Package ‘anacor’

February 19, 2015

Type Package
Title Simple and Canonical Correspondence Analysis.
Version 1.0-5
Date 2014-09-07
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, scatterplot3d, colorspace, fda,
Depends R (>= 3.0.2), rgl, splines
URL http://r-forge.r-project.org/projects/psychor/
License GPL-2
LazyData yes
LazyLoad yes
ByteCompile yes
NeedsCompilation yes
Repository CRAN
Date/Publication 2014-09-08 06:58:20

R topics documented:

anacor-package ........................................ 2
anacor .................................................. 3
bitting .................................................. 5
burtTable ............................................. 6
expandFrame .......................................... 7
galton .................................................. 8
glass ................................................... 8
maxwell ................................................. 9
Description

This package performs simple and canonical CA 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.

Details

The main function of the package is `anacor()` which performs simple CA. The input dataset \( \text{tab} \) is a (incomplete) frequency table. Various scaling options in terms of standard, centroid, Benzecri, and Goodman scaling are provided. They can be specified using the `scaling` argument. If covariates for row/column scores are provided, canonical CA will be computed. Utility functions to create various types of indicator matrices are implemented.

Author(s)

Jan de Leeuw, Patrick Mair

Maintainer: Jan de Leeuw <deleeuw@stat.ucla.edu>

References


See Also

`anacor`
Examples

data(tocher)
res <- anacor(tocher)
res
summary(res)

anacor Simple and Canonical Correspondence Analysis

Description

This function performs simple and canonical CA for 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 = TRUE, eps = 1e-06)

## S3 method for class 'anacor'
predict(x,...)
## 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","centroid","Benzecri","Goodman".
ellipse If TRUE, confidence ellipses are computed.
eps Convergence criterion for reconstitution algorithm.
x Object of class "anacor" in print.anacor.
obj Object of class "anacor" in summary.anacor.

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. Furthermore, scores can be scaled around their centroid. Goodman scaling is based on Fisher-Maung decomposition.

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

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

Author(s)

Jan de Leeuw, Patrick Mair

References


See Also

plot.anacor
bitterling

Examples

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

Examples
data(bitterling)

burtTable(data)  Creates Burt Matrix

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
**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**

```r
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**

```r
## 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**  
*Galton’s RFF data*

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

**Usage**
```r
data(galton)
```

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

**References**

**Examples**
```r
data(galton)
```

---

**glass**  
*Glass data*

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

**Usage**
```r
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

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

Examples
data(maxwell)
## maybe str(maxwell) ; plot(maxwell) ...
mkIndiList

Converts Data Frame to Indicator Matrix

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

```r
## 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

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**

2-D and 3D plots for anacor solution

**Description**

These functions produce various 2- and 3-dimensional plots for objects of class "anacor"

**Usage**

```r
## S3 method for class 'anacor'
plot(x, plot.type, plot.dim = c(1, 2), legpos = "top", arrows = FALSE, conf = 0.95, wlines = 0, xlab, ylab, main, type, xlim, ylim, cex.axis, ..., )

## S3 method for class 'anacor'
plot3d(x, plot.type, plot.dim = c(1, 2, 3), col.r = "RED", col.c = "BLUE", arrows = TRUE, xlab, ylab, zlab, main, ...)

## S3 method for class 'anacor'
plot3dstatic(x, plot.type, plot.dim = c(1, 2, 3), col.r = "RED", col.c = "BLUE", arrows = TRUE, main, xlab, ylab, zlab, xlim, ylim, zlim, ...)
```

**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.
- **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        If TRUE the points in "jointplot", "rowplot", and "colplot" are connected with the origin.
wlines       For "graphplot" only: If 0, all lines are of the same thickness. For values > 0 line thickness indicates the strength of the pull.
col.r         Color for row score plots
col.c         Color for column score plots
xlab          Label x-axis.
ylab          Label y-axis.
zlab          Label z-axis.
xlim          Scale x-axis.
ylim          Scale y-axis.
zlim          Scale z-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
sleeping

References


See Also

anacor

Examples

## 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


Examples

```r
data(sleeping)
sleeping
```

<table>
<thead>
<tr>
<th>spider</th>
<th>Hunting spider data</th>
</tr>
</thead>
</table>

Description

Abundance of hunting spiders in a Dutch dune area.

Usage

```r
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.
- **Zorасpin**: 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.
- **Ref1Lux**: Reflection of the soil surface with cloudless sky.
References


Examples

data(srole)
str(srole)

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


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**

```r
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**


**Examples**

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

*Topic datasets
  bitterling, 5
  galton, 8
  glass, 8
  maxwell, 9
  sleeping, 13
  spider, 14
  srole, 15
  tocher, 16

*Topic hplot
  plot.anacor, 11

*Topic models
  anacor, 3
  burtTable, 6
  expandFrame, 7
  mkIndilist, 10

*Topic package
  anacor-package, 2

anacor, 2, 3, 13
anacor-package, 2

bitterling, 5
burtTable, 6, 7, 10
expandFrame, 6, 7, 10
galton, 8
glass, 8

maxwell, 9
mkIndilist, 6, 7, 10
plot.anacor, 4, 11
plot3d.anacor (plot.anacor), 11
plot3dstatic (plot.anacor), 11
print.anacor (anacor), 3

sleeping, 13
spider, 14
srole, 15

summary.anacor (anacor), 3
tocher, 16