

Package ‘ToolsForCoDa’

June 18, 2020

Type Package

Title Multivariate Tools for Compositional Data Analysis

Version 1.0.5

Date 2020-06-18

Author Jan Graffelman <jan.graffelman@upc.edu>

Maintainer Jan Graffelman <jan.graffelman@upc.edu>

Depends R (>= 1.8.0), MASS, calibrate, robCompositions

Description Provides functions for multivariate analysis with compositional data. Includes a function for doing compositional canonical correlation analysis. This analysis requires two data matrices of compositions, which can be adequately transformed and used as entries in a specialized program for canonical correlation analysis, that is able to deal with singular covariance matrices. The methodology is described in Graffelman et al. (2017) <doi:10.1101/144584>. A function for log-ratio principal component analysis with condition number computations has been added to the package.

License GPL (>= 2)

URL www.R-project.org, <http://www-eio.upc.edu/~jan>

NeedsCompilation no

Repository CRAN

Date/Publication 2020-06-18 17:30:02 UTC

R topics documented:

| | |
|--------------------------|---|
| Artificial | 2 |
| bentonites | 2 |
| canocov | 3 |
| cen | 5 |
| clrmat | 6 |
| largest.kappas | 6 |
| lrpca | 7 |
| tr | 8 |

| | |
|--------------|----------|
| Index | 9 |
|--------------|----------|

| | |
|------------|--|
| Artificial | <i>Two sets of 3-part compositions</i> |
|------------|--|

Description

The list object `Artificial` contains two data frames of 3-part compositions. The data refer to the example in Section 3.1 of Graffelman et al. (2017)

Usage

```
data(Artificial)
```

Format

A list containing two data frames containing 100 observations.

Source

Laird, N. M. and Lange, C. Table 7.11, p. 124

References

Graffelman, J., Pawlowsky-Glahn, V., Egozcue, J.J. and Buccianti, A. (2017) Compositional Canonical Correlation Analysis.

| | |
|------------|---|
| bentonites | <i>Isotopic and chemical compositions of bentonites</i> |
|------------|---|

Description

The data consists of 14 geological samples from the US with their major oxide composition (SiO₂, Al₂O₃, Fe₂O₃, MnO, MgO, CaO, K₂O, Na₂O and H₂O+) and delta Deuterium and delta-18-Oxygen (dD,d18O).

Usage

```
data("bentonites")
```

Format

A data frame with 14 observations on the following 11 variables.

Si a numeric vector

Al a numeric vector

Fe a numeric vector

Mn a numeric vector

Mg a numeric vector

Ca a numeric vector

K a numeric vector

Na a numeric vector

H2O a numeric vector

dD a numeric vector

d18O a numeric vector

Source

Cadrin, A.A.J (1995), Tables 1 and 2. Reyment, R. A. and Savazzi, E. (1999), pp. 220-222.

References

Cadrin, A.A.J., Kyser, T.K., Caldwell, W.G.E. and Longstaffe, F.J. (1995) Isotopic and chemical compositions of bentonites as paleoenvironmental indicators of the Cretaceous Western Interior Seaway Palaeogeography, Palaeoclimatology, Palaeoecology 119 pp. 301–320.

Reyment, R. A. and Savazzi, E. (1999) Aspects of Multivariate Statistical Analysis in Geology, Elsevier Science B.V., Amsterdam. <https://doi.org/10.1016/B978-044482568-1/50012-4>

Examples

```
data(bentonites)
```

canocov

Canonical correlation analysis.

Description

Function canocov performs a canonical correlation analysis. It operates on raw data matrices, which are only centered in the program. It uses generalized inverses and can deal with structurally singular covariance matrices.

Usage

```
canocov(X, Y)
```

Arguments

| | |
|---|--|
| X | The n times p X matrix of observations |
| Y | The n times q Y matrix of observations |

Details

canocov computes the solution by a singular value decomposition of the transformed between set covariance matrix.

Value

Returns a list with the following results

| | |
|--------|--|
| ccor | the canonical correlations |
| A | canonical weights of the X variables |
| B | canonical weights of the Y variables |
| U | canonical X variates |
| V | canonical Y variates |
| Fs | biplot markers for X variables (standard coordinates) |
| Gs | biplot markers for Y variables (standard coordinates) |
| Fp | biplot markers for X variables (principal coordinates) |
| Gp | biplot markers for Y variables (principal coordinates) |
| Rxu | canonical loadings, (correlations X variables, canonical X variates) |
| Rxv | canonical loadings, (correlations X variables, canonical Y variates) |
| Ryu | canonical loadings, (correlations Y variables, canonical X variates) |
| Ryv | canonical loadings, (correlations Y variables, canonical Y variates) |
| Sxu | covariance X variables, canonical X variates |
| Sxv | covariance X variables, canonical Y variates |
| Syu | covariance Y variables, canonical X variates |
| Syv | covariance Y variables, canonical Y variates |
| fitRxy | goodness of fit of the between-set correlation matrix |
| fitXs | adequacy coefficients of X variables |
| fitXp | redundancy coefficients of X variables |
| fitYs | adequacy coefficients of Y variables |
| fitYp | redundancy coefficients of Y variables |

Author(s)

Jan Graffelman <jan.graffelman@upc.edu>

| | |
|--------|---|
| clrmat | <i>Centred log-ratio transformation</i> |
|--------|---|

Description

Program clrmat calculates the centred log-ratio transformation for a matrix of compositions.

Usage

```
clrmat(X)
```

Arguments

X A matrix of compositions

Value

A matrix containing the transformed data

Author(s)

Jan Graffelman <jan.graffelman@upc.edu>

Examples

```
data(Artificial)
Xsim.com <- Artificial$Xsim.com
Xclr <- clrmat(Xsim.com)
```

| | |
|----------------|--|
| largest.kappas | <i>Calculate condition indices for subcompositions</i> |
|----------------|--|

Description

Function largest.kappas calculates the condition numbers for all subcompositions of a given size, for a particular compositional data set.

Usage

```
largest.kappas(Xcom, nparts = 3, sizetoplist = 10)
```

Arguments

Xcom A data matrix with compositions in rows
nparts The number of parts for the subcompositions to be analysed.
sizetoplist The length of the list of the "best" subcompositions

Details

Log-ratio PCA is executed for each subcomposition, and the resulting eigenvalues and eigenvectors are stored.

Value

A data frame with an ordered list of subcompositions

Author(s)

Jan Graffelman (jan.graffelman@upc.edu)

Examples

```
X <- matrix(runif(600), ncol=6)
Xcom <- X/rowSums(X)
Results <- largest.kappas(Xcom)
```

lrpca

Logratio principal component analysis with condition indices

Description

Function lrpca performs logratio principal component analysis. It returns the variance decomposition, principal components, biplot coordinates and a table with condition indices.

Usage

```
lrpca(Xcom)
```

Arguments

Xcom A matrix with compositions in its rows

Details

Calculations are based on the singular value decomposition of the clr transformed compositions.

Value

| | |
|-----------|--|
| Fp | matrix with principal components |
| Fs | matrix with standardized principal components |
| Gp | biplot markers for parts (principal coordinates) |
| Gs | biplot markers for parts (standard coordinates) |
| La | eigenvalues |
| D | singular values |
| decom | table with variance decomposition |
| kappalist | table with condition indices and eigenvectors |

Author(s)

Jan Graffelman (jan.graffelman@upc.edu)

See Also

[princomp](#),

Examples

```
data(bentonites)
Ben <- bentonites[,1:8]
Ben.com <- Ben/rowSums(Ben)
out.lrpca <- lrpca(Ben.com)
```

| | |
|----|--------------------------------------|
| tr | <i>Compute the trace of a matrix</i> |
|----|--------------------------------------|

Description

tr computes the trace of a matrix.

Usage

```
tr(X)
```

Arguments

X a (square) matrix

Value

the trace (a scalar)

Author(s)

Jan Graffelman (jan.graffelman@upc.edu)

Examples

```
X <- matrix(runif(25),ncol=5)
print(X)
print(tr(X))
```


Index

*Topic **algebra**

tr, 8

*Topic **array**

tr, 8

*Topic **datasets**

Artificial, 2

bentonites, 2

*Topic **multivariate**

canocov, 3

cen, 5

clrmat, 6

largest.kappas, 6

lrpca, 7

Artificial, 2

bentonites, 2

cancor, 5

canocov, 3

cen, 5

clrmat, 6

largest.kappas, 6

lrpca, 7

princomp, 8

tr, 8