or more).
plot	Logical, if TRUE, dendrograms with various agglomeration metrics for factors (two or more) are displayed. With "index" in formula, a dendrogram considering the observation is displayed.
min_group_size	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded. For "index", this value is always 1.
method	The agglomeration method for calculating dissimilarities between observations. Available methods are "ward.D", "ward.D2", "single", "complete", "average", "mcquitty", "median" or "centroid".
max_index_sample	A number of random samples from a dataset for which you want to calculate the dissimilarity for index mode, useful with very large dataset.
automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
na_removal	Logical, if TRUE, missing value removal on rows is performed.
grouping_stat	When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations).

Value

According to the option chosen in formula, with "index" the Canberra distances matrix will be printed; instead, by specifying variables, the Canberra distances matrix or matrices (two or more) between each pair of groups and, optionally, the plot or plots (two or more) will be printed.

Note

If "index" is selected with variables, only distances between rows are calculated. Therefore, this snippet: "ccanberra(mtcars, ~am + carb + index)" will print distances only considering "index". Optionally, rows with NA values are omitted.

Examples

```
# Example with the CO2 dataset

table(CO2$Plant)

ccanberra(CO2, ~Plant,
  plot = TRUE,
  grouping_stat = 'median',
  automatic_encoding = TRUE)

# Example with the airquality dataset

summary(airquality)

ccanberra(airquality, ~index,
  plot = TRUE,
  na_removal = TRUE,
  max_index_sample = 40)
```

cchebyshev

Calculate the Chebyshev distances for each pair of factors or for the index.

Description

This function takes a dataframe and a variable or variables (two or more) in input, and returns a matrix or matrices (two or more) with the Chebyshev distances about the factors inside them. You can also select "index" to calculate the Chebyshev distances between each row.

Usage

```
cchebyshev(
  dataset,
  formula,
  plot = TRUE,
  min_group_size = 3,
  method = "average",
  max_index_sample = NULL,
  automatic_encoding = FALSE,
  na_removal = FALSE,
  grouping_stat = "mean"
)
```

Arguments

dataset A dataframe.

formula	The index of the dataframe, otherwise a variable or variables (two or more) with factors which you want to calculate the Chebyshev distances matrix or matrices (two or more).
plot	Logical, if TRUE, dendrograms with various agglomeration metrics for factors (two or more) are displayed. With "index" in formula, a dendrogram considering the observation is displayed.
min_group_size	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded. For "index", this value is always 1.
method	The agglomeration method for calculating dissimilarities between observations. Available methods are "ward.D", "ward.D2", "single", "complete", "average", "mcquitty", "median" or "centroid".
max_index_sample	A number of random samples from a dataset for which you want to calculate the dissimilarity for index mode, useful with very large dataset.
automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
na_removal	Logical, if TRUE, missing value removal on rows is performed.
grouping_stat	When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations).

Value

According to the option chosen in formula, with "index" the Chebyshev distances matrix will be printed; instead, by specifying variables, the Chebyshev distances matrix or matrices (two or more) between each pair of groups and, optionally, the plot or plots (two or more) will be printed.

Note

If "index" is selected with variables, only distances between rows are calculated. Therefore, this snippet: "cchebyshev(mtcars, ~am + carb + index)" will print distances only considering "index". Optionally, rows with NA values are omitted.

Examples

```
# Example with the CO2 dataset

table(CO2$Plant)

cchebyshev(CO2, ~Plant,
            plot = TRUE,
            grouping_stat = 'mean',
            na_removal = TRUE,
            automatic_encoding = TRUE)

# Example with the airquality dataset
```

```
summary(airquality)

cchebyshev(airquality, ~index,
  plot = TRUE,
  na_removal = TRUE,
  max_index_sample = 40)
```

ccosine	<i>Calculate the Cosine dissimilarities for each pair of factors or for the index.</i>
---------	--

Description

This function takes a dataframe and a variable or variables (two or more) in input, and returns a matrix or matrices (two or more) with the Cosine dissimilarity about the factors inside them. You can also select "index" to calculate the Cosine distances between each row.

Usage

```
ccosine(
  dataset,
  formula,
  plot = TRUE,
  min_group_size = 3,
  method = "average",
  max_index_sample = NULL,
  automatic_encoding = FALSE,
  na_removal = FALSE,
  grouping_stat = "mean"
)
```

Arguments

dataset	A dataframe.
formula	The index of the dataframe, otherwise a variable or variables (two or more) with factors which you want to calculate the Cosine dissimilarity matrix or matrices (two or more).
plot	Logical, if TRUE, dendrograms with various agglomeration metrics for factors (two or more) are displayed. With "index" in formula, a dendrogram considering the observation is displayed.
min_group_size	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded. For "index", this value is always 1.

method	The agglomeration method for calculating dissimilarities between observations. Available methods are "ward.D", "ward.D2", "single", "complete", "average", "mcquitty", "median" or "centroid".
max_index_sample	A number of random samples from a dataset for which you want to calculate the dissimilarity for index mode, useful with very large dataset.
automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
na_removal	Logical, if TRUE, missing value removal on rows is performed.
grouping_stat	When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations).

Value

According to the option chosen in formula, with "index" the Cosine dissimilarity matrix will be printed; instead, by specifying variables, the Cosine dissimilarities matrix or matrices (two or more) between each pair of groups and, optionally, the plot or plots (two or more) will be printed.

Note

If "index" is selected with variables, only distances between rows are calculated. Therefore, this snippet: "ccosine(mtcars, ~am + carb + index)" will print distances only considering "index". Optionally, rows with NA values are omitted.

Examples

```
# Example with the CO2 dataset

table(CO2$Plant)

ccosine(CO2, ~Plant,
  plot = TRUE,
  grouping_stat = 'median',
  automatic_encoding = TRUE)

# Example with the airquality dataset

summary(airquality)

ccosine(airquality, ~index,
  plot = TRUE,
  na_removal = TRUE,
  max_index_sample = 40)
```

ceulide	<i>Calculate the Euclidean distances for each pair of factors or for the index.</i>
---------	---

Description

This function takes a dataframe and a variable or variables (two or more) in input, and returns a matrix or matrices (two or more) with the Euclidean distances about each pair of factors inside them. You can also select "index" to calculate the Euclidean distances between each row.

Usage

```
ceulide(
  dataset,
  formula,
  plot = TRUE,
  min_group_size = 3,
  method = "average",
  max_index_sample = NULL,
  automatic_encoding = FALSE,
  na_removal = FALSE,
  grouping_stat = "mean"
)
```

Arguments

dataset	A dataframe.
formula	The index of the dataframe, otherwise a variable or variables (two or more) with factors which you want to calculate the Euclidean distances matrix or matrices (two or more).
plot	Logical, if TRUE, dendrograms with various agglomeration metrics for factors (two or more) are displayed. With "index" in formula, a dendrogram considering the observation is displayed.
min_group_size	Minimum group size to maintain. The default value is 3, therefore factors, inside variables, with less than 3 observations will be discarded. For "index", this value is always 1.
method	The agglomeration method for calculating dissimilarities between observations. Available methods are "ward.D", "ward.D2", "single", "complete", "average", "mcquitty", "median" or "centroid".
max_index_sample	A number of random samples from a dataset for which you want to calculate the dissimilarity for index mode, useful with very large dataset.
automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).

na_removal	Logical, if TRUE, missing value removal on rows is performed.
grouping_stat	When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations).

Value

According to the option chosen in formula, with "index" the Euclidean distance matrix will be printed; instead, by specifying variables, the Euclidean distances matrix or matrices (two or more) between each pair of groups and, optionally, the plot or plots (two or more) will be printed.

Note

If "index" is selected with variables, only distances between rows are calculated. Therefore, this snippet: "ceucide(mtcars, ~am + carb + index)" will print distances only considering "index". Optionally, rows with missing values are omitted.

Examples

```
# Example with the CO2 dataset

table(CO2$Plant)

ceucide(CO2, ~Plant,
        plot = TRUE,
        grouping_stat = 'mean',
        na_removal = TRUE,
        automatic_encoding = TRUE)

# Example with the airquality dataset

summary(airquality)

ceucide(airquality, ~index,
        plot = TRUE,
        na_removal = TRUE,
        max_index_sample = 40)
```

chamming	<i>Calculate the Hamming distances for each pair of factors or for the index.</i>
----------	---

Description

This function takes a dataframe and a variable or variables (two or more) in input, and returns a matrix or matrices (two or more) with the Hamming distances about the factors inside them. You can also select "index" to calculate the Hamming distances between each row.

Usage

```
chamming(
  dataset,
  formula,
  plot = TRUE,
  min_group_size = 3,
  method = "average",
  max_index_sample = NULL,
  automatic_encoding = FALSE,
  na_removal = FALSE,
  grouping_stat = "mean"
)
```

Arguments

dataset	A dataframe.
formula	The index of the dataframe, otherwise a variable or variables (two or more) with factors which you want to calculate the Hamming distances matrix or matrices (two or more).
plot	Logical, if TRUE, dendrograms with "median" and "centroid" agglomeration metrics for factors (two or more) are displayed. With "index" in formula, a dendrogram considering the observation is displayed.
min_group_size	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded. For "index", this value is always 1.
method	The agglomeration method for calculating dissimilarities between observations. Available methods are "ward.D", "ward.D2", "single", "complete", "average", "mcquitty", "median" or "centroid".
max_index_sample	A number of random samples from a dataset for which you want to calculate the dissimilarity for index mode, useful with very large dataset.
automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
na_removal	Logical, if TRUE, missing value removal on rows is performed.
grouping_stat	When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations).

Value

According to the option chosen in formula, with "index" the Hamming distances matrix will be printed; instead, by specifying variables, the Hamming distances matrix or matrices (two or more) between each pair of factors and, optionally, the plot or plots (two or more) will be printed.

Note

If "index" is selected with variables, only distances between rows are calculated. Therefore, this snippet: "chamming(mtcars, ~am + carb + index)" will print the distances only considering "index". Optionally, rows with NA values are omitted.

Examples

```
# Example with the CO2 dataset

table(CO2$Plant)

chamming(CO2, ~Plant,
  plot = TRUE,
  grouping_stat = 'median',
  automatic_encoding = TRUE)

# Example with the airquality dataset

summary(airquality)

chamming(airquality, ~index,
  plot = TRUE,
  na_removal = TRUE,
  max_index_sample = 40)
```

chellinger

Calculate the Hellinger distances for each pair of factors or for the index.

Description

This function takes a dataframe and a variable or variables (two or more) in input, and returns a matrix or matrices (two or more) with the Hellinger distances about the factors inside them. You can also select "index" to calculate the Hellinger distances between each row.

Usage

```
chellinger(
  dataset,
  formula,
  plot = TRUE,
  min_group_size = 3,
  method = "average",
  max_index_sample = NULL,
  automatic_encoding = FALSE,
  na_removal = FALSE,
  grouping_stat = "mean"
)
```

Arguments

dataset	A dataframe.
formula	The index of the dataframe, otherwise a variable or variables (two or more) with factors which you want to calculate the Hellinger distances matrix or matrices (two or more).
plot	Logical, if TRUE, dendrograms with various agglomeration metrics for factors (two or more) are displayed. With "index" in formula, a dendrogram considering the observation is displayed.
min_group_size	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded. For "index", this value is always 1.
method	The agglomeration method for calculating dissimilarities between observations. Available methods are "ward.D", "ward.D2", "single", "complete", "average", "mcquitty", "median" or "centroid".
max_index_sample	A number of random samples from a dataset for which you want to calculate the distance for index mode, useful with very large dataset.
automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
na_removal	Logical, if TRUE, missing value removal on rows is performed.
grouping_stat	When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations).

Value

According to the option chosen in formula, with "index" the Hellinger distances matrix will be printed; instead, by specifying variables, the Hellinger distances matrix or matrices (two or more) between each pair of groups and, optionally, the plot or plots (two or more) will be printed.

Note

If "index" is selected with variables, only distances between rows are calculated. Therefore, this snippet: "chellinger(mtcars, ~am + carb + index)" will print the distances only considering "index". Optionally, rows with NA values are omitted.

Examples

```
# Example with the CO2 dataset

table(CO2$Plant)

chellinger(CO2, ~Plant,
  plot = TRUE,
  grouping_stat = 'mean',
  na_removal = TRUE,
```

```

        automatic_encoding = TRUE)

# Example with the airquality dataset

summary(airquality)

chellinger(airquality, ~index,
  plot = TRUE,
  na_removal = TRUE,
  max_index_sample = 40)

```

cjaccard	<i>Calculate the Jaccard distances for each pair of factors or for the index.</i>
----------	---

Description

This function takes a dataframe and a variable or variables (two or more) in input, and returns a matrix or matrices (two or more) with the Jaccard distance about the factors inside them. You can also select "index" to calculate the Jaccard distances between each row.

Usage

```

cjaccard(
  dataset,
  formula,
  plot = TRUE,
  min_group_size = 3,
  method = "average",
  max_index_sample = NULL,
  automatic_encoding = FALSE,
  na_removal = FALSE,
  grouping_stat = "mean"
)

```

Arguments

dataset	A dataframe.
formula	The index of the dataframe, otherwise a variable or variables (two or more) with factors which you want to calculate the Jaccard distances matrix or matrices (two or more).
plot	Logical, if TRUE, dendrograms with various agglomeration metrics for factors (two or more) are displayed. With "index" in formula, a dendrogram considering the observation is displayed.
min_group_size	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded. For "index", this value is always 1.

method	The agglomeration method for calculating distances between observations. Available methods are "median" or "centroid".
max_index_sample	A number of random samples from a dataset for which you want to calculate the distance for index mode, useful with very large dataset.
automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
na_removal	Logical, if TRUE, missing value removal on rows is performed.
grouping_stat	When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations).

Value

According to the option chosen in formula, with "index" the Jaccard distances matrix will be printed; instead, by specifying variables, the Jaccard distances matrix or matrices (two or more) between each pair of groups and, optionally, the plot or plots (two or more) will be printed.

Note

If "index" is selected with variables, only dissimilarities between rows are calculated. Therefore, this snippet: "cjaccard(mtcars, ~am + carb + index)" will print dissimilarities only considering "index". Optionally, rows with NA values are omitted.

Examples

```
# Example with the CO2 dataset

table(CO2$Plant)

cjaccard(CO2, ~Plant,
  plot = TRUE,
  grouping_stat = 'mean',
  na_removal = TRUE,
  automatic_encoding = TRUE)

# Example with the airquality dataset

summary(airquality)

cjaccard(airquality, ~index,
  plot = TRUE,
  na_removal = TRUE,
  max_index_sample = 40)
```

ckendall	<i>Calculate the Kendall correlation or dissimilarity for each pair of factors or for the index.</i>
----------	--

Description

This function takes a dataframe and a variable or variables (two or more) in input, and returns a matrix or matrices (two or more) with the Kendall correlation or dissimilarity about the factors inside them. You can also select "index" to calculate this statistic between each row.

Usage

```
ckendall(
  dataset,
  formula,
  plot = TRUE,
  min_group_size = 3,
  method = "average",
  max_index_sample = NULL,
  automatic_encoding = FALSE,
  na_removal = FALSE,
  grouping_stat = TRUE,
  dissimilarity = TRUE,
  z_test = TRUE
)
```

Arguments

dataset	A dataframe.
formula	The index of the dataframe, otherwise a variable or variables (two or more) with factors which you want to calculate the Kendall correlation or dissimilarity matrix or matrices (two or more).
plot	Logical, if TRUE, dendrograms with various agglomeration metrics for factors (two or more) are displayed. With "index" in formula, a dendrogram considering the observation is displayed.
min_group_size	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded. For "index", this value is always 1.
method	The agglomeration method for calculating dissimilarities between observations. Available methods are "ward.D", "ward.D2", "single", "complete", "average", "mcquitty", "median" or "centroid".
max_index_sample	A number of random samples from a dataset for which you want to calculate the statistic for index mode, useful with very large dataset.

automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
na_removal	Logical, if TRUE, missing value removal on rows is performed.
grouping_stat	Logical, if TRUE, the Kendall correlation or dissimilarity between factors is printed.
dissimilarity	If TRUE, default, the dissimilarity is displayed.
z_test	If TRUE, statistical significance using the standard normal distribution is performed.

Value

According to the option chosen in formula, with "index" the Kendall matrix will be printed; instead, by specifying variables, the Kendall correlations matrix or matrices (two or more) between each pair of groups and, optionally, the plot or plots (two or more) will be printed.

Note

If "index" is selected with variables, only correlations between rows are calculated. Therefore, this snippet: "ckendall(mtcars, ~am + carb + index)" will print correlations only considering "index". Optionally, rows with NA values can be omitted and negative values are transformed in 0.

Examples

```
# Example with the CO2 dataset

table(CO2$Plant)

ckendall(CO2, ~Plant,
  plot = TRUE,
  grouping_stat = 'median',
  automatic_encoding = TRUE)

# Example with the airquality dataset

summary(airquality)

ckendall(airquality, ~index,
  plot = TRUE,
  na_removal = TRUE,
  max_index_sample = 40)
```

cmahalanobis

Calculate the Mahalanobis distances for each pair of factors or for the index.

Description

This function takes a dataframe and a variable or variables (two or more) in input, and returns a matrix or matrices (two or more) with the Mahalanobis distances about each pair of factors inside them. You can also select "index" to calculate the Mahalanobis distances between each row.

Usage

```
cmahalanobis(
  dataset,
  formula,
  plot = TRUE,
  min_group_size = 3,
  method = "average",
  pvalues_chisq = TRUE,
  max_index_sample = NULL,
  automatic_encoding = FALSE,
  na_removal = FALSE,
  grouping_stat = "mean"
)
```

Arguments

dataset	A dataframe.
formula	The index of the dataframe, otherwise a variable or variables (two or more) with factors which you want to calculate the Mahalanobis distances matrix or matrices (two or more).
plot	Logical, if TRUE, dendrograms with various agglomeration metrics for factors (two or more) are displayed. With "index" in formula, a dendrogram considering the observation is displayed.
min_group_size	Minimum group size to maintain. The default value is 3, therefore factors, inside variables, with less than 3 observations will be discarded. For "index", this value is always 1.
method	The agglomeration method for calculating dissimilarities between observations. Available methods are "ward.D", "ward.D2", "single", "complete", "average", "mcquitty", "median" or "centroid".
pvalues_chisq	If TRUE, print the result of the chi-squared test on squared distances. The distances with "pvalues_chisq = FALSE" are not squared; instead, with "pvalues_chisq = TRUE", the squared Mahalanobis distances with corresponding p_values will be printed. Default is FALSE.
max_index_sample	A number of random samples from a dataset for which you want to calculate the dissimilarity for index mode, useful with very large dataset.
automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
na_removal	Logical, if TRUE, missing value removal on rows is performed.

grouping_stat When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations).

Value

According to the option chosen in formula and in `pvalues_chisq`, with `"index"` and `"pvalues_chisq = TRUE"` the squared Mahalanobis distance matrix will be printed with corresponding pvalues; instead, with `"index"` and `"pvalues_chisq = FALSE"`, only the Mahalanobis distances (not squared) will be printed. By specifying variables, the Mahalanobis distances matrix or matrices (two or more) between each pair of factors and, optionally, the plot or plots (two or more) will be printed.

Note

If `"index"` is selected with variables, only distances between rows are calculated. Therefore, this snippet: `"cmanhattan(mtcars, ~am + carb + index)"` will print distances and plot only considering `"index"`. Optionally, rows with NA values are omitted.

Examples

```
# Example with the CO2 dataset

table(CO2$Plant)

cmanhattan(CO2, ~Plant, pvalues_chisq = TRUE,
            plot = TRUE,
            grouping_stat = 'mean',
            na_removal = TRUE,
            automatic_encoding = TRUE)

# Example with the airquality dataset

summary(airquality)

cmanhattan(airquality, ~index,
            plot = TRUE,
            na_removal = TRUE,
            max_index_sample = 40)
```

cmanhattan

Calculate the Manhattan distances for each pair of factors or for the index.

Description

This function takes a dataframe and a variable or variables (two or more) in input, and returns a matrix or matrices (two or more) with the Manhattan distances about the factors inside them. You can also select `"index"` to calculate the Manhattan distances between each row.

Usage

```
cmanhattan(
  dataset,
  formula,
  plot = TRUE,
  min_group_size = 3,
  method = "average",
  max_index_sample = NULL,
  automatic_encoding = FALSE,
  na_removal = FALSE,
  grouping_stat = "mean"
)
```

Arguments

dataset	A dataframe.
formula	The index of the dataframe, otherwise a variable or variables (two or more) with factors which you want to calculate the Manhattan distances matrix or matrices (two or more).
plot	Logical, if TRUE, dendrograms with various agglomeration metrics for factors (two or more) are displayed. With "index" in formula, a dendrogram considering the observation is displayed.
min_group_size	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded. For "index", this value is always 1.
method	The agglomeration method for calculating dissimilarities between observations. Available methods are "ward.D", "ward.D2", "single", "complete", "average", "mcquitty", "median" or "centroid".
max_index_sample	A number of random samples from a dataset for which you want to calculate the dissimilarity for index mode, useful with very large dataset.
automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
na_removal	Logical, if TRUE, missing value removal on rows is performed.
grouping_stat	When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations).

Value

According to the option chosen in formula, with "index" the Manhattan distances matrix will be printed; instead, by specifying variables, the Manhattan distances matrix or matrices (two or more) between each pair of groups and, optionally, the plot or plots (two or more) will be printed.

Note

If "index" is selected with variables, only distances between rows are calculated. Therefore, this snippet: "cmanhattan(mtcars, ~am + carb + index)" will print the distances only considering "index". Optionally, rows with NA values are omitted.

Examples

```
# Example with the CO2 dataset

table(CO2$Plant)

cmanhattan(CO2, ~Plant,
  plot = TRUE,
  grouping_stat = 'median',
  automatic_encoding = TRUE)

# Example with the airquality dataset

summary(airquality)

cmanhattan(airquality, ~index,
  plot = TRUE,
  na_removal = TRUE,
  max_index_sample = 40)
```

cminkowski	<i>Calculate the Minkowski distances for each pair of factors or for the index.</i>
------------	---

Description

This function takes a dataframe and a variable or variables (two or more) in input, and returns a matrix or matrices (two or more) with the Minkowski distances about the factors inside them. You can also select "index" to calculate the Minkowski distances between each row.

Usage

```
cminkowski(
  dataset,
  formula,
  p = 3,
  plot = TRUE,
  min_group_size = 3,
  method = "average",
  max_index_sample = NULL,
  automatic_encoding = FALSE,
  na_removal = FALSE,
  grouping_stat = "mean"
)
```

Arguments

dataset	A dataframe.
formula	The index of the dataframe, otherwise a variable or variables (two or more) with factors which you want to calculate the Minkowski distances matrix or matrices (two or more).
p	Order of the Minkowski distance.
plot	Logical, if TRUE, dendrograms with various agglomeration metrics for factors (two or more) are displayed. With "index" in formula, a dendrogram considering the observation is displayed.
min_group_size	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded. For "index", this value is always 1.
method	The agglomeration method for calculating dissimilarities between observations. Available methods are "ward.D", "ward.D2", "single", "complete", "average", "mcquitty", "median" or "centroid".
max_index_sample	A number of random samples from a dataset for which you want to calculate the dissimilarity for index mode, useful with very large dataset.
automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
na_removal	Logical, if TRUE, missing value removal on rows is performed.
grouping_stat	When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations).

Value

According to the option chosen in formula, with "index" the Minkowski distances matrix will be printed; instead, by specifying variables, the Minkowski distances matrix or matrices (two or more) between each pair of groups and, optionally, the plot or plots (two or more) will be printed.

Note

When $p < 1$, the Minkowski distance is a "dissimilarity" measure. When $p \geq 1$, the triangle inequality property is satisfied and we say "Minkowski distance". If "index" is selected with variables, only distances between rows are calculated. Therefore, this snippet: `"cminkowski(mtcars, ~am + carb + index)"` will print distances only considering "index". Optionally, rows with NA values are omitted.

Examples

```
# Example with the CO2 dataset

table(CO2$Plant)
```

```
cminkowski(CO2, ~Plant, p = 4,
            plot = TRUE,
            grouping_stat = 'mean',
            na_removal = TRUE,
            automatic_encoding = TRUE)

# Example with the airquality dataset

summary(airquality)

cminkowski(airquality, ~index, p = 3,
            plot = TRUE,
            na_removal = TRUE,
            max_index_sample = 40)
```

cpearson	<i>Calculate the Pearson correlation or dissimilarity for each pair of factors or for the index.</i>
----------	--

Description

This function takes a dataframe and a variable or variables (two or more) in input, and returns a matrix or matrices (two or more) with the Pearson correlation or dissimilarity about the factors inside them. You can also select "index" to calculate this statistic between each row.

Usage

```
cpearson(
  dataset,
  formula,
  plot = TRUE,
  min_group_size = 3,
  method = "average",
  max_index_sample = NULL,
  automatic_encoding = FALSE,
  na_removal = FALSE,
  grouping_stat = TRUE,
  dissimilarity = TRUE,
  t_test = TRUE
)
```

Arguments

dataset	A dataframe.
formula	The index of the dataframe, otherwise a variable or variables (two or more) with factors which you want to calculate the Pearson correlation or dissimilarity matrix or matrices (two or more).

plot	Logical, if TRUE, dendrograms with various agglomeration metrics for factors (two or more) are displayed. With "index" in formula, a dendrogram considering the observation is displayed.
min_group_size	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded. For "index", this value is always 1.
method	The agglomeration method for calculating dissimilarities between observations. Available methods are "ward.D", "ward.D2", "single", "complete", "average", "mcquitty", "median" or "centroid".
max_index_sample	A number of random samples from a dataset for which you want to calculate the statistic for index mode, useful with very large dataset.
automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
na_removal	Logical, if TRUE, missing value removal on rows is performed.
grouping_stat	Logical, if TRUE, the Pearson correlation or dissimilarity between factors is printed.
dissimilarity	If TRUE, default, the dissimilarity Pearson's R is displayed.
t_test	If TRUE, statistical significance using the t-student distribution is performed.

Value

According to the option chosen in formula, with "index" the Pearson' R matrix will be printed; instead, by specifying variables, the Pearson correlations matrix or matrices (two or more) between each pair of groups and, optionally, the plot or plots (two or more) will be printed.

Note

If "index" is selected with variables, only correlations between rows are calculated. Therefore, this snippet: "cpearson(mtcars, ~am + carb + index)" will print correlations only considering "index". Optionally, rows with NA values can be omitted and negative values are transformed in 0.

Examples

```
# Example with the CO2 dataset

table(CO2$Plant)

cpearson(CO2, ~Plant,
  plot = TRUE,
  grouping_stat = 'median',
  automatic_encoding = TRUE)

# Example with the airquality dataset

summary(airquality)
```

```
cpearson(airquality, ~index,
  plot = TRUE,
  na_removal = TRUE,
  max_index_sample = 40)
```

csorensendice	<i>Calculate the Sorensen-Dice dissimilarities for each pair of factors or for the index.</i>
---------------	---

Description

This function takes a dataframe and a variable or variables (two or more) in input, and returns a matrix or matrices (two or more) with the Sorensen-Dice dissimilarities about the factors inside them. You can also select "index" to calculate the Sorensen-Dice dissimilarities between each row.

Usage

```
csorensendice(
  dataset,
  formula,
  plot = TRUE,
  min_group_size = 3,
  method = "average",
  max_index_sample = NULL,
  automatic_encoding = FALSE,
  na_removal = FALSE,
  grouping_stat = "mean"
)
```

Arguments

dataset	A dataframe.
formula	The index of the dataframe, otherwise a variable or variables (two or more) with factors which you want to calculate the Sorensen-Dice dissimilarities matrix or matrices (two or more).
plot	Logical, if TRUE, dendrograms with various agglomeration metrics for factors (two or more) are displayed. With "index" in formula, a dendrogram considering the observation is displayed.
min_group_size	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded. For "index", this value is always 1.
method	The agglomeration method for calculating dissimilarities between observations. Available methods are "median" or "centroid".
max_index_sample	A number of random samples from a dataset for which you want to calculate the dissimilarity for index mode, useful with very large dataset.

automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
na_removal	Logical, if TRUE, missing value removal on rows is performed.
grouping_stat	When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations).

Value

According to the option chosen in formula, with "index" the Sorensen-Dice dissimilarities matrix will be printed; instead, by specifying variables, the Sorensen-Dice dissimilarities matrix or matrices (two or more) between each pair of groups and, optionally, the plot or plots (two or more) will be printed.

Note

If "index" is selected with variables, only dissimilarities between rows are calculated. Therefore, this snippet: "csorensendice(mtcars, ~am + carb + index)" will print dissimilarities only considering "index". Optionally, rows with NA values are omitted.

Examples

```
# Example with the CO2 dataset

table(CO2$Plant)

csorensendice(CO2, ~Plant,
  plot = TRUE,
  grouping_stat = 'median',
  automatic_encoding = TRUE)

# Example with the airquality dataset

summary(airquality)

csorensendice(airquality, ~index,
  plot = TRUE,
  na_removal = TRUE,
  max_index_sample = 40)
```

Description

This function takes a dataframe and a variable or variables (two or more) in input, and returns a matrix or matrices (two or more) with the Spearman correlation or dissimilarity about the factors inside them. You can also select "index" to calculate this statistic between each row.

Usage

```
cspearman(
  dataset,
  formula,
  plot = TRUE,
  min_group_size = 3,
  method = "average",
  max_index_sample = NULL,
  automatic_encoding = FALSE,
  na_removal = FALSE,
  grouping_stat = TRUE,
  dissimilarity = TRUE,
  t_test = TRUE
)
```

Arguments

dataset	A dataframe.
formula	The index of the dataframe, otherwise a variable or variables (two or more) with factors which you want to calculate the Spearman correlation or dissimilarity matrix or matrices (two or more).
plot	Logical, if TRUE, dendrograms with various agglomeration metrics for factors (two or more) are displayed. With "index" in formula, a dendrogram considering the observation is displayed.
min_group_size	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded. For "index", this value is always 1.
method	The agglomeration method for calculating dissimilarities between observations. Available methods are "ward.D", "ward.D2", "single", "complete", "average", "mcquitty", "median" or "centroid".
max_index_sample	A number of random samples from a dataset for which you want to calculate the statistic for index mode, useful with very large dataset.
automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
na_removal	Logical, if TRUE, missing value removal on rows is performed.
grouping_stat	Logical, if TRUE, the Spearman correlation or dissimilarity between factors is printed.
dissimilarity	If TRUE, default, the dissimilarity is displayed.
t_test	If TRUE, statistical significance using the t-student distribution is performed.

Value

According to the option chosen in formula, with "index" the Spearman matrix will be printed; instead, by specifying variables, the Spearman correlations matrix or matrices (two or more) between each pair of groups and, optionally, the plot or plots (two or more) will be printed.

Note

If "index" is selected with variables, only correlations between rows are calculated. Therefore, this snippet: "cspearman(mtcars, ~am + carb + index)" will print correlations only considering "index". Optionally, rows with NA values can be omitted and negative values are transformed in 0.

Examples

```
# Example with the CO2 dataset

table(CO2$Plant)

cspearman(CO2, ~Plant,
  plot = TRUE,
  grouping_stat = 'median',
  automatic_encoding = TRUE)

# Example with the airquality dataset

summary(airquality)

cspearman(airquality, ~index,
  plot = TRUE,
  na_removal = TRUE,
  max_index_sample = 40)
```

generate_report_cbhattacharyya

Generate a Microsoft Word document about the Bhattacharyya dissimilarities matrix or matrices (two or more) and the p-values matrix or matrices (two or more).

Description

This function takes a dataframe, a factor or factors (two or more) and returns a Microsoft Word document about the Bhattacharyya dissimilarities matrix or matrices (two or more) and the p-values matrix or matrices (two or more).

Usage

```
generate_report_cbhattacharyya(
  dataset,
  formula,
```

```

    pvalue.method = "permutation",
    seed = NULL,
    min_group_size = 3,
    num_replicas = 1000,
    automatic_encoding = TRUE,
    na_removal = TRUE,
    grouping_stat = "mean"
  )

```

Arguments

dataset	A dataframe.
formula	A variable or variables (two or more) with factors which you want to calculate the Bhattacharyya dissimilarities matrix or matrices (two or more) and the p_values matrix or matrices (two or more).
pvalue.method	A p_value method used to calculate the matrix or matrices (two or more), the default value is "permutation". Another method is "bootstrap".
seed	Optionally, set a seed for "bootstrap" or "permutation".
min_group_size	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded.
num_replicas	Number of permutations or bootstraps trials, the default value is 1000.
automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
na_removal	Logical, if TRUE, missing value removal on rows is performed.
grouping_stat	When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations).

Value

A Microsoft Word document about the Bhattacharyya dissimilarities matrix or matrices (two or more) and the p_values matrix or matrices (two or more).

Note

About pvalues, this function leverages on an early stopping procedure in which the resulting matrix is printed also if the specified number of replicas is not reached; if every 500 replicas the maximum difference between each p_value does not exceed 0.0001, the function will print the entire matrix, else it continues.

Examples

```

generate_report_cbhattacharyya(CO2, ~Plant + Type,
  pvalue.method = "permutation",
  seed = 122,
  num_replicas = 10,

```

```

grouping_stat = 'median',
automatic_encoding = TRUE)

generate_report_cbhattacharyya(airquality, ~Ozone,
  pvalue.method = 'bootstrap',
  na_removal = TRUE, num_replicas = 10)

```

```
generate_report_cbraycurtis
```

Generate a Microsoft Word document about the Bray-Curtis dissimilarities matrix or matrices (two or more) and the p-values matrix or matrices (two or more).

Description

This function takes a dataframe, a factor or factors (two or more) and returns a Microsoft Word document about the Bray-Curtis dissimilarities matrix or matrices (two or more) and the p-values matrix or matrices (two or more).

Usage

```

generate_report_cbraycurtis(
  dataset,
  formula,
  pvalue.method = "permutation",
  seed = NULL,
  min_group_size = 3,
  num_replicas = 1000,
  automatic_encoding = TRUE,
  na_removal = TRUE,
  grouping_stat = "mean"
)

```

Arguments

dataset	A dataframe.
formula	A variable or variables (two or more) with factors which you want to calculate the Bray-Curtis dissimilarities matrix or matrices (two or more) and the p_values matrix or matrices (two or more).
pvalue.method	A p_value method used to calculate the matrix or matrices (two or more), the default value is "permutation". Another method is "bootstrap".
seed	Optionally, set a seed for 'bootstrap' or 'permutation'.
min_group_size	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded.
num_replicas	Number of permutations or bootstraps trials, the default value is 1000.

automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
na_removal	Logical, if TRUE, missing value removal on rows is performed.
grouping_stat	When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations).

Value

A Microsoft Word document about the Bray-Curtis dissimilarities matrix or matrices (two or more) and the p_values matrix or matrices (two or more).

Note

About pvalues, this function leverages on an early stopping procedure in which the resulting matrix is printed also if the specified number of replicas is not reached; if every 500 replicas the maximum difference between each p_value does not exceed 0.0001, the function will print the entire matrix, else it continues.

Examples

```
generate_report_cbraycurtis(CO2, ~Plant + Type,
  pvalue.method = "permutation",
  seed = 122,
  num_replicas = 10,
  grouping_stat = 'median',
  automatic_encoding = TRUE)
```

```
generate_report_cbraycurtis(airquality, ~Ozone,
  pvalue.method = 'bootstrap',
  na_removal = TRUE, num_replicas = 10)
```

```
generate_report_ccanberra
```

Generate a Microsoft Word document about the Canberra distances matrix or matrices (two or more) and the p-values matrix or matrices (two or more).

Description

This function takes a dataframe, a factor or factors (two or more) and returns a Microsoft Word document about the Canberra distances matrix or matrices (two or more) and the p-values matrix or matrices (two or more).

Usage

```
generate_report_ccanberra(
  dataset,
  formula,
  pvalue.method = "permutation",
  seed = NULL,
  min_group_size = 3,
  num_replicas = 1000,
  automatic_encoding = TRUE,
  na_removal = TRUE,
  grouping_stat = "mean"
)
```

Arguments

dataset	A dataframe.
formula	A variable or variables (two or more) with factors which you want to calculate the Canberra distances matrix or matrices (two or more) and the p_values matrix or matrices (two or more).
pvalue.method	A p_value method used to calculate the matrix or matrices (two or more), the default value is "permutation". Another method is "bootstrap".
seed	Optionally, set a seed for "bootstrap" and "permutation".
min_group_size	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded.
num_replicas	Number of permutations or bootstraps trials, the default value is 1000.
automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
na_removal	Logical, if TRUE, missing value removal on rows is performed.
grouping_stat	When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations).

Value

A Microsoft Word document about the Canberra distances matrix or matrices (two or more) and the p_values matrix or matrices (two or more).

Note

About pvalues, this function leverages on an early stopping procedure in which the resulting matrix is printed also if the specified number of replicas is not reached; if every 500 replicas the maximum difference between each p_value does not exceed 0.0001, the function will print the entire matrix, else it continues.

Examples

```
generate_report_ccanberra(CO2, ~Plant + Type,
  pvalue.method = "permutation",
  seed = 122,
  num_replicas = 10,
  grouping_stat = 'median',
  automatic_encoding = TRUE)

generate_report_ccanberra(airquality, ~Ozone,
  pvalue.method = 'bootstrap',
  na_removal = TRUE, num_replicas = 10)
```

```
generate_report_cchebyshev
```

Generate a Microsoft Word document about the Chebyshev distances matrix or matrices (two or more) and the p-values matrix or matrices (two or more).

Description

This function takes a dataframe, a factor or factors (two or more) and returns a Microsoft Word document about the Chebyshev distances matrix or matrices (two or more) and the p-values matrix or matrices (two or more).

Usage

```
generate_report_cchebyshev(
  dataset,
  formula,
  pvalue.method = "permutation",
  seed = NULL,
  min_group_size = 3,
  num_replicas = 1000,
  automatic_encoding = TRUE,
  na_removal = TRUE,
  grouping_stat = "mean"
)
```

Arguments

dataset	A dataframe.
formula	A variable or variables (two or more) with factors which you want to calculate the Chebyshev distances matrix or matrices (two or more) and the p_values matrix or matrices (two or more).
pvalue.method	A p_value method used to calculate the matrix or matrices (two or more), the default value is "permutation". Another method is "bootstrap".

seed	Optionally, set a seed for "bootstrap" and "permutation".
min_group_size	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded.
num_replicas	Number of permutations or bootstraps trials, the default value is 1000.
automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
na_removal	Logical, if TRUE, missing value removal on rows is performed.
grouping_stat	When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations).

Value

A Microsoft Word document about the Chebyshev distances matrix or matrices (two or more) and the p_values matrix or matrices (two or more).

Note

About pvalues, this function leverages on an early stopping procedure in which the resulting matrix is printed also if the specified number of replicas is not reached; if every 500 replicas the maximum difference between each p_value does not exceed 0.0001, the function will print the entire matrix, else it continues.

Examples

```
generate_report_cchebyshev(CO2, ~Plant + Type,
  pvalue.method = "permutation",
  seed = 122,
  num_replicas = 10,
  grouping_stat = 'median',
  automatic_encoding = TRUE)

generate_report_cchebyshev(airquality, ~Ozone,
  pvalue.method = 'bootstrap',
  na_removal = TRUE, num_replicas = 10)
```

generate_report_ccosine

Generate a Microsoft Word document about the Cosine dissimilarities matrix or matrices (two or more) and the p-values matrix or matrices (two or more).

Description

This function takes a dataframe, a factor or factors (two or more) and returns a Microsoft Word document about the Cosine distances matrix or matrices (two or more) and the p-values matrix or matrices (two or more).

Usage

```
generate_report_ccosine(
  dataset,
  formula,
  pvalue.method = "permutation",
  seed = NULL,
  min_group_size = 3,
  num_replicas = 1000,
  automatic_encoding = TRUE,
  na_removal = TRUE,
  grouping_stat = "mean"
)
```

Arguments

dataset	A dataframe.
formula	A variable or variables (two or more) with factors which you want to calculate the Cosine distances matrix or matrices (two or more) and the p_values matrix or matrices (two or more).
pvalue.method	A p_value method used to calculate the matrix or matrices (two or more), the default value is "permutation". Another method is "bootstrap".
seed	Optionally, set a seed for "bootstrap" or "permutation".
min_group_size	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded.
num_replicas	Number of permutations or bootstraps trials, the default value is 1000.
automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
na_removal	Logical, if TRUE, missing value removal on rows is performed.
grouping_stat	When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations).

Value

A Microsoft Word document about the Cosine distances matrix or matrices (two or more) and the p_values matrix or matrices (two or more).

Note

About pvalues, this function leverages on an early stopping procedure in which the resulting matrix is printed also if the specified number of replicas is not reached; if every 500 replicas the maximum difference between each p_value does not exceed 0.0001, the function will print the entire matrix, else it continues.

Examples

```
generate_report_ccosine(CO2, ~Plant + Type,
  pvalue.method = "permutation",
  seed = 122,
  num_replicas = 10,
  grouping_stat = 'median',
  automatic_encoding = TRUE)
```

```
generate_report_ccosine(airquality, ~Ozone,
  pvalue.method = 'bootstrap',
  na_removal = TRUE, num_replicas = 10)
```

```
generate_report_ceuclidean
```

Generate a Microsoft Word document about the Euclidean distances matrix or matrices and the p-values matrix or matrices.

Description

This function takes a dataframe, a factor or factors (two or more) and returns a Microsoft Word document about the Euclidean distances matrix or matrices (two or more) and the p-values matrix or matrices (two or more).

Usage

```
generate_report_ceuclidean(
  dataset,
  formula,
  pvalue.method = "permutation",
  seed = NULL,
  min_group_size = 3,
  num_replicas = 1000,
  automatic_encoding = TRUE,
  na_removal = TRUE,
  grouping_stat = "mean"
)
```

Arguments

dataset	A dataframe.
formula	A variable or variables (two or more) with factors which you want to calculate the Euclidean distances matrix or matrices (two or more) and the p_values matrix or matrices (two or more).
pvalue.method	A p_value method used to calculate the matrix or matrices (two or more), the default value is "permutation". Another method is "bootstrap".
seed	Optionally, set a seed for "bootstrap" or "permutation".
min_group_size	Minimum group size to maintain. The default value is 3, therefore factors, inside variables, with less than 3 observations will be discarded.
num_replicas	Number of permutations or bootstraps trials, the default value is 1000.
automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
na_removal	Logical, if TRUE, missing value removal on rows is performed.
grouping_stat	When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations).

Value

A Microsoft Word document about the Euclidean distances matrix or matrices (two or more) and the p_values matrix or matrices (two or more).

Note

About pvalues, this function leverages on an early stopping procedure in which the resulting matrix is printed also if the specified number of replicas is not reached; if every 500 replicas the maximum difference between each p_value does not exceed 0.0001, the function will print the entire matrix, else it continues.

Examples

```
generate_report_ceuclide(CO2, ~Plant + Type,
  pvalue.method = "permutation",
  seed = 122,
  num_replicas = 10,
  grouping_stat = 'median',
  automatic_encoding = TRUE)

generate_report_ceuclide(airquality, ~Ozone,
  pvalue.method = 'bootstrap',
  na_removal = TRUE, num_replicas = 10)
```

generate_report_chamming

Generate a Microsoft Word document about the Hamming distances matrix or matrices (two or more) and the p-values matrix or matrices (two or more).

Description

This function takes a dataframe, a factor or factors (two or more) and returns a Microsoft Word document about the Hamming distances matrix or matrices (two or more) and the p-values matrix or matrices (two or more).

Usage

```
generate_report_chamming(
  dataset,
  formula,
  pvalue.method = "permutation",
  seed = NULL,
  min_group_size = 3,
  num_replicas = 1000,
  automatic_encoding = TRUE,
  na_removal = TRUE,
  grouping_stat = "mean"
)
```

Arguments

dataset	A dataframe.
formula	A variable or variables (two or more) with factors which you want to calculate the Hamming distances matrix or matrices (two or more) and the p_values matrix or matrices (two or more).
pvalue.method	A p_value method used to calculate the matrix or matrices (two or more), the default value is "permutation". Another method is "bootstrap".
seed	Optionally, set a seed for "bootstrap" and "permutation".
min_group_size	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded.
num_replicas	Number of permutations or bootstraps trials, the default value is 1000.
automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
na_removal	Logical, if TRUE, missing value removal on rows is performed.
grouping_stat	When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations).

Value

A Microsoft Word document about the Hamming distances matrix or matrices (two or more) and the p_values matrix or matrices (two or more).

Note

About pvalues, this function leverages on an early stopping procedure in which the resulting matrix is printed also if the specified number of replicas is not reached; if every 500 replicas the maximum difference between each p_value does not exceed 0.0001, the function will print the entire matrix, else it continues.

Examples

```
generate_report_chamming(CO2, ~Plant + Type,
  pvalue.method = "permutation",
  seed = 122,
  num_replicas = 10,
  grouping_stat = 'median',
  automatic_encoding = TRUE)
```

```
generate_report_chamming(airquality, ~Ozone,
  pvalue.method = 'bootstrap',
  na_removal = TRUE, num_replicas = 10)
```

generate_report_chellinger

Generate a Microsoft Word document about the Hellinger distances matrix or matrices (two or more) and the p-values matrix or matrices (two or more).

Description

This function takes a dataframe, a factor or factors (two or more) and returns a Microsoft Word document about the Hellinger distances matrix or matrices (two or more) and the p-values matrix or matrices (two or more).

Usage

```
generate_report_chellinger(
  dataset,
  formula,
  pvalue.method = "permutation",
  seed = NULL,
  min_group_size = 3,
  num_replicas = 1000,
  automatic_encoding = TRUE,
  na_removal = TRUE,
```

```
    grouping_stat = "mean"
  )
```

Arguments

dataset	A dataframe.
formula	A variable or variables (two or more) with factors which you want to calculate the Hellinger distances matrix or matrices (two or more) and the p_values matrix or matrices (two or more).
pvalue.method	A p_value method used to calculate the matrix or matrices (two or more), the default value is "permutation". Another method is "bootstrap".
seed	Optionally, set a seed for 'bootstrap' and 'permutation'.
min_group_size	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded.
num_replicas	Number of permutations or bootstraps trials, the default value is 1000.
automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
na_removal	Logical, if TRUE, missing value removal on rows is performed.
grouping_stat	When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations).

Value

A Microsoft Word document about the Hellinger distances matrix or matrices (two or more) and the p_values matrix or matrices (two or more).

Note

About pvalues, this function leverages on an early stopping procedure in which the resulting matrix is printed also if the specified number of replicas is not reached; if every 500 replicas the maximum difference between each p_value does not exceed 0.0001, the function will print the entire matrix, else it continues.

Examples

```
generate_report_chellinger(CO2, ~Plant + Type,
  pvalue.method = "permutation",
  seed = 122,
  num_replicas = 10,
  grouping_stat = 'median',
  automatic_encoding = TRUE)

generate_report_chellinger(airquality, ~Ozone,
  pvalue.method = 'bootstrap',
  na_removal = TRUE, num_replicas = 10)
```

```
generate_report_cjaccard
```

Generate a Microsoft Word document about the Jaccard distances matrix or matrices (two or more) and the p-values matrix or matrices (two or more).

Description

This function takes a dataframe, a factor and returns a Microsoft Word document about the Jaccard distances matrix or matrices and the p-values matrix or matrices.

Usage

```
generate_report_cjaccard(
  dataset,
  formula,
  pvalue.method = "permutation",
  seed = NULL,
  min_group_size = 3,
  num_replicas = 1000,
  automatic_encoding = TRUE,
  na_removal = TRUE,
  grouping_stat = "mean"
)
```

Arguments

dataset	A dataframe.
formula	A variable or variables (two or more) with factors which you want to calculate the Jaccard distances matrix or matrices and the p_values matrix or matrices.
pvalue.method	A p_value method used to calculate the matrix, the default value is "permutation". Another method is "bootstrap".
seed	Optionally, set a seed for "bootstrap" or "permutation".
min_group_size	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded.
num_replicas	Number of permutations or bootstraps trials, the default value is 1000.
automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
na_removal	Logical, if TRUE, missing value removal on rows is performed.
grouping_stat	When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations).

Value

A Microsoft Word document about the Jaccard distance matrix or matrices and the p_values matrix or matrices.

Note

About pvalues, this function leverages on an early stopping procedure in which the resulting matrix is printed also if the specified number of replicas is not reached; if every 500 replicas the maximum difference between each p_value does not exceed 0.0001, the function will print the entire matrix, else it continues.

Examples

```
generate_report_cjaccard(CO2, ~Plant + Type,  
  pvalue.method = "permutation",  
  seed = 122,  
  num_replicas = 10,  
  grouping_stat = 'median',  
  automatic_encoding = TRUE)
```

```
generate_report_cjaccard(airquality, ~Ozone,  
  pvalue.method = 'bootstrap',  
  na_removal = TRUE, num_replicas = 10)
```

```
generate_report_cmahalanobis
```

Generate a Microsoft Word document about the Mahalanobis distances matrix or matrices and the p-values matrix or matrices.

Description

This function takes a dataframe, a factor or factors (two or more) and returns a Microsoft Word document about the Mahalanobis distances matrix or matrices (two or more) and the p-values matrix or matrices.

Usage

```
generate_report_cmahalanobis(  
  dataset,  
  formula,  
  pvalue.method = "permutation",  
  seed = NULL,  
  min_group_size = 3,  
  pvalues_chisq = FALSE,  
  num_replicas = 1000,  
  automatic_encoding = TRUE,  
  na_removal = TRUE,
```

```
    grouping_stat = "mean"
  )
```

Arguments

dataset	A dataframe.
formula	A variable or variables (two or more) with factors which you want to calculate the Mahalanobis distances matrix or matrices (two or more) and the p_values matrix or matrices (two or more).
pvalue.method	A p_value method used to calculate the matrix or matrices (two or more), the default value is "permutation". Another method is "bootstrap".
seed	Optionally, set a seed for "bootstrap" or "permutation".
min_group_size	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded.
pvalues_chisq	If TRUE, print the result of the chi-squared test on squared distances. The resulting distances with "pvalues_chisq = FALSE" are not squared; instead, with "pvalues_chisq = TRUE", the squared Mahalanobis distance matrix with corresponding p_values will be printed. Default is FALSE.
num_replicas	Number of permutations or bootstraps trials, the default value is 1000.
automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
na_removal	Logical, if TRUE, missing value removal on rows is performed.
grouping_stat	When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations).

Value

A Microsoft Word document about the Mahalanobis distances matrix or matrices (two or more) and the p_values matrix or matrices (two or more).

Note

About pvalues, this function leverages on an early stopping procedure in which the resulting matrix is printed also if the specified number of replicas is not reached; if every 500 replicas the maximum difference between each p_value does not exceed 0.0001, the function will print the entire matrix, else it continues.

Examples

```
generate_report_cmahalanobis(CO2, ~Plant + Type,
  pvalues_chisq = TRUE,
  pvalue.method = "permutation",
  seed = 122,
  num_replicas = 10,
  grouping_stat = 'median',
```

```

    automatic_encoding = TRUE)

generate_report_cmahalanobis(airquality, ~Ozone,
  pvalues_chisq = FALSE,
  pvalue.method = 'bootstrap',
  na_removal = TRUE, num_replicas = 10)

```

generate_report_cmanhattan

Generate a Microsoft Word document about the Manhattan distances matrix or matrices (two or more) and the p-values matrix or matrices (two or more).

Description

This function takes a dataframe, a factor or factors (two or more) and returns a Microsoft Word document about the Manhattan distances matrix or matrices (two or more) and the p-values matrix or matrices (two or more).

Usage

```

generate_report_cmanhattan(
  dataset,
  formula,
  pvalue.method = "permutation",
  seed = NULL,
  min_group_size = 3,
  num_replicas = 1000,
  automatic_encoding = TRUE,
  na_removal = TRUE,
  grouping_stat = "mean"
)

```

Arguments

dataset	A dataframe.
formula	A variable or variables (two or more) with factors which you want to calculate the Manhattan distances matrix or matrices (two or more) and the p_values matrix or matrices (two or more).
pvalue.method	A p_value method used to calculate the matrix or matrices (two or more), the default value is "permutation". Another method is "bootstrap".
seed	Optionally, set a seed for "bootstrap" and "permutation".
min_group_size	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded.
num_replicas	Number of permutations or bootstraps trials, the default value is 1000.

automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
na_removal	Logical, if TRUE, missing value removal on rows is performed.
grouping_stat	When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations).

Value

A Microsoft Word document about the Manhattan distances matrix or matrices (two or more) and the p_values matrix or matrices (two or more).

Note

About pvalues, this function leverages on an early stopping procedure in which the resulting matrix is printed also if the specified number of replicas is not reached; if every 500 replicas the maximum difference between each p_value does not exceed 0.0001, the function will print the entire matrix, else it continues.

Examples

```
generate_report_cmanhattan(CO2, ~Plant + Type,
  pvalue.method = "permutation",
  seed = 122,
  num_replicas = 10,
  grouping_stat = 'median',
  automatic_encoding = TRUE)
```

```
generate_report_cmanhattan(airquality, ~Ozone,
  pvalue.method = 'bootstrap',
  na_removal = TRUE, num_replicas = 10)
```

```
generate_report_cminkowski
```

Generate a Microsoft Word document about the Minkowski dissimilarities/distances matrix or matrices (two or more) and the p-values matrix or matrices (two or more).

Description

This function takes a dataframe, a factor or factors (two or more) and returns a Microsoft Word document about the Minkowski dissimilarities/distances matrix or matrices (two or more) and the p-values matrix or matrices (two or more).

Usage

```
generate_report_cminkowski(
  dataset,
  formula,
  p = 3,
  pvalue.method = "permutation",
  seed = NULL,
  min_group_size = 3,
  num_replicas = 1000,
  automatic_encoding = TRUE,
  na_removal = TRUE,
  grouping_stat = "mean"
)
```

Arguments

dataset	A dataframe.
formula	A variable or variables (two or more) with factors which you want to calculate the Minkowski dissimilarities/distances matrix or matrices (two or more) and the p_values matrix or matrices (two or more).
p	Order of the Minkowski dissimilarities/distances. The default value is 3.
pvalue.method	A p_value method used to calculate the matrix or matrices (two or more), the default value is "permutation". Another method is "bootstrap".
seed	Optionally, set a seed for "bootstrap" and "permutation".
min_group_size	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded.
num_replicas	Number of permutations or bootstraps trials, the default value is 1000.
automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
na_removal	Logical, if TRUE, missing value removal on rows is performed.
grouping_stat	When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations).

Details

When $p < 1$, the Minkowski distance is a "dissimilarity" measure. When $p \geq 1$, the triangle inequality property is satisfied and we say "Minkowski distance".

Value

A Microsoft Word document about the Minkowski dissimilarities/distances matrix or matrices (two or more) and the p_values matrix or matrices (two or more).

Note

About pvalues, this function leverages on an early stopping procedure in which the resulting matrix is printed also if the specified number of replicas is not reached; if every 500 replicas the maximum difference between each p_value does not exceed 0.0001, the function will print the entire matrix, else it continues.

Examples

```
generate_report_cminkowski(CO2, ~Plant + Type,
  p = 3,
  pvalue.method = "permutation",
  seed = 122,
  num_replicas = 10,
  grouping_stat = 'median',
  automatic_encoding = TRUE)

generate_report_cminkowski(airquality, ~Ozone,
  p = 4,
  pvalue.method = 'bootstrap',
  na_removal = TRUE, num_replicas = 10)
```

generate_report_csorensendice

Generate a Microsoft Word document about the Sorensen-Dice dissimilarity matrix or matrices (two or more) and the p-values matrix or matrices (two or more).

Description

This function takes a dataframe, a factor or factors (two or more) and returns a Microsoft Word document about the Sorensen-Dice dissimilarities matrix or matrices (two or more) and the p-values matrix or matrices (two or more).

Usage

```
generate_report_csorensendice(
  dataset,
  formula,
  pvalue.method = "permutation",
  seed = NULL,
  min_group_size = 3,
  num_replicas = 1000,
  automatic_encoding = TRUE,
  na_removal = TRUE,
  grouping_stat = "mean"
)
```

Arguments

dataset	A dataframe.
formula	A variable or variables (two or more) with factors which you want to calculate the Sorensen-Dice dissimilarities matrix or matrices (two or more) and the p_values matrix or matrices (two or more).
pvalue.method	A p_value method used to calculate the matrix or matrices (two or more), the default value is "permutation". Another method is "bootstrap".
seed	Optionally, set a seed for "bootstrap" or "permutation".
min_group_size	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded.
num_replicas	Number of permutations or bootstraps trials, the default value is 1000.
automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
na_removal	Logical, if TRUE, missing value removal on rows is performed.
grouping_stat	When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations).

Value

A Microsoft Word document about the Sorensen-Dice dissimilarities matrix or matrices (two or more) and the p_values matrix or matrices (two or more).

Note

About pvalues, this function leverages on an early stopping procedure in which the resulting matrix is printed also if the specified number of replicas is not reached; if every 500 replicas the maximum difference between each p_value does not exceed 0.0001, the function will print the entire matrix, else it continues.

Examples

```
generate_report_csorensendice(CO2, ~Plant + Type,
  pvalue.method = "permutation",
  seed = 122,
  num_replicas = 10,
  grouping_stat = 'median',
  automatic_encoding = TRUE)

generate_report_csorensendice(airquality, ~Ozone,
  pvalue.method = 'bootstrap',
  na_removal = TRUE, num_replicas = 10)
```

pvaluescbatt	<i>Calculate the p_values matrix or matrices (two or more) for each factor inside variable or variables (two or more), using Bhattacharyya dissimilarities as a base.</i>
--------------	---

Description

Using the Bhattacharyya dissimilarity for the dissimilarities calculation, this function takes a dataframe, a variable or variables (two or more), a p_value method such as "bootstrap" and "permutation" and returns the p_values matrix or matrices (two or more) between each pair of factors and a plot or plots (two or more) if the user select TRUE or leaves the parameter without argument.

Usage

```
pvaluescbatt(
  dataset,
  formula,
  pvalue.method = "permutation",
  seed = NULL,
  min_group_size = 3,
  num_replicas = 1000,
  automatic_encoding = FALSE,
  na_removal = FALSE,
  grouping_stat = "mean"
)
```

Arguments

dataset	A dataframe.
formula	A variable or variables (two or more) with factors which you want to calculate Bhattacharyya dissimilarities matrix or matrices (two or more).
pvalue.method	A p_value method used to calculate the matrix or matrices (two or more), the default value is "permutation". Another method is "bootstrap".
seed	Optionally, set a seed for "bootstrap" or "permutation".
min_group_size	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded.
num_replicas	Number of permutations or bootstraps trials, the default value is 1000.
automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
na_removal	Logical, if TRUE, missing value removal on rows is performed.
grouping_stat	When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations). Then, we find p-values with the resulting distances.

Value

A list containing a matrix or matrices (two or more) of p_values and, optionally, the plot.

Note

This function leverages on an early stopping procedure in which the resulting matrix is printed also if the specified number of replicas is not reached; if every 500 replicas the maximum difference between each p_value does not exceed 0.0001, the function will print the entire matrix, else it continues.

Examples

```
pvaluesc batt(CO2, ~Plant + Type,
  pvalue.method = "permutation",
  seed = 122,
  num_replicas = 50,
  grouping_stat = 'median',
  automatic_encoding = TRUE)

pvaluesc batt(airquality, ~Ozone,
  pvalue.method = 'bootstrap',
  na_removal = TRUE,
  grouping_stat = 'mean', num_replicas = 50)
```

pvaluescbrcu	<i>Calculate the p_values matrix or matrices (two or more) for each factor inside variable or variables (two or more), using Bray-Curtis dissimilarity as a base.</i>
--------------	---

Description

Using the Bray-Curtis dissimilarity for the dissimilarities calculation, this function takes a dataframe, a variable or variables (two or more), a p_value method such as "bootstrap" and "permutation" and returns the p_values matrix or matrices (two or more) between each pair of factors and a plot or plots (two or more) if the user select TRUE or leaves the parameter without argument.

Usage

```
pvaluescbrcu(
  dataset,
  formula,
  pvalue.method = "permutation",
  seed = NULL,
  min_group_size = 3,
  num_replicas = 1000,
  automatic_encoding = FALSE,
  na_removal = FALSE,
```

```
    grouping_stat = "mean"
  )
```

Arguments

dataset	A dataframe.
formula	A variable or variables (two or more) with factors which you want to calculate the Bray-Curtis dissimilarities matrix or matrices (two or more).
pvalue.method	A p_value method used to calculate the matrix or matrices (two or more), the default value is "permutation". Another method is "bootstrap".
seed	Optionally, set a seed for 'bootstrap' or 'permutation'.
min_group_size	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded.
num_replicas	Number of permutations or bootstraps trials, the default value is 1000.
automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
na_removal	Logical, if TRUE, missing value removal on rows is performed.
grouping_stat	When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations). Then, we find p-values with the resulting distances.

Value

A list containing a matrix or matrices (two or more) of p_values and, optionally, the plot.

Note

This function leverages on an early stopping procedure in which the resulting matrix is printed also if the specified number of replicas is not reached; if every 500 replicas the maximum difference between each p_value does not exceed 0.0001, the function will print the entire matrix, else it continues.

Examples

```
pvaluescbrcu(CO2, ~Plant + Type,
  pvalue.method = "permutation",
  seed = 122,
  num_replicas = 50,
  grouping_stat = 'median',
  automatic_encoding = TRUE)

pvaluescbrcu(airquality, ~Ozone,
  pvalue.method = 'bootstrap',
  na_removal = TRUE, num_replicas = 50)
```

pvaluesccanb	<i>Calculate the p_values matrix or matrices (two or more) for each factor inside variable or variables (two or more), using Canberra distances as a base.</i>
--------------	--

Description

Using the Canberra distance for the distances calculation, this function takes a dataframe, a variable or variables (two or more), a p_value method such as "bootstrap" and "permutation" and returns the p_values matrix or matrices (two or more) between each pair of factors and a plot or plots (two or more) if the user select TRUE or leaves the parameter without argument.

Usage

```
pvaluesccanb(
  dataset,
  formula,
  pvalue.method = "permutation",
  seed = NULL,
  min_group_size = 3,
  num_replicas = 1000,
  automatic_encoding = FALSE,
  na_removal = FALSE,
  grouping_stat = "mean"
)
```

Arguments

dataset	A dataframe.
formula	A variable or variables (two or more) with factors which you want to calculate the Canberra distances matrix or matrices (two or more).
pvalue.method	A p_value method used to calculate the matrix or matrices (two or more), the default value is "permutation". Another method is "bootstrap".
seed	Optionally, set a seed for "bootstrap" and "permutation".
min_group_size	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded.
num_replicas	Number of permutations or bootstraps trials, the default value is 1000.
automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
na_removal	Logical, if TRUE, missing value removal on rows is performed.
grouping_stat	When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations). Then, we find p-values with the resulting distances.

Value

A list containing a matrix or matrices (two or more) of p_values.

Note

This function leverages on an early stopping procedure in which the resulting matrix is printed also if the specified number of replicas is not reached; if every 500 replicas the maximum difference between each p_value does not exceed 0.0001, the function will print the entire matrix, else it continues.

Examples

```
pvaluesccanb(CO2, ~Plant + Type,
  pvalue.method = "permutation",
  seed = 122,
  num_replicas = 50,
  grouping_stat = 'median',
  automatic_encoding = TRUE)

pvaluesccanb(airquality, ~Ozone,
  pvalue.method = 'bootstrap',
  na_removal = TRUE, num_replicas = 50)
```

pvaluesccheb	<i>Calculate the p_values matrix or matrices (two or more) for each factor inside variable or variables (two or more), using Chebyshev distances as a base.</i>
--------------	---

Description

Using the Chebyshev distance for the distances calculation, this function takes a dataframe, a variable or variables (two or more), a p_value method such as "bootstrap" and "permutation" and returns the p_values matrix or matrices (two or more) between each pair of factors and a plot or plots (two or more) if the user select TRUE or leaves the parameter without argument.

Usage

```
pvaluesccheb(
  dataset,
  formula,
  pvalue.method = "permutation",
  seed = NULL,
  min_group_size = 3,
  num_replicas = 1000,
  automatic_encoding = FALSE,
  na_removal = FALSE,
  grouping_stat = "mean"
)
```

Arguments

dataset	A dataframe.
formula	A variable or variables (two or more) with factors which you want to calculate the Chebyshev distances matrix or matrices (two or more).
pvalue.method	A p_value method used to calculate the matrix or matrices (two or more), the default value is "permutation". Another method is "bootstrap".
seed	Optionally, set a seed for "bootstrap" and "permutation".
min_group_size	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded.
num_replicas	Number of permutations or bootstraps trials, the default value is 1000.
automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
na_removal	Logical, if TRUE, missing value removal on rows is performed.
grouping_stat	When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations). Then, we find p-values with the resulting distances.

Value

A list containing a matrix or matrices (two or more) of p_values.

#' @note This function leverages on an early stopping procedure in which the resulting matrix is printed also if the specified number of replicas is not reached; if every 500 replicas the maximum difference between each p_value does not exceed 0.0001, the function will print the entire matrix, else it continues.

Examples

```
pvaluesccheb(CO2, ~Plant + Type,
  pvalue.method = "permutation",
  seed = 122,
  num_replicas = 50,
  grouping_stat = 'median',
  automatic_encoding = TRUE)

pvaluesccheb(airquality, ~Ozone,
  pvalue.method = 'bootstrap',
  na_removal = TRUE, num_replicas = 50)
```

pvaluesccosi	<i>Calculate the p_values matrix or matrices (two or more) for each factor inside variable or variables (two or more), using Cosine dissimilarities as a base.</i>
--------------	--

Description

Using the Cosine distance, this function takes a dataframe, a variable or variables (two or more), a p_value method such as "bootstrap" and "permutation" and returns the p_values matrix or matrices (two or more) between each pair of factors and a plot or plots (two or more) if the user select TRUE or leaves the parameter without argument.

Usage

```
pvaluesccosi(
  dataset,
  formula,
  pvalue.method = "permutation",
  seed = NULL,
  min_group_size = 3,
  num_replicas = 1000,
  automatic_encoding = FALSE,
  na_removal = FALSE,
  grouping_stat = "mean"
)
```

Arguments

dataset	A dataframe.
formula	A variable or variables (two or more) with factors which you want to calculate the Cosine distances matrix or matrices (two or more).
pvalue.method	A p_value method used to calculate the matrix or matrices (two or more), the default value is "permutation". Another method is "bootstrap".
seed	Optionally, set a seed for "bootstrap" or "permutation".
min_group_size	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded.
num_replicas	Number of permutation or bootstrap replicas to employ, the default value is 1000.
automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
na_removal	Logical, if TRUE, missing value removal on rows is performed.
grouping_stat	When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations). Then, we find p-values with the resulting distances.

Value

A list containing a matrix or matrices (two or more) of p_values and, optionally, the plot.

Note

This function leverages on an early stopping procedure in which the resulting matrix is printed also if the specified number of replicas is not reached; if every 500 replicas the maximum difference between each p_value does not exceed 0.0001, the function will print the entire matrix, else it continues.

Examples

```
pvaluesccosi(CO2, ~Plant + Type,
  pvalue.method = "permutation",
  seed = 122,
  num_replicas = 50,
  grouping_stat = 'median',
  automatic_encoding = TRUE)

pvaluesccosi(airquality, ~Ozone,
  pvalue.method = 'bootstrap',
  na_removal = TRUE, num_replicas = 50)
```

pvaluesceuc1	<i>Calculate the p_values matrix or matrices (two or more) for each pair of factors inside variable or variables (two or more), using Euclidean distances as a base.</i>
--------------	--

Description

Using the Euclidean distance for the distances calculation, this function takes a dataframe, a variable or variables (two or more), a p_value method such as "bootstrap" and "permutation" and returns the p_values matrix or matrices (two or more) between each pair of factors and a plot or plots (two or more) if the user select TRUE or leaves the parameter without argument.

Usage

```
pvaluesceuc1(
  dataset,
  formula,
  pvalue.method = "permutation",
  seed = NULL,
  min_group_size = 3,
  num_replicas = 1000,
  automatic_encoding = FALSE,
  na_removal = FALSE,
  grouping_stat = "mean"
)
```

Arguments

<code>dataset</code>	A dataframe.
<code>formula</code>	A variable or variables (two or more) with factors which you want to calculate the Euclidean distances matrix or matrices (two or more).
<code>pvalue.method</code>	A p_value method used to calculate the matrix or matrices (two or more), the default value is "permutation". Another method is "bootstrap".
<code>seed</code>	Optionally, set a seed for "bootstrap" and "permutation" methods.
<code>min_group_size</code>	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded.
<code>num_replicas</code>	Number of permutations or bootstraps trials, the default value is 1000.
<code>automatic_encoding</code>	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
<code>na_removal</code>	Logical, if TRUE, missing value removal on rows is performed.
<code>grouping_stat</code>	When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations). Then, we find p-values with the resulting distances.

Value

A list containing a matrix or matrices (two or more) of p_values.

#' @note This function leverages on an early stopping procedure in which the resulting matrix is printed also if the specified number of replicas is not reached; if every 500 replicas the maximum difference between each p_value does not exceed 0.0001, the function will print the entire matrix, else it continues.

Examples

```
pvaluesceuc1(CO2, ~Plant + Type,
  pvalue.method = "permutation",
  seed = 122,
  num_replicas = 50,
  grouping_stat = 'median',
  automatic_encoding = TRUE)

pvaluesceuc1(airquality, ~Ozone,
  pvalue.method = 'bootstrap',
  na_removal = TRUE, num_replicas = 50)
```

pvalueschamm	<i>Calculate the p_values matrix or matrices (two or more) for each factor inside variable or variables (two or more), using Hamming distances as a base.</i>
--------------	---

Description

Using the Hamming distance for the distances calculation, this function takes a dataframe, a variable or variables (two or more), a p_value method such as "bootstrap" and "permutation" and returns the p_values matrix or matrices (two or more) between each pair of factors.

Usage

```
pvalueschamm(
  dataset,
  formula,
  pvalue.method = "permutation",
  seed = NULL,
  min_group_size = 3,
  num_replicas = 1000,
  automatic_encoding = FALSE,
  na_removal = FALSE,
  grouping_stat = "median"
)
```

Arguments

dataset	A dataframe.
formula	A variable or variables (two or more) with factors which you want to calculate the Hamming distances matrix or matrices (two or more).
pvalue.method	A p_value method used to calculate the matrix or matrices (two or more), the default value is "permutation". Another method is "bootstrap".
seed	Optionally, set a seed for "bootstrap" and "permutation".
min_group_size	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded.
num_replicas	Number of permutations or bootstraps trials, the default value is 1000.
automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
na_removal	Logical, if TRUE, missing value removal on rows is performed.
grouping_stat	When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations). Then, we find p-values with the resulting distances.

Value

A list containing a matrix or matrices (two or more) of p_values.

Note

This function leverages on an early stopping procedure in which the resulting matrix is printed also if the specified number of replicas is not reached; if every 500 replicas the maximum difference between each p_value does not exceed 0.0001, the function will print the entire matrix, else it continues.

Examples

```
pvalueschamm(CO2, ~Plant + Type,
  pvalue.method = "permutation",
  seed = 122,
  num_replicas = 50,
  grouping_stat = 'median',
  automatic_encoding = TRUE)

pvalueschamm(airquality, ~Ozone,
  pvalue.method = 'bootstrap',
  na_removal = TRUE, num_replicas = 50)
```

pvalueschell	<i>Calculate the p_values matrix or matrices (two or more) for each factor inside variable or variables (two or more), using Hellinger distances as a base.</i>
--------------	---

Description

Using the Hellinger distance for the distances calculation, this function takes a dataframe, a variable or variables (two or more), a p_value method such as "bootstrap" and "permutation" and returns the p_values matrix or matrices (two or more) between each pair of factors and a plot or plots (two or more) if the user select TRUE or leaves the parameter without argument.

Usage

```
pvalueschell(
  dataset,
  formula,
  pvalue.method = "permutation",
  seed = NULL,
  min_group_size = 3,
  num_replicas = 1000,
  automatic_encoding = FALSE,
  na_removal = FALSE,
```

```
    grouping_stat = "mean"
  )
```

Arguments

dataset	A dataframe.
formula	A variable or variables (two or more) with factors which you want to calculate the Hellinger distances matrix or matrices (two or more).
pvalue.method	A p_value method used to calculate the matrix or matrices (two or more), the default value is "permutation". Another method is "bootstrap".
seed	Optionally, set a seed for 'bootstrap' and 'permutation'.
min_group_size	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded.
num_replicas	Number of permutations or bootstraps trials, the default value is 1000.
automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
na_removal	Logical, if TRUE, missing value removal on rows is performed.
grouping_stat	When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations). Then, we find p-values with the resulting distances.

Value

A list containing a matrix or matrices (two or more) of p_values and, optionally, the plot.

Note

This function leverages on an early stopping procedure in which the resulting matrix is printed also if the specified number of replicas is not reached; if every 500 replicas the maximum difference between each p_value does not exceed 0.0001, the function will print the entire matrix, else it continues.

Examples

```
pvalueschell(CO2, ~Plant + Type,
  pvalue.method = "permutation",
  seed = 122,
  num_replicas = 50,
  grouping_stat = 'median',
  automatic_encoding = TRUE)

pvalueschell(airquality, ~Ozone,
  pvalue.method = 'bootstrap',
  na_removal = TRUE, num_replicas = 50)
```

pvaluescjacc	<i>Calculate the p_values matrix or matrices (two or more) for each factor inside variable or variables (two or more), using Jaccard distances as a base.</i>
--------------	---

Description

Using the Jaccard distance for the distance calculation, this function takes a dataframe, a variable or variables (two or more), a p_value method such as "bootstrap" and "permutation" and returns the p_values matrix or matrices (two or more) between each pair of factors and a plot or plots (two or more) if the user select TRUE or leaves the parameter without argument.

Usage

```
pvaluescjacc(
  dataset,
  formula,
  pvalue.method = "permutation",
  seed = NULL,
  min_group_size = 3,
  num_replicas = 1000,
  automatic_encoding = FALSE,
  na_removal = FALSE,
  grouping_stat = "mean"
)
```

Arguments

dataset	A dataframe.
formula	A variable or variables (two or more) with factors which you want to calculate the Jaccard distance matrix or matrices (two or more).
pvalue.method	A p_value method used to calculate the matrix or matrices (two or more), the default value is "permutation". Another method is "bootstrap".
seed	Optionally, set a seed for "bootstrap" or "permutation".
min_group_size	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded.
num_replicas	Number of permutations or bootstraps trials, the default value is 1000.
automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
na_removal	Logical, if TRUE, missing value removal on rows is performed.
grouping_stat	When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations). Then, we find p-values with the resulting distances.

Value

A list containing a matrix or matrices (two or more) of p_values and, optionally, the plot.

Note

This function leverages on an early stopping procedure in which the resulting matrix is printed also if the specified number of replicas is not reached; if every 500 replicas the maximum difference between each p_value does not exceed 0.0001, the function will print the entire matrix, else it continues

Examples

```
pvaluescjacc(CO2, ~Plant + Type,
  pvalue.method = "permutation",
  seed = 122,
  num_replicas = 50,
  grouping_stat = 'median',
  automatic_encoding = TRUE)

pvaluescjacc(airquality, ~Ozone,
  pvalue.method = 'bootstrap',
  na_removal = TRUE, num_replicas = 50)
```

pvaluescmaha	<i>Calculate the p_values matrix or matrices (two or more) for each pair of factors inside variable or variables (two or more), using Mahalanobis distances as a base.</i>
--------------	--

Description

Using the Mahalanobis distance for the distances calculation, this function takes a dataframe, a variable or variables (two or more), a p_value method such as bootstrap" and "permutation" and returns the p_values matrix or matrices (two or more) between each pair of factors.

Usage

```
pvaluescmaha(
  dataset,
  formula,
  pvalue.method = "permutation",
  seed = NULL,
  min_group_size = 3,
  num_replicas = 1000,
  automatic_encoding = FALSE,
  na_removal = FALSE,
  grouping_stat = "median"
)
```

Arguments

<code>dataset</code>	A dataframe.
<code>formula</code>	A variable or variables (two or more) with factors which you want to calculate the Mahalanobis distances matrix or matrices (two or more).
<code>pvalue.method</code>	A p_value method used to calculate the matrix or matrices (two or more), the default value is "permutation". Another method is "bootstrap".
<code>seed</code>	Optionally, set a seed for "bootstrap" and "permutation" methods.
<code>min_group_size</code>	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded.
<code>num_replicas</code>	Number of permutations or bootstraps trials, the default value is 1000.
<code>automatic_encoding</code>	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
<code>na_removal</code>	Logical, if TRUE, missing value removal on rows is performed.
<code>grouping_stat</code>	When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations). Then, we find p-values with the resulting distances.

Value

A list containing a matrix or matrices (two or more) of p_values.

Note

This function leverages on an early stopping procedure in which the resulting matrix is printed also if the specified number of replicas is not reached; if every 500 replicas the maximum difference between each p_value does not exceed 0.0001, the function will print the entire matrix, else it continues.

Examples

```
pvaluescmaha(CO2, ~Plant + Type,
  pvalue.method = "permutation",
  seed = 122,
  num_replicas = 50,
  grouping_stat = 'median',
  automatic_encoding = TRUE)

pvaluescmaha(airquality, ~Ozone,
  pvalue.method = 'bootstrap',
  na_removal = TRUE, num_replicas = 50)
```

pvaluescmanh	<i>Calculate the p_values matrix or matrices (two or more) for each factor inside variable or variables (two or more), using Manhattan distances as a base.</i>
--------------	---

Description

Using the Manhattan distance for the distances calculation, this function takes a dataframe, a variable or variables (two or more), a p_value method such as "bootstrap" and "permutation" and returns the p_values matrix or matrices (two or more) between each pair of factors and a plot or plots (two or more) if the user select TRUE or leaves the parameter without argument.

Usage

```
pvaluescmanh(
  dataset,
  formula,
  pvalue.method = "permutation",
  seed = NULL,
  min_group_size = 3,
  num_replicas = 1000,
  automatic_encoding = FALSE,
  na_removal = FALSE,
  grouping_stat = "median"
)
```

Arguments

dataset	A dataframe.
formula	A variable or variables (two or more) with factors which you want to calculate the Manhattan distances matrix or matrices (two or more).
pvalue.method	A p_value method used to calculate the matrix or matrices (two or more), the default value is "permutation". Another method is "bootstrap".
seed	Optionally, set a seed for "bootstrap" and "permutation".
min_group_size	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded.
num_replicas	Number of permutations or bootstraps trials, the default value is 1000.
automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
na_removal	Logical, if TRUE, missing value removal on rows is performed.
grouping_stat	When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations). Then, we find p-values with the resulting distances.

Value

A list containing a matrix or matrices (two or more) of p_values.

Note

This function leverages on an early stopping procedure in which the resulting matrix is printed also if the specified number of replicas is not reached; if every 500 replicas the maximum difference between each p_value does not exceed 0.0001, the function will print the entire matrix, else it continues.

Examples

```
pvaluescmanh(CO2, ~Plant + Type,
  pvalue.method = "permutation",
  seed = 122,
  num_replicas = 50,
  grouping_stat = 'median',
  automatic_encoding = TRUE)

pvaluescmanh(airquality, ~Ozone,
  pvalue.method = 'bootstrap',
  na_removal = TRUE, num_replicas = 50)
```

pvaluescmink

Calculate the p_values matrix or matrices (two or more) for each factor inside variable or variables (two or more), using Minkowski dissimilarities/distances as a base.

Description

Using the Minkowski dissimilarity/distance for the dissimilarities/distances calculation, this function takes a dataframe, a variable or variables (two or more), a p_value method such as "bootstrap" and "permutation" and returns the p_values matrix or matrices (two or more) between each pair of factors and a plot or plots (two or more) if the user select TRUE or leaves the parameter without argument.

Usage

```
pvaluescmink(
  dataset,
  formula,
  pvalue.method = "permutation",
  seed = NULL,
  min_group_size = 3,
  num_replicas = 1000,
  p = 3,
  automatic_encoding = FALSE,
```

```

    na_removal = FALSE,
    grouping_stat = "median"
  )

```

Arguments

dataset	A dataframe.
formula	A variable or variables (two or more) with factors which you want to calculate the Minkowski dissimilarities/distances matrix or matrices (two or more).
pvalue.method	A p_value method used to calculate the matrix or matrices (two or more), the default value is "permutation". Another method is "bootstrap".
seed	Optionally, set a seed for "bootstrap" and "permutation".
min_group_size	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded.
num_replicas	Number of permutations or bootstraps trials, the default value is 1000.
p	Order of the Minkowski dissimilarities/distances. The default value is 3.
automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
na_removal	Logical, if TRUE, missing value removal on rows is performed.
grouping_stat	When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations). Then, we find p-values with the resulting distances.

Value

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

Note

When $p < 1$, the Minkowski distance is a "dissimilarity" measure. When $p \geq 1$, the triangle inequality property is satisfied and we say "Minkowski distance". This function leverages on an early stopping procedure in which the resulting matrix is printed also if the specified number of replicas is not reached; if every 500 replicas the maximum difference between each p_value does not exceed 0.0001, the function will print the entire matrix, else it continue.

Examples

```

pvaluescmink(CO2, ~Plant + Type,
  pvalue.method = "permutation",
  seed = 122,
  num_replicas = 50,
  p = 3,
  grouping_stat = 'median',
  automatic_encoding = TRUE)

pvaluescmink(airquality, ~Ozone,
  pvalue.method = 'bootstrap',

```

```
p = 3,
na_removal = TRUE, num_replicas = 50)
```

pvaluescore	<i>Calculate the p_values matrix or matrices (two or more) for each factor inside variable or variables (two or more), using Sorensen-Dice dissimilarities as a base.</i>
-------------	---

Description

Using the Sorensen-Dice dissimilarity for the dissimilarities calculation, this function takes a dataframe, a variable or variables (two or more), a p_value method such as "bootstrap" and "permutation" and returns the p_values matrix or matrices (two or more) between each pair of factors and a plot or plots (two or more) if the user select TRUE or leaves the parameter without argument.

Usage

```
pvaluescore(
  dataset,
  formula,
  pvalue.method = "permutation",
  seed = NULL,
  min_group_size = 3,
  num_replicas = 1000,
  automatic_encoding = FALSE,
  na_removal = FALSE,
  grouping_stat = "mean"
)
```

Arguments

dataset	A dataframe.
formula	A variable or variables (two or more) with factors which you want to calculate the Sorensen-Dice dissimilarities matrix or matrices (two or more).
pvalue.method	A p_value method used to calculate the matrix or matrices (two or more), the default value is "permutation". Another method is "bootstrap".
seed	Optionally, set a seed for "bootstrap" and "permutation".
min_group_size	Minimum group size to maintain. The default value is 3, therefore groups, inside variables, with less than 3 observations will be discarded.
num_replicas	Number of permutations or bootstraps trials, the default value is 1000.
automatic_encoding	Logical, if TRUE, names inside factor variables will be transformed in numbers with ordinal order (1,2,...).
na_removal	Logical, if TRUE, missing value removal on rows is performed.
grouping_stat	When a factor variable is specified, calculate the specified grouping statistic for each factor. Available methods are: mean (arithmetic mean), median and SDS (standard deviations). Then, we find p-values with the resulting distances.

Value

A list containing a matrix or matrices (two or more) of `p_values` and, optionally, the plot.

Note

This function leverages on an early stopping procedure in which the resulting matrix is printed also if the specified number of replicas is not reached; if every 500 replicas the maximum difference between each `p_value` does not exceed 0.0001, the function will print the entire matrix, else it continues.

Examples

```
pvaluescore(CO2, ~Plant + Type,
            pvalue.method = "permutation",
            seed = 122,
            num_replicas = 50,
            grouping_stat = 'median',
            automatic_encoding = TRUE)

pvaluescore(airquality, ~Ozone,
            pvalue.method = 'bootstrap',
            na_removal = TRUE, num_replicas = 50)
```

Index

cbhattacharyya, [3](#)
cbraycurtis, [4](#)
ccanberra, [6](#)
cchebyshev, [8](#)
ccosine, [10](#)
ceuclyde, [12](#)
chamming, [13](#)
chellinger, [15](#)
cjaccard, [17](#)
ckendall, [19](#)
cmahalanobis, [20](#)
cmanhattan, [22](#)
cminkowski, [24](#)
cpearson, [26](#)
csorensendice, [28](#)
cspearman, [29](#)

generate_report_cbhattacharyya, [31](#)
generate_report_cbraycurtis, [33](#)
generate_report_ccanberra, [34](#)
generate_report_cchebyshev, [36](#)
generate_report_ccosine, [37](#)
generate_report_ceuclyde, [39](#)
generate_report_chamming, [41](#)
generate_report_chellinger, [42](#)
generate_report_cjaccard, [44](#)
generate_report_cmahalanobis, [45](#)
generate_report_cmanhattan, [47](#)
generate_report_cminkowski, [48](#)
generate_report_csorensendice, [50](#)

pvaluescbatt, [52](#)
pvaluescbrcu, [53](#)
pvaluesccanb, [55](#)
pvaluesccheb, [56](#)
pvaluesccosi, [58](#)
pvaluesceuclyde, [59](#)
pvalueschamm, [61](#)
pvalueschell, [62](#)
pvaluescjacc, [64](#)
pvaluescmaha, [65](#)
pvaluescmanh, [67](#)
pvaluescmink, [68](#)
pvaluescsore, [70](#)