# Package 'ClustBlock'

April 23, 2020

**Title** Clustering of Datasets

**Version** 2.2.0

**Maintainer** Fabien Llobell <fabien.llobell@oniris-nantes.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>, Llobell, Vigneau & Qannari (2019) <doi: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>, Llobell, Giacalone, Labenne & Qannari (2019) <doi:10.1016/j.foodqual.2019.05.017>) are the core of this package.

**Depends** R (>= 3.4.0)

**License** GPL-3

**Encoding** UTF-8

**LazyData** true

**Imports** FactoMineR

**Suggests** ClustVarLV

**Date** 2020-04-23

**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** 2020-04-23 13:40:05 UTC

# R topics documented:

---

| catatis | *Perform the CATATIS method on different blocks of binary data from a CATA experiment* |
|---|---|

---

## Description

CATATIS method on binary blocks. Additional outputs are also computed

## Usage

```
catatis(Data,nblo,NameBlocks=NULL, NameVar=NULL, Graph=TRUE, Graph_weights=TRUE)
```

## 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 |
|---|---|
| Graph | logical. Show the graphical representation? Default: TRUE |
| Graph_weights | logical. Should the barplot of the weights be plotted? Default: TRUE |

## 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
- 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.

## See Also

plot.catatis, summary.catatis, cluscata, change_cata_format

## Examples

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

---

change_cata_format          *Change format of CATA datasets to perform CATATIS or CLUSCATA*
                            *function*

---

## Description

CATATIS and CLUSCATA operate on data where the blocks of binary variables are merged hori-
zontally. 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,... Format=2 is for
data merged vertically with the dataset for the first product, then the second...

## 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

---

choc                        *chocolates data*

---

## 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 of binary variables from a CATA experiment* |

---

## Description

Hierarchical clustering of blocks of binary data 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)

## Usage

```
cluscata(Data, nblo, NameBlocks=NULL, NameVar=NULL, Noise_cluster=FALSE,
        Itermax=30, Graph_dend=TRUE, Graph_bar=TRUE, printlevel=FALSE,
        gpmax=min(6, nblo-1), Testonlyoneclust=TRUE, alpha=0.05, nperm=50)
```

## 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 |
|---|---|
| gpmax | logical. What is maximum number of clusters to consider? Default: min(6, nblo-1) |
| Testonlyoneclust | |
| | logical. Test if there is more than one cluster? Default: TRUE |
| 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 gpmax 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

## 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)
```

---

| cluscata_kmeans | *Compute the CLUSCATA partitionning algorithm on different blocks of binary variables from a CATA experiment. Can be performed using a multi start strategy or initial partition provided by the user* |
|---|---|

---

## 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=40, rho=0, NameBlocks=NULL, NameVar=NULL,
                Itermax=30, Graph_groups=TRUE, print_attempt=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: 40 |
| 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 partitionning 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 |

**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)
summary(cl_km)
```

---

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, gpmax=min(6, length(Blocks)-1), Testonlyoneclust=TRUE,
  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 |

gpmax                    logical.  What is maximum number of clusters to consider?  Default:  min(6,
                         length(Blocks)-1)

Testonlyoneclust
                         logical. Test if there is more than one cluster? Default: TRUE

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 gpmax 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 consolida-
  tion).
- 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,
                    gpmax=min(6, ncol(Data)-1), Testonlyoneclust=TRUE,
                    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 clumn 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 |
| gpmax | 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: TRUE |
| 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 gpmax 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=40, 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 clumn 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: 40 |
| 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 senso-metrics. 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 | *Compute the CLUSTATIS partitionning algorithm on different blocks of quantitative variables. Can be performed using a multi start strategy or initial partition provided by the user* |
|---|---|

---

## Description

Partitionning 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=40, 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: 40 |
| 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 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 |
| 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)
```

## 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

## 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

| | |
|---|---|
| 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 |

### Value

a list with:

- answer: the answer of the test
- pval: pvalue of the 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](), [change_cata_format]()

### Examples

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

---

plot.catatis                    *Displays the CATATIS graphs*

---

### Description

This function plots the CATATIS map and CATATIS weights

### Usage

```
## S3 method for class 'catatis'
plot(x, Graph=TRUE, Graph_weights=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 |
| 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](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

| | |
|---|---|
| x | object of class 'cluscata'. |
| ngroups | number of groups to consider. Ignored for cluscata_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) |
| 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 CLUSCATA graphs

## See Also

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

## Examples

```
## Not run:
 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))

## End(Not run)
```

---

plot.clustatis          *Displays the CLUSTATIS graphs*

---

## 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))
```

---

plot.statis                     *Displays the STATIS graphs*

---

## 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, 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 |
| 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 clumn 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)
```

---

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.statis *Print the STATIS results*

### Description

Print the STATIS results

### Usage

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

### Arguments

x            object of class 'statis'

...          further arguments passed to or from other methods

### See Also

[statis](statis)

## smoo *smoothies data*

### 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 | *Performs the STATIS method on different blocks of quantitative variables* |
|--------|---------------------------------------------------------------------------|

## 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 od 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)
#with variables scaling
st2=statis(Data=smoo, Blocks=rep(2,24),NameBlocks = NameBlocks, Graph_weights=FALSE, scale=TRUE)
```

---

statis_FreeSort            *Performs the STATIS method on Free Sorting data*

---

## 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 clumn of the Data. If NULL, the names are S1,...Sm. Default: NULL |
| Graph_obj | logical. Show the graphical representation od 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 eachb 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                *Show the CATATIS results*

---

**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          *Show the CLUSCATA results*

---

**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                    *Show the STATIS results*

## 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