Package 'MCAvariants'

September 7, 2020

Type Package

Title Multiple Correspondence Analysis Variants

Version 2.5

Date 2020-09-07

Author Rosaria Lombardo and Eric J Beh

Maintainer Rosaria Lombardo <rosaria.lombardo@unicampania.it>

Description Provides two variants of multiple correspondence analysis (ca): multiple ca and ordered multiple ca via orthogonal polynomials of Emerson.

Depends R (> 3.0.1), methods, tools, ggplot2, ggrepel, gridExtra, plotly

LazyData true

License GPL (> 2)

URL https://www.R-project.org

NeedsCompilation no

Repository CRAN

Date/Publication 2020-09-07 14:50:03 UTC

R topics documented:

`alligator.dat` *Alligator in American lakes*

Description

The data set is a three-way contingency table. It consists of 2 rows (alligators'size), 5 columns (alligators'food) by 4 tubes (alligators'lake). The table should be converted in reduced code table, using the function `tableconvert` for getting `alligatormca`.

Usage

```
data(alligator.dat)
```

Format

A data frame with 300 alligators on the following 3 variables.

Size A numeric vector of categories ranging from 1 to 2 (small and large).

Food A numeric vector of categories ranging from 1 to 5 (type of food: fish, invertebrate, reptile, bird, other.

Lake a numeric vector of categories ranging from 1 to 4 for the four American lakes: Hancock, Oklawaha, Trafford, George.

Source

Agresti (2007), p. 270
Agresti A and Gottard A 2007 Independence in multiway contingency tables: S.n. roys breakthroughs and later developments. Journal of Statistical Planning and Inference, 137:3126–3226.

Examples

```
data(alligator.dat)
dim(alligator.dat)
dimnames(alligator.dat)
```

caplot3d *Three dimensional correspondence plot*

Description

This function is used in the plot function `plot.CAvariants` when the logical parameter is `plot3d = TRUE`. It produces a 3-dimensional visualization of the association.

Usage

```
caplot3d(coordR, coordC, inertiaper, firstaxis = 1, lastaxis = 2, thirdaxis = 3)
```

Arguments

`coordR`	The row principal or standard coordinates.
`coordC`	The column principal or standard coordinates.
`inertiaper`	The percentage of the total inertia explained inertia by each dimension.
`firstaxis`	The first axis number. By default, `firstaxis = 1`.
`lastaxis`	The second axis number. By default, `lastaxis = 2`.
`thirdaxis`	The third axis number. By default, `thirdaxis = 3`.

Note

This function depends on the `R` library `plotly`.

Author(s)

Rosaria Lombardo and Eric J. Beh

References

Beh EJ and Lombardo R 2014 Correspondence Analysis: Theory, Practice and New Strategies. Wiley.
Lombardo R Beh EJ 2016 Variants of Simple Correspondence Analysis. The R Journal, 8 (2), 167–184.

insertval2 *Secondary function to code data*

Description

Secondary function to code data in complete disjunctive form

Usage

```
insertval2(x, nmod)
```

Arguments

x	Data matrix in reduced coding (primitive coding)
nmod	number of categories of each variable

Details

It helps to return a matrix from reduced coding in complete disjunctive coding

Author(s)

Rosaria Lombardo

References

Lombardo R and Meulman JJ (2010) Journal of Classification, 27, 191-210.
Beh EJ Lombardo R (2014) Correspondence Analysis, Theory, Practice and New Strategies. Wiley

mcabasic *Classical multiple correspondence analysis*

Description

This function is used in the main function `MCAvariants` when the input parameter is `catype="mca"`.

Usage

```
mcabasic(xo, np, nmod, tmod, rows, idr, idc, idcv)
```

Arguments

xo	The starting table of variables in reduced code.
np	The column number of the starting table (coincident with the variable number).
nmod	The number of variable catgories of each variable.
tmod	The total number of variable catgories.
rows	The row number of the starting table (coincident with the individual number).
idr	The row labels of the data table.
idc	The column labels of the data table.
idcv	The labels of the categories of each variable.

Note

This function belongs to the R object class called mcabasicresults.

Author(s)

Rosaria Lombardo

References

Lombardo R and Meulman JJ (2010) Journal of Classification, 27, 191-210.
Beh EJ Lombardo R (2014) Correspondence Analysis, Theory, Practice and New Strategies. Wiley

---

mcafun *Classical multiple correspondence analysis*

Description

This function is used in the secondary function mcabasic when the input parameter of MCAvariants is catype="mca". It performs the singular value decomposition of the weighted super-indicator matrix and compute principal axes, coordinates, weights of rows and columns and total inertia.

Usage

```
mcafun(XO, Burt, np, idr, idc, nmod)
```

Arguments

XO	The super-indicator data table.
Burt	The Burt data table.
np	The number of categorical variables.
idr	The row labels of data table.
idc	The column labels of data table.
nmod	The category number of each variable.

Author(s)

Rosaria Lombardo

References

Lombardo R and Meulman JJ (2010) Journal of Classification, 27, 191-210.
Beh EJ Lombardo R (2014) Correspondence Analysis, Theory, Practice and New Strategies. Wiley

---

MCAvariants *Classic and Ordered Multiple Correspondence Analysis*

Description

It performs Classic Multiple Correspondence analysis for nominal variables (setting `catype = "mca"`) and Ordered Multiple Correspondence analysis via orthogonal polynomials (setting `catype="omca"`). When the categorical variables are nominal and ordinal, you can specify writing `FALSE` or `TRUE` in the input parameter `vordered`.

Usage

```
MCAvariants(Xtable, catype = "omca",np = 5, vordered=c(TRUE,TRUE,TRUE,TRUE,TRUE))
```

Arguments

`Xtable`	The two-way contingency table.
`catype`	The input parameter for specifying what variant of multiple correspondence analysis is considered. By default, `catype = "mca"`.
`np`	The input parameter for specifying the number of categorical variables. By default, `np = 5`.
`vordered`	The flag parameter for specifying what variable is ordered, the ordered variables should be in column close each other. By default, all the five variables are ordered: `vordered = c(FALSE,FALSE,TRUE,TRUE,TRUE)`.

Value

Description of the output returned

`Xtable`	The starting table of variables in reduced (primitive) code.
`rows`	The row number of the starting table.
`cols`	The column number of the starting table (coincident with the variable number).
`rowlabels`	The label of the row individuals.
`columnlabels`	The label of the column variable categories.
`Rprinccoord`	The coordinates of individuals.
`Cprinccoord`	The category variable coordinates.

`inertiaXsum` The total inertia when multiple correspondence analysis is performed on the indicator table.

`inertiaBurtsum` The total inertia when multiple correspondence analysis is performed on the Burt table.

`inertias` Benzecri's Adjusted Inertia values, percentages and cumulative values.

`inertiasAdjusted` The adjusted inertia values.

`catype` The kind of multiple correspondence analysis chosen, classical or ordered, that is `catype` is "mca" or "omca".

`printdims` The dimension of a matrix in print. By default it is equal to 3.

`comp` The polynomial components of inertia when `catype` is "omca". The total inertia is partitioned in terms of polynomial components.

`componentpvalue1` The p-value of the polynomial components of total inertia, when `catype` is "omca".

`degreef` The degree of freedom of polynomial components of total inertia when, `catype` is "omca".

Note

This function recalls internally two other functions, depending on the setting of the input parameter `catype`, it recalls multiple correspondence analysis or ordered multiple correspondence analysis. It gives the output object necessary for printing and plotting the results. These two important functions are `print.MCAvariants` and `plot.MCAvariants`.

Author(s)

Rosaria Lombardo and Eric J Beh

References

Lombardo R and Meulman JJ (2010) Journal of Classification, 27, 191-210.
Beh EJ Lombardo R (2014) Correspondence Analysis, Theory, Practice and New Strategies. Wiley

Examples

```
data(satisfaction)
MCAvariants(satisfaction, catype = "omca",np=5,vordered=c(TRUE,TRUE,TRUE,TRUE,TRUE))
MCAvariants(satisfaction, catype = "mca",np=5)
```

miocount *The counting function*

Description

The function that counts the number of individuals in each clusters automatically generated in ordered multiple correpsondence analysis.

Usage

```
miocount(x)
```

Arguments

x	The coordinates of axes

Note

This function is used in the function omcabasic when in the main function MCAvariants the input parameter is catype="omca".

Author(s)

Rosaria Lombardo

References

Lombardo R and Meulman JJ (2010) Journal of Classification, 27, 191-210.
Beh EJ Lombardo R (2014) Correspondence Analysis, Theory, Practice and New Strategies. Wiley

omcabasic *Ordered multiple correspondence analysis via orthogonal polynomials*

Description

This function is used in the main function MCAvariants when the input parameter is catype="omca". It requires that all categorical variables are ordered variables. It performs the hybrid decomposition of the weighted super-indicator matrix and compute polynomial axes, coordinates, weights of rows and columns and total inertia.

Usage

```
omcabasic(xo,np , nmod , tmod , rows, idr, idc, idcv,vordered)
```

Arguments

xo	The starting table of variables in reduced code.
np	The column number of the starting table (coincident with the variable number).
nmod	The number of variable catgories of each variable.
tmod	The total number of variable catgories.
rows	The row number of the starting table (coincident with the individual number).
idr	The row labels of the data table.
idc	The column labels of the data table.
idcv	The labels of the categories of each variable.
vordered	The flag parameter for specifying what variable is ordered. By default, all the five variables are ordered: `vordered = c(FALSE,FALSE,TRUE,TRUE,TRUE)`.

Note

This function belongs to the R object class called `mcabasicresults`.

Author(s)

Rosaria Lombardo

References

Lombardo R and Meulman JJ (2010) Journal of Classification, 27, 191-210.
Beh EJ Lombardo R (2014) Correspondence Analysis, Theory, Practice and New Strategies. Wiley

---

orthopoly *Orthogonal polynomials*

Description

This function is called from the function `omca`. It allows the analyst to compute the orthogonal polynomials of each ordered categorical variable. The number of the polynomials is equal to the variable category less one. The function computes the polynomial transformation of the ordered categorical variable.

Usage

```
orthopoly(marginals, scores)
```

Arguments

scores	The ordered scores of an ordered variable. By default `mj=NULL`, the natural scores (1,2,...) are computed.
marginals	The marginals, relative frequencies of the ordered variable.

Value

Describe the value returned

B the matrix of the orthogonal polynomials without the trivial polynomial.

Note

Note that the sum of the marginals of the ordered variables should be one. At the end, the various polynomial matrices will be stored in a super-diagonal matrix.

Author(s)

Rosaria Lombardo and Eric J Beh

References

Beh EJ and Lombardo R 2014 Correspondence analysis, Theory, Practice and New Strategies. Wiley.

Examples

```
orthopoly(marginals=c(.1,.2,.3,.2,.2), scores=c(1,2,3,4,5))
```

---

plot.MCAvariants *Main plot function for classical and ordered multiple correspondence analysis*

Description

This function allows the analyst to produce the suitable graphical displays with respect to the classical and ordered multiple correspondence analysis. The main plot function called from the main function `MCAvariants`. It produces classical graphical displays for `catype = "mca"` and `catype = "omca"`.

Usage

```
## S3 method for class 'MCAvariants'
plot(x, catype = "mca", firstaxis = 1, lastaxis = 2, thirdaxis = 3, cex = 0.8,
cex.lab = 0.8, prop = 1, plot3d = FALSE, plotind= FALSE, M=2,...)
```

Arguments

x	Represents the set of the output parameters of the main function MCAvariants of the R object class mcacorporateris.
catype	The input parameter specifying what variant of correspondence analysis is requested.
firstaxis	The dimension reflected along the horizontal axis.
lastaxis	The dimension reflected along the vertical axis.
thirdaxis	The third axis number when plot3d = TRUE. By default, thirdaxis = 3.
cex	The parameter that specifies the size of character labels of points in graphical displays. By default, it is equal to 1.
cex.lab	The parameter cex.lab that specifies the size of character labels of axes in graphical displays. By default, cex.lab = 0.8.
prop	The scaling parameter for specifying the limits of the plotting area. By default, it is equal to 1.
plot3d	The logical parameter specifies whether a 3D plot is to be included in the output or not. By default, plot3d = FALSE.
plotind	The logical parameter specifies whether a plot of individuals is to be included in the output or not. By default, plotind = FALSE.
M	The number of axes M considered when portraying the elliptical confidence regions. By default, it is equal to M = 2.
...	Further arguments passed to or from other methods.

Details

It produces classical graphical displays. Further when catype is equal to "omca", the individual clusters are portrayed.

Author(s)

Rosaria Lombardo and Eric J Beh

References

Lombardo R and Meulman JJ (2010) Journal of Classification, 27, 191-210.
Beh EJ Lombardo R (2014) Correspondence Analysis, Theory, Practice and New Strategies. Wiley

Examples

```
data(satisfaction)
ris1=MCAvariants(satisfaction, catype = "mca")
plot(ris1)
print(ris1)
ris2=MCAvariants(satisfaction, catype = "omca")
plot(ris2)
print(ris2)
```

`print.MCAvariants` *Main printing function*

Description

This function prints results of classical or ordered multiple correspondence analysis. The input parameter is the name of the output of the main function `MCAvariants`.

Usage

```
## S3 method for class 'MCAvariants'
print(x, printdims = 2,...)
```

Arguments

`x`	The output of the main function `CAvariants`.
`printdims`	The number of dimensions, `printdims`, that are used to generate the correspondence plot and for summarising the numerical output of the analysis. By default, `printdims = 2`.
`...`	Further arguments passed to or from other methods.

Details

This function uses another function (called `printwithaxes`) for specifying the number of matrix dimensions to print.

Value

The value of output returned depends on the kind of multiple correspondence analysis performed.

`DataTable`	The Burt data table.
`Row coordinates`	Rows in principal coordinates: the first 10.
`Column coordinates`	Column in principal coordinates.
`Polynomials`	Polynomial functions of each variable. When `catype` is `omca`.
`Linear Percentage of Clusters`	The percentage of individuals belonging to each cluster. When `catype` is `omca`.
`Polynomial Components of Total Inertia`	The decomposition of total inertia via orthogonal polynomials. When `catype` is `omca`.
`Degree of Freedom`	Degree of Freedom of Polynomial Component. When `catype` is `omca`.
`Inertia values`	Inertia values of super-indicator and Burt table.
`Benzecri's Inertia values`	Adjusted Inertia values, percentages and cumulative.

`Total Degree of Freedom`
: The degree of freedom of total inertia.

`Total inertia of X`
: Total inertia of Super-Indicator table

`Total inertia of B`
: Total inertia of BURT table.

`Chi-square values`
: Chi-square values of BURT Inertia.

`Total Chi-square values`
: Chi-square values of total Inertia of Burt table.

Author(s)

Rosaria Lombardo

References

Lombardo R and Meulman JJ (2010) Journal of Classification, 27, 191-210.
Beh EJ Lombardo R (2014) Correspondence Analysis, Theory, Practice and New Strategies. Wiley

---

`printwithaxes` *Secondary printing function*

Description

The function is called from the main print function `printmcacorporateris`. It adds the names to objects.

Usage

```
printwithaxes(res, thenames)
```

Arguments

`res`	An `R` object.
`thenames`	A character vector of up to the same length as `x`.

Note

It is called from `printmcacorporateris`.

Author(s)

Rosaria Lombardo

References

Lombardo R and Meulman JJ (2010) Journal of Classification, 27, 191-210.
Beh EJ Lombardo R (2014) Correspondence Analysis, Theory, Practice and New Strategies. Wiley

satisfaction *Patient Satisfaction*

Description

The data set consists of 235 rows and 5 columns. The rows represent the individuals (patients in an hospital) and the columns concern the five variables of satisfaction (Tangibility, Reliability, Capacity of Response, Capacity of Assurance and Empathy

Usage

```
data(satisfaction)
```

Format

A data frame with 235 observations on the following 5 variables.

`TANG` a numeric vector of ordered categories ranging from 1 to 5.

`REL` a numeric vector of ordered categories ranging from 1 to 5.

`CRES` a numeric vector of ordered categories ranging from 1 to 5.

`CASS` a numeric vector of ordered categories ranging from 1 to 5.

`EMPAT` a numeric vector of ordered categories ranging from 1 to 5.

Source

Della Valle E (2010) Della Valle E 2010 Analisi Multidimensionale dei Dati: la Qualit\'a del Lavoro nelle Cooperative Sociali. Unpublished thesis, Seconda Universit\'a degli Studi di Napoli, Capua Italy.

Examples

```
data(satisfaction)
dim(satisfaction)
dimnames(satisfaction)
```

tableconvert *Convert contingency table in table of reduced code*

Description

This simple piece of R code converts a two-way or three-way contingency table into what is required to analyse MCA (table of reduced code: n by number of variables).

Usage

```
tableconvert(N)
```

Arguments

N	A two-way or three-way contingency table to convert in a table n by np, where np is the number of the categorical variables.

Author(s)

Rosaria Lombardo and Eric J Beh

References

Beh EJ Lombardo R (2014) Correspondence Analysis, Theory, Practice and New Strategies. Wiley

Examples

```
alligatormca<-tableconvert(alligator.dat)
dimnames(alligatormca)<-list(paste("a", 1:300,sep = ""),c("Size","Food","Lake"))
```

Index