Package 'GDAtools'

May 18, 2020

Type Package

Title A Toolbox for Geometric Data Analysis and More

Version 1.5

Date 2020-05-14

Author Nicolas Robette

Maintainer Nicolas Robette <nicolas.robette@uvsq.fr>

Imports FactoMineR, nleqslv, nnet, ggplot2, ggrepel, RColorBrewer,
rlang, moreparty

Suggests cluster, WeightedCluster

Description Contains functions for 'specific' Multiple Correspondence Analysis,
Class Specific Analysis, Multiple Factor Analysis, 'standardized' MCA, computing and plot-
ting structuring factors and concentration ellipses,
inductive tests and others tools for Geometric Data Analy-
sis (Le Roux & Rouanet (2005) <doi:10.1007/1-4020-2236-0>). It also provides functions
for the translation of logit models coefficients into percentages (Deau-
vieau (2010) <doi:10.1177/0759106309352586>), weighted contingency tables, an association
measure for contingency tables (``Percentages of Maximum Deviation from Indepen-
dence'', aka PEM, see Ci-
bois (1993) <doi:10.1177/075910639304000103>) and some tools to measure bivariate associa-
tions between variables
(phi, Cramér V, correlation coefficient, eta-squared...).

License GPL (>= 2)

URL http://nicolas.robette.free.fr/outils_eng.html

NeedsCompilation no

Repository CRAN

Date/Publication 2020-05-17 22:40:05 UTC

R topics documented:

---

assoc.catcont	*Bivariate association*

Description

Measures the association between a categorical variable and a continuous variable

Usage

```
assoc.catcont(x,y)
```

Arguments

x	the categorical variable (must be a factor)
y	the continuous variable (must be a factor)

Value

A list with the following elements :

eta.squared	eta-squared between the two variables
cor.coeff	correlation coefficient between the two variables, for each level of the categorical variable

Author(s)

Nicolas Robette

References

Rakotomalala R., 'Comprendre la taille d'effet (effect size)', [http://eric.univ-lyon2.fr/~ricco/cours/slides/effect_size.pdf]

See Also

assoc.twocat, condesc, catdesc

Examples

```
data(Taste)
getindexcat(Taste[,1:5])
mca <- speMCA(Taste[,1:5],excl=c(3,6,9,12,15))
assoc.catcont(Taste$Gender,mca$ind$coord[,1])
```

---

assoc.twocat	*Bivariate association*

Description

Cross-tabulation and measures of association between two categorical variables

Usage

```
assoc.twocat(x,y,w=rep.int(1,length(x)),na=TRUE)
```

Arguments

x	the first categorical variable (must be a factor)
y	the second categorical variable (must be a factor)
w	an optional numeric vector of weights (by default, a vector of 1 for uniform weights)
na	logical. If TRUE (default), 'NA' are treated as a category. If FALSE, they are ignored

Value

A list with the following elements :

freq	cross-tabulation
prop	percentages
rprop	raw percentages
cprop	column percentages
v2.cramer	Cramer's V2 between the two variables
phi	the matrix of the phi values for each pair of levels

Author(s)

Nicolas Robette

References

Rakotomalala R., 'Comprendre la taille d'effet (effect size)', http://eric.univ-lyon2.fr/~ricco/cours/slides/effect_size.pdf

See Also

[assoc.catcont](), [condesc](), [catdesc]()

Examples

```
data(Music)
assoc.twocat(Music$Jazz,Music$Age)
```

---

burt   *Computes a Burt table*

Description

Computes a Burt table from a data frame composed of categorical variables.

Usage

```
burt(data)
```

Arguments

data   data frame with n rows (individuals) and p columns (categorical variables)

Details

A Burt table is a symmetric table that is used in correspondence analysis. It shows the frequencies for all combinations of categories of pairs of variables.

Value

Returns a square matrix. Its dimension is equal to the total number of categories in the data frame.

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

[dichotom](#)

Examples

```
## Burt table of variables in columns 1 to 5
## in the 'Music' example data set
data(Music)
burt(Music[,1:5])
```

---

catdesc   *Bivariate associations*

Description

Measures the association between a categorical variable and some continuous and/or categorical variables

Usage

```
catdesc(y,x,min.phi=NULL)
```

Arguments

y	the categorical variable to describe (must be a factor)
x	a data frame with continuous and/or categorical variables
min.phi	for the relationship between y and a categorical variable, only associations higher or equal to min.phi will be displayed. If NULL (default), they are all displayed.

Value

A list of the following items :

variables	associations between y and the variables in x, computed with BivariateAssoc function from package moreparty
bylevel	a list with one element for each level of y

Each element in bylevel has the following items :

categories	a data frame with categorical variables from x and associations measured by phi
continuous.var	a data frame with continuous variables from x and associations measured by correlation coefficients

Author(s)

Nicolas Robette

References

Rakotomalala R., 'Comprendre la taille d'effet (effect size)', [http://eric.univ-lyon2.fr/~ricco/cours/slides/effect_size.pdf]

See Also

[catdes](), [BivariateAssoc]()

Examples

```
data(Taste)
getindexcat(Taste[,1:5])
mca <- speMCA(Taste[,1:5],excl=c(3,6,9,12,15))
hc <- hclust(dist(mca$ind$coord))
clust <- as.factor(cutree(hc,5))
catdesc(clust, Taste[,c('Gender','Age')])
```

---

conc.ellipse	*Adds concentration ellipses to a correspondence analysis graph.*

Description

Adds concentration ellipses to the individuals factor map of a correspondence analysis.

Usage

```
conc.ellipse(resmca, var, sel = 1:length(levels(varb)), axes = c(1, 2),
 col = rainbow(length(sel)), pcol = rainbow(length(sel)), pcex = 0.2,
 lty = 1, lwd = 1, tcex = 1, text.lab = TRUE)
```

Arguments

resmca	object of class 'MCA', 'speMCA', 'csMCA', 'multiMCA' or 'stMCA'
var	supplementary variable to plot
sel	numeric vector of indexes of the categories to plot (by default, ellipses are plotted for every categories)
axes	length 2 vector specifying the components to plot (default is c(1,2))
col	vector of colors for the ellipses of plotted categories (by default, rainbow palette is used)
pcol	vector of colors for the points at the center of ellipses of plotted categories (by default, rainbow palette is used)
pcex	numerical value giving the amount by which points at the center of ellipses should be magnified (default is 0.2)
lty	line type for ellipses (default is 1)
lwd	line width for the ellipses (default is 1)
tcex	numerical value giving the amount by which labels at the center of ellipses should be magnified (default is 0.2)
text.lab	whether the labels at the center of ellipses should be displayed (default is TRUE)

Details

This function has to be used after the cloud of individuals has been drawn.

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

[plot.speMCA](), [plot.csMCA](), [plot.multiMCA](), [plot.stMCA]()

Examples

```
## Performs specific MCA (excluding 'NA' categories) of 'Taste' example data set,
## plots the cloud of categories
## and adds concentration ellipses for gender variable
data(Taste)
mca <- speMCA(Taste[,1:11],excl=c(3,6,9,12,15,18,21,24,27,30,33))
plot(mca,type='i')
conc.ellipse(mca,Taste$Gender)

## Draws a blue concentration ellipse for men only
plot(mca,type='i')
conc.ellipse(mca,Taste$Gender,sel=1,col='blue')
```

---

condesc   *Bivariate associations*

Description

Measures the association between a continuous variable and some continuous and/or categorical variables

Usage

```
condesc(y,x,min.cor=NULL)
```

Arguments

y	the continuous variable to describe
x	a data frame with continuous and/or categorical variables
min.cor	for the relationship between y and a categorical variable, only associations higher or equal to min.cor will be displayed. If NULL (default), they are all displayed.

Value

A list of the following items :

variables	associations between y and the variables in x, computed with BivariateAssoc function from package moreparty
categories	a data frame with categorical variables from x and associations measured by correlation coefficients

Author(s)

Nicolas Robette

References

Rakotomalala R., 'Comprendre la taille d'effet (effect size)', [http://eric.univ-lyon2.fr/~ricco/cours/slides/effect_size.pdf]

See Also

[condes](#), [BivariateAssoc](#)

Examples

```
data(Taste)
getindexcat(Taste[,1:5])
mca <- speMCA(Taste[,1:5],excl=c(3,6,9,12,15))
condesc(mca$ind$coord[,1], Taste[,c('Gender','Age')])
```

---

contrib	*Computes contributions for a correspondence analysis*

Description

From MCA results, computes contributions of categories and variables to the axes and the overall cloud.

Usage

```
contrib(resmca)
```

Arguments

resmca	object of class 'MCA', 'speMCA' or 'csMCA'

Details

The contribution of a point to an axis depends both on the distance from the point to the origin point along the axis and on the weight of the point. The contributions of points to axes are the main aid to interpretation (see Le Roux and Rouanet, 2004 and 2010).

Value

A list of data frames:

ctr	Data frame with the contributions of categories to axes
var.ctr	Data frame with the contributions of variables to axes
ctr.cloud	Data frame with the contributions of categories to the overall cloud
vctr.cloud	Data frame with the contributions of variables to the overall cloud

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

MCA, speMCA, csMCA, varsup

Examples

```
## Performs a specific MCA on the 'Music' example data set
## and compute contributions
data(Music)
mca <- speMCA(Music[,1:5],excl=c(3,6,9,12,15))
contrib(mca)
```

---

csMCA   *Performs a 'class specific' MCA*

Description

Performs a 'class specific' Multiple Correspondence Analysis, i.e. a variant of MCA consisting in analyzing a subcloud of individuals.

Usage

```
csMCA(data, subcloud = rep(TRUE, times = nrow(data)), excl = NULL, ncp = 5,
row.w = rep(1, times = nrow(data)))
```

Arguments

data	data frame with n rows (individuals) and p columns (categorical variables)
subcloud	a vector of logical values and length n. The subcloud of individuals analyzed with 'class specific' MCA is made of the individuals with value TRUE.
excl	numeric vector indicating the indexes of the 'junk' categories (default is NULL). See getindexcat to identify these indexes.
ncp	number of dimensions kept in the results (default is 5)
row.w	an optional numeric vector of row weights (by default, a vector of 1 for uniform row weights)

Details

This variant of MCA is used to study a subset of individuals with reference to the whole set of individuals, i.e. to determine the specific features of the subset. It consists in proceeding to the search of the principal axes of the subcloud associated with the subset of individuals (see Le Roux and Rouanet, 2004 and 2010).

Value

Returns an object of class 'csMCA', i.e. a list including:

eig	a list of vectors containing all the eigenvalues, the percentage of variance, the cumulative percentage of variance, the modified rates and the cumulative modified rates
call	a list with informations about input data
ind	a list of matrices containing the results for the individuals (coordinates, contributions)
var	a list of matrices containing all the results for the categories and variables (weights, coordinates, square cosine, categories contributions to axes and cloud, test values (v.test), square correlation ratio (eta2), variable contributions to axes and cloud

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

getindexcat, plot.csMCA, varsup, contrib, modif.rate, dimdesc.MCA, speMCA, MCA

Examples

```
## Performs a 'class specific' MCA on 'Music' example data set
## ignoring every 'NA' (i.e. 'not available') categories,
## and focusing on the subset of women.
data(Music)
female <- Music$Gender=='Women'
mca <- csMCA(Music[,1:5],subcloud=female,excl=c(3,6,9,12,15))
plot(mca)
```

---

dichotom   *Dichotomizes the variables in a data frame*

Description

Dichotomizes the variables in a data frame exclusively composed of categorical variables

Usage

```
dichotom(data, out = "numeric")
```

Arguments

data	data frame of categorical variables
out	character string defining the format for dichotomized variables in the output data frame. Format may be 'numeric' or 'factor' (default is 'numeric').

Value

Returns a data frame with dichotomized variables. The number of columns is equal to the total number of categories in the input data.

Author(s)

Nicolas Robette

Examples

```
## Dichotomizes 'Music' example data frame
data(Music)
dic <- dichotom(Music[,1:5])
str(dic)

## with output variables in factor format
dic <- dichotom(Music[,1:5], out='factor')
str(dic)
```

---

dimcontrib   *Describes the contributions to axes for MCA and variants of MCA*

Description

Identifies the categories and individuals that contribute the most to each dimension obtained by a Multiple Correspondence Analysis. It allows to analyze variants of MCA, such as 'specific' MCA or 'class specific' MCA.

Usage

```
dimcontrib(resmca, dim = c(1,2), best = TRUE)
```

Arguments

resmca	object of class `'MCA'`, `'speMCA'`, or `'csMCA'`
dim	dimensions to describe (default is c(1,2))
best	if FALSE, displays all the categories; if TRUE (default), displays only categories and individuals with contributions higher than average

Details

Contributions are sorted and assigned a positive or negative sign according to the corresponding categories or individuals' coordinates, so as to facilitate interpretation.

Value

Returns a list:

var	a list of categories' contributions to axes
ind	a list of individuals' contributions to axes

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

[dimdesc](), [dimdesc.MCA](), [dimeta2](), [condes](), [speMCA](), [csMCA]()

Examples

```
## Performs a specific MCA on 'Music' example data set
## ignoring every 'NA' (i.e. 'not available') categories,
## and then describes the contributions to axes.
data(Music)
getindexcat(Music[,1:5])
mca <- speMCA(Music[,1:5],excl=c(3,6,9,12,15))
dimcontrib(mca)
```

---

dimdesc.MCA   *Describes the dimensions of MCA and variants of MCA*

Description

Identifies the variables and the categories that are the most characteristic according to each dimension obtained by a Factor Analysis. It is inspired by dimdesc function in FactoMineR package (see Husson et al, 2010), but allows to analyze variants of MCA, such as 'specific' MCA or 'class specific' MCA.

Usage

```
dimdesc.MCA(resmca, ncp = 3, proba = 0.05)
```

Arguments

resmca	object of class 'MCA', 'speMCA', 'csMCA' or 'stMCA'
ncp	number of dimensions to describe (default is 3)
proba	the significance threshold considered to characterize the dimension (default is 0.05)

Details

The statistical indicator used for variables is square correlation ratio (R2) and the one used for categories is test-value (v.test).

Value

Returns a list of ncp lists including:

quali	the description of the dimensions by the categorical variables (the variables are sorted)
category	the description of the dimensions by each category of all the categorical variables (the categories are sorted)

Author(s)

Nicolas Robette

References

Husson, F., Le, S. and Pages, J. (2010). *Exploratory Multivariate Analysis by Example Using R*, Chapman and Hall.

See Also

[dimdesc](), [condes](), [speMCA](), [csMCA]()

Examples

```
## Performs a specific MCA on 'Music' example data set
## ignoring every 'NA' (i.e. 'not available') categories,
## and then describe the dimensions.
data(Music)
getindexcat(Music[,1:5])
mca <- speMCA(Music[,1:5],excl=c(3,6,9,12,15))
dimdesc.MCA(mca,proba=0.2)
```

---

dimeta2	*Describes the eta2 of a list of supplementary variables for the axes of MCA and variants of MCA*

Description

Computes eta2 for a list of supplementary variables. It allows to analyze variants of MCA, such as 'specific' MCA or 'class specific' MCA.

Usage

```
dimeta2(resmca, l, n, dim = 1:resmca$call$ncp)
```

Arguments

resmca	object of class 'MCA', 'speMCA', 'csMCA' or 'stMCA'
l	a list of supplementary variables
n	a vector of names for the supplementary variables. The vector's length must be equal to 'l''s length
dim	the axes for which eta2 are computed. Default is 1:ncp

Value

Returns a data frame with supplementary variables as rows and axes as columns.

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

[dimdesc](), [dimdesc.MCA](), [dimcontrib](), [condes](), [speMCA](), [csMCA]()

Examples

```
## Performs a specific MCA on 'Music' example data set
## ignoring every 'NA' (i.e. 'not available') categories,
## and then describes the eta2 for Gender and Age (axes 1 and 2).
data(Music)
getindexcat(Music[,1:5])
mca <- speMCA(Music[,1:5],excl=c(3,6,9,12,15))
dimeta2(mca,list(Music$Gender,Music$Age),c('Gender','Age'),dim=1:2)
```

---

dimvtest	*Describes the test-values of a list of supplementary variables for the axes of MCA and variants of MCA*

Description

Computes test-values for a list of supplementary variables. It allows to analyze variants of MCA, such as 'specific' MCA or 'class specific' MCA.

Usage

```
dimvtest(resmca, l, n, dim = 1:resmca$call$ncp)
```

Arguments

resmca	object of class 'MCA', 'speMCA', 'csMCA' or 'stMCA'
l	a list of supplementary variables
n	a vector of names for the supplementary variables. The vector's length must be equal to 'l''s length
dim	the axes for which eta2 are computed. Default is 1:ncp

Details

Test-values are ordered and only those higher than 2.58 (or lower than -2.58) are kept.

Value

Returns a list of data frames giving the test-values of the supplementary categories for the different axes.

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

dimdesc, dimdesc.MCA, dimeta2, dimcontrib, condes, speMCA, csMCA

Examples

```
## Performs a specific MCA on 'Music' example data set
## ignoring every 'NA' (i.e. 'not available') categories,
## and then describes the test-values for Gender and Age (axes 1 and 2).
data(Music)
getindexcat(Music[,1:5])
mca <- speMCA(Music[,1:5],excl=c(3,6,9,12,15))
dimvtest(mca,list(Music$Gender,Music$Age),c('Gender','Age'),dim=1:2)
```

| getindexcat | *Returns the names of the categories in a data frame* |

Description

Returns a vector of names corresponding the the categories in a data frame exclusively composed of categorical variables.

Usage

```
getindexcat(data)
```

Arguments

data   data frame of categorical variables

Details

This function may be useful prior to a 'specific' MCA, to identify the indexes of the 'junk' categories to exclude.

Value

Returns a character vector with the names of the categories of the variables in the data frame

Author(s)

Nicolas Robette

See Also

speMCA, csMCA

Examples

```
data(Music)
getindexcat(Music[,1:5])
mca <- speMCA(Music[,1:5],excl=c(3,6,9,12,15))
```

---

ggadd_ellipses   *Adds concentration ellipses to a MCA cloud of individuals*

Description

Adds concentration ellipses for a categorical variable to a MCA cloud of individuals, using the
ggplot2 framework.

Usage

```
ggadd_ellipses(p, resmca, var, sel=1:nlevels(var), axes=c(1,2),
label=TRUE, col=NULL, legend='right', level=0.86, alpha=0.5)
```

Arguments

p	ggplot object with the cloud of variables
resmca	object of class 'MCA', 'speMCA' or 'csMCA'.
var	Factor. The categorical variable used to plot ellipses.
sel	numeric vector of indexes of the categories to plot (by default, ellipses are plotted for every categories)
axes	numeric vector of length 2, specifying the components (axes) to plot. Default is c(1,2).
label	Logical. Should the labels of teh categories be plotted at the center of ellipses ? Default is TRUE.
col	Colors for the ellipses and labels of the categories. Can be the name of a palette from the RcolorBrewer package, 'bw' for a black and white palette (uses scale_color_grey()), a character vector of colors for a custom palette, or the name of a color for a single color. If NULL (default), the default palette of ggplot2 is used.

legend	the position of legends ("none", "left", "right", "bottom", "top", or two-element numeric vector). Default is right.
level	The level at which to draw an ellipse (see stat_ellipse). Default is 0.86.
alpha	Transparency of the points of individuals. Default is 0.5.

Value

a ggplot object

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

[ggcloud_variables](), [ggcloud_indiv](), [ggadd_ellipses](), [ggadd_interaction]()

Examples

```
## Performs a specific MCA on 'Music' example data set
## ignoring every 'NA' (i.e. 'not available') categories,
## draws the cloud of categories
## and adds concentration ellipses for Age.
data(Music)
getindexcat(Music[,1:5])
mca <- speMCA(Music[,1:5],excl=c(3,6,9,12,15))
p <- ggcloud_indiv(mca, col='lightgrey')
ggadd_supvar(p, mca, Music$Age)
```

---

ggadd_interaction	*Adds the interaction between two categorical supplementary variables to a MCA cloud of variables*

Description

Adds the interaction between two categorical supplementary variables to a MCA cloud of variables, using the ggplot2 framework.

Usage

```
ggadd_interaction(p, resmca, v1, v2, sel1=1:nlevels(v1), sel2=1:nlevels(v2),
axes=c(1,2), col=NULL, textsize=5, legend='right')
```

Arguments

p	ggplot object with the cloud of variables
resmca	object of class `'MCA'`, `'speMCA'` or `'csMCA'`.
v1	Factor. The first categorical supplementary variable.
v2	Factor. The second categorical supplementary variable.
sel1	Numeric vector of indexes of the categories of the first supplementary variable to be used in interaction. By default, every categories are used.
sel2	Numeric vector of indexes of the categories of the second supplementary variable to be used in interaction. By default, every categories are used.
axes	numeric vector of length 2, specifying the components (axes) to plot. Default is c(1,2).
col	Colors for the ellipses and labels of the categories. Can be the name of a palette from the RcolorBrewer package, 'bw' for a black and white palette (uses `scale_color_grey()`), a character vector of colors for a custom palette, or the name of a color for a single color. If NULL (default), the default palette of ggplot2 is used.
textsize	Size of the labels of categories. Default is 5.
legend	the position of legends ("none", "left", "right", "bottom", "top", or two-element numeric vector). Default is right.

Value

a ggplot object

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

[ggcloud_variables](#), [ggadd_supvar](#), [ggcloud_indiv](#), [ggadd_ellipses](#)

Examples

```
## Performs a specific MCA on 'Music' example data set
## ignoring every 'NA' (i.e. 'not available') categories,
## draws the cloud of categories
## and adds the interaction between Gender and Age.
data(Music)
```

```
getindexcat(Music[,1:5])
mca <- speMCA(Music[,1:5],excl=c(3,6,9,12,15))
p <- ggcloud_variables(mca, palette='lightgrey', shapes=FALSE)
ggadd_interaction(p, mca, Music$Gender, Music$Age)
```

---

ggadd_supvar	*Adds a categorical supplementary variable to a MCA cloud of variables*

Description

Adds a categorical supplementary variable to a MCA cloud of variables, using the ggplot2 framework.

Usage

```
ggadd_supvar(p, resmca, var, sel=1:nlevels(var), axes=c(1,2), col='black',
shape=1, prop=NULL, textsize=3, segment=TRUE, vname=NULL)
```

Arguments

p	ggplot object with the cloud of variables
resmca	object of class `'MCA'`, `'speMCA'` or `'csMCA'`.
var	Factor. The categorical supplementary variable. It does not need to have been used at the MCA step.
sel	Numeric vector of indexes of the categories of the supplementary variable to be added to the plot. By default, labels are plotted for every categories.
axes	numeric vector of length 2, specifying the components (axes) to plot. Default is c(1,2).
col	Character. Color of the shapes and labels of the categories. Default is black.
shape	Symbol to be used in addition the the labels of categories (default is 1). If NULL, only labels are plotted.
prop	If NULL, the size of the labels (if shape=NULL) or the shapes (otherwise) is constant. If 'n', the size is proportional the the weights of categories; if 'vtest1', the size is proportional to the test values of the categories on the first dimension of the plot; if 'vtest2', the size is proportional to the test values of the categories on the second dimension of the plot; if 'cos1', the size is proportional to the cosines of the categories on the first dimension of the plot; if 'cos2', the size is proportional to the cosines of the categories on the second dimension of the plot; if 'cos12', the size is proportional to the total cosines of the categories on the two dimensions of the plot.
textsize	Size of the labels of categories if shape is not NULL, or if shape=NULL and prop=NULL. Default is 3.
segment	Logical. Should one add lines between categories ? Default is TRUE.
vname	A character string to be used as a prefix for the labels of the categories. If NULL (default), no prefix is added.

Value

a ggplot object

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

[ggcloud_variables](), [ggcloud_indiv](), [ggadd_ellipses](), [ggadd_interaction]()

Examples

```
## Performs a specific MCA on 'Music' example data set
## ignoring every 'NA' (i.e. 'not available') categories,
## draws the cloud of categories
## and adds Age as a supplementary variable.
data(Music)
getindexcat(Music[,1:5])
mca <- speMCA(Music[,1:5],excl=c(3,6,9,12,15))
p <- ggcloud_variables(mca, palette='lightgrey', shapes=FALSE)
ggadd_supvar(p, mca, Music$Age)
```

---

ggcloud_indiv   *Plots MCA cloud of individuals*

Description

Plots a Multiple Correspondence Analysis cloud of individuals, using ggplots functions.

Usage

```
ggcloud_indiv(resmca, type='i', points='all', axes=1:2,
col='dodgerblue4', palette='Set2', alpha=0.6, repel=FALSE)
```

Arguments

resmca	object of class `'MCA'`, `'speMCA'` or `'csMCA'`.
type	If 'i', points are plotted; if 'inames', labels of individuals are plotted.
points	character string. If 'all' all points are plotted (default); if 'besth' only those who contribute most to horizontal axis are plotted; if 'bestv' only those who contribute most to vertical axis are plotted; if 'best' only those who contribute most to horizontal or vertical axis are plotted.
axes	numeric vector of length 2, specifying the components (axes) to plot. Default is c(1,2).
col	If a factor, points or labels are colored according to their category regarding this factor. If a string with color name, every points or labels have the same color. Default is 'dodgerblue4'.
palette	If `col` is a factor, can be the name of a RColorBrewer palette, 'bw' for a black and white palette (uses `scale_color_grey()`) or a character vector with a custom palette. If NULL, ggplot2 palette is used. Default is 'Set2'.
alpha	Transparency of the points or labels of individuals. Default is 0.6.
repel	Logical. When type='inames', should labels of individuals be repeled ? Default is FALSE.

Value

a ggplot object

Author(s)

Anton Perdoncin, Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

[ggcloud_variables](#), [ggadd_ellipses](#), [ggadd_supvar](#), [ggadd_interaction](#)

Examples

```
## Performs a specific MCA on 'Music' example data set
## ignoring every 'NA' (i.e. 'not available') categories,
## and then draws the cloud of individuals.
data(Music)
getindexcat(Music[,1:5])
mca <- speMCA(Music[,1:5],excl=c(3,6,9,12,15))
ggcloud_indiv(mca)
ggcloud_indiv(mca, col=Music$Gender, palette='Set2')
```

---

ggcloud_variables   *Plots MCA cloud of variables*

Description

Plots a Multiple Correspondence Analysis cloud of variables, using ggplots functions.

Usage

```
ggcloud_variables(resmca, axes=c(1,2), points='all', shapes=TRUE,
prop=NULL, textsize=3, shapesize=3, palette=NULL, alpha=1,
segment.alpha=0.5, vlab=TRUE, sep='.', legend='right')
```

Arguments

resmca	object of class `'MCA'`, `'speMCA'` or `'csMCA'`.
axes	numeric vector of length 2, specifying the components (axes) to plot. Default is c(1,2).
points	character string. If 'all' all categories are plotted (default); if 'besth' only those who contribute most to horizontal axis are plotted; if 'bestv' only those who contribute most to vertical axis are plotted; if 'best' only those who contribute most to horizontal or vertical axis are plotted.
shapes	Logical. Should shapes be plotted for categories (in addition to labels) ? Default is TRUE.
prop	If NULL, the size of the labels (if shapes=FALSE) or the shapes (if shapes=TRUE) is constant. If 'n', the size is proportional the the weights of categories; if 'ctr1', the size is proportional to the contributions of the categories on the first dimension of the plot; if 'ctr2', the size is proportional to the contributions of the categories on the second dimension of the plot; if 'ctr.cloud', the size is proportional to the total contributions of the categories on the whole cloud; if 'cos1', the size is proportional to the cosines of the categories on the first dimension of the plot; if 'cos2', the size is proportional to the cosines of the categories on the second dimension of the plot; if 'cos12', the size is proportional to the total cosines of the categories on the two dimensions of the plot.
textsize	Size of the labels of categories if shapes=TRUE, or if shapes=FALSE and prop=NULL. Default is 3.
shapesize	Size if the shapes of categories if shapes=TRUE and prop=FALSE. Default is 3.
palette	Colors for the shapes and labels of the categories, differentiated by variable. Can be the name of a palette from the RcolorBrewer package, 'bw' for a black and white palette (uses `scale_color_grey()`), a character vector of colors for a custom palette, or the name of a color for a single color. If NULL (default), the default palette of `ggplot2` is used..
alpha	Transparency of the shapes and labels of categories. Default is 1.
segment.alpha	Transparency of the line segment beside labels of categories. Default is 0.5.

vlab	Logical. Should the variable names be used as a prefix for the labels of the categories. Default is TRUE.
sep	Character string used as a separator if vlab=TRUE.
legend	the position of legends ("none", "left", "right", "bottom", "top", or two-element numeric vector). Default is right.

Value

a ggplot object

Author(s)

Anton Perdoncin, Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

[ggcloud_indiv](), [ggadd_supvar](), [ggadd_ellipses](), [ggadd_interaction]()

Examples

```
## Performs a specific MCA on 'Music' example data set
## ignoring every 'NA' (i.e. 'not available') categories,
## and then draws the cloud of categories.
data(Music)
getindexcat(Music[,1:5])
mca <- speMCA(Music[,1:5],excl=c(3,6,9,12,15))
ggcloud_variables(mca)
ggcloud_variables(mca, points='best', prop='n', palette='Set2')
```

---

homog.test	*Computes a homogeneity test for a categorical supplementary variable*

Description

From MCA results, computes a homogeneity test for a categorical supplementary variable, i.e. characterizes the homogeneity of several subclouds.

Usage

```
homog.test(resmca, var)
```

Arguments

resmca	object of class 'MCA', 'speMCA', 'csMCA', 'stMCA' or 'multiMCA'
var	the categorical supplementary variable. It does not need to have been used at the MCA step.

Value

Returns a list of square matrices, one per MCA dimension. Each matrix gives the test statistic for any pair of categories.

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

[speMCA](#), [csMCA](#), [stMCA](#), [multiMCA](#), [textvarsup](#)

Examples

```
## Performs a specific MCA on 'Music' example data set
## ignoring every 'NA' (i.e. 'not available') categories,
## and then computes a homogeneity test for age supplementary variable.
data(Music)
getindexcat(Music)
mca <- speMCA(Music[,1:5],excl=c(3,6,9,12,15))
homog.test(mca,Music$Age)
```

---

indsup	*Computes statistics for supplementary individuals*

Description

From MCA results, computes statistics (coordinates, squared cosines) for supplementary individuals.

Usage

```
indsup(resmca, supdata)
```

Arguments

resmca	object of class `'MCA'`, `'speMCA'` or `'csMCA'`
supdata	data frame with the supplementary individuals. It must have the same factors as the data frame used as input for the initial MCA.

Value

Returns a list:

coord	matrix of individuals' coordinates
cos2	matrix of individuals' square cosines

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

textindsup, speMCA, csMCA, varsup

Examples

```
## Performs a specific MCA on 'Music' example data set
## ignoring every 'NA' (i.e. 'not available') categories,
## and then computes statistics for supplementary individuals.
data(Music)
getindexcat(Music)
mca <- speMCA(Music[3:nrow(Music),1:5],excl=c(3,6,9,12,15))
indsup(mca,Music[1:2,1:5])
```

---

medoids	*Computes the medoids of clusters*

Description

Computes the medoids of a cluster solution.

Usage

```
medoids(D, cl)
```

Arguments

D	square distance matrix (n rows * n columns, i.e. n individuals) or `dist` object
cl	vector with the clustering solution (its length should be n)

Details

Medoids are representative objects of a cluster whose average dissimilarity to all the objects in the cluster is minimal. Medoids are always members of the data set (contrary to means or centroids).

Value

Returns a numeric vector with the indexes of medoids.

Author(s)

Nicolas Robette

References

Kaufman, L. and Rousseeuw, P.J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley, New York.

Anja Struyf, Mia Hubert & Peter J. Rousseeuw (1996). "Clustering in an Object-Oriented Environment". *Journal of Statistical Software*.

See Also

dist, cluster, hclust, cutree , [pam](pam)

Examples

```
## Performs of classification of the 'Music' example data set,
## a clustering in 3 groups
## and then computes the medoids.
data(Music)
temp <- dichotom(Music[,1:5])
d <- dist(temp)
clus <- cutree(hclust(d),3)
medoids(d,clus)
```

---

modif.rate	*Computes the modified rates of variance of a correspondence analysis*

Description

Computes the modified rates of variance of a correspondence analysis.

Usage

```
modif.rate(resmca)
```

Arguments

resmca   object of class `'MCA'`, `'speMCA'` or `'csMCA'`

Details

As MCA clouds often have a high dimensionality, the variance rates of the first principle axes may be quite low, which makes them hard to interpret. Benzecri (1992, p.412) proposed to use *modified rates* to better appreciate the relative importance of the principal axes.

Value

Returns a data frame with 2 variables:

mrate   Numeric vector of modified rates

cum.mrate   Numeric vector of cumulative modified rates

Author(s)

Nicolas Robette

References

Benzecri J.P., *Correspondence analysis handbook*, New-York: Dekker (1992).

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

MCA, speMCA, csMCA

Examples

```
## Computes the modified rates of variance
## of the MCA of 'Music' example data set
data(Music)
mca <- speMCA(Music[,1:5])
modif.rate(mca)
```

---

multiMCA   *Performs Multiple Factor Analysis*

Description

Performs Multiple Factor Analysis, drawing on the work of Escoffier and Pages (1994). It allows the use of MCA variants (e.g. specific MCA or class specific MCA) as inputs.

Usage

```
multiMCA(l_mca, ncp = 5, compute.rv = FALSE)
```

Arguments

l_mca	a list of objects of class MCA, speMCA or csMCA
ncp	number of dimensions kept in the results (default is 5)
compute.rv	whether RV coefficients should be computed or not (default is FALSE, which makes the function execute faster)

Details

This function binds individual coordinates from every MCA in l_mca argument, weights them by the first eigenvalue, and the resulting data frame is used as input for Principal Component Analysis (PCA).

Value

Returns an object of class 'multiMCA', i.e. a list:

eig	a list of numeric vector for eigenvalues, percentage of variance and cumulative percentage of variance
var	a list of matrices with results for input MCAs components (coordinates, correlation between variables and axes, square cosine, contributions)
ind	a list of matrices with results for individuals (coordinates, square cosine, contributions)
call	a list with informations about input data
VAR	a list of matrices with results for categories and variables in the input MCAs (coordinates, square cosine, test-values, variances)
my.mca	lists the content of the objects in l_mca argument
RV	a matrix of RV coefficients

Author(s)

Nicolas Robette

References

Escofier, B. and Pages, J. (1994) "Multiple Factor Analysis (AFMULT package)". *Computational Statistics and Data Analysis*, 18, 121-140.

See Also

`plot.multiMCA`, `varsup`, `speMCA`, `csMCA`, `MFA`, `PCA`

Examples

```
## Performs a specific MCA on music variables of 'Taste' example data set,
## another one on movie variables of 'Taste' example data set,
## and then a Multiple Factor Analysis.
data(Taste)
getindexcat(Taste[,1:5])
mca1 <- speMCA(Taste[,1:5],excl=c(3,6,9,12,15))
getindexcat(Taste[,6:11])
mca2 <- speMCA(Taste[,6:11],excl=c(3,6,9,12,15,18))
mfa <- multiMCA(list(mca1,mca2))
plot.multiMCA(mfa)
```

Music	*Music (data)*

Description

The data concerns tastes for music of a set of 500 individuals. It contains 5 variables of likes for music genres (french pop, rap, rock, jazz and classical), 2 about music listening and 2 additional variables (gender and age).

Usage

```
data(Music)
```

Format

A data frame with 500 observations and the following 7 variables:

FrenchPop is a factor with levels No, Yes, NA

Rap is a factor with levels No, Yes, NA

Rock is a factor with levels No, Yes, NA

Jazz is a factor with levels No, Yes, NA

Classical is a factor with levels No, Yes, NA

Gender is a factor with levels Men, Women

Age is a factor with levels 15-24, 25-49, 50+

OnlyMus is a factor with levels Daily, Often, Rare, Never, indicating how often one only listens to music.

Daily is a factor with levels No, Yes indicating if one listens to music every day.

Details

'NA' stands for 'not available'

Examples

```
data(Music)
str(Music)
```

---

pem   *Computes the local and global Percentages of Maximum Deviation*
                             from Independance (PEM)

Description

Computes the local and global Percentages of Maximum Deviation from Independance (PEM) of a contingency table.

Usage

```
pem(x)
```

Arguments

x   Contingency table. Accepted formats are matrices and 'table' objects.

Details

The Percentage of Maximum Deviation from Independence (PEM) is an association measure for contingency tables (see Cibois, 1993). It is an alternative to khi2, Cramer coefficient, etc.

Value

Returns a list:

peml   Matrix with local percentages of maximum deviation from independance

pemg   Numeric value, i.e. the global percentage of maximum deviation from independance

Author(s)

Nicolas Robette

References

Cibois P., 1993, Le PEM, pourcentage de l'ecart maximum : un indice de liaison entre modalites d'un tableau de contingence, *Bulletin de methodologie sociologique*, n40, p.43-63. http://cibois.pagesperso-orange.fr/bms93.pdf

See Also

table, chisq.test, assocstats

Examples

```
## Computes the PEM for the contingency table
## of jazz and age variables
## from the 'Music' example data set
data(Music)
x <- table(Music$Jazz,Music$Age)
pem(x)
```

---

plot.csMCA	*Plots 'class specific' MCA results*

Description

Plots a 'class specific' Multiple Correspondence Analysis (resulting from csMCA function), i.e. the clouds of individuals or categories.

Usage

```
## S3 method for class 'csMCA'
plot(x, type = "v", axes = 1:2, points = "all", col = "dodgerblue4", app = 0, ...)
```

Arguments

x	object of class 'csMCA'
type	character string: 'v' to plot the categories (default), 'i' to plot individuals' points, 'inames' to plot individuals' names
axes	numeric vector of length 2, specifying the components (axes) to plot (c(1,2) is default)
points	character string. If 'all' all points are plotted (default); if 'besth' only those who contribute most to horizontal axis are plotted; if 'bestv' only those who contribute most to vertical axis are plotted; if 'best' only those who contribute most to horizontal or vertical axis are plotted.
col	color for the points of the individuals or for the labels of the categories (default is 'dodgerblue4')
app	numerical value. If 0 (default), only the labels of the categories are plotted and their size is constant; if 1, only the labels are plotted and their size is proportional to the weights of the categories; if 2, points (triangles) and labels are plotted, and points size is proportional to the weight of the categories.
...	further arguments passed to or from other methods, such as cex, cex.main, ...

Details

A category is considered to be one of the most contributing to a given axis if its contribution is
higher than the average contribution, i.e. 100 divided by the total number of categories.

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Ap-
plications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured
Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

csMCA, textvarsup, conc.ellipse

Examples

```
## Performs a class specific MCA on 'Music' example data set
## ignoring every 'NA' (i.e. 'not available') categories
## and focusing on the subset of women,
## and then draws the cloud of categories.
data(Music)
female <- Music$Gender=='Women'
getindexcat(Music[,1:5])
mca <- csMCA(Music[,1:5],subcloud=female,excl=c(3,6,9,12,15))
plot(mca)
plot(mca,axes=c(2,3),points='best',col='darkred',app=1)
```

---

  plot.multiMCA   *Plots Multiple Factor Analysis*

Description

Plots Multiple Factor Analysis data, resulting from multiMCA function.

Usage

```
## S3 method for class 'multiMCA'
plot(x, type = "v", axes = c(1, 2), points = "all", threshold = 2.58,
groups = 1:x$call$ngroups, col = rainbow(x$call$ngroups), app = 0, ...)
```

Arguments

x	object of class `'multiMCA'`
type	character string: 'v' to plot the categories (default), 'i' to plot individuals' points, 'inames' to plot individuals' names
axes	numeric vector of length 2, specifying the components (axes) to plot (c(1,2) is default)
points	character string. If 'all' all points are plotted (default); if 'besth' only those who are the most correlated to horizontal axis are plotted; if 'bestv' only those who are the most correlated to vertical axis are plotted; if 'best' only those who are the most coorelated to horizontal or vertical axis are plotted.
threshold	numeric value. V-test minimal value for the selection of plotted categories.
groups	numeric vector specifying the groups of categories to plot. By default, every groups of categories will be plotted
col	a color for the points of the individuals or a vector of colors for the labels of the groups of categories (by default, rainbow palette is used)
app	numerical value. If 0 (default), only the labels of the categories are plotted and their size is constant; if 1, only the labels are plotted and their size is proportional to the weights of the categories; if 2, points (triangles) and labels are plotted, and points size is proportional to the weight of the categories.
...	further arguments passed to or from other methods, such as cex, cex.main, ...

Details

A category is considered to be one of the most correlated to a given axis if its test-value is higher then 2.58 (which corresponds to a 0.05 threshold).

Author(s)

Nicolas Robette

References

Escofier, B. and Pages, J. (1994) "Multiple Factor Analysis (AFMULT package)". *Computational Statistics and Data Analysis*, 18, 121-140.

See Also

[multiMCA](), [textvarsup](), [speMCA](), [csMCA](), [MFA]()

Examples

```
## Performs a specific MCA on music variables of 'Taste' example data set,
## another one on movie variables of 'Taste' example data set,
## and then a Multiple Factor Analysis and plots the results.
data(Taste)
mca1 <- speMCA(Taste[,1:5],excl=c(3,6,9,12,15))
mca2 <- speMCA(Taste[,6:11],excl=c(3,6,9,12,15,18))
```

```
mfa <- multiMCA(list(mca1,mca2))
plot.multiMCA(mfa,col=c('darkred','darkblue'))
plot.multiMCA(mfa,groups=2,app=1)
```

---

plot.speMCA   *Plots 'specific' MCA results*

Description

Plots a 'specific' Multiple Correspondence Analysis (resulting from speMCA function), i.e. the clouds of individuals or categories.

Usage

```
## S3 method for class 'speMCA'
plot(x, type = "v", axes = 1:2, points = "all", col = "dodgerblue4", app = 0, ...)
```

Arguments

x   object of class 'speMCA'

type   character string: 'v' to plot the categories (default), 'i' to plot individuals' points, 'inames' to plot individuals' names

axes   numeric vector of length 2, specifying the components (axes) to plot (c(1,2) is default)

points   character string. If 'all' all points are plotted (default); if 'besth' only those who contribute most to horizontal axis are plotted; if 'bestv' only those who contribute most to vertical axis are plotted; if 'best' only those who contribute most to horizontal or vertical axis are plotted.

col   color for the points of the individuals or for the labels of the categories (default is 'dodgerblue4')

app   numerical value. If 0 (default), only the labels of the categories are plotted and their size is constant; if 1, only the labels are plotted and their size is proportional to the weights of the categories; if 2, points (triangles) and labels are plotted, and points size is proportional to the weight of the categories.

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

Details

A category is considered to be one of the most contributing to a given axis if its contribution is higher than the average contribution, i.e. 100 divided by the total number of categories.

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

speMCA, textvarsup, conc.ellipse

Examples

```
## Performs a specific MCA on 'Music' example data set
## ignoring every 'NA' (i.e. 'not available') categories,
## and then draws the cloud of categories.
data(Music)
getindexcat(Music[,1:5])
mca <- speMCA(Music[,1:5],excl=c(3,6,9,12,15))
plot(mca)
plot(mca,axes=c(2,3),points='best',col='darkred',app=1)
```

---

plot.stMCA   *Plots 'standardized' MCA results*

Description

Plots a 'standardized' Multiple Correspondence Analysis (resulting from stMCA function), i.e. the clouds of individuals or categories.

Usage

```
## S3 method for class 'stMCA'
plot(x, type = "v", axes = 1:2, points = "all", threshold = 2.58, groups=NULL,
                       col = "dodgerblue4", app = 0, ...)
```

Arguments

x	object of class 'stMCA'
type	character string: 'v' to plot the categories (default), 'i' to plot individuals' points, 'inames' to plot individuals' names
axes	numeric vector of length 2, specifying the components (axes) to plot (c(1,2) is default)
points	character string. If 'all' all points are plotted (default); if 'besth' only those who are the most correlated to horizontal axis are plotted; if 'bestv' only those who are the most correlated to vertical axis are plotted; if 'best' only those who are the most coorelated to horizontal or vertical axis are plotted.

threshold   numeric value. V-test minimal value for the selection of plotted categories.

groups   only if x$call$input.mca = 'multiMCA', i.e. if the MCA standardized to x object
                     was a `multiMCA` object. Numeric vector specifying the groups of categories to
                     plot. By default, every groups of categories will be plotted

col   color for the points of the individuals or for the labels of the categories (default
                     is 'dodgerblue4')

app   numerical value. If 0 (default), only the labels of the categories are plotted and
                     their size is constant; if 1, only the labels are plotted and their size is proportional
                     to the weights of the categories; if 2, points (triangles) and labels are plotted, and
                     points size is proportional to the weight of the categories.

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

Details

A category is considered to be one of the most correlated to a given axis if its test-value is higher
then 2.58 (which corresponds to a 0.05 threshold).

Author(s)

Nicolas Robette

References

Robette, Bry and Roueff, 2014, "Un dialogue de sourds dans le theatre statistique? Analyse ge-
ometrique des donnees et effets de structure", *forthcoming*

See Also

stMCA, textvarsup, conc.ellipse

Examples

```
## Performs a standardized MCA on 'Music' example data set
## ignoring every 'NA' (i.e. 'not available') categories
## and controlling for age,
## and then draws the cloud of categories.
data(Music)
mca <- speMCA(Music[,1:5],excl=c(3,6,9,12,15))
stmca <- stMCA(mca,control=list(Music$Age))
plot(stmca)
plot(stmca,axes=c(2,3),points='best',col='darkred',app=1)
```

---

prop.wtable | *Transforms a (possibly weighted) contingency table into percentages*

Description

Computes a contingency table from one or two vectors, with the possibility of specifying weights, and then computes the percentages.

Usage

```
prop.wtable(var1,var2=NULL,w=rep.int(1,length(var1)),dir=0,digits=1,mar=TRUE,na=TRUE)
```

Arguments

var1	an object which can be interpreted as factor
var2	an optional object which can be interpreted as factor
w	an optional numeric vector of weights (by default, a vector of 1 for uniform weights)
dir	integer: 0 for percentages, 1 for row percentages and 2 for column percentages (default is 0)
digits	integer indicating the number of decimal places (default is 1)
mar	logical. If TRUE (default), margins are computed
na	logical. If TRUE (default), 'NA' are treated as a category. If FALSE, they are ignored

Value

Returns a contingency table expressed in percentages in matrix format.

Author(s)

Nicolas Robette

See Also

[wtable](#), [prop.table](#)

Examples

```
## Computes a contingency table
## (expressed in percentages)
## of jazz and age variables
## from the 'Music' example data set
## with or without weights
data(Music)
prop.wtable(Music$Jazz)
prop.wtable(Music$Jazz,Music$Gender)
```

```
prop.wtable(Music$Jazz,Music$Gender,dir=1)
prop.wtable(Music$Jazz,Music$Gender,dir=2)

weight <- rep(c(0,0.5,1,1.5,2), length.out=nrow(Music))
prop.wtable(Music$Jazz,w=weight)
prop.wtable(Music$Jazz,Music$Age,weight)
prop.wtable(Music$Jazz,Music$Age,weight,dir=1)
prop.wtable(Music$Jazz,Music$Age,weight,dir=2)
```

---

speMCA   *Performs a 'specific' MCA*

Description

Performs a 'specific' Multiple Correspondence Analysis, i.e. a variant of MCA that allows to treat undesirable categories as passive categories.

Usage

```
speMCA(data, excl = NULL, ncp = 5, row.w = rep(1, times = nrow(data)))
```

Arguments

data	data frame with n rows (individuals) and p columns (categorical variables)
excl	numeric vector indicating the indexes of the "junk" categories (default is NULL). See "getindexcat" to identify these indexes.
ncp	number of dimensions kept in the results (default is 5)
row.w	an optional numeric vector of row weights (by default, a vector of 1 for uniform row weights)

Details

Undesirable categories may be of several kinds: infrequent categories (say, <5 percents), heterogeneous categories (e.g. 'others') or uninterpretable categories (e.g. 'not available'). In these cases, 'specific' MCA may be useful to ignore these categories for the determination of distances between individuals (see Le Roux and Rouanet, 2004 and 2010).

Value

Returns an object of class 'speMCA', i.e. a list including:

eig	a list of vectors containing all the eigenvalues, the percentage of variance, the cumulative percentage of variance, the modified rates and the cumulative modified rates
call	a list with informations about input data
ind	a list of matrices containing the results for the individuals (coordinates, contributions)

var a list of matrices containing all the results for the categories and variables (weights, coordinates, square cosine, categories contributions to axes and cloud, test values (v.test), square correlation ratio (eta2), variable contributions to axes and cloud

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

getindexcat, plot.speMCA, varsup, contrib, modif.rate, dimdesc.MCA, MCA, csMCA

Examples

```
## Performs a specific MCA on 'Music' example data set
## ignoring every 'NA' (i.e. 'not available') categories.
data(Music)
getindexcat(Music[,1:5])
mca <- speMCA(Music[,1:5],excl=c(3,6,9,12,15))
str(mca)
```

---

stMCA *Performs a 'standardized' MCA*

Description

Performs a 'standardized' Multiple Correspondence Analysis, i.e it takes MCA results and forces all the dimensions to be orthogonal to a supplementary 'control' variable.

Usage

```
stMCA(resmca, control)
```

Arguments

resmca an object of class 'MCA', 'speMCA', 'csMCA' or 'multiMCA'

control a list of 'control' variables

Details

Standardized MCA unfolds in several steps. First, for each dimension of an input MCA, individual coordinates are used as dependent variable in a linear regression model and the 'control' variable is included as covariate in the same model. The residuals from every models are retained and bound together. The resulting data frame is composed of continuous variables and its number of columns is equal to the number of dimensions in the input MCA. Lastly, this data frame is used as input in a Principal Component Analysis.

Value

Returns an object of class "stMCA". This object will be similar to resmca argument, still it does not comprehend modified rates, categories contributions and variables contributions.

Author(s)

Nicolas Robette

References

Robette, Bry and Roueff, 2014, "Un dialogue de sourds dans le theatre statistique? Analyse geometrique des donnees et effets de structure", [http://nicolas.robette.free.fr/publis.html], *forthcoming*.

See Also

[plot.stMCA](), [MCA](), [speMCA](), [csMCA](), [multiMCA](), [PCA]()

Examples

```
## Performs a specific MCA on 'Music' example data set
## ignoring every 'NA' (i.e. 'not available') categories,
## and then performs a 'standardized' MCA controlling for age.
data(Music)
mca <- speMCA(Music[,1:5],excl=c(3,6,9,12,15))
plot(mca)
textvarsup(mca,Music$Age,col='red')
stmca <- stMCA(mca,control=list(Music$Age))
plot(stmca)
textvarsup(stmca,Music$Age,col='red')
```

---

tabcontrib	*Displays the categories contributing most to axes for MCA and variants of MCA*

Description

Identifies the categories that contribute the most to a given dimension of a Multiple Correspondence Analysis and organizes them into a fancy table. It allows to analyze variants of MCA, such as 'specific' MCA or 'class specific' MCA.

Usage

```
tabcontrib(resmca, dim = 1)
```

Arguments

resmca	object of class `'MCA'`, `'speMCA'`, or `'csMCA'`
dim	dimension to describe (default is 1st dimension)

Details

Best contributions - i.e. higher than average - are assigned a positive or negative sign according to the corresponding categories' coordinates, so as to facilitate interpretation. Then they are sorted and organized according to the most contributing variables.

Value

Returns a data frame with the following columns:

var	the names of the most contributing variables
moda	the names of the most contributing cateories
ctr1	'negative' contributions, i.e. corresponding to categories with coordinates lower than zero
ctr2	'positive' contributions, i.e. corresponding to categories with coordinates higher than zero
weight	weight of the categories
ctrtot	sum of the best contributions for a given variable
cumctrtot	cumulated contributions

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

[dimcontrib](), [dimdesc](), [dimdesc.MCA](), [dimeta2](), [condes](), [speMCA](), [csMCA]()

Examples

```
## Performs a specific MCA on 'Music' example data set
## ignoring every 'NA' (i.e. 'not available') categories,
## and then describes the contributions to axes.
data(Music)
getindexcat(Music[,1:5])
mca <- speMCA(Music[,1:5],excl=c(3,6,9,12,15))
tabcontrib(mca,1)
tabcontrib(mca,2)
```

Taste   *Taste (data)*

Description

The data concerns tastes for music and movies of a set of 500 individuals. It contains 5 variables of likes for music genres (french pop, rap, rock, jazz and classical), 6 variables of likes for movie genres (comedy, crime, animation, science fiction, love, musical) and 2 additional variables (gender and age).

Usage

```
data(Taste)
```

Format

A data frame with 500 observations and the following 13 variables:

FrenchPop is a factor with levels No, Yes, NA

Rap is a factor with levels No, Yes, NA

Rock is a factor with levels No, Yes, NA

Jazz is a factor with levels No, Yes, NA

Classical is a factor with levels No, Yes, NA

Comedy is a factor with levels No, Yes, NA

Crime is a factor with levels No, Yes, NA

Animation is a factor with levels No, Yes, NA

SciFi is a factor with levels No, Yes, NA

Love is a factor with levels No, Yes, NA

Musical is a factor with levels No, Yes, NA

Gender is a factor with levels Men, Women

Age is a factor with levels 15-24, 25-49, 50+

Details

'NA' stands for 'not available'

Examples

```
data(Taste)
str(Taste)
```

| textindsup | *Adds supplementary individuals to a MCA graph* |

Description

Adds supplementary individuals to a MCA graph of the cloud of the individuals.

Usage

```
textindsup(resmca, supdata, axes = c(1, 2), col = "darkred")
```

Arguments

resmca	object of class `'MCA'`, `'speMCA'`, or `'csMCA'`
supdata	data frame with the supplementary individuals. It must have the same factors as the data frame used as input for the initial MCA.
axes	numeric vector of length 2, specifying the dimensions (axes) to plot (default is c(1,2))
col	color for the labels of the categories (default is 'darkred')

Author(s)

Nicolas Robette

See Also

[indsup](#), [plot.speMCA](#), [plot.csMCA](#)

Examples

```
## Performs a specific MCA on 'Music' example data set
## ignoring every 'NA' (i.e. 'not available') categories,
## plots the cloud of individuals,
## and then adds supplementary individuals.
data(Music)
getindexcat(Music)
mca <- speMCA(Music[3:nrow(Music),1:5],excl=c(3,6,9,12,15))
plot(mca,type='i')
textindsup(mca,Music[1:2,1:5])
```

---

textvarsup   *Adds a categorical supplementary variable to a MCA graph*

Description

Adds a categorical supplementary variable to a MCA graph of the cloud of categories.

Usage

```
textvarsup(resmca, var, sel = 1:nlevels(var), axes = c(1, 2),
                  col = "black", app = 0, vname = NULL)
```

Arguments

resmca	object of class 'MCA', 'speMCA', 'csMCA', 'stMCA' or 'multiMCA'
var	the categorical supplementary variable. It does not need to have been used at the MCA step.
sel	numeric vector of indexes of the categories of the supplementary variable to be added to the plot (by default, labels are plotted for every categories)
axes	numeric vector of length 2, specifying the dimensions (axes) to plot (default is c(1,2))
col	color for the labels of the categories (default is black)
app	numerical value. If 0 (default), only the labels are plotted and their size is constant; if 1, only the labels are plotted and their size is proportional to the weights of the categories; if 2, points (triangles) and labels are plotted, and points size is proportional to the weight of the categories.
vname	a character string to be used as a prefix for the labels of the categories (null by default)

Author(s)

Nicolas Robette

See Also

[plot.speMCA](), [plot.csMCA](), [plot.stMCA](), [plot.multiMCA](), [varsup]()

Examples

```
## Performs a specific MCA on 'Music' example data set
## ignoring every 'NA' (i.e. 'not available') categories,
## plots the cloud of categories,
## and then adds gender and age supplementary categories.
data(Music)
getindexcat(Music)
mca <- speMCA(Music[,1:5],excl=c(3,6,9,12,15))
```

```
plot(mca,col='gray')
textvarsup(mca,Music$Gender,col='darkred')
textvarsup(mca,Music$Age,sel=c(1,3),col='orange',vname='age',app=1)
```

---

translate.logit   *Translate logit regression coefficients into percentages*

Description

Performs a logit regression and then computes the effects of covariates expressed in percentages (through two methods: 'pure' effects and 'experimental' effects; see Deauvieau, 2010)

Usage

```
translate.logit(formula,data,nit=0)
```

Arguments

formula	an object of class formula (or one that can be coerced to that class): a symbolic description of the model to be fitted. Every variables have to be factors.
data	a data frame containing the variables in the model
nit	number of bootstrap iterations for confidence interval computation. Default is 0, i.e. no confidence interval is computed.

Details

This function works with binomial as well as multinomial regression models. If the dependant variable has two factors, glm is used, if it has more than two factors multinom function (from nnet package) is used. The function expresses the regression coefficients as percentages through three distinct methods: raw percentages, 'pure effects' percentages and 'experimental effects' percentages (see Deauvieau, 2010).

Value

The function returns a list:

reg	An object of class glm or nnet (depending on the number of factors of the dependent variable)
summary	The results of summary function applied to reg element
percents	A matrix or a list of matrices (depending on the number of factors of the dependent variable) with regression coefficients expressed as percentages
boot.ci	A matrix or a list of matrices (depending on the number of factors of the dependent variable) with confidence intervals computed with bootstrap

Author(s)

Nicolas Robette

References

Deauvieau, J. (2010), 'Comment traduire sous forme de probabilites les resultats d'une modelisation logit ?', Bulletin of Sociological Methodology / Bulletin de Methodologie Sociologique 105(1), 5-23.

Deauvieau, J. (2011), 'Est-il possible et souhaitable traduire sous forme de probabilites un coefficient logit ? Reponse aux remarques formulees par Marion Selz a propos de mon article paru dans le BMS en 2010', Bulletin of Sociological Methodology / Bulletin de Methodologie Sociologique 112(1), 32-42.

See Also

[glm](), [multinom]()

Examples

```
## An example for binomial logit regression
data(Music)
translate.logit(Daily ~ Gender + Age, Music)

## An example for multinomial logit regression
translate.logit(OnlyMus ~ Gender + Age, Music)
```

---

varsup	*Computes statistics for a categorical supplementary variable*

Description

From MCA results, computes statistics (weights, coordinates, contributions, test-values, variances) for a categorical supplementary variable.

Usage

```
varsup(resmca, var)
```

Arguments

resmca	object of class `'MCA'`, `'speMCA'`, `'csMCA'`, `'stMCA'` or `'multiMCA'`
var	the categorical supplementary variable. It does not need to have been used at the MCA step.

Value

Returns a list:

weight	numeric vector of categories weights
coord	data frame of categories coordinates
cos2	data frame of categories square cosine

var	data frame of categories within variances, variance between and within categories and variable square correlation ratio (eta2)
v.test	data frame of categories test-values

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

speMCA, csMCA, multiMCA, textvarsup

Examples

```
## Performs a specific MCA on 'Music' example data set
## ignoring every 'NA' (i.e. 'not available') categories,
## and then computes statistics for age supplementary variable.
data(Music)
getindexcat(Music)
mca <- speMCA(Music[,1:5],excl=c(3,6,9,12,15))
varsup(mca,Music$Age)
```

---

wtable	*Computes a (possibly weighted) contingency table*

Description

Computes a contingency table from one or two vectors, with the possibility of specifying weights.

Usage

```
wtable(var1,var2=NULL,w=rep.int(1,length(var1)),digits=0,mar=TRUE,na=TRUE)
```

Arguments

var1	an object which can be interpreted as factor
var2	an optional object which can be interpreted as factor
w	an optional numeric vector of weights (by default, a vector of 1 for uniform weights)
digits	integer indicating the number of decimal places (default is 0)

mar	logical. If TRUE (default), margins are computed
na	logical. If TRUE (default), 'NA' are treated as a category. If FALSE, they are ignored

Value

Returns a contingency table in matrix format.

Author(s)

Nicolas Robette

See Also

table, prop.wtable

Examples

```
## Computes a contingency table
## of jazz and age variables
## from the 'Music' example data set
## with or without weights
data(Music)
wtable(Music$Jazz)
wtable(Music$Jazz,Music$Age)

weight <- rep(c(0,0.5,1,1.5,2), length.out=nrow(Music))
wtable(Music$Jazz,w=weight,digits=1)
wtable(Music$Jazz,Music$Age,weight,1)
```

Index

51