Package 'qha'

October 13, 2022

Version 0.0.8

Date 2016-09-08

Title Qualitative Harmonic Analysis

Author Martha Lucia Corrales <martha.corrales@usa.edu.co> and
Campo Elias Pardo <cepardot@unal.edu.co>

Maintainer Campo Elias Pardo <cepardot@unal.edu.co>

Depends R (>= 3.2.1), ade4,FactoClass,FactoMineR

Description Multivariate description of the state changes of a qualitative variable by
Correspondence Analysis and Clustering. See:
Deville, J.C., & Saporta, G. (1983).
Correspondence analysis, with an extension towards nominal time series.
Journal of econometrics, 22(1-2), 169-189.
Corrales, M.L., & Pardo, C.E. (2015) <doi:10.15332/s2027-3355.2015.0001.01>.
Analisis de datos longitudinales cualitativos con analisis de correspondencias y clasificacion.
Comunicaciones en Estadistica, 8(1), 11-32.

License GPL (>= 2)

Encoding latin1

NeedsCompilation no

Repository CRAN

Date/Publication 2016-09-10 01:29:30

R topics documented:

`classification` *Cluster Analysis*

Description

It makes cluster analysis of the data using the nfcl first factorial coordinates

Usage

```
classification(datos,numclass=2,numaxes=5,pesos=NULL)
```

Arguments

`datos`	object of type data frame or matrix with coordinates of factorial analysis
`numclass`	number of class. Default 2
`numaxes`	number of dimensions for the classification. Default 5
`pesos`	a vector of row weights. Default NULL

Details

The data given by datos is clustered by the Ward method using the function ward.cluster of the package FactoClass. This classification is optimized across the method kmeans.

Value

An object of class "kmeans", a list with components: cluster, centers, withinss and size. To more information see the function kmeans.

Author(s)

Martha Lucia Corrales <martha.corrales@usa.edu.co> & Campo Elias Pardo <cepardot@unal.edu.co>

Examples

```
data(ratingTV)


# Fuzzy Correspondence Analysis
fuzzyb<-fuzzy(datos=ratingTV$tab, vectorc=c(15,15,15,15,15,15))
6

#Cluster analysis
clases<-classification(datos=fuzzyb$li,numclass=7, numaxes=6)
clases$cluster
clases$centers
clases$size
```

`combination` *Combination of Qualitative Harmonic and Multiple Factor Analyses*

Description

It realizes the combination among the Qualitative Harmonic Analysis and the Multiply Factorial Analysis for the treatment of longitudinal qualitative variables.

Usage

```
combination(datos, vectorc, ilustra = NULL, ilustc=NULL, ilust.type=NULL,pesos=NULL)
```

Arguments

`datos`	object of type dataframe or matrix containing positive values (counts or proportions)
`vectorc`	a vector containing the number of categories for each fuzzy variable
`ilustra`	object of type matrix or data frame with the illustrative variables .Default = NULL)
`ilustc`	a vector containing the number of variables in each ilustrative group
`ilust.type`	the type of variable in each ilustrative group: "c" for quantitative variables, "s" for quantitative variables scales to unit variance, "n" for qualitative variables. By default all variables are qualitative)
`pesos`	a vector containing the weights of the rows

Details

The AAC presented by the first time in 1979 by Deville and Saporta, is an exploratory method of longitudinal categorical information which aim is to construct the information of the individuals of a sample, during a certain period of time. These individuals are described by the chronology of a succession of changes of stages among a set of possible stages. Deville and Saporta demonstrated that information of this type can be analyzed of form similar to what would make in a process escalar (harmonic analysis).

When in the AAC it is chosen for a recodification that indicates the proportion of time that an individual remains in a category, with regard to a specific period of time, there is obtained a table that can turn as a table of diffuse codification, that is to say, the individuals take different values of the same variable with different degrees of association; and this one can be analyzed across an extension of the analysis of multiple correspondences: the diffuse analysis of correspondences, proposed by Chevenet, Doledec and Chessel in 1994.

The analysis factorial multiple appears as a method factorial for the description of the same set of individuals across several groups of variables. In the longitudinal studies, the groups are formed according to the moment of the time in which the variables have been measured. It is obtained then, a table with combinations of transverse cuts in the time.

The function combination, considers the use of both methods to analyze longitudinal categorical data.

Value

Returns the individuals factor map, the variables factor map and the groups factor map. The function combination returns an object type MFA of the package FactoMineR.

Author(s)

Martha Lucia Corrales <martha.corrales@usa.edu.co> & Campo Elias Pardo <cepardot@unal.edu.co>

Examples

```
# data(ratingTV)
# AAC_FMA<-combination(datos=ratingTV$tab,vectorc=c(15,15,15,15,15,15),
#        ilustra=ratingTV$ilus,ilustc=c(4))
#10
#AAC_FMA$separate.analyses
#AAC_FMA$eig
#AAC_FMA$group
#AAC_FMA$summary.quali
#AAC_FMA$partial.axes
```

---

duration *Calculate of duration time*

Description

It calculates the duration time for every individual in the category of the longitudinal variable

Usage

```
duration(datos, units = "auto")
```

Arguments

datos	a data frame containing ID, Start-Time,End-Time, Modality of the Longitudinal Variable
units	unis time:"secs", "mins", "hours", "days", "weeks", "months", "years"

Details

It calculates the duration time to the table containing ID, START_TIME,END_TIME,MODALIDAD

Value

datos	a data frame containing ID, Modality of the Longitudinal Variable, Duration time

Author(s)

Martha Lucia Corrales <martha.corrales@usa.edu.co> & Campo Elias Pardo <cepardot@unal.edu.co>

durationtotableA *Groups by Time Periods*

Description

It realizes the transformation of the table that contains the durations in every modality of the longitudinal variable in the table that it groups for periods of time

Usage

```
durationtotableA(x,periodos)
```

Arguments

x	a data frame containing ID, Modality of the Longitudinal Variable, Duration
periodos	a vector containing the duration of each period time to do a groups

Details

This function transforms the table that contains the durations passed by the individuals in every modality of a longitudinal variable, in the table that groups these modalities in periods of time. That is to say, in the table that contains the quantity of time that an individual remains in a category of the longitudinal variable, in every period of time

Value

Conteo	a data frame with categories by number of time units

Author(s)

Martha Lucia Corrales <martha.corrales@usa.edu.co> & Campo Elias Pardo <cepardot@unal.edu.co>

Examples

```
#data(zonas)
#tableA1<-durationtotableA(zonas$zona1,c(360,240,60,150,160,230))
```

fuzzy *Qualitative Harmonic Analysis*

Description

It realizes qualitative harmonic analysis like a fuzzy correspondence analysis

Usage

```
fuzzy(datos,vectorc,nf=5,pesos=NULL)
```

Arguments

datos	object of type data frame or matrix
vectorc	a vector containing the number of categories for each fuzzy variable
nf	an integer indicating the number of kept axes. Default nf=5
pesos	a vector of row weights

Details

The AAC presented by the first time in 1979 by Deville and Saporta, is an exploratory method of longitudinal categorical information which aim is to construct the information of the individuals of a sample, during a certain period of time. These individuals are described by the chronology of a succession of changes of stages among a set of possible stages. Deville and Saporta demonstrated that information of this type can be analyzed of form similar to what would make in a process escalar (harmonic analysis).

In its practical application it is the accomplishment of an analysis of correspondences

When in the AAC it is chosen for a recodification that indicates the proportion of time that an individual remains in a category, with regard to a specific period of time, there is obtained a table that can turn as a table of diffuse codification, that is to say, the individuals take different values of the same variable with different degrees of association; and this one can be analyzed across an extension of the analysis of multiple correspondences: the diffuse analysis of correspondences, proposed by Chevenet, Dol?dec and Chessel in 1994.

Value

fuzzya	object of type dudi

Author(s)

Martha Lucia Corrales <martha.corrales@usa.edu.co> & Campo Elias Pardo <cepardot@unal.edu.co>

Examples

```
data(ratingTV)

# Fuzzy Correspondence Analysis with weights
fuzzya<-fuzzy(datos=ratingTV$tab, vectorc=c(15,15,15,15,15,15), pesos=ratingTV$weight)
6


# Fuzzy Correspondence Analysis without weights
fuzzyb<-fuzzy(datos=ratingTV$tab, vectorc=c(15,15,15,15,15,15))
6

# First factorial plane with FactoClass
#plot(fuzzya,1,2,Tcol=FALSE,ucal=10)
#plot(fuzzyb,1,2,Tcol=FALSE,ucal=10)
```

---

hqa *Combination of Qualitative Harmonic and Multiple Factor Analyses and Clustering*

Description

It realizes the combination between two methods for the processing of longitudinal categorical data: the Qualitative Harmonic and the Multiple Factorial Analysis.

Usage

```
hqa(base, conteos=FALSE, units=NULL, durat=FALSE, periodos=NULL, pesos = NULL,
    ilustra=NULL,ilustc = NULL,ilust.type = NULL, nfact=5, nfcl=5, k.clust=NULL,
    combinat=TRUE, vector, tableclass=FALSE, clasifica=TRUE)
```

Arguments

base	object of type data frame or matrix
conteos	TRUE if you want to do data frame ID,MOD,DURATION. Default TRUE
units	time: "secs", "mins", "hours", "days", "weeks", "months", "years". Default = NULL
durat	TRUE if you want to calculate the DURATION by the function. Default FALSE
periodos	a vector containing the duration of each period of time
pesos	a vector of row weights
ilustra	object of type data frame or matrix with the illustrative variables. Default NULL
ilustc	a vector containing the number of variables in each ilustrative group

ilust.type	the type of variable in each ilustrative group: "c" for quantitative variables, "s" for quantitative variables scales to unit variance, "n" for qualitative variables. By default all variables are qualitative
nfact	number of axes to use into the factorial analysis . Default nfact=5
nfcl	number of axes to use in the classification. Default nfcl=5
k.clust	number of classes to work. Default k.clust= NULL
combinat	TRUE if you want to do combination HQA and MFA, FALSE if you want to do only AAC. Default TRUE)
vector	a vector containing the number of categories for each fuzzy variable
tableclass	TRUE if you want a function to Suggest you the number of axes to use in the classificacion. Default FALSE
clasifica	TRUE if you want to do the classificacion. Default TRUE

Details

A new statistical methodology is proposed in order to analyze longitudinal categorical data. This methodology considers the use of two methods: Qualitative Harmonic and Multiple Factor Analysis. The analysis is complemented by an analysis of classification using the first coordinates factoriales of the data.

Value

HQA	An object of type dudi or MFA
Clases	An object of class "kmeans"
Active	characterization of the classes considering the longitudinal active variable
Ilust	characterization of the classes considering the ilustratives variables

Author(s)

Martha Lucia Corrales <martha.corrales@usa.edu.co> & Campo Elias Pardo <cepardot@unal.edu.co>

References

Corrales, M. L., & Pardo, C. E. (2015). Analisis de datos longitudinales cualitativos con analisis de correspondencias y clasificacion. Comunicaciones en Estadistica, 8(1), 11-32. http://dx.doi.org/10.15332/s2027-3355.2015.0001.01

Examples

```
#data(ratingTV)
#rating <- hqa(base=ratingTV$tab,ilustra=ratingTV$ilus, #vector=c(15,15,15,15,15,15), ilustc=c(4))
#10
#rating$HQA
#rating$Clases
#rating$Active
#rating$Ilust
```

ratingTV *Rating TV in Colombia*

Description

An application for the study of Colombian rating television data for a given day and the resulting typology of television viewers.

Usage

```
data(ratingTV)
```

Format

Object of class `list` with: `$tab` a `data.frame` of 1306 rows and 90 columns; `$ilus` a `data.frame` of 1306 rows and 4 colums of sociodemographic variables; and a `vector` `weights` with the 1306 weights of the rows.

Source

The Data were provided by professor Juan de la Rosa Ramos

References

Corrales, M.L., & Pardo, C.E. (2015). Analisis de datos longitudinales cualitativos con analisis de correspondencias y clasificacion. Comunicaciones en Estadistica, 8(1), 11-32. http://dx.doi.org/10.15332/s2027-3355.2015.0001.01

tableclass *Correspondence Analysis Factorial Planes*

Description

It suggests the number of axes to using in a cluster analysis

Usage

```
tableclass(tabla,numaxes=2,numclass=2)
```

Arguments

`tabla`	object of type data frame or matrix with coordinates of factorial analysis
`numaxes`	number of dimensions for the classification. Default 2
`numclass`	number of class. Default 2

Details

To understand the differences between elements of differents class, the planes factoriales are insufficient. It is necessary, therefore, to proceed to the classification of the individuals in the vectorial space of the first factors of the analysis. To decide the number of axes to using, Barbary (1996) proposed a procedure which consists of realizing the classification several times changing the number of axes and using as criterion the number of individuals of the class with major quantity of individuals. The analysis to interpreting is that one that originates the minor number of individuals in the majority class.

The function tableclass realizes the classification several times y it proposes a number of axes to the classification.

Value

`tableclas` object of type matrix

Author(s)

Martha Lucia Corrales <martha.corrales@usa.edu.co> & Campo Elias Pardo <cepardot@unal.edu.co>

Examples

```
data(ratingTV)

# Fuzzy Correspondence Analysis
fuzzyb<-fuzzy(datos=ratingTV$tab, vectorc=c(15,15,15,15,15,15))
20

tabla_clases<-tableclass(fuzzyb$li, numaxes=4, numclass=7)
tabla_clases
```

`zonas` *Zonas*

Description

An application for the study of Colombian rating television data for a given day and the resulting typology of television viewers.

Usage

```
data(zonas)
```

Format

Object whit class `data.frame` of 1306 rows and 90 columns.

Source

The Data were provided by professor Juan de la Rosa Ramos

References

Corrales, M.L., & Pardo, C.E. (2015). Analisis de datos longitudinales cualitativos con analisis de correspondencias y clasificacion. Comunicaciones en Estadistica, 8(1), 11-32. http://dx.doi.org/10.15332/s2027-3355.2015.0001.01

Index