

Package ‘ClustBlock’

October 12, 2022

Title Clustering of Datasets

Version 3.0.0

Maintainer Fabien Llobell <fllobell@hotmail.fr>

Description Hierarchical and partitioning algorithms of blocks of variables. The partitioning algorithm includes an option called noise cluster to set aside atypical blocks of variables. The CLUSTATIS method (for quantitative blocks) (Llobell, Cariou, Vigneau, Labenne & Qannari (2020) <[doi:10.1016/j.foodqual.2018.05.013](https://doi.org/10.1016/j.foodqual.2018.05.013)>, Llobell, Vigneau & Qannari (2019) <[doi:10.1016/j.foodqual.2019.02.017](https://doi.org/10.1016/j.foodqual.2019.02.017)>) and the CLUSCATA method (for Check-All-That-Apply data) (Llobell, Cariou, Vigneau, Labenne & Qannari (2019) <[doi:10.1016/j.foodqual.2018.09.006](https://doi.org/10.1016/j.foodqual.2018.09.006)>, Llobell, Giacalone, Labenne & Qannari (2019) <[doi:10.1016/j.foodqual.2019.05.017](https://doi.org/10.1016/j.foodqual.2019.05.017)>) are the core of this package. The CATATIS methods allows to compute some indices and tests to control the quality of CATA data. Multivariate analysis and clustering of subjects for quantitative multi-block data, CATA, Free Sorting and JAR experiments are available.

Depends R (>= 3.4.0)

License GPL-3

Encoding UTF-8

LazyData true

Imports FactoMineR

Suggests ClustVarLV

Date 2022-09-07

RoxygenNote 6.1.1

NeedsCompilation no

Author Fabien Llobell [aut, cre] (Oniris/XLSTAT),
Evelyne Vigneau [ctb] (Oniris),
Veronique Cariou [ctb] (Oniris),
El Mostafa Qannari [ctb] (Oniris)

Repository CRAN

Date/Publication 2022-09-07 22:23:08 UTC

R topics documented:

ClustBlock-package	3
catatis	4
catatis_jar	5
change_cata_format	7
change_cata_format2	8
cheese	9
choc	9
cluscata	10
cluscata_jar	12
cluscata_kmeans	14
cluscata_kmeans_jar	16
clustatis	18
clustatis_FreeSort	20
clustatis_FreeSort_kmeans	22
clustatis_kmeans	24
consistency_cata	26
consistency_cata_panel	27
PB_tuna	28
plot.catatis	28
plot.cluscata	29
plot.clustatis	31
plot.statis	32
preprocess_FreeSort	33
preprocess_JAR	34
print.catatis	35
print.cluscata	35
print.clustatis	36
print.statis	36
simil_groups_cata	37
smoo	38
statis	38
statis_FreeSort	40
straw	41
summary.catatis	42
summary.cluscata	43
summary.clustatis	44
summary.statis	45

Description

Hierarchical and partitioning algorithms of blocks of variables. The CLUSTATIS method and the CLUSCATA method are the core of this package. The CATATIS methods allows to compute some indices and tests to control the quality of CATA data. Multivariate analysis and clustering of subjects for quantitative multiblock data, CATA, Free Sorting and JAR experiments are available.

Details

Package:	ClustBlock
Type:	Package
Version:	3.0.0
First version Date:	2019-03-06
Last version Date:	2022-09-07

Author(s)

Fabien Llobell, Evelyne Vigneau, Veronique Cariou, El Mostafa Qannari
Maintainer: <f1llobell@hotmail.fr>

References

- Llobell, F., Cariou, V., Vigneau, E., Labenne, A., & Qannari, E. M. (2020). Analysis and clustering of multiblock datasets by means of the STATIS and CLUSTATIS methods. Application to sensometrics. *Food Quality and Preference*, 79, 103520.
- Llobell, F., Vigneau, E., & Qannari, E. M. (2019). Clustering datasets by means of CLUSTATIS with identification of atypical datasets. Application to sensometrics. *Food Quality and Preference*, 75, 97-104.
- Llobell, F., Cariou, V., Vigneau, E., Labenne, A., & Qannari, E. M. (2019). A new approach for the analysis of data and the clustering of subjects in a CATA experiment. *Food quality and preference*, 72, 31-39.
- Llobell, F., Giacalone, D., Labenne, A., & Qannari, E. M. (2019). Assessment of the agreement and cluster analysis of the respondents in a CATA experiment. *Food Quality and Preference*, 77, 184-190.
- Llobell, F., & Qannari, E. M. (2020). CLUSTATIS: Cluster analysis of blocks of variables. *Electronic Journal of Applied Statistical Analysis*, 13(2), 436-453.
- Llobell, F. (2020). Classification de tableaux de données, applications en analyse sensorielle (Doctoral dissertation, Nantes, Ecole nationale vétérinaire).

catatis	<i>Perform the CATATIS method on different blocks from a CATA experiment</i>
---------	--

Description

CATATIS method. Additional outputs are also computed. Non-binary data are accepted and weights can be tested.

Usage

```
catatis(Data,nblo,NameBlocks=NULL, NameVar=NULL, Graph=TRUE, Graph_weights=TRUE,
        Test_weights=FALSE, nperm=100)
```

Arguments

Data	data frame or matrix where the blocks of binary variables are merged horizontally. If you have a different format, see change_cata_format
nblo	integer. Number of blocks (subjects).
NameBlocks	string vector. Name of each block (subject). Length must be equal to the number of blocks. If NULL, the names are S1,...Sm. Default: NULL
NameVar	string vector. Name of each variable (attribute, the same names for each subject). Length must be equal to the number of attributes. If NULL, the colnames of the first block are taken. Default: NULL
Graph	logical. Show the graphical representation? Default: TRUE
Graph_weights	logical. Should the barplot of the weights be plotted? Default: TRUE
Test_weights	logical. Should the the weights be tested? Default: FALSE
nperm	integer. Number of permutation for the weight tests. Default: 100

Value

a list with:

- S: the S matrix: a matrix with the similarity coefficient among the subjects
- compromise: a matrix which is the compromise of the subjects (akin to a weighted average)
- weights: the weights associated with the subjects to build the compromise
- weights_tests: the weights tests results
- lambda: the first eigenvalue of the S matrix
- overall error: the error for the CATATIS criterion
- error_by_sub: the error by subject (CATATIS criterion)
- error_by_prod: the error by product (CATATIS criterion)
- s_with_compromise: the similarity coefficient of each subject with the compromise
- homogeneity: homogeneity of the subjects (in percentage)

- CA: the results of correspondance analysis performed on the compromise dataset
- eigenvalues: the eigenvalues associated to the correspondance analysis
- inertia: the percentage of total variance explained by each axis of the CA
- scalefactors: the scaling factors of each subject
- nb_1: the number of 1 in each block, i.e. the number of checked attributes by subject.
- param: parameters called

References

Llobell, F., Cariou, V., Vigneau, E., Labenne, A., & Qannari, E. M. (2019). A new approach for the analysis of data and the clustering of subjects in a CATA experiment. *Food Quality and Preference*, 72, 31-39. Bonnet, L., Ferney, T., Riedel, T., Qannari, E.M., Llobell, F. (September 14, 2022) .Using CATA for sensory profiling: assessment of the panel performance. Eurosense, Turku, Finland.

See Also

[plot.catatis](#), [summary.catatis](#), [cluscata](#), [change_cata_format](#), [change_cata_format2](#)

Examples

```
data(straw)
res.cat=catatis(straw, nblo=114)
summary(res.cat)
plot(res.cat)

#Vertical format with sessions
data("PB_tuna")
chang=change_cata_format2(PB_tuna, nprod= 6, nattr= 27, nsub = 12, nsess= 3)
res.cat2=catatis(Data= chang$Datafinal, nblo = 12, NameBlocks = chang$NameSub, Test_weights=TRUE)

#Vertical format without sessions
Data=PB_tuna[1:66,2:30]
chang2=change_cata_format2(Data, nprod= 6, nattr= 27, nsub = 11, nsess= 1)
res.cat3=catatis(Data= chang2$Datafinal, nblo = 11, NameBlocks = chang2$NameSub)
```

catatis_jar

Perform the CATATIS method on Just About Right data.

Description

CATATIS method adapted to JAR data.

Usage

```
catatis_jar(Data, nprod, nsub, levelsJAR=3, beta=0.1, Graph=TRUE, Graph_weights=TRUE,
Test_weights=FALSE, nperm=100)
```

Arguments

Data	data frame where the first column is the Assessors, the second is the products and all other columns the JAR attributes with numbers (1 to 3 or 1 to 5, see levelsJAR)
nprod	integer. Number of products.
nsub	integer. Number of subjects.
levelsJAR	integer. 3 or 5 levels. If 5, the data will be transformed in 3 levels.
beta	numerical. Parameter for agreement between JAR and other answers. Between 0 and 0.5.
Graph	logical. Show the graphical representation? Default: TRUE
Graph_weights	logical. Should the barplot of the weights be plotted? Default: TRUE
Test_weights	logical. Should the the weights be tested? Default: FALSE
nperm	integer. Number of permutation for the weight tests. Default: 100

Value

a list with:

- S: the S matrix: a matrix with the similarity coefficient among the subjects
- compromise: a matrix which is the compromise of the subjects (akin to a weighted average)
- weights: the weights associated with the subjects to build the compromise
- weights_tests: the weights tests results
- lambda: the first eigenvalue of the S matrix
- overall error: the error for the CATATIS criterion
- error_by_sub: the error by subject (CATATIS criterion)
- error_by_prod: the error by product (CATATIS criterion)
- s_with_compromise: the similarity coefficient of each subject with the compromise
- homogeneity: homogeneity of the subjects (in percentage)
- CA: the results of correspondance analysis performed on the compromise dataset
- eigenvalues: the eigenvalues associated to the correspondance analysis
- inertia: the percentage of total variance explained by each axis of the CA
- scalefactors: the scaling factors of each subject
- nb_1: Can be ignored
- param: parameters called

References

Llobell, F., Vigneau, E. & Qannari, E. M. ((September 14, 2022). Multivariate data analysis and clustering of subjects in a Just about right task. Eurosense, Turku, Finland.

See Also

[catatis](#), [plot.catatis](#), [summary.catatis](#), [cluscata_jar](#), [preprocess_JAR](#), [cluscata_kmeans_jar](#)

Examples

```
data(cheese)
res.cat=catatis_jar(Data=cheese, nprod=8, nsub=72, levelsJAR=5)
summary(res.cat)
#plot(res.cat)
```

change_cata_format	<i>Change format of CATA datasets to perform CATATIS or CLUSCATA function</i>
--------------------	---

Description

CATATIS and CLUSCATA operate on data where the blocksvariables are merged horizontally. If you have a different format, you can use this function to change the format. Format=1 is for data merged vertically with the dataset of the first subject, then the second,... with products in same order Format=2 is for data merged vertically with the dataset for the first product, then the second... with subjects in same order

Unlike change_cata_format2, you don't need to specify products and subjects, just make sure they are in the right order.

Usage

```
change_cata_format(Data, nprod, nattr, nsub, format=1, NameProds=NULL, NameAttr=NULL)
```

Arguments

Data	data frame or matrix. Correspond to your data
nprod	integer. Number of products
nattr	integer. Number of attributes
nsub	integer. Number of subjects.
format	integer (1 or 2). See the description
NameProds	string vector with the names of the products (length must be nprod)
NameAttr	string vector with the names of attributes (length must be nattr)

Value

The arranged data for CATATIS and CLUSCATA function

See Also

[catatis](#), [cluscata](#), [change_cata_format2](#)

change_cata_format2 *Change format of CATA datasets to perform the package functions*

Description

CATATIS and CLUSCATA operate on data where the blocks of variables are merged horizontally. If you have a vertical format, you can use this function to change the format. The first column must contain the sessions, the second the subjects, the third the products and the others the attributes. If you don't have sessions, then the first column must contain the subjects and the second the products. Unlike change_cata_format function, you can enter data with sessions and/or mixed data in terms of products/subjects. However, you have to set columns to indicate this beforehand.

Usage

```
change_cata_format2(Data, nprod, nattr, nsub, nsess)
```

Arguments

Data	data frame or matrix. Correspond to your data
nprod	integer. Number of products
nattr	integer. Number of attributes
nsub	integer. Number of subjects.
nsess	integer. Number of sessions

Value

The arranged data for CATATIS and CLUSCATA function and the subjects names in the correct order.

See Also

[catatis](#), [cluscata](#), [change_cata_format](#)

Examples

```
#Vertical format with sessions
data("PB_tuna")
chang=change_cata_format2(PB_tuna, nprod= 6, nattr= 27, nsub = 12, nsess= 3)
res.cat2=catatis(Data= chang$Datafinal, nblo = 12, NameBlocks = chang$NameSub)

#Vertical format without sessions
Data=PB_tuna[1:66,2:30]
chang2=change_cata_format2(Data, nprod= 6, nattr= 27, nsub = 11, nsess= 1)
res.cat3=catatis(Data= chang2$Datafinal, nblo = 11, NameBlocks = chang2$NameSub)
res.clu3=cluscata(Data= chang2$Datafinal, nblo = 11, NameBlocks = chang2$NameSub)
```

cheese	<i>cheese Just About Right data</i>
--------	-------------------------------------

Description

cheese Just About Right data

Usage

```
data(cheese)
```

Format

JAR data. A data frame with Assessors, Products and JAR attributes. 10 products, 9 attributes and 72 subjects.

References

Luc, A., Lê, S., Philippe, M., Qannari, E. M., & Vigneau, E. (2022). Free JAR experiment: Data analysis and comparison with JAR task. *Food Quality and Preference*, 98, 104453.

Examples

```
data(cheese)
```

choc	<i>chocolates data</i>
------	------------------------

Description

chocolates data

Usage

```
data(choc)
```

Format

Free sorting data. A data frame with 14 rows (the chocolates) and 25 columns (the subjects). The numbers indicate the groups to which the products (rows) are assigned.

References

Courcoux, P., Qannari, E. M., Taylor, Y., Buck, D., & Greenhoff, K. (2012). Taxonomic free sorting. *Food Quality and Preference*, 23(1), 30-35.

Examples

```
data(choc)
```

cluscata

*Perform a cluster analysis of blocks from a CATA experiment***Description**

Hierarchical clustering of blocks from a CATA experiment. Each cluster of blocks is associated with a compromise computed by the CATATIS method. The hierarchical clustering is followed by a partitioning algorithm (consolidation). Non-binary data are accepted.

Usage

```
cluscata(Data, nblo, NameBlocks=NULL, NameVar=NULL, Noise_cluster=FALSE,
         Itermax=30, Graph_dend=TRUE, Graph_bar=TRUE, printlevel=FALSE,
         gpmx=min(6, nblo-2), Testonlyoneclust=FALSE, alpha=0.05,
         nperm=50, Warnings=FALSE)
```

Arguments

Data	data frame or matrix where the blocks of binary variables are merged horizontally. If you have a different format, see change_cata_format
nblo	numerical. Number of blocks (subjects).
NameBlocks	string vector. Name of each block (subject). Length must be equal to the number of blocks. If NULL, the names are S1,...Sm. Default: NULL
NameVar	string vector. Name of each variable (attribute, the same names for each subject). Length must be equal to the number of attributes. If NULL, the colnames of the first block are taken. Default: NULL
Noise_cluster	logical. Should a noise cluster be computed? Default: FALSE
Itermax	numerical. Maximum of iteration for the partitioning algorithm. Default:30
Graph_dend	logical. Should the dendrogram be plotted? Default: TRUE
Graph_bar	logical. Should the barplot of the difference of the criterion and the barplot of the overall homogeneity at each merging step of the hierarchical algorithm be plotted? Default: TRUE
printlevel	logical. Print the number of remaining levels during the hierarchical clustering algorithm? Default: FALSE
gpmx	logical. What is maximum number of clusters to consider? Default: min(6, nblo-2)
Testonlyoneclust	logical. Test if there is more than one cluster? Default: FALSE
alpha	numerical between 0 and 1. What is the threshold to test if there is more than one cluster? Default: 0.05
nperm	numerical. How many permutations are required to test if there is more than one cluster? Default: 50
Warnings	logical. Display warnings about the fact that none of the subjects in some clusters checked an attribute or product? Default: FALSE

Value

Each partitionK contains a list for each number of clusters of the partition, K=1 to gpmx with:

- group: the clustering partition after consolidation. If Noise_cluster=TRUE, some subjects could be in the noise cluster ("K+1")
- rho: the threshold for the noise cluster
- homogeneity: homogeneity index (
- s_with_compromise: similarity coefficient of each subject with its cluster compromise
- weights: weight associated with each subject in its cluster
- compromise: the compromise of each cluster
- CA: list. the correspondance analysis results on each cluster compromise (coordinates, contributions...)
- inertia: percentage of total variance explained by each axis of the CA for each cluster
- s_all_cluster: the similarity coefficient between each subject and each cluster compromise
- criterion: the CLUSCATA criterion error
- param: parameters called
- type: parameter passed to other functions

There is also at the end of the list:

- dend: The CLUSCATA dendrogram
- cutree_k: the partition obtained by cutting the dendrogram in K clusters (before consolidation).
- overall_homogeneity_ng: percentage of overall homogeneity by number of clusters before consolidation (and after if there is no noise cluster)
- diff_crit_ng: variation of criterion when a merging is done before consolidation (and after if there is no noise cluster)
- test_one_cluster: decision and pvalue to know if there is more than one cluster
- param: parameters called
- type: parameter passed to other functions

References

Llobell, F., Cariou, V., Vigneau, E., Labenne, A., & Qannari, E. M. (2019). A new approach for the analysis of data and the clustering of subjects in a CATA experiment. *Food Quality and Preference*, 72, 31-39.

Llobell, F., Giacalone, D., Labenne, A., Qannari, E.M. (2019). Assessment of the agreement and cluster analysis of the respondents in a CATA experiment. *Food Quality and Preference*, 77, 184-190.

See Also

[plot.cluscata](#), [summary.cluscata](#), [catatis](#), [cluscata_kmeans](#), [change_cata_format](#), [change_cata_format2](#)

Examples

```

data(straw)
#with 40 subjects
res=cluscata(Data=straw[,1:(16*40)], nblo=40)
#plot(res, ngroups=3, Graph_dend=FALSE)
summary(res, ngroups=3)
#With noise cluster
res2=cluscata(Data=straw[,1:(16*40)], nblo=40, Noise_cluster=TRUE,
Graph_dend=FALSE, Graph_bar=FALSE)
#with all subjects
res=cluscata(Data=straw, nblo=114, printlevel=TRUE)

#Vertical format
Data=PB_tuna[1:66,2:30]
chang2=change_cata_format2(Data, nprod= 6, nattr= 27, nsub = 11, nsess= 1)
res3=cluscata(Data= chang2$Datafinal, nblo = 11, NameBlocks = chang2$NameSub)

```

cluscata_jar

Perform a cluster analysis of subjects in a JAR experiment.

Description

Hierarchical clustering of subjects from a JAR experiment. Each cluster of subjects is associated with a compromise computed by the CATATIS method. The hierarchical clustering is followed by a partitioning algorithm (consolidation).

Usage

```

cluscata_jar(Data, nprod, nsub, levelsJAR=3, beta=0.1, Noise_cluster=FALSE,
Itermax=30, Graph_dend=TRUE, Graph_bar=TRUE, printlevel=FALSE,
gpmx=min(6, nsub-2), Testonlyoneclust=FALSE, alpha=0.05,
nperm=50, Warnings=FALSE)

```

Arguments

Data	data frame where the first column is the Assessors, the second is the products and all other columns the JAR attributes with numbers (1 to 3 or 1 to 5, see levelsJAR)
nprod	integer. Number of products.
nsub	integer. Number of subjects.
levelsJAR	integer. 3 or 5 levels. If 5, the data will be transformed in 3 levels.
beta	numerical. Parameter for agreement between JAR and other answers. Between 0 and 0.5.

Noise_cluster	logical. Should a noise cluster be computed? Default: FALSE
Itermax	numerical. Maximum of iteration for the partitioning algorithm. Default:30
Graph_dend	logical. Should the dendrogram be plotted? Default: TRUE
Graph_bar	logical. Should the barplot of the difference of the criterion and the barplot of the overall homogeneity at each merging step of the hierarchical algorithm be plotted? Default: TRUE
printlevel	logical. Print the number of remaining levels during the hierarchical clustering algorithm? Default: FALSE
gpmx	logical. What is maximum number of clusters to consider? Default: min(6, nblo-2)
Testonlyoneclust	logical. Test if there is more than one cluster? Default: FALSE
alpha	numerical between 0 and 1. What is the threshold to test if there is more than one cluster? Default: 0.05
nperm	numerical. How many permutations are required to test if there is more than one cluster? Default: 50
Warnings	logical. Display warnings about the fact that none of the subjects in some clusters checked an attribute or product? Default: FALSE

Value

Each partitionK contains a list for each number of clusters of the partition, K=1 to gpmx with:

- group: the clustering partition after consolidation. If Noise_cluster=TRUE, some subjects could be in the noise cluster ("K+1")
- rho: the threshold for the noise cluster
- homogeneity: homogeneity index (
- s_with_compromise: similarity coefficient of each subject with its cluster compromise
- weights: weight associated with each subject in its cluster
- compromise: the compromise of each cluster
- CA: list. the correspondance analysis results on each cluster compromise (coordinates, contributions...)
- inertia: percentage of total variance explained by each axis of the CA for each cluster
- s_all_cluster: the similarity coefficient between each subject and each cluster compromise
- criterion: the CLUSCATA criterion error
- param: parameters called
- type: parameter passed to other functions

There is also at the end of the list:

- dend: The CLUSCATA dendrogram
- cutree_k: the partition obtained by cutting the dendrogram in K clusters (before consolidation).

- overall_homogeneity_ng: percentage of overall homogeneity by number of clusters before consolidation (and after if there is no noise cluster)
- diff_crit_ng: variation of criterion when a merging is done before consolidation (and after if there is no noise cluster)
- test_one_cluster: decision and pvalue to know if there is more than one cluster
- param: parameters called
- type: parameter passed to other functions

References

Llobell, F., Vigneau, E. & Qannari, E. M. ((September 14, 2022). Multivariate data analysis and clustering of subjects in a Just about right task. Eurosense, Turku, Finland.

See Also

[plot.cluscata](#), [summary.cluscata](#), [catatis_jar](#), [preprocess_JAR](#), [cluscata_kmeans_jar](#)

Examples

```
data(cheese)
res=cluscata_jar(Data=cheese, nprod=8, nsub=72, levelsJAR=5)
#plot(res, ngroups=4, Graph_dend=FALSE)
summary(res, ngroups=4)
```

cluscata_kmeans	<i>Compute the CLUSCATA partitionning algorithm on different blocks from a CATA experiment. Can be performed using a multi start strategy or initial partition provided by the user</i>
-----------------	---

Description

Partitionning of binary Blocks from a CATA experiment. Each cluster is associated with a compromise computed by the CATATIS method. Moreover, a noise cluster can be set up.

Usage

```
cluscata_kmeans(Data,nblo, clust, nstart=100, rho=0, NameBlocks=NULL, NameVar=NULL,
                Itermax=30, Graph_groups=TRUE, print_attempt=FALSE, Warnings=FALSE)
```

Arguments

Data	data frame or matrix where the blocks of binary variables are merged horizontally. If you have a different format, see change_cata_format
nblo	numerical. Number of blocks (subjects).
clust	numerical vector or integer. Initial partition or number of starting partitions if integer. If numerical vector, the numbers must be 1,2,3,...,number of clusters
nstart	numerical. Number of starting partitions. Default: 100
rho	numerical between 0 and 1. Threshold for the noise cluster. If 0, there is no noise cluster. Default: 0
NameBlocks	string vector. Name of each block. Length must be equal to the number of blocks. If NULL, the names are S1,...Sm. Default: NULL
NameVar	string vector. Name of each variable (attribute, the same names for each subject). Length must be equal to the number of attributes. If NULL, the colnames of the first block are taken. Default: NULL
Itermax	numerical. Maximum of iterations by partitioning algorithm. Default: 30
Graph_groups	logical. Should each cluster compromise graphical representation be plotted? Default: TRUE
print_attempt	logical. Print the number of remaining attempts in multi-start case? Default: FALSE
Warnings	logical. Display warnings about the fact that none of the subjects in some clusters checked an attribute or product? Default: FALSE

Value

a list with:

- group: the clustering partition. If $\rho > 0$, some subjects could be in the noise cluster ("K+1")
- rho: the threshold for the noise cluster
- homogeneity: percentage of homogeneity of the subjects in each cluster and the overall homogeneity
- s_with_compromise: Similarity coefficient of each subject with its cluster compromise
- weights: weight associated with each subject in its cluster
- compromise: The compromise of each cluster
- CA: The correspondance analysis results on each cluster compromise (coordinates, contributions...)
- inertia: percentage of total variance explained by each axis of the CA for each cluster
- s_all_cluster: the similarity coefficient between each subject and each cluster compromise
- param: parameters called
- criterion: the CLUSCATA criterion error
- type: parameter passed to other functions

References

Llobell, F., Cariou, V., Vigneau, E., Labenne, A., & Qannari, E. M. (2019). A new approach for the analysis of data and the clustering of subjects in a CATA experiment. *Food Quality and Preference*, 72, 31-39.

Llobell, F., Giacalone, D., Labenne, A., Qannari, E.M. (2019). Assessment of the agreement and cluster analysis of the respondents in a CATA experiment. *Food Quality and Preference*, 77, 184-190.

See Also

[plot.cluscata](#) , [summary.cluscata](#), [catatis](#), [cluscata](#), [change_cata_format](#)

Examples

```
data(straw)
cl_km=cluscata_kmeans(Data=straw[,1:(16*40)], nblo=40, clust=3)
#plot(cl_km, Graph_groups=FALSE, Graph_weights = TRUE)
summary(cl_km)
```

`cluscata_kmeans_jar` *Perform a cluster analysis of subjects in a JAR experiment.*

Description

Partitionning of subject from a JAR experiment. Each cluster is associated with a compromise computed by the CATATIS method. Moreover, a noise cluster can be set up.

Usage

```
cluscata_kmeans_jar(Data, nprod, nsub, levelsJAR=3, beta=0.1, clust, nstart=100, rho=0,
  Itermax=30, Graph_groups=TRUE, print_attempt=FALSE, Warnings=FALSE)
```

Arguments

Data	data frame where the first column is the Assessors, the second is the products and all other columns the JAR attributes with numbers (1 to 3 or 1 to 5, see levelsJAR)
nprod	integer. Number of products.
nsub	integer. Number of subjects.
levelsJAR	integer. 3 or 5 levels. If 5, the data will be transformed in 3 levels.
beta	numerical. Parameter for agreement between JAR and other answers. Between 0 and 0.5.

clust	numerical vector or integer. Initial partition or number of starting partitions if integer. If numerical vector, the numbers must be 1,2,3,...,number of clusters
nstart	numerical. Number of starting partitions. Default: 100
rho	numerical between 0 and 1. Threshold for the noise cluster. If 0, there is no noise cluster. Default: 0
Itermax	numerical. Maximum of iterations by partitioning algorithm. Default: 30
Graph_groups	logical. Should each cluster compromise graphical representation be plotted? Default: TRUE
print_attempt	logical. Print the number of remaining attempts in multi-start case? Default: FALSE
Warnings	logical. Display warnings about the fact that none of the subjects in some clusters checked an attribute or product? Default: FALSE

Value

a list with:

- group: the clustering partition. If $\rho > 0$, some subjects could be in the noise cluster ("K+1")
- rho: the threshold for the noise cluster
- homogeneity: percentage of homogeneity of the subjects in each cluster and the overall homogeneity
- s_with_compromise: Similarity coefficient of each subject with its cluster compromise
- weights: weight associated with each subject in its cluster
- compromise: The compromise of each cluster
- CA: The correspondance analysis results on each cluster compromise (coordinates, contributions...)
- inertia: percentage of total variance explained by each axis of the CA for each cluster
- s_all_cluster: the similarity coefficient between each subject and each cluster compromise
- param: parameters called
- criterion: the CLUSCATA criterion error
- type: parameter passed to other functions

References

Llobell, F., Vigneau, E. & Qannari, E. M. ((September 14, 2022). Multivariate data analysis and clustering of subjects in a Just about right task. Eurosense, Turku, Finland.

See Also

[plot.cluscata](#), [summary.cluscata](#), [catatis_jar](#), [preprocess_JAR](#), [cluscata_jar](#)

Examples

```
data(cheese)
res=cluscata_kmeans_jar(Data=cheese, nprod=8, nsub=72, levelsJAR=5, clust=4)
#plot(res)
summary(res)
```

clustatis

Perform a cluster analysis of blocks of quantitative variables

Description

Hierarchical clustering of quantitative Blocks followed by a partitioning algorithm (consolidation). Each cluster of blocks is associated with a compromise computed by the STATIS method. Moreover, a noise cluster can be set up.

Usage

```
clustatis(Data,Blocks,NameBlocks=NULL,Noise_cluster=FALSE,scale=FALSE,
  Itermax=30, Graph_dend=TRUE, Graph_bar=TRUE,
  printlevel=FALSE, gpmx=min(6, length(Blocks)-2), Testonlyoneclust=FALSE,
  alpha=0.05, nperm=50)
```

Arguments

Data	data frame or matrix. Correspond to all the blocks of variables merged horizontally
Blocks	numerical vector. The number of variables of each block. The sum must be equal to the number of columns of Data
NameBlocks	string vector. Name of each block. Length must be equal to the length of Blocks vector. If NULL, the names are B1,...Bm. Default: NULL
Noise_cluster	logical. Should a noise cluster be computed? Default: FALSE
scale	logical. Should the data variables be scaled? Default: FALSE
Itermax	numerical. Maximum of iteration for the partitioning algorithm. Default: 30
Graph_dend	logical. Should the dendrogram be plotted? Default: TRUE
Graph_bar	logical. Should the barplot of the difference of the criterion and the barplot of the overall homogeneity at each merging step of the hierarchical algorithm be plotted? Default: TRUE
printlevel	logical. Print the number of remaining levels during the hierarchical clustering algorithm? Default: FALSE
gpmx	logical. What is maximum number of clusters to consider? Default: min(6, length(Blocks)-2)

Testonlyoneclust	logical. Test if there is more than one cluster? Default: FALSE
alpha	numerical between 0 and 1. What is the threshold to test if there is more than one cluster? Default: 0.05
nperm	numerical. How many permutations are required to test if there is more than one cluster? Default: 50

Value

Each partitionK contains a list for each number of clusters of the partition, K=1 to gpmx with:

- group: the clustering partition of datasets after consolidation. If Noise_cluster=TRUE, some blocks could be in the noise cluster ("K+1")
- rho: the threshold for the noise cluster
- homogeneity: homogeneity index (
- rv_with_compromise: RV coefficient of each block with its cluster compromise
- weights: weight associated with each block in its cluster
- comp_RV: RV coefficient between the compromises associated with the various clusters
- compromise: the W compromise of each cluster
- coord: the coordinates of objects of each cluster
- inertia: percentage of total variance explained by each axis for each cluster
- rv_all_cluster: the RV coefficient between each block and each cluster compromise
- criterion: the CLUSTATIS criterion error
- param: parameters called in the consolidation
- type: parameter passed to other functions

There is also at the end of the list:

- dend: The CLUSTATIS dendrogram
- cutree_k: the partition obtained by cutting the dendrogram for K clusters (before consolidation).
- overall_homogeneity_ng: percentage of overall homogeneity by number of clusters before consolidation (and after if there is no noise cluster)
- diff_crit_ng: variation of criterion when a merging is done before consolidation (and after if there is no noise cluster)
- test_one_cluster: decision and pvalue to know if there is more than one cluster
- param: parameters called
- type: parameter passed to other functions

References

- Llobell, F., Cariou, V., Vigneau, E., Labenne, A., & Qannari, E. M. (2018). Analysis and clustering of multiblock datasets by means of the STATIS and CLUSTATIS methods. Application to sensometrics. *Food Quality and Preference*, in Press.
- Llobell, F., Vigneau, E., Qannari, E. M. (2019). Clustering datasets by means of CLUSTATIS with identification of atypical datasets. Application to sensometrics. *Food Quality and Preference*, 75, 97-104.

See Also

[plot.clustatis](#), [summary.clustatis](#), [clustatis_kmeans](#), [statis](#)

Examples

```
data(smoo)
NameBlocks=paste0("S",1:24)
cl=clustatis(Data=smoo,Blocks=rep(2,24),NameBlocks = NameBlocks)
#plot(cl, ngroups=3, Graph_dend=FALSE)
summary(cl)
#with noise cluster
cl2=clustatis(Data=smoo,Blocks=rep(2,24),NameBlocks = NameBlocks,
Noise_cluster=TRUE, Graph_dend=FALSE, Graph_bar=FALSE)
```

clustatis_FreeSort *Perform a cluster analysis of free sorting data*

Description

Hierarchical clustering of free sorting data followed by a partitioning algorithm (consolidation). Each cluster of blocks is associated with a compromise computed by the STATIS method. Moreover, a noise cluster can be set up.

Usage

```
clustatis_FreeSort(Data, NameSub=NULL, Noise_cluster=FALSE,Itermax=30,
Graph_dend=TRUE, Graph_bar=TRUE, printlevel=FALSE,
gpmx=min(6, ncol(Data)-1), Testonlyoneclust=FALSE,
alpha=0.05, nperm=50)
```

Arguments

Data	data frame or matrix. Corresponds to all variables that contain subjects results. Each column corresponds to a subject and gives the groups to which the products (rows) are assigned
NameSub	string vector. Name of each subject. Length must be equal to the number of column of the Data. If NULL, the names are S1,...Sm. Default: NULL
Noise_cluster	logical. Should a noise cluster be computed? Default: FALSE
Itermax	numerical. Maximum of iteration for the partitioning algorithm. Default: 30
Graph_dend	logical. Should the dendrogram be plotted? Default: TRUE
Graph_bar	logical. Should the barplot of the difference of the criterion and the barplot of the overall homogeneity at each merging be plotted? Default: FALSE
printlevel	logical. Print the number of remaining levels during the hierarchical clustering algorithm? Default: FALSE

gpmx	logical. What is maximum number of clusters to consider? Default: min(6, ncol(Data)-1)
Testonlyoneclust	logical. Test if there is more than one cluster? Default: FALSE
alpha	numerical between 0 and 1. What is the threshold to test if there is more than one cluster? Default: 0.05
nperm	numerical. How many permutations are required to test if there is more than one cluster? Default: 50

Value

Each partitionK contains a list for each number of clusters of the partition, K=1 to gpmx with:

- group: the clustering partition of subjects after consolidation. If Noise_cluster=TRUE, some subjects could be in the noise cluster ("K+1")
- rho: the threshold for the noise cluster
- homogeneity: homogeneity index (
- rv_with_compromise: RV coefficient of each block with its cluster compromise
- weights: weight associated with each subject in its cluster
- comp_RV: RV coefficient between the compromises associated with the various clusters
- compromise: the W compromise of each cluster
- coord: the coordinates of objects of each cluster
- inertia: percentage of total variance explained by each axis for each cluster
- rv_all_cluster: the RV coefficient between each subject and each cluster compromise
- criterion: the CLUSTATIS criterion error
- param: parameters called in the consolidation
- type: parameter passed to other functions

There is also at the end of the list:

- dend: The CLUSTATIS dendrogram
- cutree_k: the partition obtained by cutting the dendrogram for K clusters (before consolidation).
- overall_homogeneity_ng: percentage of overall homogeneity by number of clusters before consolidation (and after if there is no noise cluster)
- diff_crit_ng: variation of criterion when a merging is done before consolidation (and after if there is no noise cluster)
- test_one_cluster: decision and pvalue to know if there is more than one cluster
- param: parameters called
- type: parameter passed to other functions

References

Llobell, F., Cariou, V., Vigneau, E., Labenne, A., & Qannari, E. M. (2018). Analysis and clustering of multiblock datasets by means of the STATIS and CLUSTATIS methods. Application to sensometrics. *Food Quality and Preference*, in Press.

Llobell, F., Vigneau, E., Qannari, E. M. (2019). Clustering datasets by means of CLUSTATIS with identification of atypical datasets. Application to sensometrics. *Food Quality and Preference*, 75, 97-104.

See Also

[clustatis](#), [preprocess_FreeSort](#), [summary.clustatis](#), [plot.clustatis](#)

Examples

```
data(choc)
res.clu=clustatis_FreeSort(choc)
plot(res.clu, Graph_dend=FALSE)
summary(res.clu)
```

clustatis_FreeSort_kmeans

Compute the CLUSTATIS partitionning algorithm on free sorting data

Description

Partitionning algorithm for Free Sorting data. Each cluster is associated with a compromise computed by the STATIS method. Moreover, a noise cluster can be set up.

Usage

```
clustatis_FreeSort_kmeans(Data, NameSub=NULL, clust, nstart=100, rho=0, Itermax=30,
  Graph_groups=TRUE, Graph_weights=FALSE, print_attempt=FALSE)
```

Arguments

Data	data frame or matrix. Corresponds to all variables that contain subjects results. Each column corresponds to a subject and gives the groups to which the products (rows) are assigned
NameSub	string vector. Name of each subject. Length must be equal to the number of column of the Data. If NULL, the names are S1,...Sm. Default: NULL
clust	numerical vector or integer. Initial partition or number of starting partitions if integer. If numerical vector, the numbers must be 1,2,3,...,number of clusters
nstart	integer. Number of starting partitions. Default: 100
rho	numerical between 0 and 1. Threshold for the noise cluster. Default:0
Itermax	numerical. Maximum of iterations by partitionning algorithm. Default: 30

Graph_groups	logical. Should each cluster compromise be plotted? Default: TRUE
Graph_weights	logical. Should the barplot of the weights in each cluster be plotted? Default: FALSE
print_attempt	logical. Print the number of remaining attempts in the multi-start case? Default: FALSE

Value

a list with:

- group: the clustering partition. If $\rho > 0$, some subjects could be in the noise cluster ("K+1")
- rho: the threshold for the noise cluster
- homogeneity: percentage of homogeneity of the subjects in each cluster and the overall homogeneity
- rv_with_compromise: RV coefficient of each subject with its cluster compromise
- weights: weight associated with each subject in its cluster
- comp_RV: RV coefficient between the compromises associated with the various clusters
- compromise: the W compromise of each cluster
- coord: the coordinates of objects of each cluster
- inertia: percentage of total variance explained by each axis for each cluster
- rv_all_cluster: the RV coefficient between each subject and each cluster compromise
- criterion: the CLUSTATIS criterion error
- param: parameters called
- type: parameter passed to other functions

References

- Llobell, F., Cariou, V., Vigneau, E., Labenne, A., & Qannari, E. M. (2018). Analysis and clustering of multiblock datasets by means of the STATIS and CLUSTATIS methods. Application to sensometrics. *Food Quality and Preference*, in Press.
- Llobell, F., Vigneau, E., Qannari, E. M. (2019). Clustering datasets by means of CLUSTATIS with identification of atypical datasets. Application to sensometrics. *Food Quality and Preference*, 75, 97-104.

See Also

[clustatis_FreeSort](#), [preprocess_FreeSort](#), [summary.clustatis](#), [plot.clustatis](#)

Examples

```
data(choc)
res.clu=clustatis_FreeSort_kmeans(choc, clust=2)
plot(res.clu, Graph_groups=FALSE, Graph_weights=TRUE)
summary(res.clu)
```

clustatis_kmeans	<i>Compute the CLUSTATIS partitioning algorithm on different blocks of quantitative variables. Can be performed using a multi start strategy or initial partition provided by the user</i>
------------------	--

Description

Partitioning algorithm for quantitative variables. Each cluster is associated with a compromise computed by the STATIS method. Moreover, a noise cluster can be set up.

Usage

```
clustatis_kmeans(Data, Blocks, clust, nstart=100, rho=0, NameBlocks=NULL,
  Itermax=30, Graph_groups=TRUE, Graph_weights=FALSE,
  scale=FALSE, print_attempt=FALSE)
```

Arguments

Data	data frame or matrix. Correspond to all the blocks of variables merged horizontally
Blocks	numerical vector. The number of variables of each block. The sum must be equal to the number of columns of Data
clust	numerical vector or integer. Initial partition or number of starting partitions if integer. If numerical vector, the numbers must be 1,2,3,...,number of clusters
nstart	integer. Number of starting partitions. Default: 100
rho	numerical between 0 and 1. Threshold for the noise cluster. Default:0
NameBlocks	string vector. Name of each block. Length must be equal to the length of Blocks vector. If NULL, the names are B1,...,Bm. Default: NULL
Itermax	numerical. Maximum of iterations by partitioning algorithm. Default: 30
Graph_groups	logical. Should each cluster compromise be plotted? Default: TRUE
Graph_weights	logical. Should the barplot of the weights in each cluster be plotted? Default: FALSE
scale	logical. Should the data variables be scaled? Default: FALSE
print_attempt	logical. Print the number of remaining attempts in the multi-start case? Default: FALSE

Value

a list with:

- group: the clustering partition. If $\rho > 0$, some blocks could be in the noise cluster ("K+1")
- rho: the threshold for the noise cluster
- homogeneity: percentage of homogeneity of the blocks in each cluster and the overall homogeneity

- `rv_with_compromise`: RV coefficient of each block with its cluster compromise
- `weights`: weight associated with each block in its cluster
- `comp_RV`: RV coefficient between the compromises associated with the various clusters
- `compromise`: the W compromise of each cluster
- `coord`: the coordinates of objects of each cluster
- `inertia`: percentage of total variance explained by each axis for each cluster
- `rv_all_cluster`: the RV coefficient between each block and each cluster compromise
- `criterion`: the CLUSTATIS criterion error
- `param`: parameters called
- `type`: parameter passed to other functions

References

Llobell, F., Cariou, V., Vigneau, E., Labenne, A., & Qannari, E. M. (2018). Analysis and clustering of multiblock datasets by means of the STATIS and CLUSTATIS methods. Application to sensometrics. *Food Quality and Preference*, in Press.

Llobell, F., Vigneau, E., Qannari, E. M. (2019). Clustering datasets by means of CLUSTATIS with identification of atypical datasets. Application to sensometrics. *Food Quality and Preference*, 75, 97-104.

See Also

[plot.clustatis](#), [clustatis](#), [summary.clustatis](#), [statis](#)

Examples

```
data(smoo)
NameBlocks=paste0("S",1:24)
#with multi-start
cl_km=clustatis_kmeans(Data=smoo,Blocks=rep(2,24),NameBlocks = NameBlocks, clust=3)
#with an initial partition
cl=clustatis(Data=smoo,Blocks=rep(2,24),NameBlocks = NameBlocks,
Graph_dend=FALSE)
partition=cl$cutree_k$partition3
cl_km2=clustatis_kmeans(Data=smoo,Blocks=rep(2,24),NameBlocks = NameBlocks,
clust=partition, Graph_weights=FALSE, Graph_groups=FALSE)
graphics.off()
```

consistency_cata	<i>Test the consistency of each attribute in a CATA experiment</i>
------------------	--

Description

Permutation test on the agreement between subjects for each attribute in a CATA experiment

Usage

```
consistency_cata(Data,nblo, nperm=100, alpha=0.05, printAttrTest=FALSE)
```

Arguments

Data	data frame or matrix. Correspond to all the blocks of variables merged horizontally
nblo	numerical. Number of blocks (subjects).
nperm	numerical. How many permutations are required? Default: 100
alpha	numerical between 0 and 1. What is the threshold? Default: 0.05
printAttrTest	logical. Print the number of remaining attributes to be tested? Default: FALSE

Value

a list with:

- consist: the consistent attributes
- no_consist: the inconsistent attributes
- pval: pvalue for each test

References

Llobell, F., Giacalone, D., Labenne, A., Qannari, E.M. (2019). Assessment of the agreement and cluster analysis of the respondents in a CATA experiment. *Food Quality and Preference*, 77, 184-190.

See Also

[consistency_cata_panel](#), [change_cata_format](#), [change_cata_format2](#)

Examples

```
data(straw)
#with only 40 subjects
consistency_cata(Data=straw[,1:(16*40)], nblo=40)
#with all subjects
consistency_cata(Data=straw, nblo=114, printAttrTest=TRUE)
```

`consistency_cata_panel`*Test the consistency of the panel in a CATA experiment*

Description

Permutation test on the agreement between subjects in a CATA experiment

Usage

```
consistency_cata_panel(Data, nblo, nperm=100, alpha=0.05)
```

Arguments

<code>Data</code>	data frame or matrix. Correspond to all the blocks of variables merged horizontally
<code>nblo</code>	numerical. Number of blocks (subjects).
<code>nperm</code>	numerical. How many permutations are required? Default: 100
<code>alpha</code>	numerical between 0 and 1. What is the threshold? Default: 0.05

Value

a list with:

- `answer`: the answer of the test
- `pval`: pvalue of the test
- `dis`: distance between the homogeneity and the median of the permutations

References

Llobell, F., Giacalone, D., Labenne, A., Qannari, E.M. (2019). Assessment of the agreement and cluster analysis of the respondents in a CATA experiment. *Food Quality and Preference*, 77, 184-190. Bonnet, L., Ferney, T., Riedel, T., Qannari, E.M., Llobell, F. (September 14, 2022) .Using CATA for sensory profiling: assessment of the panel performance. Eurosense, Turku, Finland.

See Also

[consistency_cata](#), [change_cata_format](#), [change_cata_format2](#)

Examples

```
data(straw)
#with all subjects
consistency_cata_panel(Data=straw, nblo=114)
```

PB_tuna	<i>Plant based tuna data</i>
---------	------------------------------

Description

Plant based tuna data

Usage

```
data(PB_tuna)
```

Format

CATA data with sessions. A data frame with the sessions, the panelists, the products and CATA attributes.

References

Bonnet, L., Ferney, T., Riedel, T., Qannari, E.M., Llobell, F. (September 14, 2022) .Using CATA for sensory profiling: assessment of the panel performance. Eurosense, Turku, Finland.

Examples

```
data(PB_tuna)
```

plot.catatis	<i>Displays the CATATIS graphs</i>
--------------	------------------------------------

Description

This function plots the CATATIS map and CATATIS weights

Usage

```
## S3 method for class 'catatis'
plot(x, Graph=TRUE, Graph_weights=TRUE, Graph_eig=TRUE,
     axes=c(1,2), tit="CATATIS", cex=1, col.obj="blue", col.attr="red", ...)
```

Arguments

x	object of class 'catatis'
Graph	logical. Show the graphical representation? Default: TRUE
Graph_weights	logical. Should the barplot of the weights be plotted? Default: TRUE
Graph_eig	logical. Should the barplot of the eigenvalues be plotted? Only with Graph=TRUE. Default: TRUE

axes	numerical vector (length 2). Axes to be plotted
tit	string. Title for the graphical representation. Default: 'CATATIS'
cex	numerical. Numeric character expansion factor; multiplied by par("cex") yields the final character size. NULL and NA are equivalent to 1.0.
col.obj	numerical or string. Color for the objects points. Default: "blue"
col.attr	numerical or string. Color for the attributes points. Default: "red"
...	further arguments passed to or from other methods

Value

the CATATIS map

See Also

[catatis](#)

Examples

```
data(straw)
res.cat=catatis(straw, nblo=114)
plot(res.cat, Graph_weights=FALSE, axes=c(1,3))
```

plot.cluscata *Displays the CLUSCATA graphs*

Description

This function plots dendrogram, variation of the merging criterion, weights and CATATIS map of each cluster

Usage

```
## S3 method for class 'cluscata'
plot(x, ngroups=NULL, Graph_groups=TRUE, Graph_dend=TRUE,
     Graph_bar=FALSE, Graph_weights=FALSE, axes=c(1,2), cex=1,
     col.obj="blue", col.attr="red", ...)
```

Arguments

<code>x</code>	object of class 'cluscata'.
<code>ngroups</code>	number of groups to consider. Ignored for <code>cluscata_kmeans</code> results. Default: recommended number of clusters
<code>Graph_groups</code>	logical. Should each cluster compromise graphical representation be plotted? Default: TRUE
<code>Graph_dend</code>	logical. Should the dendrogram be plotted? Default: TRUE
<code>Graph_bar</code>	logical. Should the barplot of the difference of the criterion and the barplot of the overall homogeneity at each merging step of the hierarchical algorithm be plotted? Also available after consolidation if <code>Noise_cluster=FALSE</code> . Default: FALSE
<code>Graph_weights</code>	logical. Should the barplot of the weights in each cluster be plotted? Default: FALSE
<code>axes</code>	numerical vector (length 2). Axes to be plotted. Default: <code>c(1,2)</code>
<code>cex</code>	numerical. Numeric character expansion factor; multiplied by <code>par("cex")</code> yields the final character size. NULL and NA are equivalent to 1.0.
<code>col.obj</code>	numerical or string. Color for the objects points. Default: "blue"
<code>col.attr</code>	numerical or string. Color for the attributes points. Default: "red"
<code>...</code>	further arguments passed to or from other methods

Value

the CLUSCATA graphs

See Also

[cluscata](#), [cluscata_kmeans](#)

Examples

```
data(straw)
res=cluscata(Data=straw[,1:(16*40)], nblo=40)
plot(res, ngroups=3, Graph_dend=FALSE)
plot(res, ngroups=3, Graph_dend=FALSE, Graph_bar=FALSE, Graph_weights=FALSE, axes=c(1,3))
```

plot.clustatis	<i>Displays the CLUSTATIS graphs</i>
----------------	--------------------------------------

Description

This function plots dendrogram, variation of the merging criterion, weights and STATIS map of each cluster

Usage

```
## S3 method for class 'clustatis'
plot(x, ngroups=NULL, Graph_groups=TRUE, Graph_dend=TRUE,
     Graph_bar=FALSE, Graph_weights=FALSE, axes=c(1,2), col=NULL, cex=1, font=1, ...)
```

Arguments

x	object of class 'clustatis'.
ngroups	number of groups to consider. Ignored for clustatis_kmeans results. Default: recommended number of clusters
Graph_groups	logical. Should each cluster compromise graphical representation be plotted? Default: TRUE
Graph_dend	logical. Should the dendrogram be plotted? Default: TRUE
Graph_bar	logical. Should the barplot of the difference of the criterion and the barplot of the overall homogeneity at each merging step of the hierarchical algorithm be plotted? Also available after consolidation if Noise_cluster=FALSE. Default: FALSE
Graph_weights	logical. Should the barplot of the weights in each cluster be plotted? Default: FALSE
axes	numerical vector (length 2). Axes to be plotted. Default: c(1,2)
col	vector. Color for each object. Default: rainbow(nrow(Data))
cex	numerical. Numeric character expansion factor; multiplied by par("cex") yields the final character size. NULL and NA are equivalent to 1.0.
font	numerical. Integer specifying font to use for text. 1=plain, 2=bold, 3=italic, 4=bold italic, 5=symbol. Default: 1
...	further arguments passed to or from other methods

Value

the CLUSTATIS graphs

See Also

[clustatis](#), [clustatis_kmeans](#)

Examples

```
data(smoo)
NameBlocks=paste0("S",1:24)
cl=clustatis(Data=smoo,Blocks=rep(2,24),NameBlocks = NameBlocks)
plot(cl, ngroups=3, Graph_dend=FALSE)
plot(cl, ngroups=3, Graph_dend=FALSE, axes=c(1,3))
graphics.off()
```

plot.statis	<i>Displays the STATIS graphs</i>
-------------	-----------------------------------

Description

This function plots the STATIS map and STATIS weights

Usage

```
## S3 method for class 'statis'
plot(x, axes=c(1,2), Graph_obj=TRUE,
Graph_weights=TRUE, Graph_eig=TRUE, tit="STATIS", col=NULL, cex=1, font=1,
xlim=NULL, ylim=NULL, ...)
```

Arguments

x	object of class 'statis'
axes	numerical vector (length 2). Axes to be plotted. Default: c(1,2)
Graph_obj	logical. Should the compromise graphical representation be plotted? Default: TRUE
Graph_weights	logical. Should the barplot of the weights be plotted? Default: TRUE
Graph_eig	logical. Should the barplot of the eigenvalues be plotted? Only with Graph_obj=TRUE. Default: TRUE
tit	string. Title for the objects graphical representation. Default: 'STATIS'
col	vector. Color for each object. If NULL, col=rainbow(nrow(Data)). Default: NULL
cex	numerical. Numeric character expansion factor; multiplied by par("cex") yields the final character size. NULL and NA are equivalent to 1.0.
font	numerical. Integer specifying font to use for text. 1=plain, 2=bold, 3=italic, 4=bold italic, 5=symbol. Default: 1
xlim	numerical vector (length 2). Minimum and maximum for x coordinates.
ylim	numerical vector (length 2). Minimum and maximum for y coordinates.
...	further arguments passed to or from other methods

Value

the STATIS graphs

See Also

[statis](#)

Examples

```
data(smoo)
NameBlocks=paste0("S",1:24)
st=statis(Data=smoo,Blocks=rep(2,24),NameBlocks = NameBlocks)
plot(st, axes=c(1,3), Graph_weights=FALSE)
```

preprocess_FreeSort *Preprocessing for Free Sorting Data*

Description

For Free Sorting Data, this preprocessing is needed.

Usage

```
preprocess_FreeSort(Data, NameSub=NULL)
```

Arguments

Data	data frame or matrix. Corresponds to all variables that contain subjects results. Each column corresponds to a subject and gives the groups to which the products (rows) are assigned
NameSub	string vector. Name of each subject. Length must be equal to the number of column of the Data. If NULL, the names are S1,...Sm. Default: NULL

Value

A list with:

- new_Data: the Data transformed
- Blocks: the number of groups for each subject
- NameBlocks: the name of each subject

References

Llobell, F., Cariou, V., Vigneau, E., Labenne, A., & Qannari, E. M. (2018). Analysis and clustering of multiblock datasets by means of the STATIS and CLUSTATIS methods. Application to sensometrics. Food Quality and Preference, in Press.

See Also

[clustatis](#), [clustatis_FreeSort](#)

Examples

```
data(choc)
prepro=preprocess_FreeSort(choc)
```

```
preprocess_JAR
```

Preprocessing for Just About Right Data

Description

For JAR data, this preprocessing is needed.

Usage

```
preprocess_JAR(Data, nprod, nsub, levelsJAR=3, beta=0.1)
```

Arguments

Data	data frame where the first column is the Assessors, the second is the products and all other columns the JAR attributes with numbers (1 to 3 or 1 to 5, see levelsJAR)
nprod	integer. Number of products.
nsub	integer. Number of subjects.
levelsJAR	integer. 3 or 5 levels. If 5, the data will be transformed in 3 levels.
beta	numerical. Parameter for agreement between JAR and other answers. Between 0 and 0.5.

Value

A list with:

- Datafinal: the Data transformed
- NameSub: the name of each subject in the right order

References

Llobell, F., Vigneau, E. & Qannari, E. M. (September 14, 2022). Multivariate data analysis and clustering of subjects in a Just about right task. Eurosense, Turku, Finland.

See Also

[catatis_jar](#), [cluscata_jar](#), [cluscata_kmeans_jar](#)

Examples

```
data(cheese)
prepro=preprocess_JAR(cheese, nprod=8, nsub=72, levelsJAR=5)
```

print.catatis *Print the CATATIS results*

Description

Print the CATATIS results

Usage

```
## S3 method for class 'catatis'
print(x, ...)
```

Arguments

x object of class 'catatis'
... further arguments passed to or from other methods

See Also

[catatis](#)

print.cluscata *Print the CLUSCATA results*

Description

Print the CLUSCATA results

Usage

```
## S3 method for class 'cluscata'
print(x, ...)
```

Arguments

x object of class 'cluscata'
... further arguments passed to or from other methods

See Also

[cluscata](#), [cluscata_kmeans](#)

print.clustatis *Print the CLUSTATIS results*

Description

Print the CLUSTATIS results

Usage

```
## S3 method for class 'clustatis'  
print(x, ...)
```

Arguments

x object of class 'clustatis'
... further arguments passed to or from other methods

See Also

[clustatis](#), [clustatis_kmeans](#)

print.stat1s *Print the STATIS results*

Description

Print the STATIS results

Usage

```
## S3 method for class 'stat1s'  
print(x, ...)
```

Arguments

x object of class 'stat1s'
... further arguments passed to or from other methods

See Also

[stat1s](#)

simil_groups_cata	<i>Testing the difference in perception between two predetermined groups of subjects in a CATA experiment</i>
-------------------	---

Description

Test adapted to CATA data to determine whether two predetermined groups of subjects have a different perception or not. For example, men and women.

Usage

```
simil_groups_cata(Data, groups, one=1, two=2, nperm=50, Graph=TRUE,
  alpha= 0.05, printl=FALSE)
```

Arguments

Data	data frame or matrix. Correspond to all the blocks of variables merged horizontally
groups	categorical vector. The groups of each subject . The length must be the number of subjects.
one	string. Name of the group 1 in groups vector.
two	string. Name of the group 2 in groups vector.
nperm	numerical. How many permutations are required? Default: 50
Graph	logical. Should the CATATIS graph of each group be plotted? Default: TRUE
alpha	numerical between 0 and 1. What is the threshold of the test? Default: 0.05
printl	logical. Print the number of remaining permutations during the algorithm? Default: FALSE

Value

a list with:

- decision: the decision of the test
- pval: pvalue of the test

References

Llobell, F., Giacalone, D., Jaeger, S.R. & Qannari, E. M. (2021). CATA data: Are there differences in perception? JSM conference.

Llobell, F., Giacalone, D., Jaeger, S.R. & Qannari, E. M. (2021). CATA data: Are there differences in perception? AgroStat conference.

Examples

```
data(straw)
groups=sample(1:2, 114, replace=TRUE)
simil_groups_cata(straw, groups, one=1, two=2)
```

smoo	<i>smoothies data</i>
------	-----------------------

Description

smoothies data

Usage

```
data(smoo)
```

Format

Projective mapping (or Napping) data. A data frame with 8 rows (the number of smoothies) and 48 columns (the number of consumers * 2). For each consumer, we have the coordinates of the products on the sheet of paper.

References

Francois Husson, Sebastien Le and Marine Cadoret (2017). *SensoMineR: Sensory Data Analysis*. R package version 1.23. <https://CRAN.R-project.org/package=SensoMineR>

Examples

```
data(smoo)
```

statis	<i>Performs the STATIS method on different blocks of quantitative variables</i>
--------	---

Description

STATIS method on quantitative blocks. SUPplementary outputs are also computed

Usage

```
statis(Data,Blocks,NameBlocks=NULL,Graph_obj=TRUE, Graph_weights=TRUE, scale=FALSE)
```

Arguments

Data	data frame or matrix. Correspond to all the blocks of variables merged horizontally
Blocks	numerical vector. The number of variables of each block. The sum must be equal to the number of columns of Data
NameBlocks	string vector. Name of each block. Length must be equal to the length of Blocks vector. If NULL, the names are B1,...Bm. Default: NULL
Graph_obj	logical. Show the graphical representation of the objects? Default: TRUE
Graph_weights	logical. Should the barplot of the weights be plotted? Default: TRUE
scale	logical. Should the data variables be scaled? Default: FALSE

Value

a list with:

- RV: the RV matrix: a matrix with the RV coefficient between blocks of variables
- compromise: a matrix which is the compromise of the blocks (akin to a weighted average)
- weights: the weights associated with the blocks to build the compromise
- lambda: the first eigenvalue of the RV matrix
- overall error : the error for the STATIS criterion
- error_by_conf: the error by configuration (STATIS criterion)
- rv_with_compromise: the RV coefficient of each block with the compromise
- homogeneity: homogeneity of the blocks (in percentage)
- coord: the coordinates of each object
- eigenvalues: the eigenvalues of the svd decomposition
- inertia: the percentage of total variance explained by each axis
- error_by_obj: the error by object (STATIS criterion)
- scalefactors: the scaling factors of each block
- proj_config: the projection of each object of each configuration on the axes: presentation by configuration
- proj_objects: the projection of each object of each configuration on the axes: presentation by object

References

- Lavit, C., Escoufier, Y., Sabatier, R., Traissac, P. (1994). The act (statis method). Computational 462 Statistics & Data Analysis, 18 (1), 97-119.
- Llobell, F., Cariou, V., Vigneau, E., Labenne, A., & Qannari, E. M. (2018). Analysis and clustering of multiblock datasets by means of the STATIS and CLUSTATIS methods. Application to sensometrics. Food Quality and Preference, in Press.

See Also

[plot.statis](#), [clustatis](#)

Examples

```

data(smoo)
NameBlocks=paste0("S",1:24)
st=statis(Data=smoo, Blocks=rep(2,24),NameBlocks = NameBlocks)
#plot(st, axes=c(1,3))
summary(st)
#with variables scaling
st2=statis(Data=smoo, Blocks=rep(2,24),NameBlocks = NameBlocks, Graph_weights=FALSE, scale=TRUE)

```

statis_FreeSort	<i>Performs the STATIS method on Free Sorting data</i>
-----------------	--

Description

STATIS method on Free Sorting data. A lot of supplementary informations are also computed

Usage

```
statis_FreeSort(Data, NameSub=NULL, Graph_obj=TRUE, Graph_weights=TRUE)
```

Arguments

Data	data frame or matrix. Corresponds to all variables that contain subjects results. Each column corresponds to a subject and gives the groups to which the products (rows) are assigned
NameSub	string vector. Name of each subject. Length must be equal to the number of column of the Data. If NULL, the names are S1,...Sm. Default: NULL
Graph_obj	logical. Show the graphical representation of the objects? Default: TRUE
Graph_weights	logical. Should the barplot of the weights be plotted? Default: TRUE

Value

a list with:

a list with:

- RV: the RV matrix: a matrix with the RV coefficient between subjects
- compromise: a matrix which is the compromise of the subjects (akin to a weighted average)
- weights: the weights associated with the subjects to build the compromise
- lambda: the first eigenvalue of the RV matrix
- overall error : the error for the STATIS criterion
- error_by_conf: the error by configuration (STATIS criterion)
- rv_with_compromise: the RV coefficient of each subject with the compromise

- homogeneity: homogeneity of the subjects (in percentage)
- coord: the coordinates of each object
- eigenvalues: the eigenvalues of the svd decomposition
- inertia: the percentage of total variance explained by each axis
- error_by_obj: the error by object (STATIS criterion)
- scalefactors: the scaling factors of each subject
- proj_config: the projection of each object of each subject on the axes: presentation by subject
- proj_objects: the projection of each object of each subject on the axes: presentation by object

References

- Lavit, C., Escoufier, Y., Sabatier, R., Traissac, P. (1994). The act (statis method). Computational 462 Statistics & Data Analysis, 18 (1), 97-119.\
- Llobell, F., Cariou, V., Vigneau, E., Labenne, A., & Qannari, E. M. (2018). Analysis and clustering of multiblock datasets by means of the STATIS and CLUSTATIS methods. Application to sensometrics. Food Quality and Preference, in Press.

See Also

[preprocess_FreeSort](#), [clustatis_FreeSort](#)

Examples

```
data(choc)
res.sta=statis_FreeSort(choc)
```

straw

strawberries data

Description

strawberries data

Usage

```
data(straw)
```

Format

CATA data. A data frame with 6 rows (the number of strawberries) and 1824 columns (the number of consumers (114) * the number of attributes (16)). For each consumer, each attribute and each product, there is 1 if the attribute has been checked by the consumer for the product, and 0 if not.

References

Ares, G., & Jaeger, S. R. (2013). Check-all-that-apply questions: Influence of attribute order on sensory product characterization. *Food Quality and Preference*, 28(1), 141-153.

Examples

```
data(straw)
```

summary.catatis	<i>Show the CATATIS results</i>
-----------------	---------------------------------

Description

This function shows the CATATIS results

Usage

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

Arguments

object	object of class 'catatis'.
...	further arguments passed to or from other methods

Value

a list with:

- homogeneity: homogeneity of the subjects (in percentage)
- weights: the weights associated with the subjects to build the compromise
- eigenvalues: the eigenvalues associated to the correspondance analysis
- inertia: the percentage of total variance explained by each axis of the CA

See Also

[catatis](#)

summary.cluscata	<i>Show the CLUSCATA results</i>
------------------	----------------------------------

Description

This function shows the cluscata results

Usage

```
## S3 method for class 'cluscata'  
summary(object, ngroups=NULL, ...)
```

Arguments

object	object of class 'cluscata'.
ngroups	number of groups to consider. Ignored for cluscata_kmeans results. Default: recommended number of clusters
...	further arguments passed to or from other methods

Value

the CLUSCATA principal results

a list with:

- group: the clustering partition
- homogeneity: homogeneity index (
- weights: weight associated with each subject in its cluster
- rho: the threshold for the noise cluster
- test_one_cluster: decision and pvalue to know if there is more than one cluster

See Also

[cluscata](#), [cluscata_kmeans](#)

summary.clustatis *Show the CLUSTATIS results*

Description

This function shows the clustatis results

Usage

```
## S3 method for class 'clustatis'  
summary(object, ngroups=NULL, ...)
```

Arguments

object	object of class 'clustatis'.
ngroups	number of groups to consider. Ignored for clustatis_kmeans results. Default: recommended number of clusters
...	further arguments passed to or from other methods

Value

the CLUSTATIS principal results

a list with:

- group: the clustering partition
- homogeneity: homogeneity index (
- weights: weight associated with each block in its cluster
- rho: the threshold for the noise cluster
- test_one_cluster: decision and pvalue to know if there is more than one cluster

See Also

[clustatis](#) , [clustatis_kmeans](#)

summary.statis	<i>Show the STATIS results</i>
----------------	--------------------------------

Description

This function shows the STATIS results

Usage

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

Arguments

object	object of class 'statis'.
...	further arguments passed to or from other methods

Value

a list with:

- homogeneity: homogeneity of the blocks (in percentage)
- weights: the weights associated with the blocks to build the compromise
- eigenvalues: the eigenvalues of the svd decomposition
- inertia: the percentage of total variance explained by each axis

See Also

[statis](#)

Index

- * **Blocks**
 - ClustBlock-package, 3
 - * **CATA**
 - catatis, 4
 - change_cata_format, 7
 - change_cata_format2, 8
 - cluscata, 10
 - cluscata_kmeans, 14
 - ClustBlock-package, 3
 - consistency_cata, 26
 - consistency_cata_panel, 27
 - plot.catatis, 28
 - plot.cluscata, 29
 - print.catatis, 35
 - print.cluscata, 35
 - simil_groups_cata, 37
 - summary.catatis, 42
 - summary.cluscata, 43
 - * **Check-All-That-Apply**
 - ClustBlock-package, 3
 - * **Cluster analysis**
 - ClustBlock-package, 3
 - * **Clustering**
 - ClustBlock-package, 3
 - * **Free Sorting**
 - ClustBlock-package, 3
 - * **FreeSorting**
 - clustatis_FreeSort, 20
 - clustatis_FreeSort_kmeans, 22
 - preprocess_FreeSort, 33
 - statis_FreeSort, 40
 - * **JAR**
 - catatis_jar, 5
 - cluscata_jar, 12
 - cluscata_kmeans_jar, 16
 - ClustBlock-package, 3
 - preprocess_JAR, 34
 - * **Just About Right**
 - ClustBlock-package, 3
 - * **Noise Cluster**
 - ClustBlock-package, 3
 - * **Profiling**
 - ClustBlock-package, 3
 - * **Projective mapping**
 - ClustBlock-package, 3
 - * **Sensory**
 - ClustBlock-package, 3
 - * **datasets**
 - cheese, 9
 - choc, 9
 - PB_tuna, 28
 - smoo, 38
 - straw, 41
 - * **quantitative**
 - clustatis, 18
 - clustatis_kmeans, 24
 - plot.clustatis, 31
 - plot.statis, 32
 - print.clustatis, 36
 - print.statis, 36
 - statis, 38
 - summary.clustatis, 44
 - summary.statis, 45
- catatis, 4, 6–8, 11, 16, 29, 35, 42
catatis_jar, 5, 14, 17, 35
change_cata_format, 4, 5, 7, 8, 10, 11, 15, 16, 26, 27
change_cata_format2, 5, 7, 8, 11, 26, 27
cheese, 9
choc, 9
cluscata, 5, 7, 8, 10, 16, 30, 36, 43
cluscata_jar, 6, 12, 17, 35
cluscata_kmeans, 11, 14, 30, 36, 43
cluscata_kmeans_jar, 6, 14, 16, 35
clustatis, 18, 22, 25, 31, 34, 36, 39, 44
clustatis_FreeSort, 20, 23, 34, 41
clustatis_FreeSort_kmeans, 22
clustatis_kmeans, 20, 24, 31, 36, 44

ClustBlock (ClustBlock-package), 3
ClustBlock-package, 3
consistency_cata, 26, 27
consistency_cata_panel, 26, 27

PB_tuna, 28
plot.catatis, 5, 6, 28
plot.cluscata, 11, 14, 16, 17, 29
plot.clustatis, 20, 22, 23, 25, 31
plot.statis, 32, 39
preprocess_FreeSort, 22, 23, 33, 41
preprocess_JAR, 6, 14, 17, 34
print.catatis, 35
print.cluscata, 35
print.clustatis, 36
print.statis, 36

simil_groups_cata, 37
smoo, 38
statis, 20, 25, 33, 36, 38, 45
statis_FreeSort, 40
straw, 41
summary.catatis, 5, 6, 42
summary.cluscata, 11, 14, 16, 17, 43
summary.clustatis, 20, 22, 23, 25, 44
summary.statis, 45