Package ‘FactoMineR’

March 29, 2013

Version 1.24

Date 2013-03-12

Title Multivariate Exploratory Data Analysis and Data Mining with R

Author Francois Husson, Julie Josse, Sebastien Le, Jeremy Mazet

Maintainer Francois Husson <francois.husson@agrocampus-ouest.fr>

Depends car, ellipse, lattice, cluster, scatterplot3d, leaps

Suggests missMDA, flashClust

Description an R package for exploratory data analysis

License GPL (>= 2)

URL http://factominer.free.fr

Encoding latin1

NeedsCompilation no

Repository CRAN

Date/Publication 2013-03-29 18:47:43

R topics documented:

AovSum ................................................................. 3
autoLab ............................................................... 4
CA ................................................................. 5
cdes ............................................................. 6
children ........................................................... 8
coeffRV ........................................................... 8
condes ........................................................... 9
coord.ellipse .................................................... 10
decathlon ......................................................... 12
descfreq ........................................................ 12
R topics documented:

dimdsc ................................................................. 13
DMFA ................................................................. 14
estim_ncp .............................................................. 16
FAMD ................................................................. 17
footsize ............................................................... 18
geomorphology ...................................................... 19
GPA ................................................................. 19
graph.var ............................................................. 21
HCPC ................................................................. 22
HMFA ................................................................. 24
hobbies ............................................................... 26
JO ................................................................. 27
MCA ................................................................. 27
MFA ................................................................. 29
milk ................................................................. 32
mortality ............................................................. 33
PCA ................................................................. 34
plot.CA .............................................................. 36
plot.catdes .......................................................... 38
plot.DMFA ........................................................... 39
plot.FAMD .......................................................... 40
plot.GPA ............................................................ 42
plot.HCPC ........................................................... 43
plot.HMFA ........................................................... 44
plot.MCA ............................................................ 45
plot.MFA ............................................................ 47
plot.PCA ............................................................ 50
plot.spMCA .......................................................... 53
plotellipses ......................................................... 54
plotGPApartial ....................................................... 56
plotMFApartial ....................................................... 57
poison .............................................................. 59
poison.text .......................................................... 60
poulet .............................................................. 60
prefpls ............................................................... 61
print.CA ............................................................. 62
print.FAMD .......................................................... 63
print.GPA ............................................................ 63
print.HCPC .......................................................... 64
print.HMFA .......................................................... 65
print.MCA ............................................................ 65
print.MFA ............................................................ 66
print.PCA ............................................................ 67
print.spMCA .......................................................... 68
reconst ............................................................ 68
RegBest .............................................................. 69
senso .............................................................. 70
simule .............................................................. 71
Analysis of variance with the contrasts sum (the sum of the coefficients is 0)

Description

Analysis of variance with the contrasts sum (the sum of the coefficients is 0)
Test for all the coefficients
Handle missing values

Usage

AovSum(formula, data, na.action = na.omit, ...)

Arguments

formula the formula for the model 'y~x1+x2+x1:x2'
data a data-frame
na.action (where relevant) information returned by model.frame on the special handling of NAs.
... other arguments, cf the function lm

Value

Retourne des objets
Ftest a table with the F-tests
Ttest a table with the t-tests

Author(s)

Francois Husson <husson@agrocampus-ouest.fr>
autoLab

Function to better position the labels on the graphs

Description

Function to better position the labels on the graphs.

Usage

```r
autoLab(x, y = NULL, labels = seq(along = x), cex = 1,
        method = c("SANN", "GA"),
        allowSmallOverlap = FALSE,
        trace = FALSE, shadotext = FALSE,
        doPlot = TRUE, ...)
```

Arguments

- `x` the x-coordinates
- `y` the y-coordinates
- `labels` the labels
- `cex` cex
- `method` not used
- `allowSmallOverlap` boolean
- `trace` boolean
- `shadotext` boolean
- `doPlot` boolean
- `...` further arguments passed to or from other methods

Examples

```r
## Example two-way anova
data(senso)
res = aovsum (Score~ Product + Day , data=senso)
res

## Example two-way anova with interaction
data(senso)
res2 = aovsum (Score~ Product + Day + Product : Day, data=senso)
res2

## Example ancova
data(footsize)
res3 = aovsum (footsize ~ size + sex + size : sex, data=footsize)
res3
```

See Also

- `aov`, `lm`
Value

See the text function

Description

Performs Correspondence Analysis (CA) including supplementary row and/or column points.

Usage

CA(X, ncp = 5, row.sup = NULL, col.sup = NULL, quanti.sup=NULL, quali.sup = NULL, graph = TRUE, axes = c(1,2), row.w = NULL)

Arguments

X a data frame with \( n \) rows and \( p \) columns, i.e. a contingency table
ncp number of dimensions kept in the results (by default 5)
row.sup a vector indicating the indexes of the supplementary rows
col.sup a vector indicating the indexes of the supplementary columns
quanti.sup a vector indicating the indexes of the supplementary continuous variables
quali.sup a vector indicating the indexes of the categorical supplementary variables
graph boolean, if TRUE a graph is displayed
axes a length 2 vector specifying the components to plot
row.w an optional row weights (by default, uniform row weights)

Value

Returns a list including:

eig a matrix containing all the eigenvalues, the percentage of variance and the cumulative percentage of variance

col a list of matrices with all the results for the column variable (coordinates, square cosine, contributions, inertia)
row a list of matrices with all the results for the row variable (coordinates, square cosine, contributions, inertia)
col.sup a list of matrices containing all the results for the supplementary column points (coordinates, square cosine)
row.sup a list of matrices containing all the results for the supplementary row points (coordinates, square cosine)
quanti.sup if quanti.sup is not NULL, a matrix containing the results for the supplementary continuous variables (coordinates, square cosine)
if quali.sup is not NULL, a list of matrices with all the results for the supplementary categorical variables (coordinates of each categories of each variables, v.test which is a criterion with a Normal distribution, square correlation ratio)

call

a list with some statistics

Returns the row and column points factor map.
The plot may be improved using the argument autolab, modifying the size of the labels or selecting some elements thanks to the plot.CA function.

Author(s)

Francois Husson <Francois.Husson@agrocampus-ouest.fr>, Jeremy Mazet

References


See Also

print.CA, summary.CA, plot.CA, dimdesc

Examples

data(children)
res.ca <- CA (children, col.sup = 6:8, row.sup = 15:18)
summary(res.ca)

catdes  Categories description

Description

Description of the categories of one factor by categorical variables and/or by quantitative variables

Usage

catdes(donnee,num.var,proba = 0.05, row.w = NULL)
Arguments

donnee  a data frame made up of at least one categorical variables and a set of quantitative variables and/or categorical variables
num.var  the indice of the variable to characterized
proba  the significance threshold considered to characterized the category (by default 0.05)
row.w  a vector of integer corresponding to an optional row weights (by default, a vector of 1 for uniform row weights)

Value

Returns a list including:

test.chi  The categorical variables which characterized the factor are listed in ascending order (from the one which characterized the most the factor to the one which significantly characterized with the proba proba

category  description of each category of the num.var by each category of all the categorical variables
quanti.var  the global description of the num.var variable by the quantititative variables with the square correlation coefficient and the p-value of the F-test in a one-way analysis of variance (assuming the hypothesis of homoscedsticity)
quanti  the description of each category of the num.var variable by the quantitative variables.

Author(s)

Francois Husson <Francois.Husson@agrocampus-ouest.fr>

References


See Also

plot.catdes, condes

Examples

data(wine)
catdes(wine, num.var=2)
### children

**Children (data)**

**Description**

The data used here is a contingency table that summarizes the answers given by different categories of people to the following question: according to you, what are the reasons that can make hesitate a woman or a couple to have children?

**Usage**

```r
data(children)
```

**Format**

A data frame with 18 rows and 8 columns. Rows represent the different reasons mentioned, columns represent the different categories (education, age) people belong to.

**Source**

Traitements Statistiques des Enquêtes (D. Grangé, L.Lebart, eds.) Dunod, 1993

**Examples**

```r
data(children)
res.ca <- CA(children, col.sup = 6:8, row.sup = 15:18)
```

### coeffsRV

**Calculate the RV coefficient and test its significance**

**Description**

Calculate the RV coefficient and test its significance.

**Usage**

```r
coeffRV(X, Y)
```

**Arguments**

- `X` a matrix with `n` rows (individuals) and `p` numerous columns (variables)
- `Y` a matrix with `n` rows (individuals) and `p` numerous columns (variables)
Details

Calculates the RV coefficient between X and Y. It returns also the standardized RV, the expectation, the variance and the skewness under the permutation distribution. These moments are used to approximate the exact distribution of the RV statistic with the Pearson type III approximation and the p-value associated to this test is given.

Value

A list containing the following components:

- **RV** the RV coefficient between the two matrices
- **RVs** the standardized RV coefficients
- **mean** the mean of the RV permutation distribution
- **variance** the variance of the RV permutation distribution
- **skewness** the skewness of the RV permutation distribution
- **p.value** the p-value associated to the test of the significativity of the RV coefficient (with the Pearson type III approximation)

Author(s)

Julie Josse, Francois Husson <Francois.Husson@agrocampus-ouest.fr>

References


Examples

```r
data(wine)
X <- wine[,3:7]
Y <- wine[,11:20]
coeffRV(X,Y)
```

---

**condes**

*Continuous variable description*

Description

Description continuous by quantitative variables and/or by categorical variables
Usage

condes(donnee, num.var, proba = 0.05)

Arguments

donnee: a data frame made up of at least one quantitative variable and a set of quantitative
variables and/or categorical variables
num.var: the number of the variable to characterized
proba: the significance threshold considered to characterized the category (by default
0.05)

Value

Returns a list including:
quanti: the description of the num.var variable by the quantitative variables. The vari-
ables are sorted in ascending order (from the one which characterized the most
to the one which significantly characterized with the proba proba)
quali: The categorical variables which characterized the continuous variables are listed
in ascending order
category: description of the continuous variable num.var by each category of all the cate-
gorical variables

Author(s)

Francois Husson <Francois.Husson@agrocampus-ouest.fr>

See Also

catdes

Examples

data(decathlon)
condes(decathlon, num.var=3)

---

coord.ellipse

Construct confidence ellipses

Description

Construct confidence ellipses

Usage

coord.ellipse (coord.simul, centre = NULL, axes = c(1, 2),
level.conf = 0.95, npoint = 100, bary = FALSE)
coord.ellipse

Arguments

coord.simul a data frame containing the coordinates of the individuals for which the confidence ellipses are constructed. This data frame can contain more than 2 variables; the variables taken into account are chosen after. The first column must be a factor which allows to associate one row to an ellipse. The simule object of the result of the simule function correspond to a data frame.

centre a data frame whose columns are the same than those of the coord.simul, and with the coordinates of the centre of each ellipse. This parameter is optional and NULL by default; in this case, the centre of the ellipses is calculated from the data

axes a length 2 vector specifying the components of coord.simul that are taken into account

level.conf confidence level used to construct the ellipses. By default, 0.95

npoint number of points used to draw the ellipses

bary boolean, if bary = TRUE, the coordinates of the ellipse around the barycentre of individuals are calculated

Value

res a data frame with (npoint times the number of ellipses) rows and three columns. The first column is the factor of coord.simul, the two others columns give the coordinates of the ellipses on the two dimensions chosen.

call the parameters of the function chosen

Author(s)

Jeremy Mazet

See Also

simule

Examples

data(decathlon)
res.pca <- PCA(decathlon, quanti.sup = 11:12, quali.sup = 13, graph=FALSE)
aa <- cbind.data.frame(decathlon[,13], res.pca$ind$coord)
bb <- coord.ellipse(aa, bary=TRUE)
plot.PCA(res.pca, habillage=13, ellipse=bb)

## To automatically draw ellipses around the barycentres of all the categorical variables
plotellipses(res.pca)
**decathlon**

*Performance in decathlon (data)*

**Description**

The data used here refer to athletes’ performance during two sporting events.

**Usage**

```r
data(decathlon)
```

**Format**

A data frame with 41 rows and 13 columns: the first ten columns corresponds to the performance of the athletes for the 10 events of the decathlon. The columns 11 and 12 correspond respectively to the rank and the points obtained. The last column is a categorical variable corresponding to the sporting event (2004 Olympic Game or 2004 Decastar)

**Source**

Département de mathématiques appliquées, Agrocampus Rennes

**Examples**

```r
data(decathlon)
res.pca <- PCA(decathlon, quanti.sup = 11:12, quali.sup=13)
```

---

**descfreq**

*Description of frequencies*

**Description**

Description of the rows of a contingency table or of groups of rows of a contingency table

**Usage**

```r
descfreq(donnee, by.quali = NULL, proba = 0.05)
```

**Arguments**

- **donnee**
  a data frame corresponding to a contingency table (quantitative data)

- **by.quali**
  a factor used to merge the data from different rows of the contingency table; by default NULL and each row is characterized

- **proba**
  the significance threshold considered to characterized the category (by default 0.05)
Value

Returns a list with the characterization of each rows or each group of the by.quali. A test corresponding to the hypergeometric distribution is performed and the probability to observe a more extreme value than the one observed is calculated. For each row (or category), each of the columns characterising the row are sorted in ascending order of p-value.

Author(s)

Francois Husson <Francois.Husson@agrocampus-ouest.fr>

References


See Also

catdes, condes, textual

Examples

data(children)
descfreq(children[1:14,1:5])  ## desc of rows
descfreq(t(children[1:14,1:5])) ## desc of columns

---

dimdesc | Dimension description

Description

This function is designed to point out the variables and the categories that are the most characteristic according to each dimension obtained by a Factor Analysis.

Usage

dimdesc(res, axes = 1:3,_proba = 0.05)

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>res</td>
<td>an object of class PCA, MCA, CA, MFA or HMFA</td>
</tr>
<tr>
<td>axes</td>
<td>a vector with the dimensions to describe</td>
</tr>
<tr>
<td>proba</td>
<td>the significance threshold considered to characterized the dimension (by default 0.05)</td>
</tr>
</tbody>
</table>
DMFA

Value

Returns a list including:

- **quanti**: the description of the dimensions by the quantitative variables. The variables are sorted.
- **quali**: the description of the dimensions by the categorical variables

Author(s)

Francois Husson <Francois.Husson@agrocampus-ouest.fr>

References


See Also

PCA, CA, MCA, MFA, HMFA

Examples

```r
data(decathlon)
res.pca <- PCA(decathlon, quanti.sup = 11:12, quali.sup=13, graph=FALSE)
dimdesc(res.pca)
```

---

**DMFA**

*Dual Multiple Factor Analysis (DMFA)*

Description

Performs Dual Multiple Factor Analysis (DMFA) with supplementary individuals, supplementary quantitative variables and supplementary categorical variables.

Usage

```r
DMFA(don, num.fact = ncol(data), scale.unit = TRUE, ncp = 5,
     quanti.sup = NULL, quali.sup = NULL, graph = TRUE, axes=c(1,2))
```

Arguments

- **don**: a data frame with *n* rows (individuals) and *p* columns (numeric variables)
- **num.fact**: the number of the categorical variable which allows to make the group of individuals
- **scale.unit**: a boolean, if TRUE (value set by default) then data are scaled to unit variance
- **ncp**: number of dimensions kept in the results (by default 5)
- **quanti.sup**: a vector indicating the indexes of the quantitative supplementary variables
quali.sup  a vector indicating the indexes of the categorical supplementary variables
graph  boolean, if TRUE a graph is displayed
axes  a length 2 vector specifying the components to plot

Value

Returns a list including:

- `eig` a matrix containing all the eigenvalues, the percentage of variance and the cumulative percentage of variance
- `var` a list of matrices containing all the results for the active variables (coordinates, correlation between variables and axes, square cosine, contributions)
- `ind` a list of matrices containing all the results for the active individuals (coordinates, square cosine, contributions)
- `ind.sup` a list of matrices containing all the results for the supplementary individuals (coordinates, square cosine)
- `quanti.sup` a list of matrices containing all the results for the supplementary quantitative variables (coordinates, correlation between variables and axes)
- `quali.sup` a list of matrices containing all the results for the supplementary categorical variables (coordinates of each categories of each variables, and v.test which is a criterion with a Normal distribution)
- `svd` the result of the singular value decomposition
- `var.partiel` a list with the partial coordinate of the variables for each group
- `cor.dim.gr` a list with the data centered by group
- `group` a list with the results for the groups (coordinate, normalized coordinates, cos2)
- `Cov` a list with the covariance matrices for each group

Returns the individuals factor map and the variables factor map.

Author(s)

Francois Husson <Francois.Husson@agrocampus-ouest.fr>

See Also

plot.DMFA, dimdesc

Examples

```r
## Example with the famous Fisher’s iris data
res.dmfa = DMFA (iris, num.fact = 5)
```
estim_ncp

Estimate the number of components in Principal Component Analysis

Description

Estimate the number of components in PCA.

Usage

```r
estim_ncp(X, ncp.min=0, ncp.max=NULL, scale=TRUE, method="GCV")
```

Arguments

- `X`: a data frame with continuous variables
- `ncp.min`: minimum number of dimensions to interpret, by default 0
- `ncp.max`: maximum number of dimensions to interpret, by default NULL which corresponds to the number of columns minus 2
- `scale`: a boolean, if TRUE (value set by default) then data are scaled to unit variance
- `method`: method used to estimate the number of components, "GCV" for the generalized cross-validation approximation or "Smooth" for the smoothing method (by default "GCV")

Value

Returns `ncp` the best number of dimensions to use (find the minimum or the first local minimum) and the mean error for each dimension tested

Author(s)

- Francois Husson <Francois.Husson@agrocampus-ouest.fr>
- Julie Josse <Julie.Josse@agrocampus-ouest.fr>

References


See Also

- PCA

Examples

```r
data(decathlon)
nb.dim <- estim_ncp(decathlon[,1:10], scale=TRUE)
```
FAMD

Factor Analysis for Mixed Data

Description

FAMD is a principal component method dedicated to explore data with both continuous and categorical variables. It can be seen roughly as a mixed between PCA and MCA. More precisely, the continuous variables are scaled to unit variance and the categorical variables are transformed into a disjunctive data table (crisp coding) and then scaled using the specific scaling of MCA. This ensures to balance the influence of both continuous and categorical variables in the analysis. It means that both variables are on a equal foot to determine the dimensions of variability. This method allows one to study the similarities between individuals taking into account mixed variables and to study the relationships between all the variables. It also provides graphical outputs such as the representation of the individuals, the correlation circle for the continuous variables and representations of the categories of the categorical variables, and also specific graphs to visualize the associations between both type of variables.

Usage

FAMD (base, ncp = 5, graph = TRUE, sup.var = NULL,
     ind.sup = NULL, axes = c(1,2), row.w = NULL, tab.comp = NULL)

Arguments

- `base`: a data frame with n rows (individuals) and p columns
- `ncp`: number of dimensions kept in the results (by default 5)
- `graph`: boolean, if TRUE a graph is displayed
- `ind.sup`: a vector indicating the indexes of the supplementary individuals
- `sup.var`: a vector indicating the indexes of the supplementary variables
- `axes`: a length 2 vector specifying the components to plot
- `row.w`: an optional row weights (by default, uniform row weights)
- `tab.comp`: object obtained from the imputeFAMD function of the missMDA package that allows to handle missing values

Value

Returns a list including:

- `eig`: a matrix containing all the eigenvalues, the percentage of variance and the cumulative percentage of variance
- `var`: a list of matrices containing all the results for the variables considered as group (R2, coordinates, square cosine, contributions, distance to the origin, the R2 between each variable and each factor)
- `ind`: a list of matrices with all the results for the individuals (coordinates, square cosine, contributions)
footsize

quali.var  a list of matrices with all the results for the categorical variables (coordinates, square cosine, contributions, v.test)
quantiVar a list of matrices with all the results for the quantitative variables (coordinates, correlation, square cosine, contributions)
call      a list with some statistics

Returns the individuals factor map.

Author(s)

Francois Husson <Francois.Husson@agrocampus-ouest.fr>, Jeremy Mazet

References


See Also

print.FAMD, summary.FAMD, plot.FAMD

Examples

## Not run:
data(geomorphology)
res <- FAMD(geomorphology)
summary(res.ca)

data(wine)
res <- FAMD(wine[,c(1,2,3:zero.noslash,31)])

## End(Not run)

footsize

Description

Dataset for the covariance analysis (a quantitative variable explained by quantitative (continuous) and qualitative (categorical) variables)

Usage

data(footsize)

Format

Dataset with 84 rows and 3 columns: footsize, size and sex
Examples

```r
data(footsize)
res3 <- AovSum(footsize ~ size + sex + size:sex, data=footsize)
res3
```

Description

The data used here concern a geomorphology analysis.

Usage

```r
data(geomorphology)
```

Format

A data frame with 75 rows and 11 columns. Rows represent the individuals, columns represent the different questions. 10 variables are quantitative and one variable is qualitative. The dataset is analysed in: http://www.sciencedirect.com/science/article/pii/S0169555X11006362

Examples

```r
data(geomorphology)
res <- FAMD(geomorphology)
plot(res, choix="ind", habillage=4)
```

GPA

*Generalised Procrustes Analysis*

Description

Performs Generalised Procrustes Analysis (GPA) that takes into account missing values.

Usage

```r
GPA(df, tolerance=10^-10, nbiteration=200, scale=TRUE, group, name.group= NULL, graph = TRUE, axes = c(1,2))
```
Arguments

- df: a data frame with \( n \) rows (individuals) and \( p \) columns (quantitative variables)
- tolerance: a threshold with respect to which the algorithm stops, i.e. when the difference between the GPA loss function at step \( n \) and \( n+1 \) is less than tolerance
- nbiter: the maximum number of iterations until the algorithm stops
- scale: a boolean, if TRUE (which is the default value) scaling is required
- group: a vector indicating the number of variables in each group
- name.group: a vector indicating the name of the groups (the groups are successively named group.1, group.2 and so on, by default)
- graph: boolean, if TRUE a graph is displayed
- axes: a length 2 vector specifying the components to plot

Details

Performs a Generalised Procrustes Analysis (GPA) that takes into account missing values: some data frames of df may have non described or non evaluated rows, i.e. rows with missing values only. The algorithm used here is the one developed by Commandeur.

Value

A list containing the following components:

- \( RV \): a matrix of RV coefficients between partial configurations
- \( RVs \): a matrix of standardized RV coefficients between partial configurations
- \( simi \): a matrix of Procrustes similarity indexes between partial configurations
- \( scaling \): a vector of isotropic scaling factors
- \( dep \): an array of initial partial configurations
- \( consensus \): a matrix of consensus configuration
- \( Xfin \): an array of partial configurations after transformations
- \( correlations \): correlation matrix between initial partial configurations and consensus dimensions
- \( PANOVA \): a list of "Procrustes Analysis of Variance" tables, per assesor (config), per product(objet), per dimension (dimension)

Author(s)

Elisabeth Morand
References


Examples

```r
## Not run:
data(wine)
res.gpa <- GPA(wine[,-(1:2)], group=c(5,3,10,9,2),
   name.group=c("olf","vis","olfag","gust","ens"))
### If you want to construct the partial points for some individuals only
plotGPApartial (res.gpa)

## End(Not run)
```

### graph.var

*Make graph of variables*

**Description**

Plot the graphs of the variables after a Factor Analysis.

**Usage**

```r
graph.var(x, axes = c(1, 2),
   xlim = NULL, ylim = NULL, col.sup = "blue",
   col.var = "black", draw="all", label=draw, lim.cos2.var = 0.1,
   cex = 1, title = NULL, new.plot = TRUE, ...)
```

**Arguments**

- `x`: an object of class PCA, MCA, MFA or HMFA
- `axes`: a length 2 vector specifying the components to plot
- `xlim`: range for the plotted ‘x’ values, defaulting to the range of the finite values of ‘x’
- `ylim`: range for the plotted ‘y’ values, defaulting to the range of the finite values of ‘y’
- `col.sup`: a color for the quantitative supplementary variables
- `col.var`: a color for the variables
draw a list of character for the variables which are drawn (by default, all the variables are drawn). You can draw all the active variables by putting "var" and/or all the supplementary variables by putting "quanti.sup" and/or a list with the names of the variables which should be drawn.

label a list of character for the variables which are labelled (by default, all the drawn variables are labelled). You can label all the active variables by putting "var" and/or all the supplementary variables by putting "quanti.sup" and/or a list with the names of the variables which should be labelled.

lim.cos2.var value of the square cosinus under the variables are not drawn.

cex cf. function par in the graphics package.

title string corresponding to the title of the graph you draw (by default NULL and a title is chosen).

new.plot boolean, if TRUE, a new graphical device is created.

... further arguments passed to or from other methods.

Value

Returns the variables factor map.

Author(s)

Francois Husson &lt;Francois.Husson@agrocampus-ouest.fr&gt;

See Also

PCA, MFA, MCA, DMFA, HMFA

Examples

data(decathlon)
res.pca <- PCA(decathlon, quanti.sup = 11:12, quali.sup = 13, graph = FALSE)
graph.var (res.pca, draw = c("var","Points"),
label = c("Long.jump", "Points"))
Usage

HCPC(res, nb.clust=0, consol=TRUE, iter.max=10, min=3, max=NULL, metric="euclidean", method="ward", order=TRUE, graph.scale="inertia", nb.par=5, graph=TRUE, proba=0.05, cluster.CA="rows", kk=Inf,...)

Arguments

res Either the result of a factor analysis, a dataframe, or a vector.
nb.clust an integer. If 0, the tree is cut at the level the user clicks on. If -1, the tree is automatically cut at the suggested level (see details). If a (positive) integer, the tree is cut with nb.clusters clusters.
consol a boolean. If TRUE, a k-means consolidation is performed.
iter.max An integer. The maximum number of iterations for the consolidation.
min an integer. The least possible number of clusters suggested.
max an integer. The higher possible number of clusters suggested; by default the minimum between 10 and the number of individuals divided by 2.
mometric The metric used to built the tree. See agnes for details.
method The method used to built the tree. See agnes for details.
order A boolean. If TRUE, clusters are ordered following their center coordinate on the first axis.
graph.scale A character string. By default "inertia" and the height of the tree corresponds to the inertia gain, else "sqrt-inertia" the square root of the inertia gain.
nb.par An integer. The number of edited paragons.
graph If TRUE, graphics are displayed. If FALSE, no graph are displayed.
proba The probability used to select axes and variables in catdes (see catdes for details).
cluster.CA A string equals to "rows" or "columns" for the clustering of Correspondence Analysis results.
kk An integer; if this integer is greater than 0, a Kmeans is performed with kk clusters and the top of the hierarchical tree is constructed from this partition. This is very useful if the number of individuals is high. Inf can be used to construct all the tree.
... Other arguments from other methods.

Details

The function first built a hierarchical tree. Then the sum of the within-cluster inertia are calculated for each partition. The suggested partition is the one with the higher relative loss of inertia (i(clusters n+1)/i(cluster n)).

The absolut loss of inertia (i(cluster n)-i(cluster n+1)) is ploted with the tree.

If the ascending clustering is constructed from a data-frame with a lot of rows (individuals), it is possible to first perform a partition with kk clusters and then construct the tree from the (weighted) kk clusters.
HMFA

Hierarchical Multiple Factor Analysis

Description

Performs a hierarchical multiple factor analysis, using an object of class list of data.frame.

Usage

\[
\text{HMFA}(X,H,\text{type} = \text{rep("s", length(H[[1]])}, \text{ncp} = 5, \text{graph} = \text{TRUE}, \\
\text{axes} = \text{c}(1,2), \text{name.group} = \text{NULL})
\]
Arguments

X  a data.frame

H  a list with one vector for each hierarchical level; in each vector the number of
  variables or the number of group constituting the group

type  the type of variables in each group in the first partition; three possibilities: "c"
  or "s" for quantitative variables (the difference is that for "s", the variables are
  scaled in the program), "n" for categorical variables; by default, all the variables
  are quantitative and the variables are scaled unit

ncp  number of dimensions kept in the results (by default 5)

graph  boolean, if TRUE a graph is displayed

axes  a length 2 vector specifying the components to plot

name.group  a list of vector containing the name of the groups for each level of the hierarchy
  (by default, NULL and the group are named L1.G1, L1.G2 and so on)

Value

Returns a list including:

  eig  a matrix containing all the eigenvalues, the percentage of variance and the cu-
        mulative percentage of variance

  group  a list with first a list of matrices with the coordinates of the groups for each
          level and second a matrix with the canonical correlation (correlation between
          the coordinates of the individuals and the partial points))

  ind  a list of matrices with all the results for the active individuals (coordinates,
        square cosine, contributions)

  quanti.var  a list of matrices with all the results for the quantitative variables (coordinates,
               correlation between variables and axes)

  quali.var  a list of matrices with all the results for the supplementary categorical variables
             (coordinates of each categories of each variables, and v.test which is a criterion
             with a Normal distribution)

  partial  a list of arrays with the coordinates of the partial points for each partition

Author(s)

  Sebastien Le, Francois Husson <Francois.Husson@agrocampus-ouest.fr>

References

  of sensory profiles, Food Quality and Preferences, 18 (6), 453-464.

See Also

  print.HMFA, plot.HMFA, dimdesc
Examples

```r
data(wine)
hierar <- list(c(2,5,3,10,9,2), c(4,2))
res.hmfa <- HMFA(wine, H = hierar, type=c("n",rep("s",5)))
```

Description

The data used here concern a questionnaire on hobbies. We asked to 8403 individuals how answer questions about their hobbies (18 questions). The following four variables were used to label the individuals: sex (male, female), age (15-25, 26-35, 36-45, 46-55, 56-65, 66-75, 76-85, 86-100), marital status (single, married, widowed, divorced, remarried), profession (manual labourer, unskilled worker, technician, foreman, senior management, employee, other). And finally, a quantitative variable indicates the number of hobbies practised out of the 18 possible choices.

Usage

```r
data(tea)
```

Format

A data frame with 8403 rows and 23 columns. Rows represent the individuals, columns represent the different questions. The first 18 questions are active ones, and the 4 following ones are supplementary categorical variables and the 23th is a supplementary quantitative variable (the number of activities)

Examples

```r
data(hobbies)
## Not run:
res.mca <- MCA(hobbies,quali.sup=19:22,quanti.sup=23,method="Burt")
plot(res.mca,invisible=c("ind","quali.sup"),hab="quali") ### active var. only
plot(res.mca,invisible=c("var","quali.sup"),cex=.5,label="none") ### individuals only
plot(res.mca,invisible=c("ind","var"),hab="quali") ### supp. qualitative var. only
dimdesc(res.mca)
plotellipses(res.mca,keepvar=1:4)
## End(Not run)
```
Number of medals in athletism during olympic games per country

Description

This data frame is a contingency table with the athletism events (in row) and the countries (in columns). Each cell gives the number of medals obtained during the 5 olympics games from 1992 to 2008 (Barcelona 1992, Atlanta 1996, Sydney 2000, Athens 2004, Beijing 2008).

Usage

data(JO)

Format

A data frame with the 24 events in athletism and in column the 58 countries who obtained at least one medal.

Examples

```r
## Not run:
data(JO)
res.ca <- CA(JO)
res.ca <- CA(JO, axes = 3:4)
## End(Not run)
```

Multiple Correspondence Analysis (MCA)

Description

Performs Multiple Correspondence Analysis (MCA) with supplementary individuals, supplementary quantitative variables and supplementary categorical variables. Missing values are treated as an additional level, categories which are rare can be ventilated.

Usage

```r
MCA(X, ncp = 5, ind.sup = NULL, quanti.sup = NULL, quali.sup = NULL, graph = TRUE, level.ventil = 0, axes = c(1,2), row.w = NULL, method="Indicator", na.method="NA", tab.disj=NULL)
```
Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>a data frame with ( n ) rows (individuals) and ( p ) columns (categorical variables)</td>
</tr>
<tr>
<td>ncp</td>
<td>number of dimensions kept in the results (by default 5)</td>
</tr>
<tr>
<td>ind.sup</td>
<td>a vector indicating the indexes of the supplementary individuals</td>
</tr>
<tr>
<td>quanti.sup</td>
<td>a vector indicating the indexes of the quantitative supplementary variables</td>
</tr>
<tr>
<td>quali.sup</td>
<td>a vector indicating the indexes of the categorical supplementary variables</td>
</tr>
<tr>
<td>graph</td>
<td>boolean, if TRUE a graph is displayed</td>
</tr>
<tr>
<td>level.ventil</td>
<td>a proportion corresponding to the level under which the category is ventilated; by default, 0 and no ventilation is done</td>
</tr>
<tr>
<td>axes</td>
<td>a length 2 vector specifying the components to plot</td>
</tr>
<tr>
<td>row.w</td>
<td>an optional row weights (by default, a vector of 1 for uniform row weights)</td>
</tr>
<tr>
<td>method</td>
<td>a string corresponding to the name of the method used: &quot;Indicator&quot; (by default) is the CA on the Indicator matrix, &quot;Burt&quot; is the CA on the Burt table. For Burt and the Indicator, the graph of the individuals and the graph of the categories are given</td>
</tr>
<tr>
<td>na.method</td>
<td>a string corresponding to the name of the method used if there are missing values; available methods are &quot;NA&quot; or &quot;Average&quot; (by default, &quot;NA&quot;)</td>
</tr>
<tr>
<td>tab.disj</td>
<td>optional data.frame corresponding to the disjunctive table used for the analysis; it corresponds to a disjunctive table obtained from imputation method (see package missMDA).</td>
</tr>
</tbody>
</table>

Value

Returns a list including:

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>eig</td>
<td>a matrix containing all the eigenvalues, the percentage of variance and the cumulative percentage of variance</td>
</tr>
<tr>
<td>var</td>
<td>a list of matrices containing all the results for the active variables (coordinates, square cosine, contributions, v.test, square correlation ratio)</td>
</tr>
<tr>
<td>ind</td>
<td>a list of matrices containing all the results for the active individuals (coordinates, square cosine, contributions)</td>
</tr>
<tr>
<td>ind.sup</td>
<td>a list of matrices containing all the results for the supplementary individuals (coordinates, square cosine)</td>
</tr>
<tr>
<td>quanti.sup</td>
<td>a matrix containing the coordinates of the supplementary quantitative variables (the correlation between a variable and an axis is equal to the variable coordinate on the axis)</td>
</tr>
<tr>
<td>quali.sup</td>
<td>a list of matrices with all the results for the supplementary categorical variables (coordinates of each categories of each variables, square cosine and v.test which is a criterion with a Normal distribution, square correlation ratio)</td>
</tr>
<tr>
<td>call</td>
<td>a list with some statistics</td>
</tr>
</tbody>
</table>

Returns the graphs of the individuals and categories and the graph with the variables. The plots may be improved using the argument autolab, modifying the size of the labels or selecting some elements thanks to the `plot.MCA` function.
Author(s)
Francois Husson <husson@agrocampus-ouest.fr>, Julie Josse, Jeremy Mazet

References

See Also
plotellipses, summary.MCA, print.MCA, plot.MCA, dimdesc

Examples
## Not run:
## Tea example
data(tea)
res.mca <- MCA(tea,quanti.sup=19,quali.sup=20:36)
summary(res.mca)
plot(res.mca,invisible=c("var","quali.sup","quanti.sup"),cex=0.7)
plot(res.mca,invisible=c("ind","quali.sup","quanti.sup"),cex=0.8)
plot(res.mca,invisible=c("quali.sup","quanti.sup"),cex=0.8)
dimdesc(res.mca)
plotellipses(res.mca,keepvar=1:4)
plotellipses(res.mca,keepvar="Tea")

## Hobbies example
data(hobbies)
res.mca <- MCA(hobbies,quali.sup=19:22,quanti.sup=23)
plot(res.mca,invisible=c("ind","quali.sup"),hab="quali")
plot(res.mca,invisible=c("var","quali.sup"),cex=.5,label="none")
dimdesc(res.mca)
plotellipses(res.mca,keepvar=1:4)

## Example with missing values : use the missMDA package
require(missMDA)
data(vnf.example)
completed <- imputeMCA(vnf.example,ncp=2)
res.mca <- MCA(vnf.example,tab.disj=completed$tab.disj)

## End(Not run)
### Description

Performs Multiple Factor Analysis in the sense of Escofier-Pages with supplementary individuals and supplementary groups of variables. Groups of variables can be quantitative, categorical or contingency tables. Missing values in numeric variables are replaced by the column mean. Missing values in categorical variables are treated as an additional level.

### Usage

```r
MFA (base, group, type = rep("s", length(group)), ind.sup = NULL, ncp = 5, name.group = NULL, num.group.sup = NULL, graph = TRUE, weight.col.mfa = NULL, row.w = NULL, axes = c(1, 2), tab.comp = NULL)
```

### Arguments

- **base**: a data frame with \( n \) rows (individuals) and \( p \) columns (variables)
- **group**: a list indicating the number of variables in each group
- **type**: the type of variables in each group; four possibilities: "c" or "s" for quantitative variables (the difference is that for "s" variables are scaled to unit variance), "n" for categorical variables and "f" for frequencies (from a contingency tables); by default, all variables are quantitative and scaled to unit variance
- **ind.sup**: a vector indicating the indexes of the supplementary individuals
- **ncp**: number of dimensions kept in the results (by default 5)
- **name.group**: a vector containing the name of the groups (by default, NULL and the group are named group.1, group.2 and so on)
- **num.group.sup**: the indexes of the illustrative groups (by default, NULL and no group are illustrative)
- **graph**: boolean, if TRUE a graph is displayed
- **weight.col.mfa**: vector of weights, useful for HMFA method (by default, NULL and an MFA is performed)
- **row.w**: an optional row weights (by default, a vector of 1 for uniform row weights)
- **axes**: a length 2 vector specifying the components to plot
- **tab.comp**: object obtained from the imputeMFA function of the missMDA package that allows to handle missing values

### Value

- **summary.quali**: a summary of the results for the categorical variables
- **summary.quanti**: a summary of the results for the quantitative variables
- **separate.analyses**: the results for the separate analyses
- **eig**: a matrix containing all the eigenvalues, the percentage of variance and the cumulative percentage of variance
group a list of matrices containing all the results for the groups (Lg and RV coefficients, coordinates, square cosine, contributions, distance to the origin, the correlations between each group and each factor)

rapport.inertie inertia ratio

ind a list of matrices containing all the results for the active individuals (coordinates, square cosine, contributions)

ind.sup a list of matrices containing all the results for the supplementary individuals (coordinates, square cosine)

quanti.var a list of matrices containing all the results for the quantitative variables (coordinates, correlation between variables and axes, contribution, cos2)

quali.var a list of matrices containing all the results for categorical variables (coordinates of each categories of each variables, contribution and v.test which is a criterion with a Normal distribution)

freq a list of matrices containing all the results for the frequencies (coordinates, contribution, cos2)

quanti.var.sup a list of matrices containing all the results for the supplementary quantitative variables (coordinates, correlation between variables and axes, cos2)

quali.var.sup a list of matrices containing all the results for the supplementary categorical variables (coordinates of each categories of each variables, cos2 and v.test which is a criterion with a Normal distribution)

freq.sup a list of matrices containing all the results for the supplementary frequencies (coordinates, cos2)

partial.axes a list of matrices containing all the results for the partial axes (coordinates, correlation between variables and axes, correlation between partial axes)

global.pca the result of the analysis when it is considered as a unique weighted PCA

Returns the individuals factor map, the variables factor map and the groups factor map.
The plots may be improved using the argument autolab, modifying the size of the labels or selecting some elements thanks to the plot.MFA function.

Author(s)
Francois Husson <Francois.Husson@agrocampus-ouest.fr>, J. Mazet

References

See Also
print.MFA, summary.MFA, plot.MFA, dimdesc
Examples

data(wine)
res <- MFA(wine, group=c(2,5,3,10,9,2), type=c("n","rep","s",5)),
      ncp=5, name.group=c("orig","olf","vis","olfag","gust","ens"),
      num.group.sup=c(1,6))
summary(res)
barplot(res$eig[,1],main="Eigenvalues",names.arg=1:nrow(res$eig))

## Not run:
#### Confidence ellipses around categories per variable
plotellipses(res)
plotellipses(res,keepvar="Label") ## for 1 variable

#### Interactive graph
liste = plotMFApartial(res)
plot(res,choix="ind",habillage = "Terroir")

###Example with groups of categorical variables
data (poison)
MFA(poison, group=c(2,2,5,6), type=c("s","n","n","n"),
      name.group=c("desc","desc2","symptom","eat"),
      num.group.sup=1:2)

###Example with groups of frequency tables
data(mortality)
res<-MFA(mortality,group=c(9,9),type=c("f","f"),
         name.group=c("1979","2006"))

## End(Not run)

dataset to illustrate the selection of variables in regression

data(milk)

Format

Dataset with 85 rows and 6 columns: 85 milks described by the 5 variables: density, fat content, protein, casein, dry, yield
The cause of mortality in France in 1979 and 2006

Description

The cause of mortality in France in 1979 and 2006.

Usage

data(mortality)

Format

A data frame with 62 rows (the different causes of death) and 18 columns. Each column corresponds to an age interval (15-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75-84, 85-94, 95 and more) in a year. The 9 first columns correspond to data in 1979 and the 9 last columns to data in 2006. In each cell, the counts of deaths for a cause of death in an age interval (in a year) is given.

Source

Centre d’épidemiologie sur les causes médicales

Examples

data(mortality)

## Not run:
res<-MFA(mortality,group=c(9,9),type=c("f","f"),
         name.group=c("1979","2006"))
plot(res,choix="freq",invisible="ind",habillage="group")
lines(res$freq$coord[1:9,1],mfa$freq$coord[1:9,2],col="red")
lines(res$freq$coord[10:18,1],mfa$freq$coord[10:18,2],col="green")

## End(Not run)
**Principal Component Analysis (PCA)**

**Description**

Performs Principal Component Analysis (PCA) with supplementary individuals, supplementary quantitative variables and supplementary categorical variables. Missing values are replaced by the column mean.

**Usage**

```r
PCA(X, scale.unit = TRUE, ncp = 5, ind.sup = NULL,
    quanti.sup = NULL, quali.sup = NULL, row.w = NULL,
    col.w = NULL, graph = TRUE, axes = c(1,2))
```

**Arguments**

- **X** a data frame with $n$ rows (individuals) and $p$ columns (numeric variables)
- **ncp** number of dimensions kept in the results (by default 5)
- **scale.unit** a boolean, if TRUE (value set by default) then data are scaled to unit variance
- **ind.sup** a vector indicating the indexes of the supplementary individuals
- **quanti.sup** a vector indicating the indexes of the quantitative supplementary variables
- **quali.sup** a vector indicating the indexes of the categorical supplementary variables
- **row.w** an optional row weights (by default, a vector of 1 for uniform row weights)
- **col.w** an optional column weights (by default, uniform column weights)
- **graph** boolean, if TRUE a graph is displayed
- **axes** a length 2 vector specifying the components to plot

**Value**

Returns a list including:

- **eig** a matrix containing all the eigenvalues, the percentage of variance and the cumulative percentage of variance
- **var** a list of matrices containing all the results for the active variables (coordinates, correlation between variables and axes, square cosine, contributions)
- **ind** a list of matrices containing all the results for the active individuals (coordinates, square cosine, contributions)
- **ind.sup** a list of matrices containing all the results for the supplementary individuals (coordinates, square cosine)
- **quanti.sup** a list of matrices containing all the results for the supplementary quantitative variables (coordinates, correlation between variables and axes)
quali.sup        a list of matrices containing all the results for the supplementary categorical variables (coordinates of each categories of each variables, v.test which is a criterion with a Normal distribution, and eta2 which is the square correlation coefficient between a qualitative variable and a dimension)

Returns the individuals factor map and the variables factor map.
The plots may be improved using the argument autolab, modifying the size of the labels or selecting some elements thanks to the plot.PCA function.

Author(s)
Francois Husson <Francois.Husson@agrocampus-ouest.fr>, Jeremy Mazet

References

See Also
print.PCA, summary.PCA, plot.PCA, dimdesc

Examples

data(decathlon)
res.pca <- PCA(decathlon, quanti.sup = 11:12, quali.sup=13)
## plot of the eigenvalues
## barplot(res.pca$eig[,1],main="Eigenvalues",names.arg=1:nrow(res.pca$eig))
summary(res.pca)
plot(res.pca,choix="ind",habillage=13)
dimdesc(res.pca, axes = 1:2)
## To draw ellipses around the categories of the 13th variable (which is categorical)
plotellipses(res.pca,13)

## Example with missing data
## use package missMDA
## Not run:
require(missMDA)
data(orange)
nb <- estim_ncpPCA(orange,ncp.min=0,ncp.max=5,method.cv="Kfold",nbsim=50)
imputed <- imputePCA(orange,ncp=nb$ncp)
res.pca <- PCA(imputed$completeObs)

## End(Not run)
plot.CA

**Draw the Correspondence Analysis (CA) graphs**

**Description**

Draw the Correspondence Analysis (CA) graphs.

**Usage**

```r
## S3 method for class 'CA'
plot(x, axes = c(1, 2),
    xlim = NULL, ylim = NULL,
    invisible = c("none", "row", "col", "row.sup", "col.sup", "quali.sup"),
    choix = c("CA", "quanti.sup"), col.row="blue", col.col="red",
    col.row.sup="darkblue", col.col.sup="darkred",
    col.quali.sup="magenta", col.quanti.sup="blue",
    label = c("all", "none", "row", "row.sup", "col", "col.sup", "quali.sup"),
    title = NULL, palette = NULL, autoLab = c("auto", "yes", "no"),
    new.plot=FALSE, selectRow = NULL, selectCol = NULL,
    unselect = 0.7, shadowtext = FALSE, ...)
```

**Arguments**

- `x`: an object of class CA
- `axes`: a length 2 vector specifying the components to plot
- `xlim`: range for the plotted 'x' values, defaulting to the range of the finite values of 'x'
- `ylim`: range for the plotted 'y' values, defaulting to the range of the finite values of 'y'
- `invisible`: string indicating if some points should be unlabelled ("row", "col", "row.sup", "col.sup", "quali.sup")
- `choix`: the graph to plot ("CA" for the CA map, "quanti.sup" for the supplementary quantitative variables)
- `col.row`: a color for the rows points
- `col.col`: a color for columns points
- `col.row.sup`: a color for the supplementary rows points
- `col.col.sup`: a color for supplementary columns points
- `col.quali.sup`: a color for the supplementary categorical variables
- `col.quanti.sup`: a color for the supplementary quantitative variables
- `label`: a list of character for the elements which are labelled (by default, all the elements are labelled ("row", "row.sup", "col", "col.sup", "quali.sup")
- `title`: string corresponding to the title of the graph you draw (by default NULL and a title is chosen)
palette the color palette used to draw the points. By default colors are chosen. If you want to define the colors: palette=palette(c("black","red","blue")); or you can use: palette=palette(rainbow(30)), or in black and white for example: palette=palette(gray(seq(0,.9,len=25)));

autoLab if autoLab="auto", autoLab is equal to "yes" if there are less than 50 elements and "no" otherwise; if "yes", the labels of the drawn elements are placed in a "good" way (can be time-consuming if many elements), and if "no" the elements are placed quickly but may overlap.

new.plot boolean, if TRUE, a new graphical device is created

selectRow a selection of the rows that are drawn; see the details section

selectCol a selection of the columns that are drawn; see the details section

unselect may be either a value between 0 and 1 that gives the transparency of the unselected objects (if unselect=1 the transparency is total and the elements are not drawn, if unselect=0 the elements are drawn as usual but without any label) or may be a color (for example unselect="grey60")

shadowtext boolean; if true put a shadow on the labels (rectangles are written under the labels which may lead to difficulties to modify the graph with another program)

... further arguments passed to or from other methods, such as cex, cex.main, ...

Details

The argument autoLab = "yes" is time-consuming if there are many labels that overlap. In this case, you can modify the size of the characters in order to have less overlapping, using for example cex=0.7.

The selectRow and selectCol arguments can be used in order to select a part of the elements that are drawn. For example, you can use:

selectRow = 1:5 and then the rows 1 to 5 are drawn.

select = c("name1","name5") and then the rows that have the names name1 and name5 are drawn.

select = "coord 10" and then the 10 rows (10 active and 10 supplementaries) that have the highest (squared) coordinates on the 2 chosen dimensions are drawn.

select = "contrib 10" and then the 10 rows (10 active) that have the highest contribution on the 2 dimensions of your plot are drawn.

select = "cos2 5" and then the 5 rows (5 actives and 5 supplementaries) that have the highest cos2 on the 2 dimensions of your plot are drawn.

select = "dist 8" and then the 8 rows (8 actives and 8 supplementaries) that have the highest distance to the center of gravity are drawn.

Value

Returns the individuals factor map and the variables factor map.

Author(s)

Jeremy Mazet, Francois Husson <Francois.Husson@agrocampus-ouest.fr>
plot.catdes

See Also

CA

Examples

data(children)
res.ca <- CA (children, col.sup = 6:8, row.sup = 15:18)
## select rows and columns that have a cos2 greater than 0.8
plot(res.ca,selectCol="cos2 > 0.8",selectRow="cos2 > 0.8")

Description

Plots a graph from a catdes output.

Usage

## S3 method for class 'catdes'
plot(x,col="deepskyblue",show="all",numchar=10,...)

Arguments

x A catdes object, see catdes for details.
col The color of the bars.
show a strig. If "quali", only the categorical variables are used. If "quanti", only the the quantitative variables are used. If "all", both quali and quanti are used.
numchar number of characters for the labels
... further arguments passed to or from other methods

Value

Returns choosen plot.

Author(s)

Guillaume Le Ray, Francois Husson <husson@agrocampus-ouest.fr>

See Also

catdes
plot.DMFA

Examples

```r
## Not run:
data(wine)
res.c=catdes(wine, num.var=2)
plot.catdes(res.c)
```

```r
## End(Not run)
```

### Description

Plot the graphs for a Principal Component Analysis (DMFA) with supplementary individuals, supplementary quantitative variables and supplementary categorical variables.

#### Usage

```r
## S3 method for class 'DMFA'
plot(x, axes = c(1, 2), choix = "ind", label="all",
     lim.cos2.var = 0., xlim=NULL, ylim=NULL, title = NULL,
     palette = NULL, new.plot = FALSE,
     autoLab = c("auto","yes","no"), ...) 
```

#### Arguments

- `x`: an object of class DMFA
- `axes`: a length 2 vector specifying the components to plot
- `choix`: the graph to plot ("ind" for the individuals, "var" for the variables)
- `label`: a list of character for the elements which are labelled (by default, all the elements are labelled ("ind", "ind.sup", "quali", "var", "quanti.sup"))
- `lim.cos2.var`: value of the square cosinus under the variables are not drawn
- `xlim`: range for the plotted 'x' values, defaulting to the range of the finite values of 'x'
- `ylim`: range for the plotted 'y' values, defaulting to the range of the finite values of 'y'
- `title`: string corresponding to the title of the graph you draw (by default NULL and a title is chosen)
- `palette`: the color palette used to draw the points. By default colors are chosen. If you want to define the colors : palette=palette(c("black","red","blue")); or you can use: palette=palette(rainbow(30)), or in black and white for example: palette=palette(gray(seq(0,.9,len=25)))
- `new.plot`: boolean, if TRUE, a new graphical device is created
- `autoLab`: if autoLab="auto", autoLab is equal to "yes" if there are less than 50 elements and "no" otherwise; if "yes", the labels of the drawn elements are placed in a "good" way (can be time-consuming if many elements), and if "no" the elements are placed quickly but may overlap
- `...`: further arguments passed to or from other methods
Value

Returns the individuals factor map and the variables factor map, the partial variables representation and the groups factor map.

Author(s)

Francois Husson <husson@agrocampus-ouest.fr>

See Also

DMFA

Description

It provides the graphical outputs associated with the principal component method for mixed data: FAMD.

Usage

```r
## S3 method for class 'FAMD'
plot(x, choix = c("ind","var","quanti","quali"), axes = c(1, 2),
     lab.var = TRUE, lab.ind = TRUE, habillage = "none", col.lab = FALSE,
     col.hab = NULL, invisible = NULL, lim.cos2.var = /zero.noslash.,
     xlim = NULL, ylim = NULL, title = NULL, palette=NULL, autoLab = c("auto","yes","no"), new.plot = FALSE,
     select = NULL, unselect = /zero.noslash.7, shadowtext = FALSE, ...)
```

Arguments

- **x**: an object of class FAMD
- **choix**: a string corresponding to the graph that you want to do ("ind" for the individual or categorical variables graph, "var" for all the variables (quantitative and categorical), "quanti" for the correlation circle)
- **axes**: a length 2 vector specifying the components to plot
- **lab.var**: boolean indicating if the labelled of the variables should be drawn on the map
- **lab.ind**: boolean indicating if the labelled of the individuals should be drawn on the map
- **habillage**: string corresponding to the color which are used. If "ind", one color is used for each individual else if it is the name or the position of a categorical variable, it colors according to the different categories of this variable
- **col.lab**: boolean indicating if the labelled should be colored
- **col.hab**: vector indicating the colors to use to labelled the rows or columns elements chosen in habillage
invisible list of string; for choix = "ind", the individuals can be omit (invisible = "ind"), or supplementary individuals (invisible="ind.sup") or the centeresg of gravity of the categorical variables (invisible= "quali"); if invisible = c("ind","ind.sup"), just the centers of gravity are drawn

lim.cos2.var value of the square cosinus under the variables are not drawn

xlim range for the plotted 'x' values, defaulting to the range of the finite values of 'x'

ylim range for the plotted 'y' values, defaulting to the range of the finite values of 'y'

title string corresponding to the title of the graph you draw (by default NULL and a title is chosen)

palette the color palette used to draw the points. By default colors are chosen. If you want to define the colors : palette=palette(c("black","red","blue")); or you can use: palette=palette(rainbow(30)), or in black and white for example: palette=palette(gray(seq(0,.9,len=25))

autoLab if autoLab="auto", autoLab is equal to "yes" if there are less than 50 elements and "no" otherwise; if "yes", the labels of the drawn elements are placed in a "good" way (can be time-consuming if many elements), and if "no" the elements are placed quickly but may overlap

new.plot boolean, if TRUE, a new graphical device is created

select a selection of the elements that are drawn; see the details section

unselect may be either a value between 0 and 1 that gives the transparency of the unselected objects (if unselect=1 the transparency is total and the elements are not drawn, if unselect=0 the elements are drawn as usual but without any label) or may be a color (for example unselect="grey60")

shadowtext boolean; if true put a shadow on the labels (rectangles are written under the labels which may lead to difficulties to modify the graph with another program)

... further arguments passed to or from other methods, such as cex, cex.main, ...

Value

Returns the individuals factor map and the variables factor map.

Author(s)

François Husson <François.Husson@agrocampus-ouest.fr>

See Also

FAMD

Examples

data(geomorphology)
res <- FAMD(geomorphology)
plot(res,choix="ind",habillage=4)
plot.GPA  

**Description**

Draw the General Procrustes Analysis (GPA) map.

**Usage**

```r
## S3 method for class 'GPA'
plot(x, axes = c(1, 2),
     lab.ind.moy = TRUE, habillage = "ind",
     partial = "all", chrono = FALSE, xlim = NULL, ylim = NULL,
     cex = 1, title = NULL, palette = NULL, ...)
```

**Arguments**

- `x`: an object of class GPA
- `axes`: a length 2 vector specifying the components to plot
- `lab.ind.moy`: boolean, if TRUE, the label of the mean points are drawn
- `habillage`: string corresponding to the color which are used. If "ind", one color is used for each individual; if "group" the individuals are colored according to the group
- `partial`: list of the individuals or of the center of gravity for which the partial points should be drawn (by default, partial = "none" and no partial points are drawn)
- `chrono`: boolean, if TRUE, the partial points of a same point are linked (useful when groups correspond to different moment)
- `xlim`: range for the plotted 'x' values, defaulting to the range of the finite values of 'x'
- `ylim`: range for the plotted 'y' values, defaulting to the range of the finite values of 'y'
- `cex`: cf. function `par` in the `graphics` package
- `title`: string corresponding to the title of the graph you draw (by default NULL and a title is chosen)
- `palette`: the color palette used to draw the points. By default colors are chosen. If you want to define the colors: `palette=palette(c("black","red","blue"));` or you can use: `palette=palette(rainbow(30)),` or in black and white for example: `palette=palette(gray(seq(0,.9,len=25)))`
- `...`: further arguments passed to or from other methods

**Value**

Returns the General Procrustes Analysis map.

**Author(s)**

Elisabeth Morand, Francois Husson <Francois.Husson@agrocampus-ouest.fr>
See Also

GPA

plot.HCPC

Plots for Hierarchical Classification on Principle Components (HCPC) results

Description

Plots graphs from a HCPC result: tree, barplot of inertia gains and first factor map with or without the tree, in 2 or 3 dimensions.

Usage

## S3 method for class 'HCPC'
plot(x, axes=c(1,2), choice="3D.map", rect=TRUE, draw.tree=TRUE, ind.names=TRUE, t.level="all", title=NULL, new.plot=FALSE, max.plot=15, tree.barplot=TRUE, centers.plot=FALSE, ...)

Arguments

x
A HCPC object, see HCPC for details.

axes
a two integers vector. Defines the axes of the factor map to plot.

choice
A string. "tree" plots the tree. "bar" plots bars of inertia gains. "map" plots a factor map, individuals colored by cluster. "3D.map" plots the same factor map, individuals colored by cluster, the tree above.

rect
a boolean. If TRUE, rectangles are drawn around clusters if choice ="tree".

tree.barplot
a boolean. If TRUE, the barplot of intra inertia losses is added on the tree graph.

draw.tree
A boolean. If TRUE, the tree is projected on the factor map if choice ="map".

ind.names
A boolean. If TRUE, the individuals names are added on the factor map when choice="3D.map"

t.level
Either a positive integer or a string. A positive integer indicates the starting level to plot the tree on the map when draw.tree=TRUE. If "all", the whole tree is plotted. If "centers", it draws the tree starting t the centers of the clusters.

title
a string. Title of the graph. NULL by default and a title is automatically defined

centers.plot
a boolean. If TRUE, the centers of clusters are drawn on the 3D factor maps.

new.plot
a boolean. If TRUE, the plot is done in a new window.

max.plot
The max for the bar plot

... 
Other arguments from other methods.

Value

Returns choosen plot.
Author(s)
Guillaume Le Ray, Quentin Molto, Francois Husson <husson@agrocampus-ouest.fr>

See Also
HCPC

Examples

data(iris)
# Clustering, auto nb of clusters:
res.hcpc=HCPC(iris[1:4], nb.clust=3)
# 3D graph from a different point of view:
plot.HCPC(res.hcpc, choice="3D.map", angle=60)

---

plot.HMFA

**Draw the Hierarchical Multiple Factor Analysis (HMFA) graphs**

Description

Draw the Hierarchical Multiple Factor Analysis (HMFA) graphs

Usage

```r
## S3 method for class 'HMFA'
plot(x, axes = c(1,2),num=6, choix = "ind",
     lab.grpe = TRUE, lab.var = TRUE, lab.ind.moy = TRUE,
     invisible = NULL, lim.cos2.var = 0.,
     xlim = NULL, ylim = NULL, cex = 1, title = NULL, new.plot = FALSE, ...)
```

Arguments

- `x`: an object of class HMFA
- `axes`: a length 2 vector specifying the components to plot
- `num`: number of graphs in a same windows
- `choix`: a string corresponding to the graph that you want to do ("ind" for the individual or categorical variables graph, "var" for the quantitative variables graph, "axes" for the graph of the partial axes, "group" for the groups representation)
- `lab.grpe`: boolean, if TRUE, the label of the groups are drawn
- `lab.var`: boolean, if TRUE, the label of the variables are drawn
- `lab.ind.moy`: boolean, if TRUE, the label of the mean points are drawn
- `invisible`: list of string; for choix="ind", the individuals can be omit (invisible = "ind"), or the centers of gravity of the categorical variables (invisible= "qual")
- `lim.cos2.var`: value of the square cosinus under with the points are not drawn
- `xlim`: range for the plotted 'x' values, defaulting to the range of the finite values of 'x'
plot.MCA

ylim range for the plotted 'y' values, defaulting to the range of the finite values of 'y'
cex cf. function par in the graphics package
title string corresponding to the title of the graph you draw (by default NULL and a title is chosen)
new.plot boolean, if TRUE, a new graphical device is created
...

Value

Returns the individuals factor map and the variables factor map.

Author(s)

Jérémy Mazet, François Husson <Francois.Husson@agrocampus-ouest.fr>

See Also

HMFA

Examples

data(wine)
  hierar <- list(c(2,5,3,10,9,2), c(4,2))
  res.hmfa <- HMFA(wine, H = hierar, type=c("n",rep("s",5)), graph = FALSE)
  plot(res.hmfa, invisible="quali")
  plot(res.hmfa, invisible="ind")

---

plot.MCA  Draw the Multiple Correspondence Analysis (MCA) graphs

Description

Draw the Multiple Correspondence Analysis (MCA) graphs.

Usage

## S3 method for class 'MCA'
plot(x, axes = c(1, 2), choix=c("ind","var","quanti.sup"),
  xlim = NULL, ylim = NULL,
  invisible = c("none","ind","var","ind.sup","quali.sup","quanti.sup"),
  col.ind = "blue", col.var = "red", col.quali.sup = "darkgreen",
  col.ind.sup = "darkblue", col.quanti.sup = "blue",
  label = c("all","none","ind","var","ind.sup","quali.sup","quanti.sup"),
  title = NULL, habillage = "none",
  palette = NULL, autoLab = c("auto","yes","no"), new.plot = FALSE,
  select = NULL, selectMod = NULL, unselect = 0.7,
  shadowtext = FALSE, ...)
Arguments

- **x**: an object of class MCA
- **axes**: a length 2 vector specifying the components to plot
- **choix**: the graph to plot ("ind" for the individuals and the categories, "var" for the variables, "quanti.sup" for the supplementary quantitative variables)
- **xlim**: range for the plotted 'x' values, defaulting to the range of the finite values of 'x'
- **ylim**: range for the plotted 'y' values, defaulting to the range of the finite values of 'y'
- **invisible**: string indicating if some points should not be drawn ("ind", "var", "ind.sup", "quali.sup", "quanti.sup")
- **col.ind**: a color for the individuals, if color ="none" the label is not written
- **col.var**: a color for the categories of categorical variables, if color ="none" the label is not written
- **col.quali.sup**: a color for the categorical supplementary variables, if color ="none" the label is not written
- **col.ind.sup**: a color for the supplementary individuals only if there is not habillage, if color ="none" the label is not written
- **col.quanti.sup**: a color for the supplementary quantitative variables, if color ="none" the label is not written
- **label**: print the labels of the points; "all" print all the labels; may be a vector with "ind" (for the individuals),"ind.sup" (for the supplementary individuals),"var" (for the active categories), "quali.sup" "var" (for the supplementary categories)
- **title**: string corresponding to the title of the graph you draw (by default NULL and a title is chosen)
- **habillage**: string corresponding to the color which are used. If "none", one color is used for the individual, another one for the categorical variables; if "quali", one color is used for each categorical variables; else if it is the position of a categorical variable, it colors according to the different categories of this variable
- **palette**: the color palette used to draw the points. By default colors are chosen. If you want to define the colors : palette=palette(c("black","red","blue")); or you can use: palette=palette(rainbow(30)), or in black and white for example: palette=palette(gray(seq(0,.9,len=25)))
- **autoLab**: if autoLab="auto", autoLab is equal to "yes" if there are less than 50 elements and "no" otherwise; if "yes", the labels of the drawn elements are placed in a "good" way (can be time-consuming if many elements), and if "no" the elements are placed quickly but may overlap
- **new.plot**: boolean, if TRUE, a new graphical device is created
- **select**: a selection of the elements that are drawn; see the details section
- **selectMod**: a selection of the categories that are drawn; see the details section
- **unselect**: may be either a value between 0 and 1 that gives the transparency of the unselected objects (if unselect=1 the transparency is total and the elements are not drawn, if unselect=0 the elements are drawn as usual but without any label) or may be a color (for example unselect="grey60")
- **shadowtext**: boolean; if true put a shadow on the labels (rectangles are written under the labels which may lead to difficulties to modify the graph with another program)
- **...**: further arguments passed to or from other methods, such as cex, cex.main, ...
Details

The argument `autoLab = "yes"` is time-consuming if there are many labels that overlap. In this case, you can modify the size of the characters in order to have less overlapping, using for example `cex=0.7`.

The `select` argument can be used in order to select a part of the elements (individuals if you draw the graph of individuals, or variables if you draw the graph of variables) that are drawn. For example, you can use:

- `select = 1:5` and then the elements 1:5 are drawn.
- `select = c("name1","name5")` and then the elements that have the names `name1` and `name5` are drawn.
- `select = "coord 10"` and then the 10 elements that have the highest (squared) coordinates on the 2 chosen dimensions are drawn.
- `select = "contrib 10"` and then the 10 elements that have the highest contribution on the 2 dimensions of your plot are drawn.
- `select = "cos2 5"` and then the 5 elements that have the highest cos2 on the 2 dimensions of your plot are drawn.
- `select = "dist 8"` and then the 8 elements that have the highest distance to the center of gravity are drawn.

The `selectMod` argument can be used in order to select the categories that are drawn.

Value

Returns the individuals factor map and the variables factor map.

Author(s)

Jeremy Mazet, Francois Husson <Francois.Husson@agrocampus-ouest.fr>

See Also

MCA

Examples

data (poison)
res.mca = MCA (poison, quali.sup = 3:4, quanti.sup = 1:2, graph=FALSE)
plot.MCA(res.mca,invisible=c("var","quali.sup"))
plot.MCA(res.mca,invisible="ind")
plot.MCA(res.mca,choix="var")

plot.MFA  

**Draw the Multiple Factor Analysis (MFA) graphs**

Description

Draw the Multiple Factor Analysis (MFA) graphs.
## S3 method for class 'MFA'
plot(x, axes = c(1, 2), choix = c("ind","var","group","axes","freq"),
     ellipse=NULL, ellipse.par=NULL,
     lab.grpe=TRUE, lab.var=TRUE, lab.ind=TRUE,
     lab.par=FALSE, lab.col=TRUE, habillage="ind", col.hab=NULL,
     invisible = c("none","ind","ind.sup","quanti","quanti.sup",
                   "quali","row","row.sup","col","col.sup"),
     partial = NULL, lim.cos2.var = 0.,
     chrono = FALSE, xlim = NULL, ylim = NULL,
     title = NULL, palette = NULL,
     autoLab = c("auto","yes","no"), new.plot = FALSE,
     select = NULL, unselect = 0.7, shadowtext = FALSE, ...)

Arguments

**x**
an object of class MFA

**choix**
a string corresponding to the graph that you want to do ("ind" for the individual
or categorical variables graph, "var" for the quantitative variables graph, "freq"
for the frequence or contingency tables, "axes" for the graph of the partial axes,
"group" for the groups representation)

**axes**
a length 2 vector specifying the components to plot

**ellipse**
boolean (NULL by default), if not null, draw ellipses around the individuals,
and use the results of coord.ellipse

**ellipse.par**
boolean (NULL by default), if not null, draw ellipses around the partial individ-
uals, and use the results of coord.ellipse

**lab.grpe**
boolean, if TRUE, the labels of the groups are drawn

**lab.var**
boolean, if TRUE, the labels of the variables are drawn

**lab.ind**
boolean, if TRUE, the labels of the mean points are drawn

**lab.par**
boolean, if TRUE, the labels of the partial points are drawn

**lab.col**
boolean, if TRUE, the labels of the columns for the contingency tables are drawn

**habillage**
string corresponding to the color which are used. If "ind", one color is used for
each individual; if "group" the individuals are colored according to the group;
else if it is the name or the position of a categorical variable, it colors according
to the different categories of this variable

**col.hab**
the colors to use. By default, colors are chosen

**invisible**
list of string; for choix ="ind", the individuals can be omit (invisible = "ind"), or
supplementary individuals (invisible = "ind.sup") or the centereg of gravity of the
categorical variables (invisible = "quali"); if invisible = c("ind","ind.sup"), just
the centers of gravity are drawn; if choix="var", invisible="quanti" suppress the
active variable and invisible = "quanti.sup" suppress the supplementary variables

**partial**
list of the individuals or of the center of gravity for which the partial points
should be drawn (by default, partial = NULL and no partial points are drawn)

**lim.cos2.var**
value of the square cosinus under with the points are not drawn
chrono

boolean, if TRUE, the partial points of a same point are linked (useful when groups correspond to different moment)

xlim

range for the plotted 'x' values, defaulting to the range of the finite values of 'x'

ylim

range for the plotted 'y' values, defaulting to the range of the finite values of 'y'

title

string corresponding to the title of the graph you draw (by default NULL and a title is chosen)

palette

the color palette used to draw the points. By default colors are chosen. If you want to define the colors: palette=palette(c("black","red","blue")); or you can use: palette=palette(rainbow(30)), or in black and white for example: palette=palette(gray(seq(0,.9,len=25))

autoLab

if autoLab="auto", autoLab is equal to "yes" if there are less than 50 elements and "no" otherwise; if "yes", the labels of the drawn elements are placed in a "good" way (can be time-consuming if many elements), and if "no" the elements are placed quickly but may overlap

new.plot

boolean, if TRUE, a new graphical device is created

select

a selection of the elements that are drawn; see the details section

unselect

may be either a value between 0 and 1 that gives the transparency of the unselected objects (if unselect=1 the transparency is total and the elements are not drawn, if unselect=0 the elements are drawn as usual but without any label) or may be a color (for example unselect="grey60")

shadowtext

boolean; if true put a shadow on the labels (rectangles are written under the labels which may lead to difficulties to modify the graph with another program)

...

further arguments passed to or from other methods, such as cex, cex.main, ...

Details

The argument autoLab = "yes" is time-consuming if there are many labels that overlap. In this case, you can modify the size of the characters in order to have less overlapping, using for example cex=0.7.

The select argument can be used in order to select a part of the elements (individuals if you draw the graph of individuals, or variables if you draw the graph of variables) that are drawn. For example, you can use:

select = 1:5 and then the elements 1:5 are drawn.
select = c("name1","name5") and then the elements that have the names name1 and name5 are drawn.
select = "coord 10" and then the 10 elements (individuals, variables, frequencies) that have the highest (squared) coordinates on the 2 chosen dimensions are drawn.
select = "contrib 10" and then the 10 elements (individuals, variables, frequencies) that have the highest contribution on the 2 dimensions of your plot are drawn.
select = "cos2 5" and then the 5 elements (individuals, variables, frequencies) that have the highest cos2 on the 2 dimensions of your plot are drawn.
selectMod = "contrib 5", the 5 categories that contribute the most to the two dimensions are drawn.
selectMod = "cos2 0.5", the categories that have a cos2 greater than 0.5 on the two dimensions are drawn.
selectMod = "v.test 2", the categories that have a v.test higher than the value on one of the two dimensions are drawn.
plot.PCA

Draw the Principal Component Analysis (PCA) graphs

Description

Plot the graphs for a Principal Component Analysis (PCA) with supplementary individuals, supplementary quantitative variables and supplementary categorical variables.

Usage

```r
## S3 method for class 'PCA'
plot(x, axes = c(1, 2), choix = c("ind","var"),
     ellipse = NULL, xlim = NULL, ylim = NULL, habillage="none",
     col.hab = NULL, col.ind="black", col.ind.sup="blue",
     col.quali="magenta", col.quanti.sup="blue", col.var="black",
     label = c("all","none","ind","ind.sup","quali","var","quanti.sup"),
     invisible = c("none","ind","ind.sup","quali","var","quanti.sup"),
     lim.cos2.var = 0., title = NULL, palette=NULL,
     autoLab = c("auto","yes","no"), new.plot = FALSE, select = NULL,
     unselect = 0.7, shadowtext = FALSE, ...)
```
Arguments

x  an object of class PCA
axes  a length 2 vector specifying the components to plot
choix  the graph to plot ("ind" for the individuals, "var" for the variables)
ellipse  boolean (NULL by default), if not null, draw ellipses around the individuals, and use the results of coord.ellipse
xlim  range for the plotted 'x' values, defaulting to the range of the finite values of 'x'
ylim  range for the plotted 'y' values, defaulting to the range of the finite values of 'y'
habillage  give no color for the individuals ("none"), a color for each individual ("ind"), or color the individuals among a categorical variable (give the number of the categorical variable)
col.hab  a vector with the color to use for the individuals
col.ind  a color for the individuals if there only is not habillage
col.ind.sup  a color for the supplementary individuals only if there is not habillage
col.quali  a color for the categories of categorical variable sonly if there is not habillage
col.quanti.sup  a color for the quantitative supplementary variables
col.var  a color for the variables
label  a list of character for the elements which are labelled (by default, all the elements are labelled ("ind", "ind.sup", "quali", "var", "quanti.sup"))
invisible  string indicating if some points should not be drawn ("ind", "ind.sup" or "quali" for the individual graph and "var" or "quanti.sup" for the correlation circle graph)
lim.cos2.var  value of the square cosinus under the variables are not drawn
title  string corresponding to the title of the graph you draw (by default NULL and a title is chosen)
palette  the color palette used to draw the points. By default colors are chosen. If you want to define the colors : palette=palette(c("black","red","blue")); or you can use: palette=palette(rainbow(30)), or in black and white for example: palette=palette(gray(seq(0,.9,len=25))
autoLab  if autoLab="auto", autoLab is equal to "yes" if there are less than 50 elements and "no" otherwise; if "yes", the labels of the drawn elements are placed in a "good" way (can be time-consuming if many elements), and if "no" the elements are placed quickly but may overlap
new.plot  boolean, if TRUE, a new graphical device is created
select  a selection of the elements that are drawn; see the details section
unselect  may be either a value between 0 and 1 that gives the transparency of the unselected objects (if unselect=1 the transparency is total and the elements are not drawn, if unselect=0 the elements are drawn as usual but without any label) or may be a color (for example unselect="grey60")
shadowtext  boolean; if true put a shadow on the labels (rectangles are written under the labels which may lead to difficulties to modify the graph with another program)
...  further arguments passed to or from other methods, such as cex, cex.main, ...
The argument `autoLab = "yes"` is time-consuming if there are many labels that overlap. In this case, you can modify the size of the characters in order to have less overlapping, using for example `cex=0.7`.

The `select` argument can be used in order to select a part of the elements (individuals if you draw the graph of individuals, or variables if you draw the graph of variables) that are drawn. For example, you can use:

```r
select = 1:5 and then the elements 1:5 are drawn.
select = c("name1","name5") and then the elements that have the names name1 and name5 are drawn.
select = "coord 10" and then the 10 elements that have the highest (squared) coordinates on the 2 chosen dimensions are drawn.
select = "contrib 10" and then the 10 elements that have the highest contribution on the 2 dimensions of your plot are drawn.
select = "cos2 5" and then the 5 elements that have the highest cos2 on the 2 dimensions of your plot are drawn.
select = "dist 8" and then the 8 elements that have the highest distance to the center of gravity are drawn.
```

**Value**

Returns the individuals factor map and the variables factor map.

**Author(s)**

Jeremy Mazet, Francois Husson <Francois.Husson@agrocampus-ouest.fr>

**See Also**

PCA

**Examples**

```r
data(decathlon)
res.pca <- PCA(decathlon, quanti.sup = 11:12, quali.sup = 13)
plot(res.pca, habillage = 13, col.hab=c("green","blue"))
## To automatically draw ellipses around the barycentres of all the categorical variables
plotellipses(res.pca)
## Selection of some individuals
plot(res.pca,select="contrib 7") #plot the 7 individuals that have the highest cos2
plot(res.pca,select="cos2 0.8") #plot the individuals that have a cos2 greater than 0.8
plot(res.pca,select="cos2 5") #plot the 5 individuals that have the highest cos2
plot(res.pca,choix="var",select="cos2 0.6") #plot the variables that have a cos2 greater than 0.6
```
**plot.spMCA**

*Draw the specific Multiple Correspondence Analysis (spMCA) graphs*

**Description**

Draw the specific Multiple Correspondence Analysis (spMCA) graphs.

**Usage**

```r
## S3 method for class 'spMCA'
plot(x, axes = c(1, 2), choix = c("ind", "var", "quanti.sup"),
     xlim = NULL, ylim = NULL, invisible = NULL,
     col.ind = "blue", col.var = "red", col.quali.sup = "darkgreen",
     col.ind.sup = "darkblue", col.quanti.sup = "blue",
     label = "all", cex = 1, title = NULL, habillage = "none",
     palette = NULL, autoLab = c("auto", "yes", "no"), new.plot = FALSE,
     select = NULL, selectMod = NULL, unselect = 0.7, ...)
```

**Arguments**

- `x`: an object of class spMCA
- `axes`: a length 2 vector specifying the components to plot
- `choix`: the graph to plot ("ind" for the individuals and the categories, "var" for the variables)
- `xlim`: range for the plotted 'x' values, defaulting to the range of the finite values of 'x'
- `ylim`: range for the plotted 'y' values, defaulting to the range of the finite values of 'y'
- `invisible`: string indicating if some points should not be drawn ("ind", "var")
- `col.ind`: a color for the individuals, if color = "none" the label is not written
- `col.var`: a color for the categories of categorical variables, if color = "none" the label is not written
- `col.quali.sup`: not used. a color for the categorical supplementary variables, if color = "none" the label is not written
- `col.ind.sup`: not used. a color for the supplementary individuals only if there is not habillage, if color = "none" the label is not written
- `col.quanti.sup`: not used. a color for the supplementary quantitative variables, if color = "none" the label is not written
- `label`: print the labels of the points
- `cex`: cf. function `par` in the `graphics` package
- `title`: string corresponding to the title of the graph you draw (by default NULL and a title is chosen)
- `habillage`: string corresponding to the color which are used. If "none", one color is used for the individual, another one for the categorical variables; if "quali", one color is used for each categorical variables; else if it is the position of a categorical variable, it colors according to the different categories of this variable
plotellipses

The color palette used to draw the points. By default colors are chosen. If you want to define the colors: palette=palette(c("black","red","blue")); or you can use: palette=palette(rainbow(30)), or in black and white for example: palette=palette(gray(seq(0,.9,len=25))

autoLab

if autoLab="auto", autoLab is equal to "yes" if there are less than 50 elements and "no" otherwise; if "yes", the labels of the drawn elements are placed in a "good" way (can be time-consuming if many elements), and if "no" the elements are placed quickly but may overlap.

new.plot

boolean, if TRUE, a new graphical device is created

select

a selection of the elements that are drawn; see the details section

selectMod

a selection of the categories that are drawn; see the details section

unselect

may be either a value between 0 and 1 that gives the transparency of the unselected objects (if unselect=1 the transparency is total and the elements are not drawn, if unselect=0 the elements are drawn as usual but without any label) or may be a color (for example unselect="grey60")

... further arguments passed to or from other methods, such as cex, cex.main, ...

Value

Returns the individuals factor map and the variables factor map.

Author(s)

Francois Husson <Francois.Husson@agrocampus-ouest.fr>

See Also

spMCA

Examples

data (poison)
res <- spMCA (poison[,3:8],excl=c(1,3))
plot(res,invisible="ind")
plot(res,invisible="var")
## Categories that have a cos2 less than 0.6 are not drawn
plot(res,invisible="var",selectMod="cos2 0.6")
plot(res,choix="var")

plotellipses

Draw confidence ellipses around the categories.

Description

Draw confidence ellipses around the categories.
Usage

plotellipses(model, keepvar = "all", axis = c(1, 2), means=TRUE, level = 0.95, magnify = 2, cex = 0.5, pch = 20, pch.means=15, type = c("g","p"), keepnames = TRUE, namescat = NULL, xlim=xlim, ylim=ylim, lwd=1, label="all", autoLab=c("auto","yes","no"),...)

Arguments

model an object of class MCA or PCA or MFA
keepvar a boolean or numeric vector of indexes of variables or a character vector of names of variables. If keepvar is "all","quali" or "quali.sup" variables which are plotted are all the categorical variables, only those which are used to compute the dimensions (active variables) or only the supplementary categorical variables. If keepvar is a numeric vector of indexes or a character vector of names of variables, only relevant variables are plotted.
axis a length 2 vector specifying the components to plot
means boolean which indicates if the confidence ellipses are for (the coordinates of) the means of the categories (the empirical variance is divided by the number of observations) or for (the coordinates of) the observations of the categories
level the confidence level for the ellipses
magnify numeric which control how the level names are magnified. A value of 2 means that the level names have character expansion equal to two times cex
cex cf. function par in the graphics package
pch plotting character for coordinates, cf. function par in the graphics package
pch.means plotting character for means, cf. function par in the graphics package
type cf. function xyplot in the lattice package
keepnames a boolean or numeric vector of indexes of variables or a character vector of names of variables. If keepnames is TRUE, names of levels are taken from the (modified) dataset extracted from modele, if FALSE trimming names is done. When trimming, names of levels are taken from the (modified) dataset extracted from modele, then, the corresponding number of characters of names of original variables plus 1 is removed. If keepnames is a vector of indexes or names, trimming is done on all variables excepted whose in keepnames
namescat a vector giving for each observation the value of categorical variable, each variable are stacked under each other. If NULL, names are taken from the (modified) dataset extracted from modele
xlim range for the plotted 'x' values, defaulting to the range of the finite values of 'x'
ylim range for the plotted 'y' values, defaulting to the range of the finite values of 'y'
lwd The line width, a positive number, defaulting to 1
label a list of character for the elements which are labelled (by default, "all", you can use "none", "ind", ind.sup))
autoLab if autoLab="auto", autoLab is equal to "y" if there are less than 50 elements and "no" otherwise; if "y", the labels of the drawn elements are placed in a "good" way (can be time-consuming if many elements), and if "no" the elements are placed quickly but may overlap
plotGPApartial

... further arguments passed to or from other methods

Value

Return a graph with the ellipses. If only one variable is chosen, the graph is different.

Author(s)

Pierre-Andre Cornillon, Francois Husson <Francois.Husson@agrocampus-ouest.fr>

See Also

MCA, PCA

Examples

## Not run:
data(poison)
res.mca = MCA(poison, quali.sup = 3:4, quanti.sup = 1:2)
plotellipses(res.mca)
plotellipses(res.mca,keepvar=1:4)

## End(Not run)
data(decathlon)
res.pca <- PCA(decathlon, quanti.sup = 11:12, quali.sup=13)
plotellipses(res.pca,keepvar=13)

---

plotGPApartial  

**Draw an interactive General Procrustes Analysis (GPA) map**

Description

Draw an interactive General Procrustes Analysis (GPA) map. The graph is interactive and clicking on a point will draw the partial points, if you click on a point for which the partial points are yet drawn, the partial points are deleted. To stop the interactive plot, click on the title (or in the top of the graph)

Usage

plotGPApartial(x, axes = c(1, 2),
              lab.ind.moy = TRUE, habillage = "ind",
              chrono = FALSE, draw.partial = NULL,
              xlim = NULL, ylim = NULL, cex = 1, title = NULL, palette = NULL, ...)
plotMFApartial

Arguments

- **x**: an object of class GPA
- **axes**: a length 2 vector specifying the components to plot
- **lab.ind.moy**: boolean, if TRUE, the label of the mean points are drawn
- **habillage**: string corresponding to the color which are used. If "ind", one color is used for each individual; if "group" the individuals are colored according to the group
- **chrono**: boolean, if TRUE, the partial points of a same point are linked (useful when groups correspond to different moment)
- **draw.partial**: data frame of a boolean variable for all the individuals and all the centers of gravity and with for which the partial points should be drawn (by default, NULL and no partial points are drawn)
- **xlim**: range for the plotted 'x' values, defaulting to the range of the finite values of 'x'
- **ylim**: range for the plotted 'y' values, defaulting to the range of the finite values of 'y'
- **cex**: cf. function `par` in the `graphics` package
- **title**: string corresponding to the title of the graph you draw (by default NULL and a title is chosen)
- **palette**: the color palette used to draw the points. By default colors are chosen. If you want to define the colors: `palette=palette(c("black","red","blue"));` or you can use: `palette=palette(rainbow(30)),` or in black and white for example: `palette=palette(gray(seq(0,.9,len=25)))`
- ... further arguments passed to or from other methods

Value

Returns the General Procrustes Analysis map.

Author(s)

Elisabeth Morand, Francois Husson <Francois.Husson@agrocampus-ouest.fr>

See Also

GPA

plotMFApartial  
Plot an interactive Multiple Factor Analysis (MFA) graph

Description

Draw an interactive Multiple Factor Analysis (MFA) graphs.
Usage

plotMFApartial(x, axes = c(1, 2),
    lab.ind = TRUE, lab.par = FALSE, habillage = "ind",
    chrono = FALSE, col.hab = NULL, invisible = NULL,
    draw.partial = NULL, xlim = NULL, ylim = NULL,
    cex = 1, title = NULL, palette = NULL, ...)

Arguments

- **x**: an object of class MFA
- **axes**: a length 2 vector specifying the components to plot
- **lab.ind**: boolean, if TRUE, the label of the mean points are drawn
- **lab.par**: boolean, if TRUE, the label of the partial points are drawn
- **habillage**: string corresponding to the color which are used. If "ind", one color is used for each individual; if "quali" the individuals are colored according to one categorical variable; if "group" the individuals are colored according to the group
- **chrono**: boolean, if TRUE, the partial points of a same point are linked (useful when groups correspond to different moment)
- **col.hab**: the colors to use. By default, colors are chosen
- **invisible**: list of string; for choix = "ind", the individuals can be omit (invisible = "ind"); or supplementary individuals (invisible = "ind.sup") or the center of gravity of the categorical variables (invisible = "quali"); if invisible = c("ind", "ind.sup"), just the centers of gravity are drawn
- **draw.partial**: data frame of a boolean variable for all the individuals and all the centers of gravity and with for which the partial points should be drawn (by default, NULL and no partial points are drawn)
- **xlim**: range for the plotted 'x' values, defaulting to the range of the finite values of 'x'
- **ylim**: range for the plotted 'y' values, defaulting to the range of the finite values of 'y'
- **cex**: cf. function `par` in the `graphics` package
- **title**: string corresponding to the title of the graph you draw (by default NULL and a title is chosen)
- **palette**: the color palette used to draw the points. By default colors are chosen. If you want to define the colors: palette = palette(c("black","red","blue")); or you can use: palette = palette(rainbow(30)), or in black and white for example: palette = palette(gray(seq(0,.9,len=25))
- **...**: further arguments passed to or from other methods

Value

Draw a graph with the individuals and the centers of gravity. The graph is interactive and clicking on a point will draw the partial points, if you click on a point for which the partial points are yet drawn, the partial points are deleted. To stop the interactive plot, click on the title (or in the top of the graph)
poison

Author(s)
Jeremy Mazet, Francois Husson <Francois.Husson@agrocampus-ouest.fr>

See Also
MFA, plot.MFA

Examples

```r
## Not run:
data(wine)
res.wine = MFA(wine, group=c(2,5,3,10,9,2), type=c("n", rep("s", 5)), ncp=5,
  name.group=c("orig","olf","vis","olfag","gust","ens"),
  num.group.sup=c(1,6), graph=FALSE)
liste = plotMFApartial(res.wine)
## End(Not run)
```

poison

Poison

Description

The data used here refer to a survey carried out on a sample of children of primary school who suffered from food poisoning. They were asked about their symptoms and about what they ate.

Usage

data(poison)

Format

A data frame with 55 rows and 15 columns.

Examples

data(poison)
res.mca <- MCA(poison, quanti.sup = 1:2, quali.sup=c(3,4))
Description

The data used here refer to a survey carried out on a sample of children of primary school who suffered from food poisoning. They were asked about their symptoms and about what they ate.

Usage

data(poison)

Format

A data frame with 55 rows and 3 columns (the sex, if they are sick or not, and a textual variable with their symptom and what they eat).

Examples

data(poison.text)
res.text <- textual(poison.text, num.text = 3, contingence.by = c(1,2))
## Contingence table for the sex variable, the sick variable and the couple
## of variable sick-sex
res.text2 <- textual(poison.text, num.text = 3, contingence.by = list(1,2,c(1,2)))

Description

Donnees genomiques sur les poulets

Usage

data(poulet)

Format

Un data frame avec 43 observations sur 7407 variables.

Régime  a factor with levels J16 J16R16 J16R5 J48 J48R24 N

Autre variable  variable quantitative correspondant a l'expression d’un gene
Examples

```r
## Not run:
data(poulet)
res.pca = PCA(poulet, quali.sup=1, graph=FALSE)
plot(res.pca)
plot(res.pca, habillage=1, label="quali", palette=palette(c("black","red","blue","darkgreen","purple","orange")))
dimdesc(res.pca)
## Dessine des ellipses autour des centres de gravite
aa=cbind.data.frame(poulet[,1],res.pca$ind$coord)
bb=coord.ellipse(aa,bary=TRUE)
plot.PCA(res.pca, habillage=1, ellipse=bb)
## End(Not run)
```

prefpls

**Scatter plot and additional variables with quality of representation contour lines**

Description

This function is useful to interpret the usual graphs $(x,y)$ with additional quantitative variables.

Usage

```r
prefpls(donnee, var1 = 1, var2 = 2, firstvar = 3,
        lastvar = ncol(donnee), levels = c(0.2,0.4,0.6,0.7,0.8,0.9,1),
        asp = 1, nbchar = max(nchar(colnames(donnee))), title = NULL,
        choix="var")
```

Arguments

- `donnee`: a data frame made up of quantitative variables
- `var1`: the position of the variable corresponding to the x-axis
- `var2`: the position of the variable corresponding to the y-axis
- `firstvar`: the position of the first endogenous variable
- `lastvar`: the position of the last endogenous variable (by default the last column of `donnee`)
- `levels`: a list of the levels displayed in the graph of variables
- `asp`: aspect ratio for the graph of the individuals
- `nbchar`: the number of characters used for the labels of the variables
- `title`: string corresponding to the title of the graph you draw (by default NULL and a title is chosen)
- `choix`: the graph to plot ("ind" for the individuals, "var" for the variables)

Details

This function is very useful when there is a strong correlation between two variables $x$ and $y$.
**Value**

A scatter plot of the individuals

A graph with additional variables and the quality of representation contour lines.

**Author(s)**

Francois Husson &lt;Francois.Husson@agrocampus-ouest.fr&gt;

**References**


**Examples**

```r
data(decathlon)
prefpls(decathlon[,c(11,12,1:10)])
```

---

**print.CA**

*Print the Correspondance Analysis (CA) results*

**Description**

Print the Correspondance Analysis (CA) results.

**Usage**

```r
## S3 method for class 'CA'
print(x, file = NULL, sep = ";", ...)  
```

**Arguments**

- `x` an object of class CA
- `file` A connection, or a character string naming the file to print to. If NULL (the default), the results are not printed in a file
- `sep` character string to insert between the objects to print (if the argument file is not NULL
- `...` further arguments passed to or from other methods

**Author(s)**

Jeremy Mazet, Francois Husson &lt;Francois.Husson@agrocampus-ouest.fr&gt;

**See Also**

`CA, write.infile`
**print.FAMD**

Print the Multiple Factor Analysis of mixt Data (FAMD) results.

**Description**

Print the Multiple Factor Analysis of mixt Data (FAMD) results.

**Usage**

```r
## S3 method for class 'FAMD'
print(x, file = NULL, sep = ";", ...)
```

**Arguments**

- `x`: an object of class FAMD
- `file`: A connection, or a character string naming the file to print to. If NULL (the default), the results are not printed in a file
- `sep`: character string to insert between the objects to print (if the argument file is not NULL
- `...`: further arguments passed to or from other methods

**Author(s)**

Jeremy Mazet, Francois Husson <Francois.Husson@agrocampus-ouest.fr>

**See Also**

- [FAMD](#)

---

**print.GPA**

Print the Generalized Procrustes Analysis (GPA) results.

**Description**

Print the Generalized Procrustes Analysis (GPA) results.

**Usage**

```r
## S3 method for class 'GPA'
print(x, file = NULL, sep = ";", ...)
```

**Author(s)**

Jeremy Mazet, Francois Husson <Francois.Husson@agrocampus-ouest.fr>

**See Also**

- [GPA](#)
Arguments

x an object of class GPA
file A connection, or a character string naming the file to print to. If NULL (the default), the results are not printed in a file
sep character string to insert between the objects to print (if the argument file is not NULL
...
... further arguments passed to or from other methods

Author(s)

Elisabeth Morand, Francois Husson <Francois.Husson@agrocampus-ouest.fr>

See Also

GPA, write.infile

Description

Print the Hierarchical Clustering on Principal Components (HCPC) results.

Usage

## S3 method for class 'HCPC'
print(x, file = NULL, sep = ";", ...)


**print.HMFA**  
*Print the Hierarchical Multiple Factor Analysis results*

---

**Description**

Print the Hierarchical Multiple Factor Analysis results.

**Usage**

```r
## S3 method for class 'HMFA'
print(x, file = NULL, sep = ";", ...)  
```

**Arguments**

- `x`: an object of class HMFA
- `file`: A connection, or a character string naming the file to print to. If NULL (the default), the results are not printed in a file
- `sep`: character string to insert between the objects to print (if the argument file is not NULL
- `...`: further arguments passed to or from other methods

**Author(s)**

Sebastien Le, Francois Husson <Francois.Husson@agrocampus-ouest.fr>

**See Also**

- `HMFA`, `write.infile`

---

**print.MCA**  
*Print the Multiple Correspondance Analysis (MCA) results*

---

**Description**

Print the Multiple Correspondance Analysis (spMCA) results.

**Usage**

```r
## S3 method for class 'MCA'
print(x, file = NULL, sep = ";", ...)  
```

---
Arguments

\( x \)  an object of class MCA

\( \text{file} \)  A connection, or a character string naming the file to print to. If \text{NULL} (the default), the results are not printed in a file

\( \text{sep} \)  character string to insert between the objects to print (if the argument \text{file} is not \text{NULL})

\( \ldots \)  further arguments passed to or from other methods

Author(s)

Francois Husson <Francois.Husson@agrocampus-ouest.fr>

See Also

\( \text{MCA}, \text{write.infile} \)

print.MFA \hspace{1cm} \text{Print the Multiple Factor Analysis results}

Description

Print the Multiple Factor Analysis results.

Usage

\[
\text{## S3 method for class 'MFA'} \\
\text{print}(x, \text{file} = \text{NULL}, \text{sep} = ":", \ldots)
\]

Arguments

\( x \)  an object of class MFA

\( \text{file} \)  A connection, or a character string naming the file to print to. If \text{NULL} (the default), the results are not printed in a file

\( \text{sep} \)  character string to insert between the objects to print (if the argument \text{file} is not \text{NULL})

\( \ldots \)  further arguments passed to or from other methods

Author(s)

Jeremy Mazet, Francois Husson <Francois.Husson@agrocampus-ouest.fr>

See Also

\( \text{MFA}, \text{write.infile} \)
print.PCA

Print the Principal Component Analysis (PCA) results

Description

Print the Principal Component Analysis (PCA) results.

Usage

```r
## S3 method for class 'PCA'
print(x, file = NULL, sep = ";", ...)  
```

Arguments

- **x**: an object of class PCA
- **file**: A connection, or a character string naming the file to print to. If NULL (the default), the results are not printed in a file
- **sep**: character string to insert between the objects to print (if the argument file is not NULL)
- **...**: further arguments passed to or from other methods

Author(s)

Jeremy Mazet, Francois Husson <Francois.Husson@agrocampus-ouest.fr>

See Also

- `PCA`
- `write.infile`

Examples

```r
## Not run:
data(decathlon)
res.pca <- PCA(decathlon, quanti.sup = 11:12, quali.sup = 13)
print(res.pca, file="c:/essai.csv", sep = ";")

## End(Not run)
```
### print.spMCA

*Print the specific Multiple Correspondance Analysis (spMCA) results*

**Description**

Print the specific Multiple Correspondance Analysis (spMCA) results.

**Usage**

```r
## S3 method for class 'spMCA'
print(x, file = NULL, sep = ",", ...)  
```

**Arguments**

- `x`: an object of class spMCA
- `file`: A connection, or a character string naming the file to print to. If NULL (the default), the results are not printed in a file
- `sep`: character string to insert between the objects to print (if the argument file is not NULL
- `...`: further arguments passed to or from other methods

**Author(s)**

Francois Husson <Francois.Husson@agrocampus-ouest.fr>

**See Also**

- `spMCA`, `write.infile`

---

### reconst

*Reconstruction of the data from the PCA or MFA results*

**Description**

Reconstruct the data from the PCA or MFA results.

**Usage**

```r
reconst(res, ncp=NULL)  
```

**Arguments**

- `res`: an object of class PCA or MFA
- `ncp`: number of dimensions used to reconstitute the data (by default NULL and the number of dimensions calculated for the PCA or MFA is used)
RegBest

Value

Returns a data frame with the number of individuals and the number of variables used for the PCA or MFA

Author(s)

Francois Husson <Francois.Husson@agrocampus-ouest.fr>, Julie Josse <Julie.Josse@agrocampus-ouest.fr>

See Also

PCA, MFA

Examples

data(decathlon)
res.pca <- PCA(decathlon, quanti.sup = 11:12, quali.sup=13, graph=FALSE)
rec <- reconst(res.pca,ncp=2)

RegBest

Select variables in multiple linear regression

Description

Find an optimal submodel

Usage

RegBest(y, x, int = TRUE, wt=NULL, na.action = na.omit, method=c("r2","Cp", "adjr2"), nbest=1)

Arguments

y
A response vector

x
A matrix of predictors

int
Add an intercept to the model

wt
Optional weight vector

na.action
Handling missing values

method
Calculate R-squared, adjusted R-squared or Cp to select the model. By default a the F-test on the r-square is used

nbest
number of best models for each set of explained variables (by default 1)

Value

Returns the objects

all
gives all the nbest best models for a given number of variables

best
the best model
Author(s)

Francois Husson <husson@agrocampus-ouest.fr>

See Also

lm

Examples

```r
data(milk)
res = RegBest(y=milk[,6],x=milk[,-6])
res$best
```

Description

Dataset to illustrate one-way and Two-way analysis of variance

Usage

```r
data(senso)
```

Format

Dataset with 45 rows and 3 columns: Score, Product and Day

Examples

```r
## Example of 2-way analysis of variance
data(senso)
res <- AovSum (Score~ Product + Day, data=senso)
res

## Example of 2-way analysis of variance with interaction
data(senso)
res2 <- AovSum (Score~ Product + Day + Product : Day, data=senso)
res2
```
Simulate by bootstrap

Description

Simulate by bootstrap

Usage

```
simule(data, nb.simul)
```

Arguments

data
A data frame from which the rows are the original data from which the simualte data are calculated (by the average of a bootstrap sample. The columns corresponds to the variables for which the simulation should be done. The first column must be a factor allowing to group the rows. A bootstrap simulation is done for each level of this factor.

nb.simul
The number of simulations.

Details

The simulation is independently done for each level of the factor. The number of rows can be different for each levels.

Value

- **mean**: Data.frame with all the levels of the factor variable, and for each variable, the mean of the original data.
- **simul**: Data.frame with all the levels of the factor variable, and for each variable, the nb.simul bootstrap simulations.
- **simul.mean**: Data.frame with all the levels of the factor variable, and for each variable, the mean of the simulated data.

Author(s)

Jeremy Mazet
spMCA: Specific Multiple Correspondence Analysis (spMCA)

Description

Performs Specific Multiple Correspondence Analysis (spMCA) with supplementary individuals, supplementary quantitative variables and supplementary categorical variables.

Usage

spMCA(data, excl=NULL, ncp=5)

Arguments

data: a data frame with \( n \) rows (individuals) and \( p \) columns (categorical variables)
excl: a vector with the indices of the categories to exclude
ncp: number of dimensions kept in the results (by default 5)

Value

Returns a list including:
eig: a matrix containing all the eigenvalues, the percentage of variance and the cumulative percentage of variance
ind: a list of matrices containing all the results for the active individuals (coordinates, contributions)
var: a list of matrices containing all the results for the active variables (coordinates, contribution, square cosine, v.test, eta2, contribution of the variables)
call: a list with some statistics

Returns the graphs of the individuals and categories and the graph with the variables.

Author(s)

Nicolas Robette, Francois Husson <husson@agrocampus-ouest.fr>,

References


See Also

print.spMCA, plot.spMCA
Examples

```r
## Not run:
data (poison)
res <- spMCA (poison[,3:8], excl=c(1,3))
## End(Not run)
```

summary.CA

Printing summaries of correspondence analysis objects

Description

Printing summaries of correspondence analysis objects

Usage

```r
## S3 method for class 'CA'
summary(object, nb.dec = 3, nbelements = 10, ncp = 3, align.names = TRUE, file = "", ...) 
```

Arguments

- `object`: an object of class CA
- `nb.dec`: number of decimal printed
- `nbelements`: number of elements written (rows, columns, ...); use `nbelements = Inf` if you want to have all the elements
- `ncp`: number of dimensions printed
- `align.names`: boolean, if TRUE the names of the objects are written using the same number of characters
- `file`: a connection, or a character string naming the file to print to
- `...`: further arguments passed to or from other methods, such as `cex`, `cex.main`, ...

Author(s)

François Husson <husson@agrocampus-ouest.fr>

See Also

- `CA`
**summary.FAMD**  
*Printing summaries of FAMD objects*

**Description**

Printing summaries of factor analysis on mixed data objects

**Usage**

```r
## S3 method for class 'FAMD'
summary(object, nb.dec = 3, nbelements=10, 
         ncp = 3, align.names=TRUE, file="", ...)```

**Arguments**

- **object**: an object of class FAMD
- **nb.dec**: number of decimal printed
- **nbelements**: number of elements written (rows, columns, ...); use nbelements = Inf if you want to have all the elements
- **ncp**: number of dimensions printed
- **align.names**: boolean, if TRUE the names of the objects are written using the same number of characters
- **file**: a connection, or a character string naming the file to print to
- **...**: further arguments passed to or from other methods, such as cex, cex.main, ...

**Author(s)**

François Husson <husson@agrocampus-ouest.fr>

**See Also**

- FAMD

---

**summary.MCA**  
*Printing summaries of MCA objects*

**Description**

Printing summaries of multiple correspondence analysis objects

**Usage**

```r
## S3 method for class 'MCA'
summary(object, nb.dec = 3, nbelements=10, 
         ncp = 3, align.names=TRUE, file="", ...)```
Arguments

- `object`: an object of class MCA
- `nb.dec`: number of decimal printed
- `nbelements`: number of elements written (rows, columns, ...); use `nbelements = Inf` if you want to have all the elements
- `ncp`: number of dimensions printed
- `align.names`: boolean, if TRUE the names of the objects are written using the same number of characters
- `file`: a connection, or a character string naming the file to print to
- `...`: further arguments passed to or from other methods, such as `cex`, `cex.main`, ...

Author(s)

Francois Husson <husson@agrocampus-ouest.fr>

See Also

MCA

Description

Printing summaries of multiple factor analysis objects

Usage

```r
## S3 method for class 'MFA'
summary(object, nb.dec = 3, nbelements=10, 
ncp = 3, align.names=TRUE, file="", ...)  
```

Arguments

- `object`: an object of class MFA
- `nb.dec`: number of decimal printed
- `nbelements`: number of elements written (rows, columns, ...); use `nbelements = Inf` if you want to have all the elements
- `ncp`: number of dimensions printed
- `align.names`: boolean, if TRUE the names of the objects are written using the same number of characters
- `file`: a connection, or a character string naming the file to print to
- `...`: further arguments passed to or from other methods, such as `cex`, `cex.main`, ...
summary.PCA

Author(s)

Francois Husson <husson@agrocampus-ouest.fr>

See Also

MFA

summary.PCA

Printing summaries of PCA objects

Description

Printing summaries of principal component analysis objects

Usage

## S3 method for class 'PCA'
summary(object, nb.dec = 3, nbelements=10,
ncp = 3, align.names=TRUE, file="", ...)  

Arguments

object an object of class PCA
nb.dec number of decimal printed
nbelements number of elements written (rows, columns, ...); use nbelements = Inf if you
want to have all the elements
ncp number of dimensions printed
align.names boolean, if TRUE the names of the objects are written using the same number
of characters
file a connection, or a character string naming the file to print to
... further arguments passed to or from other methods, such as cex, cex.main, ...

Author(s)

Francois Husson <husson@agrocampus-ouest.fr>

See Also

PCA
### svd.triplet

**Singular Value Decomposition of a Matrix**

**Description**

Compute the singular-value decomposition of a rectangular matrix with weights for rows and columns.

**Usage**

```r
svd.triplet(X, row.w=NULL, col.w=NULL, ncp=Inf)
```

**Arguments**

- `X` a data matrix
- `row.w` vector with the weights of each row (NULL by default and the weights are uniform)
- `col.w` vector with the weights of each column (NULL by default and the weights are uniform)
- `ncp` the number of components kept for the outputs

**Value**

- `vs` a vector containing the singular values of 'x';
- `u` a matrix whose columns contain the left singular vectors of 'x';
- `v` a matrix whose columns contain the right singular vectors of 'x'.

**See Also**

`svd`

### tab.disjonctif

**Make a disjonctif table**

**Description**

Make a disjonctif table.

**Usage**

```r
tab.disjonctif(tab)
```

**Arguments**

- `tab` a data frame with factors
The disjonctif table

**Description**

Create a disjunctive table. The missing values are replaced by the proportion of the category.

**Usage**

```r
tab.disjonctif.prop(tab, seed=NULL, row.w=NULL)
```

**Arguments**

- `tab`: a data frame with factors
- `row.w`: an optional row weights (by default, a vector of 1 for uniform row weights)
- `seed`: a single value, interpreted as an integer for the set.seed function (if seed = NULL, missing values are initially imputed by the mean of each variable)

**Value**

The disjonctif table.prop

---

**tea**

*tea (data)*

**Description**

The data used here concern a questionnaire on tea. We asked to 300 individuals how they drink tea (18 questions), what are their product’s perception (12 questions) and some personal details (4 questions).

**Usage**

```r
data(tea)
```

**Format**

A data frame with 300 rows and 36 columns. Rows represent the individuals, columns represent the different questions. The first 18 questions are active ones, the 19th is a supplementary quantitative variable (the age) and the last variables are supplementary categorical variables.
Examples

data(tea)
res.mca=MCA(tea,quanti.sup=19,quali.sup=2:36)
plot(res.mca,invisible=c("var","quali.sup","quanti.sup"),cex=0.7)
plot(res.mca,invisible=c("ind","quali.sup","quanti.sup"),cex=0.8)
plot(res.mca,invisible=c("quali.sup","quanti.sup"),cex=0.8)
dimdesc(res.mca)
plotellipses(res.mca,keepvar=1:4)

## make a hierarchical clustering: click on the tree to define the number of clusters
## HCPC(res.mca)

---

### Text mining

**Description**

Calculates the number of occurrence of each word and a contingency table

**Usage**

```r
textual(tab, num.text, contingence.by=1:ncol(tab),
   maj.in.min = TRUE, sep.word=NULL)
```

**Arguments**

- `tab`: a data frame with one textual variable
- `num.text`: indice of the textual variable
- `contingence.by`: a list with the indices of the variables for which a contingency table is calculated
  - by default, a contingency table is calculated for all the variables (except the textual one). A contingency table can also be calculated for couple of variables. If `contingence.by` is equal to `num.text`, then the contingency table is calculated for each row of the data table
- `maj.in.min`: boolean, if TRUE majuscule are transformed in minuscule
- `sep.word`: a string with all the characters which correspond to separator of words (by default, `sep.word` = "; (,?,/,:!$§=\n;){<>[])@-"")

**Value**

Returns a list including:

- `cont.table`: the contingency table with in rows the categories of the categorical variables (or the couple of categories), and in column the words, and in each cell the number of occurrence
- `nb.words`: a data.frame with all the words and for each word, the number of lists in which it is present, and the number of occurrence
Author(s)
Francois Husson <Francois.Husson@agrocampus-ouest.fr>

See Also
CA, descfreq

Examples

data(wine)
res.pca = PCA(wine, ncp=5, quali.sup = 1:2)
## Example of PCA

## Not run:
## Example of MCA
res.mca = MCA(wine, ncp=5, quanti.sup = 3:ncol(wine))

Description
The data used here refer to 21 wines of Val de Loire.

Usage

data(wine)

Format
A data frame with 21 rows (the number of wines) and 31 columns: the first column corresponds to
the label of origin, the second column corresponds to the soil, and the others correspond to sensory
descriptors.

Source
Centre de recherche INRA d'Angers

Examples

data(wine)

## Example of PCA
res.pca = PCA(wine, ncp=5, quali.sup = 1:2)

## Not run:
## Example of MCA
res.mca = MCA(wine, ncp=5, quanti.sup = 3:ncol(wine))
## Example of MFA
res.mfa = MFA(wine.group=c(2,5,3,10,9,2),type=c("n",rep("s",5)),ncp=5,
name.group=c("orig","olf","vis","olfag","gust","ens"),
num.group.sup=c(1,6),graph=FALSE)
plotellipses(res.mfa)
plotellipses(res.mfa,keepvar="Label") ## for 1 variable
## End(Not run)

---

### write.infile

**Print in a file**

**Description**

Print in a file.

**Usage**

```r
write.infile(X, file, sep=";", append = FALSE, nb.dec=4)
```

**Arguments**

- **X**: an object of class list, data.frame, matrix, ...
- **file**: A connection, or a character string naming the file to print to
- **sep**: character string to insert between the objects to print (if the argument file is not NULL)
- **append**: logical. If TRUE output will be appended to file; otherwise, it will overwrite the contents of file.
- **nb.dec**: number of decimal printed, by default 4

**Author(s)**

Francois Husson <Francois.Husson@agrocampus-ouest.fr>

**Examples**

```r
## Not run:
data(decathlon)
res.pca <- PCA(decathlon, quanti.sup = 11:12, quali.sup = 13)
write.infile(res.pca, file="c:/essai.csv", sep = ";")
## End(Not run)
```
Index

*Topic **algebra**
  
  *svd.triplet, 77*
  
  *tab.disjonctif, 77*
  
  *tab.disjonctif.prop, 78*

*Topic **datasets**
  
  *children, 8*
  
  *decathlon, 12*
  
  *footsize, 18*
  
  *geomorphology, 19*
  
  *hobbies, 26*
  
  *JO, 27*
  
  *milk, 32*
  
  *mortality, 33*
  
  *poison, 59*
  
  *poison.text, 60*
  
  *poulet, 60*
  
  *senso, 70*
  
  *tea, 78*
  
  *wine, 80*

*Topic **dplot**
  
  *autoLab, 4*
  
  *coord.ellipse, 10*
  
  *graph.var, 21*
  
  *plot.CA, 36*
  
  *plot.cadres, 38*
  
  *plot.DMFA, 39*
  
  *plot.FAMD, 40*
  
  *plot.GPA, 42*
  
  *plot.HCPC, 43*
  
  *plot.HMFA, 44*
  
  *plot.MCA, 45*
  
  *plot.MFA, 47*
  
  *plot.PCA, 50*
  
  *plot.spMCA, 53*
  
  *plotGPAsample, 56*
  
  *plotMFAsample, 57*
  
  *prefpls, 61*
  
  *simule, 71*

*Topic **models**
  
  *AovSum, 3*
  
  *RegBest, 69*

*Topic **multivariate**
  
  *CA, 5*
  
  *cadoses, 6*
  
  *coeffRv, 8*
  
  *condes, 9*
  
  *descfrec, 12*
  
  *dimdesc, 13*
  
  *DMFA, 14*
  
  *estim_ncp, 16*
  
  *FAMD, 17*
  
  *GPA, 19*
  
  *HCPC, 22*
  
  *HMFA, 24*
  
  *MCA, 27*
  
  *MFA, 29*
  
  *PCA, 34*
  
  *plotellipses, 54*
  
  *reconst, 68*
  
  *spMCA, 72*
  
  *textual, 79*

*Topic **print**
  
  *print.CA, 62*
  
  *print.FAMD, 63*
  
  *print.GPA, 63*
  
  *print.HCPC, 64*
  
  *print.HMFA, 65*
  
  *print.MCA, 65*
  
  *print.MFA, 66*
  
  *print.PCA, 67*
  
  *print.spMCA, 68*
  
  *write.infile, 81*

  *agnes, 23*
  
  *aov, 4*
  
  *AovSum, 3*
  
  *autoLab, 4*

  *CA, 5, 14, 38, 62, 73, 80*
INDEX

catdes, 6, 10, 13, 23, 24, 38
children, 8
coeffRV, 8
condes, 7, 9, 13
coord.ellipse, 10
decathlon, 12
descfreq, 12, 80
dimdesc, 6, 13, 15, 25, 29, 31, 35
DMFA, 14, 22, 40
estim.ncp, 16
FAMD, 17, 41, 63, 74
footsize, 18
geomorphology, 19
GPA, 19, 43, 57, 64
graph.var, 21
HCPC, 22, 43, 44, 64
HMFA, 14, 22, 24, 45, 65
hobbies, 26
JO, 27
lm, 4, 70
MCA, 14, 22, 27, 47, 56, 66, 75
MFA, 14, 22, 29, 50, 59, 66, 69, 76
milk, 32
mortality, 33
par, 22, 42, 45, 53, 55, 57, 58
PCA, 14, 16, 22, 34, 52, 56, 67, 69, 76
plot.CA, 6, 36
plot.catdes, 7, 38
plot.DMFA, 15, 39
plot.FAMD, 18, 40
plot.GPA, 42
plot.HCPC, 24, 43
plot.HMFA, 25, 44
plot.MCA, 28, 29, 45
plot.MFA, 31, 47, 59
plot.PCA, 35, 50
plot.spMCA, 53, 72
plot.ellipses, 29, 54
plotGPAPartial, 56
plotMFApartial, 57
poison, 59
poison.text, 60
poulet, 60
prefpls, 61
print.CA, 6, 62
print.FAMD, 18, 63
print.GPA, 63
print.HCPC, 64
print.HMFA, 25, 65
print.MCA, 29, 65
print.MFA, 31, 66
print.PCA, 35, 67
print.spMCA, 68, 72
reconst, 68
RegBest, 69
senso, 70
simule, 11, 71
spMCA, 54, 68, 72
summary.CA, 6, 73
summary.FAMD, 18, 74
summary.MCA, 29, 74
summary.MFA, 31, 75
summary.PCA, 35, 76
svd, 77
svd.triplet, 77
tab.disjonctif, 77
tab.disjonctif.prop, 78
tea, 78
textual, 13, 79
wine, 80
write.infile, 62, 64–68, 81
xyplot, 55