

Package ‘anacor’

May 1, 2017

Type Package

Title Simple and Canonical Correspondence Analysis

Version 1.1-3

Date 2017-04-29

Author Patrick Mair [aut, cre], Jan De Leeuw [aut]

Maintainer Patrick Mair <mair@fas.harvard.edu>

Description Performs simple and canonical CA (covariates on rows/columns) on a two-way frequency table (with missings) by means of SVD. Different scaling methods (standard, centroid, Benzecri, Goodman) as well as various plots including confidence ellipsoids are provided.

Imports graphics, stats, car, colorspace, fda, splines, grDevices

Depends R (>= 3.0.2)

URL <http://r-forge.r-project.org/projects/psychor/>

License GPL-2

LazyData yes

LazyLoad yes

ByteCompile yes

NeedsCompilation yes

Repository CRAN

Date/Publication 2017-05-01 16:11:20 UTC

R topics documented:

| | |
|-----------------------|---|
| anacor | 2 |
| bitterling | 4 |
| burtTable | 5 |
| expandFrame | 6 |
| galton | 7 |
| glass | 7 |
| maxwell | 8 |
| mkIndiList | 9 |

| | |
|-----------------------|----|
| plot.anacor | 10 |
| sleeping | 12 |
| spider | 13 |
| srole | 14 |
| tocher | 15 |

| | |
|--------------|-----------|
| Index | 16 |
|--------------|-----------|

anacor *Simple and Canonical Correspondence Analysis*

Description

This function performs simple and canonical CA for possibly incomplete tables based on SVD. Different scaling methods for row and column scores are provided.

Usage

```
anacor(tab, ndim = 2, row.covariates, col.covariates, scaling = c("Benzecri", "Benzecri"),
ellipse = FALSE, eps = 1e-06)
```

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

Arguments

| | |
|----------------|--|
| tab | Data frame of dimension n times m with frequencies. Missings are coded as NA. |
| ndim | Number of dimensions. |
| row.covariates | Matrix with n rows containing covariates for the row scores. |
| col.covariates | Matrix with m rows containing covariates for the column scores. |
| scaling | A vector with two elements. The first one corresponds to the method for row scaling, the second one for column scaling. Available scaling methods are "standard", "Benzecri", "Goodman". |
| ellipse | If TRUE, confidence ellipses are computed. |
| eps | Convergence criterion for reconstitution algorithm. |
| x | Object of class "anacor" in print.anacor. |
| ... | Additional arguments ignored. |

Details

Missing values in `tab` are imputed using the reconstitution algorithm. Setting `scaling` to "standard" leads to standard coordinates. Principal coordinates can be computed by means of Benzecri decomposition. Goodman scaling is based on Fisher-Maung decomposition.

For large datasets it is suggested to set `ellipse = FALSE`. If `ellipse = TRUE`, make sure that there are no rows and columns that have full 0 entries.

Value

| | |
|---------------------------------|--|
| <code>row.scores</code> | Scaled row scores. |
| <code>col.scores</code> | Scaled column scores. |
| <code>ndim</code> | Number of dimensions extracted. |
| <code>chisq</code> | Total chi-square value. |
| <code>chisq.decomp</code> | Chi-square decomposition across dimensions with p-values. |
| <code>singular.values</code> | Singular values without trivial solution. |
| <code>se.singular.values</code> | Standard errors for the singular values. |
| <code>stestmat</code> | Test results for singular values (only for <code>ellipse = TRUE</code>). |
| <code>left.singvec</code> | Left singular vectors without trivial solution. |
| <code>right.singvec</code> | Right singular vectors without trivial solution. |
| <code>eigen.values</code> | Eigenvalues for the fitted dimensionality. |
| <code>eigenvall</code> | Full vector of eigenvalues (principal inertias). |
| <code>datname</code> | Name of the dataset. |
| <code>tab</code> | Table with imputed frequencies in case of missings. |
| <code>row.covariates</code> | Matrix with row covariates. |
| <code>col.covariates</code> | Matrix with column covariates. |
| <code>scaling</code> | Scaling Method. |
| <code>bdmat</code> | List of matrices with observed and fitted Benzecri distances for rows and columns. |
| <code>rmse</code> | Root mean squared error of Bezencri distances (rows and columns). |
| <code>row.acov</code> | Covariance matrix for row scores. |
| <code>col.acov</code> | Covariance matrix for column scores. |
| <code>cancoef</code> | List containing canonical coefficients (CCA only). |
| <code>sitescores</code> | List containing the site scores (CCA only). |
| <code>isetcor</code> | List containing the intraset correlations (CCA only). |

Author(s)

Jan de Leeuw, Patrick Mair

References

De Leeuw, J. and Mair, P. (2009). Simple and Canonical Correspondence Analysis Using the R Package `anacor`. *Journal of Statistical Software*, 31(5), 1-18. <http://www.jstatsoft.org/v31/i05/>

See Also

[plot.anacor](#)

Examples

```
## simple CA on Tocher data, symmetric standard coordinates
data(tocher)
res <- anacor(tocher)
res

## simple CA on Tocher data, asymmetric coordinates
res <- anacor(tocher, scaling = c("standard", "Benzecri"))
res

## 2- and 5-dimensional solutions for bitterling data, Benzecri scaling
data(bitterling)
res1 <- anacor(bitterling, ndim = 2, scaling = c("Benzecri", "Benzecri"))
res2 <- anacor(bitterling, ndim = 5, scaling = c("Benzecri", "Benzecri"))
res1
res2

## Canonical CA on Maxwell data, Goodman scaling
data(maxwell)
res <- anacor(maxwell$table, row.covariates = maxwell$row.covariates,
scaling = c("Goodman", "Goodman"))
res
```

bitterling

Bitterling

Description

This dataset concerns reproductive behavior of male bitterlings with data derived from 13 sequences using a moving time-window of size two.

Usage

```
data(bitterling)
```

Format

A frequency tables with bitterling reproductive behavior at time point 1 (rows) and at time point 2 (columns).

jk jerking

tu turning beats

hb head butting

chs chasing

ft fleeing

qu quivering
le leading
hdp head down posture
sk skimming
sn snapping
chf chafing
ffl finflickering

References

Wiepkema, P.R. (1961). An ethological analysis of the reproductive behavior of the bitterling (*rhodeus amarus bloch*). Archives Neerlandais Zoologique, 14, 103-199.

Examples

```
data(bitterling)
```

| | |
|-----------|----------------------------|
| burtTable | <i>Creates Burt Matrix</i> |
|-----------|----------------------------|

Description

Utility function to produce a Burt matrix out of a data-frame.

Usage

```
burtTable(data)
```

Arguments

data Data frame to be converted.

See Also

[expandFrame](#), [mkIndiList](#)

Examples

```
## sleeping bags  
data(sleeping)  
sleeping_cat <- sleeping  
temp_cat <- cut(sleeping$Temperature, c(-20, -1, 7), labels = c("warm", "cold"))  
sleeping_cat$Temperature <- temp_cat  
weight_cat <- cut(sleeping$Weight, c(700, 1100, 2200), labels = c("light", "heavy"))  
sleeping_cat$Weight <- weight_cat  
price_cat <- cut(sleeping$Price, c(100, 250, 350, 700),
```

```

labels = c("cheap", "medium", "expensive"))
sleeping_cat$Price <- price_cat
sleeping_cat
burtTable(sleeping_cat)

```

expandFrame

Expand Matrix

Description

This utility function expands a matrix or data frame to an indicator supermatrix and optionally converts this to a data frame again. By default NA becomes zero and constant rows and columns are eliminated.

Usage

```
expandFrame(tab, clean = TRUE, zero = TRUE, returnFrame = TRUE)
```

Arguments

| | |
|-------------|---|
| tab | Data frame (factors). Missings are coded as NA. |
| clean | If TRUE, rows and columns with 0 margins in data frame are deleted. |
| zero | If TRUE, NA's are replaced by 0. |
| returnFrame | If TRUE, a data frame is returned; if FALSE a matrix. |

See Also

[burtTable](#), [mkIndiList](#)

Examples

```

## sleeping bags
data(sleeping)
sleeping_cat <- sleeping
temp_cat <- cut(sleeping$Temperature, c(-20, -1, 7), labels = c("warm", "cold"))
sleeping_cat$Temperature <- temp_cat
weight_cat <- cut(sleeping$Weight, c(700, 1100, 2200), labels = c("light", "heavy"))
sleeping_cat$Weight <- weight_cat
price_cat <- cut(sleeping$Price, c(100, 250, 350, 700),
labels = c("cheap", "medium", "expensive"))
sleeping_cat$Price <- price_cat
sleeping_cat
expandFrame(sleeping_cat)

```

| | |
|--------|--------------------------|
| galton | <i>Galton's RFF data</i> |
|--------|--------------------------|

Description

Records of family faculties cross-classification of midparent height and adult children height in inches.

Usage

```
data(galton)
```

Format

A frequency table with 11 times 14 height classifications in inches.

References

Galton, F. (1889). Natural Inheritance. London: MacMillan.

Examples

```
data(galton)
```

| | |
|-------|-------------------|
| glass | <i>Glass data</i> |
|-------|-------------------|

Description

Table with occupational status of fathers versus occupational status of their sons for a sample of 3497 British families.

Usage

```
data(glass)
```

Format

Rows represent occupation of fathers, columns occupation of sons.

PROF professional and high administrative
EXEC managerial and executive
HSUP higher supervisory
LSUP lower supervisory
SKIL skilled manual and routine nonmanual
SEMI semi-skilled manual
UNSK unskilled manual

References

Glass, D.V. (1954). *Social Mobility in Britain*. Glencoe: Free Press.

Examples

```
data(glass)
## maybe str(glass) ; plot(glass) ...
```

maxwell

Maxwell's data

Description

This data set is a hypothetical data set originally contrived by Maxwell (1961) for demonstrating his method of discriminant analysis. The data consist of three criterion groups, schizophrenic, manic-depressive and anxiety state, and four binary predictor variables each indicating either presence (1) or absence (0) of a certain symptom. The four symptoms are anxiety suspicion, schizophrenic type of thought disorders, and delusions of guilt. These four binary variables were factorially combined to form 16 distinct patterns of symptoms (predictor patterns), and each of these patterns is identified with a row of the table, which contains the cross-classification of 620 patients according to the 16 patterns of symptoms and the three criterion groups.

Usage

```
data(maxwell)
```

Format

A list with the frequency table as the first element and the row covariates as the second.

Details

This dataset can be used for canonical CA. The binary predictor variables can be considered as row covariates.

References

Maxwell, A.E. (1961). Canonical variate analysis when the variables are dichotomous. *Educational and Psychological Measurement*, 21,259-271.

Examples

```
data(maxwell)
## maybe str(maxwell) ; plot(maxwell) ...
```

| | |
|------------|--|
| mkIndiList | <i>Converts Data Frame to Indicator Matrix</i> |
|------------|--|

Description

This function takes a data frame, a vector of types, a list of knot vectors, and a vector of orders. It returns a list of codings for the variables, i.e., crisp indicator, numerical version, or fuzzy indicator.

Usage

```
mkIndiList(data, type = rep("C",dim(data)[2]), knots, ord)
```

Arguments

| | |
|-------|--|
| data | Data frame to be converted. |
| type | If "C", a crisp indicator is returned; if "A", a numerical version; if "F", the b-spline basis as a fuzzy indicator. |
| knots | List of knot sequences for type-F coding. |
| ord | Vector with b-spline order for type-F coding. |

Details

For the fuzzy coding, the variable values need to be provided as integers. Each list element contains a vector with knots (breaks) for each variable separately. The order is defined through the ord argument as vector (again, for each variable). See bsplines help file for more details.

See Also

[expandFrame](#), [burtTable](#)

Examples

```
## sleeping bags crisp and numeric
data(sleeping)
sleeping_cat <- sleeping
temp_cat <- cut(sleeping$Temperature, c(-20, -1, 7), labels = c("warm", "cold"))
sleeping_cat$Temperature <- temp_cat
weight_cat <- cut(sleeping$Weight, c(700, 1100, 2200), labels = c("light", "heavy"))
sleeping_cat$Weight <- weight_cat
price_cat <- cut(sleeping$Price, c(100, 250, 350, 700),
labels = c("cheap", "medium", "expensive"))
sleeping_cat$Price <- price_cat
sleeping_cat
mkIndiList(sleeping_cat)      ## crisp
mkIndiList(sleeping_cat, type = rep("A", ncol(sleeping_cat)))      ## numeric
mkIndiList(sleeping_cat, type = c("A","A","A","C"))      ## mixed

## artificial data fuzzy coding
```

```
x1 <- sample(1:6, 20, replace = TRUE)
x2 <- sample(1:3, 20, replace = TRUE)
data <- data.frame(x1,x2)
knots <- list(c(1,3,5,6), c(1,2,3))
ord <- c(2,1)
mkIndiList(data, type = c("F","F"), knots = knots, ord = ord)

## Also mixed indicator versions are possible
mkIndiList(data, type = c("C","F"), knots = knots, ord = ord)
```

plot.anacor

Plots for anacor solution

Description

These functions produce various plots for objects of class "anacor"

Usage

```
## S3 method for class 'anacor'
plot(x, plot.type = "jointplot", plot.dim = c(1,2), col.row = "cadetblue",
     col.column = "coral1", catlabels = list(label.row = TRUE, label.col = TRUE,
     col.row = "cadetblue", col.column = "coral1", cex = 0.8, pos = 3),
     legpos = "top", arrows = c(FALSE, FALSE), conf = 0.95, wlines = 0, asp = 1, pch = 20,
     xlab, ylab, main, type, xlim, ylim, cex.axis2, ...)
```

Arguments

| | |
|------------|---|
| x | Object of class "anacor" |
| plot.type | Type of plot to be produced (details see below): 2-D and 3-D for "jointplot", "rowplot", and "colplot"; 2-D for "regplot", "graphplot", "benzplot", "transplot", and "orddiag". |
| plot.dim | Vector of length 2 with Dimensions to be plotted. For "regplot" a single value should be provided, for "transplot" more than two dimensions are allowed, and for "benzplot" this argument is ignored. |
| col.row | Color row categories |
| col.column | Color column categories |
| catlabels | Various parameter settings for labels |
| legpos | Position of the legend (for "transplot" only) |
| conf | Ellipsoid confidence level for "jointplot", "rowplot", and "colplot", assuming that the ellipse where computed in anacor(). If NULL, no ellipsoids are drawn. |
| arrows | Whether arrows from the origin to the row scores (first element) or column scores (second element) should be drawn. |

| | |
|-----------|---|
| wlines | For "graphplot" only: If 0, all lines are of the same thickness. For values > 0 line thickness indicates the strength of the pull |
| asp | Aspect ratio. |
| pch | Symbol for plotting points. |
| xlab | Label x-axis. |
| ylab | Label y-axis. |
| xlim | Scale x-axis. |
| ylim | Scale y-axis. |
| main | Plot title. |
| type | Whether points, lines or both should be plotted; for "regplot" and "transplot" only. |
| cex.axis2 | For "regplot" only. The magnification to be used for the category labels in the scaled solution relative to the current setting of cex. |
| ... | Additional graphical parameters. |

Details

The following plot types are provided: "jointplot" plots row and column scores into the same device, "rowplot" and "colplot" plot the row scores and column scores, respectively, in separate devices. For these types of plots 3-dimensional versions are provided. The graph plot is an unlabeled version of the joint plot where the points are connected by lines. Options are provided (wlines) to steer the line thickness indicating the connection strength.

The regression plot ("regplot") provides two plots. First, the unscaled solution is plotted. A frequency grid for the row categories (x-axis) and column categories (y-axis) is produced. The regression line is based on the category weighted means of the relative frequencies: the blue line on the column-wise means on the x-axis and the column category on the y-axis, the red line is based on the row categories on the x-axis and the row-wise means on the y-axis. In a second device the scaled solution is plotted. The frequency grid is determined by the row scores (x-axis) and the column scores (y-axis). Now, instead of the row/column categories, the column scores (black line y-axis) and the row scores (red line x-axis) are used.

The transformation plot ("transplot") plots the row/column categories against the row/column scores. The Benzecri plot ("benzplot") plots the observed distances against the fitted distances. It is assumed that the CA result is Benzecri scaled. The ordination diagram ("orddiag") for CCA produces a joint plot and includes the column and row covariates based on intraset correlations.

Author(s)

Jan de Leeuw, Patrick Mair

References

De Leeuw, J. and Mair, P. (2009). Simple and Canonical Correspondence Analysis Using the R Package anacor. Journal of Statistical Software, 31(5), 1-18. <http://www.jstatsoft.org/v31/i05/>

See Also[anacor](#)**Examples**

```
## symmetric map
data(tocher)
res <- anacor(tocher)
plot(res, conf = NULL, main = "Symmetric Map")

## simple CA on Tocher data, asymmetric coordinates
res <- anacor(tocher, scaling = c("standard", "Benzecri"))
res

## Regression plots using Glass data
data(glass)
res <- anacor(glass)
plot(res, plot.type = "regplot", xlab = "fathers occupation", ylab = "sons occupation")

## Benzecri Plots for bitterling data
data(bitterling)
res1 <- anacor(bitterling, ndim = 2, scaling = c("Benzecri", "Benzecri"))
res2 <- anacor(bitterling, ndim = 5, scaling = c("Benzecri", "Benzecri"))
res2
plot(res1, plot.type = "benzplot", main = "Benzecri Distances (2D)")
plot(res2, plot.type = "benzplot", main = "Benzecri Distances (5D)")

## Column score plot, transformation plot, and ordination diagram for canonical CA
data(maxwell)
res <- anacor(maxwell$table, row.covariates = maxwell$row.covariates,
scaling = c("Goodman", "Goodman"))
res
plot(res, plot.type = "colplot", xlim = c(-1.5,1), conf = NULL)
plot(res, plot.type = "transplot", legpos = "topright")
plot(res, plot.type = "orddiag")
```

sleeping

Sleeping Bags

Description

This data set provides 4 variables measured on 21 sleeping bags. The variables are temperature, weight, price, and material.

Usage

sleeping

Format

A data frame of dimension 21 times 4.

References

Prediger, S. (1997). Symbolic objects in formal concept analysis. In G. Mineau, and A. Fall (eds.), Proceedings of the 2nd International Symposium on Knowledge, Retrieval, Use, and Storage for Efficiency.

Examples

```
data(sleeping)
sleeping
```

spider

Hunting spider data

Description

Abundance of hunting spiders in a Dutch dune area.

Usage

```
data(glass)
```

Format

A list of data frames containing the frequency table (28 observations) and the row covariates.

Table:

Alopacce Abundance of *Alopecosa accentuata*.

Alopcune Abundance of *Alopecosa cuneata*.

Alopfabr Abundance of *Alopecosa fabrilis*.

Arctlute Abundance of *Arctosa lutetiana*.

Arctperi Abundance of *Arctosa perita*.

Auloalbi Abundance of *Aulonia albimana*.

Pardlugu Abundance of *Pardosa lugubris*.

Pardmont Abundance of *Pardosa monticola*.

Pardnigr Abundance of *Pardosa nigriceps*.

Pardpull Abundance of *Pardosa pullata*.

Trocterr Abundance of *Trochosa terricola*.

Zoraspin Abundance of *Zora spinimana*.

Row covariates:

WaterCon Log percentage of soil dry mass.
 BareSand Log percentage cover of bare sand.
 FallTwig Log percentage cover of fallen leaves and twigs.
 CoveMoss Log percentage cover of the moss layer.
 CoveHerb Log percentage cover of the herb layer.
 ReflLux Reflection of the soil surface with cloudless sky.

References

Van der Aart, P.J.M. and Smeek-Enserink, N. (1975). Correlations between distributions of hunting spiders (Lycosidae, Ctenidae) and environmental characteristics in a dune area. *Netherlands Journal of Zoology*, 25, 1–45.

Examples

```
data(spider)
str(spider)
```

srole

Srole Data

Description

This dataset provides a cross-classification of subjects according to their mental health status and parents' socio-economic status.

Usage

```
data(srole)
```

Format

Mental health has four categories (rows): well, mild symptom formation, moderate symptom formation, and impaired. There are six categories of socio-economic status in the columns.

References

Srole, L., Langner, T.S., Michael, S.T., Opler, M.K., & Rennie, T.A.C. (1962). *Mental health in the metropolis: The midtown Manhattan study*. New York: McGraw-Hill.

Examples

```
data(srole)
## maybe str(srole) ; plot(srole) ...
```

`tocher`*Tocher's eye/hair color data.*

Description

Eye color and hair color cross-classification of 5387 Scottish school children.

Usage

```
data(tocher)
```

Format

Frequency table with eye color in the rows (blue, light, medium, dark) and hair color in the columns (fair, red, medium, dark, black).

References

Maung, K. (1941). Discriminant analysis of Tocher's eye colour data for Scottish school children. *Annals of Eugenics*, 11, 64-67.

Examples

```
data(tocher)
## maybe str(tocher) ; plot(tocher) ...
```

Index

*Topic **datasets**

- bitterling, 4
- galton, 7
- glass, 7
- maxwell, 8
- sleeping, 12
- spider, 13
- srole, 14
- tocher, 15

*Topic **hplot**

- plot.anacor, 10

*Topic **models**

- anacor, 2
- burtTable, 5
- expandFrame, 6
- mkIndiList, 9

anacor, 2, 12

bitterling, 4

burtTable, 5, 6, 9

expandFrame, 5, 6, 9

galton, 7

glass, 7

maxwell, 8

mkIndiList, 5, 6, 9

plot.anacor, 3, 10

print.anacor (anacor), 2

screepLOT (plot.anacor), 10

sleeping, 12

spider, 13

srole, 14

tocher, 15