

Package ‘robCompositions’

May 8, 2009

Type Package

Title Robust Estimation for Compositional Data.

Version 1.2.2

Date 2009-04-29

Depends utils, robustbase, car, e1071, MASS, compositions

Author Matthias Templ, Karel Hron, Peter Filzmoser

Maintainer Matthias Templ <templ@tuwien.ac.at>

Description The package includes methods for imputation of compositional data including robust methods, (robust) outlier detection for compositional data, (robust) principal component analysis for compositional data, (robust) factor analysis for compositional data and (robust) Anderson-Darling normality tests for compositional data.

License GPL-2

LazyLoad yes

Repository CRAN

Date/Publication 2009-05-08 11:22:02

R topics documented:

| | |
|-------------------------|----|
| robCompositions-package | 2 |
| aDist | 3 |
| adjust | 5 |
| adtest | 6 |
| adtestWrapper | 7 |
| aitchison359 | 9 |
| aitchison360 | 9 |
| aitchison395 | 10 |
| alrEM | 11 |
| ilr | 12 |

| | |
|---------------------------------|----|
| impCoda | 13 |
| impKNNa | 15 |
| invlr | 17 |
| machineOperators | 18 |
| outCoDa | 18 |
| pcaCoDa | 20 |
| pfa | 21 |
| plot.imp | 23 |
| plot.outCoDa | 25 |
| plot.pcaCoDa | 26 |
| print.adtestWrapper | 27 |
| print.imp | 28 |
| print.outCoDa | 29 |
| print.pcaCoDa | 30 |
| robVariation | 31 |
| summary.adtestWrapper | 32 |
| summary.imp | 33 |

| | |
|--------------|-----------|
| Index | 34 |
|--------------|-----------|

robCompositions-package

Robust Estimation for Compositional Data.

Description

This first version of the package includes advanced methods for imputation of compositional data including robust methods. The package will be enhanced with other multivariate methods for compositional data in near future.

Details

| | |
|-----------|-----------------|
| Package: | robCompositions |
| Type: | Package |
| Version: | 1.0 |
| Date: | 2008-12-15 |
| License: | GPL 2 |
| LazyLoad: | yes |

Author(s)

Peter Filzmoser, Karel Hron, Matthias Templ

Maintainer: Matthias Templ <templ@tuwien.ac.at>

References

Aitchison, J. (1986) *The Statistical Analysis of Compositional Data* Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). 416p. \

Hron, K. and Templ, M. and Filzmoser, P. (2008) Imputation of missing values for compositional data using classical and robust methods *Research Report SM-2008-4, Vienna University of Technology*, 15 pages.

See Also

[<compositions>](#)

Examples

```
## k nearest neighbor imputation
data(aitchison395)
aitchison395[1,3]
aitchison395[1,3] <- NA
impKNNa(aitchison395)$xImp[1,3]

## iterative model based imputation
data(aitchison395)
x <- aitchison395
x[1,3]
x[1,3] <- NA
xi <- impCoda(x)$xImp
xi[1,3]
s1 <- sum(x[1,-3])
impS <- sum(xi[1,-3])
xi[,3] * s1/impS

xi <- impKNNa(aitchison395)
xi
summary(xi)
plot(xi, which=1)
plot(xi, which=2)
plot(xi, which=3)
```

aDist

Aitchison distance

Description

Computes the Aitchison distance between two observations or between two data sets.

Usage

```
aDist(x, y)
```

Arguments

`x` a vector, matrix or data.frame
`y` a vector, matrix or data.frame with equal dimension as `x`

Details

This distance measure accounts for the relative scale property of the Aitchison distance. It measures the distance between two compositions if `x` and `y` are vectors and evaluate sum of the distances between `x` and `y` for each row of `x` and `y` if `x` and `y` are matrices or data frames.

It is not designed to apply it on one matrix, such as function `'acomp()'` in package `'compositions'`, but it is designed to compare different matrices.

The underlying code is written in C and allows a fast computation also for large data sets.

Value

The Aitchison distance between two compositions or between two data sets.

Author(s)

Matthias Templ

References

Aitchison, J. (1986) *The Statistical Analysis of Compositional Data* Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). 416p.

See Also

[ilr](#)

Examples

```
data(aitchison395)
x <- xOrig <- aitchison395
## Aitchison distance between the first 2 observations:
aDist(x[,1], x[,2])

## set some missing values:
x[1,3] <- x[3,5] <- x[2,4] <- x[5,3] <- x[8,3] <- NA

## impute them:
xImp <- impCoda(x, method="ltsReg")$xImp

## calculate the relative Aitchison distance between xOrig and xImp:
aDist(xOrig, xImp)
```

`adjust`*Adjusting for original scale*

Description

Results from the model based iterative methods provides the results in another scale (but the ratios are still the same). This function rescale the output to the original scale.

Usage

```
adjust(x)
```

Arguments

`x` object from class 'imp'

Details

It is self-explaining if you try the examples.

Value

The object of class 'imp' but with the adjusted imputed data.

Author(s)

Matthias Templ

References

Hron, K. and Templ, M. and Filzmoser, P. (2008) Imputation of missing values for compositional data using classical and robust methods *Research Report SM-2008-4, Vienna University of Technology*, 15 pages.

See Also

[impCoda](#)

Examples

```
data(aitchison395)
x <- aitchison395
x[1,3] <- x[2,4] <- x[3,3] <- x[3,4] <- NA
xi <- impCoda(x)
x
xi$xImp
adjust(xi)$xImp
```

`adtest`*Anderson-Darling Normality Tests*

Description

This function provides three kinds of Anderson-Darling Normality Tests (Anderson and Darling, 1952).

Usage

```
adtest(x, R = 1000, locscatt = "standard")
```

Arguments

| | |
|-----------------------|---|
| <code>x</code> | either a numeric vector, or a data.frame, or a matrix |
| <code>R</code> | Number of Monte Carlo simulations to obtain p-values |
| <code>locscatt</code> | standard for classical estimates of mean and (co)variance. robust for robust estimates using ‘covMcd()’ from package robustbase |

Details

Three version of the test are implemented (univariate, angle and radius test) and it depends on the data which test is chosen.

If the data is univariate the univariate Anderson-Darling test for normality is applied.

If the data is bivariate the angle Anderson-Darling test for normality is performed out.

If the data is multivariate the radius Anderson-Darling test for normality is used.

If ‘locscatt’ is equal to “robust” then within the procedure, robust estimates of mean and covariance are provided using ‘covMcd()’ from package robustbase.

To provide estimates for the corresponding p-values, i.e. to compute the probability of obtaining a result at least as extreme as the one that was actually observed under the null hypothesis, we use Monte Carlo techniques where we check how often the statistic of the underlying data is more extreme than statistics obtained from simulated normal distributed data with the same (column-wise-) mean(s) and (co)variance.

Value

| | |
|------------------------|---|
| <code>statistic</code> | The result of the corresponding test statistic |
| <code>method</code> | The chosen method (univariate, angle or radius) |
| <code>p.value</code> | p-value |

Note

These functions are use by [adtestWrapper](#).

Author(s)

Karel Hron, Matthias Templ

References

Anderson, T.W. and Darling, D.A. (1952) Asymptotic theory of certain goodness-of-fit criteria based on stochastic processes. *Annals of Mathematical Statistics*, **23** 193-212.

See Also

[adtestWrapper](#)

Examples

```
adtest(rnorm(100))
data(machineOperators)
x <- machineOperators
adtest(ilr(x[,1:2]))
adtest(ilr(x[,1:3]))
adtest(ilr(x))
adtest(ilr(x[,1:2]), locscatt="robust")
```

adtestWrapper

Wrapper for Anderson-Darling tests

Description

A set of Anderson-Darling tests (Anderson and Darling, 1952) are applied as proposed by Aitchison (Aichison, 1986).

Usage

```
adtestWrapper(x, alpha = 0.05, R = 1000, robustEst = FALSE)
```

Arguments

| | |
|-----------|---|
| x | compositional data of class data.frame or matrix |
| alpha | significance level |
| R | Number of Monte Carlo simulations in order to provide p-values. |
| robustEst | logical |

Details

First, the data is transformed using the ‘ilr’-transformation. After applying this transformation

- all (D-1)-dimensional marginal, univariate distributions are tested using the univariate Anderson-Darling Test for normality.
- all 0.5 (D-1)(D-2)-dimensional bivariate angle distributions are tested using the Anderson-Darling angle Test for normality.
- the (D-1)-dimensional radius distribution is tested using the Anderson-Darling radius Test for normality.

Value

| | |
|-------|--|
| res | a list including each test result |
| check | information about the rejection of the Null hypothesis |
| alpha | the underlying significance level |
| info | further information which is used by the print and summary method. |
| est | standard for standard estimations and robust for robust estimation |

Author(s)

Matthias Templ and Karel Hron

References

Anderson, T.W. and Darling, D.A. (1952) *Asymptotic theory of certain goodness-of-fit criteria based on stochastic processes* Annals of Mathematical Statistics, **23** 193-212.

Aitchison, J. (1986) *The Statistical Analysis of Compositional Data* Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). 416p.

See Also

[adtest](#), [ilr](#)

Examples

```
data(machineOperators)
a <- adtestWrapper(machineOperators, R=50) # choose higher value of R
a
summary(a)
```

aitchison359 *Arctic lake sediment data*

Description

Sand, silt, clay compositions of 39 sediment samples at different water depths in an Arctic lake. This data set can be found on page 359 of the Aitchison book (see reference).

Usage

```
data(aitchison359)
```

Format

A data frame with 39 observations on the following 3 variables.

V1 numeric vector of percentages of sand

V2 numeric vector of percentages of silt

V3 numeric vector of percentages of clay

Details

Please note, that the original data set consists of four variables. However, usually the first three are mainly used in applications.

Source

Aitchison, J. (1986) *The Statistical Analysis of Compositional Data* Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). 416p.

Examples

```
data(aitchison359)
```

aitchison360 *Another data set from the Aitchison book*

Description

AFM compositions of 23 aphyric Skye lavas. This data set can be found on page 360 of the Aitchison book (see reference).

Usage

```
data(aitchison360)
```

Format

A data frame with 23 observations on the following 3 variables.

V1 a numeric vector of percentages of Na₂O+K₂O

V2 a numeric vector of percentages of Fe₂O₃

V3 a numeric vector of percentages of MgO

Source

Aitchison, J. (1986) *The Statistical Analysis of Compositional Data* Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). 416p.

Examples

```
data(aitchison360)
```

```
aitchison395      Household expenditures data
```

Description

This data set from Aitchison (1986), p. 395, describes household expenditures (in former Hong Kong dollars) on five commodity groups.

Usage

```
data(aitchison395)
```

Format

A data frame with 20 observations on the following 5 variables.

V1 housing (including fuel and light)

V2 foodstuff

V3 alcohol

V4 tobacco

V5 other goods (including clothing, footwear and durable goods)

Details

This data set contains household expenditures on five commodity groups of 20 single men. The variables represent housing (including fuel and light), foodstuff, alcohol and tobacco, other goods (including clothing, footwear and durable goods) and services (including transport and vehicles). Thus they represent the ratios of the men's income spent on the mentioned expenditures.

Source

Aitchison, J. (1986) *The Statistical Analysis of Compositional Data* Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). 416p.

Examples

```
data(aitchison395)
aitchison395[1,3]
aitchison395[1,3] <- NA
impKNNa(aitchison395)$xImp[1,3]
```

alrEM

alr-EM Imputation for Rounded Zeros

Description

A modified EM alr-algorithm for replacing rounded zeros in compositional data sets.

Usage

```
alrEM(x, pos = ncol(x), dl = rep(0.05, ncol(x) - 1), eps = 1e-04, maxit = 50)
```

Arguments

| | |
|-------|---|
| x | Compositional data |
| pos | Position of the rationing variable for alr transformation |
| dl | Detection limit for each part |
| eps | convergence criteria |
| maxit | maximum number of iterations |

Details

Statistical analysis of compositional data including zeros runs into problems, because log-ratios cannot be applied. Usually, rounded zeros are considered as missing not at random missing values. The algorithm first applies an additive log-ratio transformation to the compositions. Then the rounded zeros are imputed using a modified EM algorithm. Details can be found in Palarea-Albaladej and Martinn-Fernandez (2008).

The major difference between their MATLAB implementation and this R function is that they use the reg.sweep methodology for regression analysis. Nevertheless, we use the standard lm function from R which based on QR-decompositions, which is more accurate than sweep.

Value

| | |
|-------|---|
| xOrig | Original data frame or matrix |
| xImp | Imputed data |
| wind | Index of the missing values in the data |
| iter | Number of iterations |
| eps | eps |

Author(s)

Matthias Templ and Karel Hron

References

Palarea-Albaladej, J. and Mart'in-Fern'andez, J.A. (2008) A modified EM alr-algorithm for replacing rounded zeros in compositional data sets. *Computers and Geosciences*, **34**, 902-917

See Also

[impCoda](#)

Examples

```
require(robCompositions)
data(aitchison359)
x <- aitchison359
## generate rounded zeros artificially:
x[x < 5] <- 0
xia <- alrEM(x, pos=2, dl=rep(5,3))
xia$xImp
```

ilr

Isometric log-ratio transformation

Description

An isometric log-ratio transformation with a special choice of the balances according to Hron et al. (2009).

Usage

```
ilr(x)
```

Arguments

x object of class data frame or matrix with positive entries

Details

This choice of the balances separates all the relative information of the part x_1 from the remaining parts. It is useful for estimating missing values in x_1 by regression of the remaining variables.

Value

The ilr transformed data.

Author(s)

Karel Hron, Matthias Templ

References

Egozcue J.J., V. Pawlowsky-Glahn, G. Mateu-Figueras and C. Barcel'ó-Vidal (2003) Isometric log-ratio transformations for compositional data analysis. *Mathematical Geology*, **35**(3) 279-300. \

Hron, K. and Templ, M. and Filzmoser, P. (2008) Imputation of missing values for compositional data using classical and robust methods *Research Report SM-2008-4, Vienna University of Technology*, 15 pages.

See Also

[invilr](#)

Examples

```
require(MASS)
Sigma <- matrix(c(5.05,4.95,4.95,5.05), ncol=2, byrow=TRUE)
z <- invilr(mvrnorm(100, mu=c(0,2), Sigma=Sigma))
require(compositions)
plot(acomp(z))
plot(invilr(z))
```

impCoda

Imputation of missing values in compositional data

Description

This function offers different methods for the imputation of missing values in compositional data. Missing values are initialized with proper values. Then iterative algorithms try to find better estimations for the former missing values.

Usage

```
impCoda(x, maxit = 10, eps = 0.5, method = "ltsReg", closed = FALSE, init = "KNN",
```

Arguments

| | |
|--------|---|
| x | data frame or matrix |
| maxit | maximum number of iterations |
| eps | convergence criteria |
| method | imputation method |
| closed | imputation of transformed data (using ilr transformation) or in the original space (closed == TRUE) |
| init | method for initializing missing values |
| k | number of nearest neighbors (if init == "KNN") |
| dl | detection limit(s), only important for the imputation of rounded zeros |
| noise | amount of adding random noise to predictors after convergency |

Details

eps: The algorithm is finished as soon as the imputed values stabilize, i.e. until the sum of Aitchison distances from the present and previous iteration changes only marginally (eps).\

method: Several different methods can be chosen, such as 'ItsReg': least trimmed squares regression is used within the iterative procedure. 'lm': least squares regression is used within the iterative procedure. 'classical': principal component analysis is used within the iterative procedure. 'ItsReg2': least trimmed squares regression is used within the iterative procedure. The imputed values are perturbed in the direction of the predictor by values drawn from a normal distribution with mean and standard deviation related to the corresponding residuals and multiplied by `noise`.

method 'roundedZero' is experimental. It imputes rounded zeros within our iterative framework.

Value

| | |
|----------|--|
| xOrig | Original data frame or matrix |
| xImp | Imputed data |
| criteria | Sum of the Aitchison distances from the present and previous iteration |
| iter | Number of iterations |
| maxit | Maximum number of iterations |
| w | Amount of imputed values |
| wind | Index of the missing values in the data |

Author(s)

Matthias Templ, Karel Hron

References

Hron, K. and Templ, M. and Filzmoser, P. (2008) Imputation of missing values for compositional data using classical and robust methods *Research Report SM-2008-4, Vienna University of Technology*, 15 pages.

See Also[impKNNa, ilr](#)**Examples**

```

data(aitchison395)
x <- aitchison395
x[1,3]
x[1,3] <- NA
xi <- impCoda(x)$xImp
xi[1,3]
s1 <- sum(x[1,-3])
impS <- sum(xi[1,-3])
xi[,3] * s1/impS

```

impKNNa

*Imputation of missing values in compositional data using knn methods***Description**

This function offers several knn methods for the imputation of missing values in compositional data.

Usage

```
impKNNa(x, method = "knn", k = 3, metric = "Aitchison", agg = "median", primitive =
```

Arguments

| | |
|-----------|---|
| x | data frame or matrix |
| method | method (at the moment, only “knn” can be used) |
| k | number of nearest neighbors chosen for imputation |
| metric | “Aichison” or “Euclidean” |
| agg | “median” or “mean”, for the aggregation of the nearest neighbors |
| primitive | if TRUE, a more enhanced search for the k -nearest neighbors is obtained (see details) |
| normknn | An adjustment of the imputed values is performed if TRUE |
| das | depricated. if TRUE, the original definition of the Aitchison distance is used to calculate distances between observations. if FALSE, a version using the clr transformation is used. |
| adj | either ‘median’ (default) or ‘sum’ can be chosen for the adjustment of the nearest neighbors, see Hron et al., 2008. |

Details

The `Aitchison` `metric` should be chosen when dealing with compositional data, the `Euclidean` `metric` otherwise. \

If `primitive == FALSE`, a sequential search for the k -nearest neighbors is applied for every missing value where all information corresponding to the non-missing cells plus the information in the variable to be imputed plus some additional information is available. If `primitive == TRUE`, a search of the k -nearest neighbors among observations is applied where in addition to the variable to be imputed any further cells are non-missing. \

If `normknn` is `TRUE` (preferred option) the imputed cells from a nearest neighbor method are adjusted with special adjustment factors (more details can be found online (see the references)).

Value

| | |
|---------------------|---|
| <code>xOrig</code> | Original data frame or matrix |
| <code>xImp</code> | Imputed data |
| <code>w</code> | Amount of imputed values |
| <code>wind</code> | Index of the missing values in the data |
| <code>metric</code> | Metric used |

Author(s)

Matthias Templ

References

Hron, K. and Templ, M. and Filzmoser, P. (2008) Imputation of missing values for compositional data using classical and robust methods *Research Report SM-2008-4, Vienna University of Technology*, 15 pages.

See Also

[impCoda](#)

Examples

```
data(aitchison395)
x <- aitchison395
x[1,3]
x[1,3] <- NA
xi <- impKNNa(x)$xImp
xi[1,3]
```

`invilr`*Inverse isometric log-ratio transformation*

Description

The inverse transformation of 'ilr()'.

Usage

```
invilr(x.ilr)
```

Arguments

`x.ilr` data frame or matrix

Value

The transformed data.

Author(s)

Karel Hron

References

Egozcue J.J., V. Pawlowsky-Glahn, G. Mateu-Figueras and C. Barcel'ó-Vidal (2003) Isometric log-ratio transformations for compositional data analysis. *Mathematical Geology*, **35**(3) 279-300 \

Hron, K. and Templ, M. and Filzmoser, P. (2008) Imputation of missing values for compositional data using classical and robust methods *Research Report SM-2008-4, Vienna University of Technology*, 15 pages.

See Also

~~objects to See Also as [help](#), ~~~

Examples

```
require(MASS)
Sigma <- matrix(c(5.05,4.95,4.95,5.05), ncol=2, byrow=TRUE)
set.seed(123)
z <- mvrnorm(100, mu=c(0,2), Sigma=Sigma)
#require(compositions)
#plot(acomp(z))
plot(invilr(z))
```

machineOperators *Machine Operators data set*

Description

The data set from Aitchison (1986), p. 382, contains compositions of eight-hour shifts of 27 machine operators. The parts represent proportions of shifts in each activity: high-quality production, low-quality production, machine setting and machine repair.

Usage

```
data(machineOperators)
```

Format

A data frame with 27 observations on the following 4 variables.

h_qproduction high-quality production

l_qproduction low-quality production

setting machine settings

repair machine repair

References

Aitchison, J. (1986) *The Statistical Analysis of Compositional Data* Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). 416p.

Examples

```
## maybe str(machineOperators) ; plot(machineOperators) ...
```

outCoDa *Outlier detection for compositional data*

Description

Outlier detection for compositional data using standard and robust statistical methods.

Usage

```
outCoDa(x, alpha = 0.05, method = "robust", h = 1/2)
```

Arguments

| | |
|--------|--|
| x | compositional data |
| alpha | significance level |
| method | either “robust” (default) or “standard” |
| h | the size of the subsets for the robust covariance estimation according the MCD-estimator for which the determinant is minimized (the default is $(n+p+1)/2$). |

Details

The outlier detection procedure is based on (robust) Mahalanobis distances after a isometric logratio transformation of the data. Observations with squared Mahalanobis distance greater equal a certain quantile of the Chi-squared distribution are marked as outliers.

If method “robust” is chosen, the outlier detection is based on the homogeneous majority of the compositional data set. If method “standard” is used, standard measures of location and scatter are applied during the outlier detection procedure.

Value

| | |
|--------------|--|
| mahalDist | resulting Mahalanobis distance |
| limit | (1 - alpha) quantile of the Chi-squared distribution |
| outlierIndex | logical vector indicating outliers and non-outliers |

Note

It is highly recommended to use the robust version of the procedure.

Author(s)

Matthias Templ, Karel Hron

References

- Egozcue J.J., V. Pawlowsky-Glahn, G. Mateu-Figueras and C. Barcel' o-Vidal (2003) Isometric logratio transformations for compositional data analysis. *Mathematical Geology*, **35**(3) 279-300. \
- Filzmoser, P., and Hron, K. (2008) Outlier detection for compositional data using robust methods. *Math. Geosciences*, **40** 233-248.\
- Rousseeuw, P.J., Van Driessen, K. (1999) A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, **41** 212-223.

See Also

~~objects to See Also as [help](#), ~~~

Examples

```
data(aitchison395)
oD <- outCoDa(aitchison395)
oD
```

pcaCoDa

Robust principal component analysis for compositional data

Description

This function performs out a robust principal component analysis for compositional data.

Usage

```
pcaCoDa(x, method = "robust")
```

Arguments

| | |
|--------|---|
| x | compositional data |
| method | either “robust” (default) or ‘standard’ |

Details

The compositional data set is transformed using the ilr transformation. Afterwards, robust principal component analysis is performed. Resulting loadings and scores are back-transformed to the clr space where the compositional biplot can be shown.

Value

| | |
|-------------------|---|
| scores | scores in clr space |
| loadings | loadings in clr space |
| eigenvalues | eigenvalues of the clr covariance matrix |
| method | method |
| princompOutputClr | output of princomp needed in plot.pcaCoDa |

Author(s)

K. Hron, P. Filzmoser, M. Templ

References

Filzmoser, P., Hron, K., Reimann, C. (2009) Principal Component Analysis for Compositional Data with Outliers. *Environmetrics*, accepted for publication.

See Also

[print.pcaCoDa](#), [plot.pcaCoDa](#)

Examples

```
data(aitchison395)
p1 <- pcaCoDa(aitchison395)
p1
plot(p1)
```

pfa

Factor analysis for compositional data

Description

Computes the principal factor analysis of the input data.

Usage

```
pfa(x, factors, data = NULL, covmat = NULL, n.obs = NA,
    subset, na.action, start = NULL,
    scores = c("none", "regression", "Bartlett"),
    rotation = "varimax", maxiter = 5, control = NULL, ...)
```

Arguments

| | |
|-----------|--|
| x | (robustly) scaled input data |
| factors | number of factors |
| data | default value is NULL |
| covmat | (robustly) computed covariance or correlation matrix |
| n.obs | number of observations |
| subset | if a subset is used |
| na.action | what to do with NA values |
| start | starting values |
| scores | which method should be used to calculate the scores |
| rotation | if a rotation should be made |
| maxiter | maximum number of iterations |
| control | default value is NULL |
| ... | arguments for creating a list |

Details

The main difference of this implementation is that uniquenesses are not longer of diagonal form.

Value

| | |
|-----------------|---|
| loadings | A matrix of loadings, one column for each factor. The factors are ordered in decreasing order of sums of squares of loadings. |
| uniqueness | uniqueness |
| correlation | correlation matrix |
| criteria | The results of the optimization: the value of the negativ log-likelihood and information of the iterations used. |
| factors | the factors |
| dof | degrees of freedom |
| method | “principal” |
| n.obs | number of observations if available, or NA |
| call | The matched call. |
| STATISTIC, PVAL | The significance-test statistic and p-value, if they can be computed |

Author(s)

Peter Filzmoser

References

C. Reimann, P. Filzmoser, R.G. Garrett, and R. Dutter (2008): *Statistical Data Analysis Explained. Applied Environmental Statistics with R*. John Wiley and Sons, Chichester, 2008.

Examples

```

data(aitchison395)
x <- aitchison395

# construct orthonormal basis:
V <- matrix(0, nrow=ncol(x), ncol=ncol(x)-1)
for (i in 1:ncol(V)) {
  V[1:i, i] <- 1/i
  V[i+1, i] <- (-1)
  V[, i] <- V[, i]*sqrt(i/(i+1))
}

z <- ilr(x) #ilr transformed data
y <- z

set.seed(200)
require(robustbase)
z.mcd <- covMcd(z)
mean_z <- z.mcd$center
mean_y <- V

var_z <- z.mcd$cov
var_y <- V

```

```

#classical scaling
y.sc <- scale(y,scale=FALSE) #only centering

#robust scaling
#y.rsc <- scale(y,mean_y,scale=FALSE) #only centering

resllogcentr <- pfa(y.sc, factors=1, scores="Bartlett", rotation="varimax")

#reslRlogcentr <- pfa(y.rsc, factors=1, covmat=var_y, scores="Bartlett", rotation="varimax")

```

plot.imp

Plot method for objects of class imp

Description

This function provides several diagnostic plots for the imputed data set in order to see how the imputed values are distributed in comparison with the original data values.

Usage

```

plot.imp(x, ..., which = 1, ord = 1:ncol(x), colcomb = "missnonmiss", plotvars = NU
  col = c("skyblue", "red"), alpha = NULL, lty = par("lty"), xaxt = "s", xaxlabels
  las = 3, interactive = TRUE, pch = c(1, 3), smooth = FALSE, reg.line = FALSE,
  legend.plot = FALSE, ask = prod(par("mfcol")) < length(which) && dev.interactive
  center = FALSE, scale = FALSE, id = FALSE, seg.l = 0.02, seg1 = TRUE)

```

Arguments

| | |
|----------|---|
| x | object of class ‘imp’ |
| ... | other parameters to be passed through to plotting functions. |
| which | if a subset of the plots is required, specify a subset of the numbers 1:3. |
| ord | determines the ordering of the variables |
| colcomb | if colcomb=“missnonmiss”, observations with missings in any variable are highlighted. Otherwise, observations with missings in any of the variables specified by colcomb are highlighted in the parallel coordinate plot. |
| plotvars | Parameter for the parallel coordinate plot. A vector giving the variables to be plotted. If NULL (the default), all variables are plotted. |
| col | a vector of length two giving the colors to be used in the plot. The second color will be used for highlighting. |
| alpha | a numeric value between 0 and 1 giving the level of transparency of the colors, or NULL. This can be used to prevent overplotting. |

| | |
|-------------|--|
| lty | a vector of length two giving the line types. The second line type will be used for the highlighted observations. If a single value is supplied, it will be used for both non-highlighted and highlighted observations. |
| xaxt | the x-axis type (see par). |
| xaxlabels | a character vector containing the labels for the x-axis. If NULL, the column names of x will be used. |
| las | the style of axis labels (see par). |
| interactive | a logical indicating whether the variables to be used for highlighting can be selected interactively (see 'Details'). |
| pch | a vector of length two giving the symbol of the plotting points. The symbol will be used for the highlighted observations. If a single value is supplied, it will be used for both non-highlighted and highlighted observations. |
| smooth | if TRUE a lowess smooth is plotted in each off-diagonal panel of the multiple scatterplot. Further detail can be found in package <i>car</i> . |
| reg.line | if not FALSE a line is plotted using the function given by this argument; e.g., using <i>rlm</i> in package <i>MASS</i> plots a robust-regression line within the multiple scatterplot. |
| legend.plot | if TRUE then a legend for the groups is plotted in the bottom-right cell of the multiple scatterplot. |
| ask | logical; if TRUE, the user is asked before each plot, see par(ask=) . |
| center | logical, indicates if the data should be centered prior plotting the ternary plot. |
| scale | logical, indicates if the data should be centered prior plotting the ternary plot. |
| id | reads the position of the graphics pointer when the (first) mouse button is pressed and returns the corresponding index of the observation. (only used by the ternary plot) |
| seg.l | length of the plotting symbol (spikes) for the ternary plot. |
| seg1 | if TRUE, the spikes of the plotting symbol are justified. |

Details

The first plot (which == 1) is a multiple scatterplot where for the imputed values another plot symbol and color is used in order to highlight them.

Plot 2 is a parallel coordinate plot in which imputed values in certain variables are highlighted. In parallel coordinate plots, the variables are represented by parallel axes. Each observation of the scaled data is shown as a line. If *interactive* is TRUE, the variables to be used for highlighting can be selected interactively. Observations which includes imputed values in any of the selected variables will be highlighted. A variable can be added to the selection by clicking on a coordinate axis. If a variable is already selected, clicking on its coordinate axis will remove it from the selection. Clicking anywhere outside the plot region quits the interactive session.

Plot 3 shows a ternary diagram in which imputed values are highlighted, i.e. those spikes of the chosen plotting symbol are colored in red for which of the values are missing in the unimputed data set.

Value

None (invisible NULL).

Author(s)

Matthias Templ

References

Aitchison, J. (1986) *The Statistical Analysis of Compositional Data* Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). 416p.

Wegman, E. J. (1990) *Hyperdimensional data analysis using parallel coordinates* Journal of the American Statistical Association 85, 664–675.

See Also

[impCoda](#), [impKNNa](#), [plot.acomp](#), [scatterplot.matrix](#)

Examples

```
data(aitchison395)
aitchison395[1,3]
aitchison395[1,3] <- NA
xi <- impKNNa(aitchison395)
xi
summary(xi)
plot(xi, which=1)
plot(xi, which=2)
plot(xi, which=3)
plot(xi, which=3, seg1=FALSE)
```

plot.outCoDa

plot method for outCoDa objects

Description

Plots the Mahalanobis distance.

Usage

```
plot.outCoDa(x, y, ...)
```

Arguments

| | |
|-----|-----------------------------|
| x | object from class 'outCoDa' |
| y | ... |
| ... | ... |

Details

The dashed line indicates the $(1 - \alpha)$ quantile of the Chi-squared distribution. Observations with Mahalanobis distance greater than this quantile could be considered as compositional outliers.

Author(s)

Matthias Templ

References

Filzmoser, P., and Hron, K. (2008) Outlier detection for compositional data using robust methods. *Math. Geosciences*, **40** 233-248.

See Also

[outCoDa](#)

Examples

```
data(aitchison395)
oD <- outCoDa(aitchison395)
oD
```

plot.pcaCoDa *Plot method*

Description

Provides robust compositional biplots.

Usage

```
plot.pcaCoDa(x, y, ...)
```

Arguments

| | |
|-----|---------------------------|
| x | object of class 'pcaCoDa' |
| y | ... |
| ... | ... |

Details

The robust compositional biplot according to Aitchison and Greenacre (2002), computed from resulting (robust) loadings and scores, is performed.

Value

~Describe the value returned If it is a LIST, use

comp1 Description of 'comp1'

comp2 Description of 'comp2'

...

Author(s)

M. Templ, K. Hron

References

Aitchison, J. and Greenacre, M. (2002). Biplots of compositional data. *Applied Statistics*, **51**, 375-392. \

Filzmoser, P., Hron, K., Reimann, C. (2009) Principal Component Analysis for Compositional Data with Outliers. *Environmetrics*, accepted for publication.

See Also

[pcaCoDa](#)

Examples

```
data(aitchison395)
p1 <- pcaCoDa(aitchison395)
p1
plot(p1)
```

```
print.adtestWrapper
```

print method for objects of class adtestWrapper

Description

Provides a short print output as shown in the examples.

Usage

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

Arguments

x object of class 'adtestWrapper'

... ...

Details

Have a look at the example, it's self-explaining.

Value

nothing

Author(s)

Matthias Templ and Karel Hron

See Also

[adtestWrapper](#), [summary.adtestWrapper](#)

Examples

```
data(machineOperators)
a <- adtestWrapper(machineOperators, R=50) # choose higher value of R
a
summary(a)
```

print.imp

Print method for objects of class imp

Description

The function returns a few information about how many missings are imputed and possible other information about the amount of iterations, for example.

Usage

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

Arguments

x an object of class 'imp'
...

Value

None (invisible NULL).

Author(s)

Matthias Templ

See Also

[impCoda](#), [impKNNa](#)

Examples

```
data(aitchison395)
aitchison395[1,3]
aitchison395[1,3] <- NA
xi <- impCoda(aitchison395)
xi
summary(xi)
plot(xi, which=1:2)
```

print.outCoDa *print method for outCoDa objects*

Description

Gives a short information of the amount of outliers in objects of class 'outCoDa'.

Usage

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

Arguments

| | |
|-----|---------------------------|
| x | object of class 'ourCoDa' |
| ... | ... |

Author(s)

Matthias Templ, Karel Hron

See Also

[outCoDa](#)

Examples

```
data(aitchison395)
oD <- outCoDa(aitchison395)
oD
```

| | |
|---------------|---|
| print.pcaCoDa | <i>Print method for pcaCoDa objects</i> |
|---------------|---|

Description

Print method for objects of class 'pcaCoDa'.

Usage

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

Arguments

| | |
|-----|---------------------------|
| x | object of class 'pcaCoDa' |
| ... | ... |

Value

Prints the (cummulative) percentages of explained variability for clr transformed data by principal component analysis.

Author(s)

M. Templ, K. Hron

See Also

[pcaCoDa](#), [plot.pcaCoDa](#)

Examples

```
data(aitchison395)  
p1 <- pcaCoDa(aitchison395)  
p1  
plot(p1)
```

| | |
|--------------|--------------------------------|
| robVariation | <i>Robust variation matrix</i> |
|--------------|--------------------------------|

Description

Estimates the variation matrix with robust methods.

Usage

```
robVariation(x)
```

Arguments

x data frame or matrix with positive entries

Details

The variation matrix is estimated for a given compositional data set. Instead of using the classical standard deviations the [MAD](#) is used.

Value

The robust variation matrix.

Author(s)

Matthias Templ

References

Aitchison, J. (1986) *The Statistical Analysis of Compositional Data* Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). 416p.

See Also

[variation](#)

Examples

```
data(aitchison395)
robVariation(aitchison395)
# require(compositions)
# variation(aitchison395)
```

```
summary.adtestWrapper
```

summary method for objects of class adtestWrapper

Description

Provides a summary as shown in the examples.

Usage

```
## S3 method for class 'adtestWrapper':  
summary(object, ...)
```

Arguments

| | |
|--------|-------------------------------------|
| object | object of class 'adtestWrapper' |
| ... | additional arguments passed through |

Details

A similar output is proposed by (Pawlowsky-Glahn, et al. (2008)). In addition to that, p-values are provided.

Value

a data frame including an information about the ilr-variables used (first column), the underlying test (second column), the test statistics (third column), the corresponding estimated p-values (fourth column) and an information about the rejection of the Null hypothesis (last column).

Author(s)

Matthias Templ and Karel Hron

References

Pawlowsky-Glahn, V., Egozcue, J.J. and Tolosana-Delgado, R. (2008), *Lecture Notes on Compositional Data Analysis* Universitat de Girona, <http://diobma.udg.edu:8080/dspace/bitstream/10256.1/297/1/CoDa-book.pdf>

See Also

[adtestWrapper](#)

Examples

```
data(machineOperators)  
a <- adtestWrapper(machineOperators, R=50) # choose higher value of R  
a  
summary(a)
```

`summary.imp`*Summary method for objects of class imp*

Description

A short comparison of the original data and the imputed data is given.

Usage

```
## S3 method for class 'imp':  
summary(object, ...)
```

Arguments

```
object      an object of class 'imp'  
...         ...
```

Details

Note that this function will be enhanced with more sophisticated methods in future versions of the package. It is very rudimental in its present form.

Value

None (invisible NULL).

Author(s)

Matthias Templ

See Also

[impCoda](#), [impKNNa](#)

Examples

```
data(aitchison395)  
aitchison395[1,3]  
aitchison395[1,3] <- NA  
xi <- impKNNa(aitchison395)  
xi  
summary(xi)  
plot(xi, which=1:2)
```

Index

- *Topic **aplot**
 - plot.imp, 23
 - plot.pcaCoDa, 26
- *Topic **arith**
 - aDist, 4
- *Topic **datasets**
 - aitchison359, 9
 - aitchison360, 9
 - aitchison395, 10
 - machineOperators, 18
- *Topic **hplot**
 - plot.imp, 23
 - plot.outCoDa, 25
- *Topic **htest**
 - adtest, 6
 - adtestWrapper, 7
- *Topic **iteration**
 - impCoda, 13
- *Topic **manip**
 - adjust, 5
 - impKNNa, 15
- *Topic **math**
 - aDist, 4
 - ilr, 12
 - invilr, 17
- *Topic **multivariate**
 - alrEM, 11
 - impCoda, 13
 - impKNNa, 15
 - outCoDa, 18
 - pcaCoDa, 20
 - pfa, 21
 - robVariation, 31
- *Topic **package**
 - robCompositions-package, 2
- *Topic **print**
 - print.adtestWrapper, 27
 - print.imp, 28
 - print.outCoDa, 29
 - print.pcaCoDa, 30
 - summary.adtestWrapper, 32
 - summary.imp, 33
- *Topic **robust**
 - impCoda, 13
 - robVariation, 31
 - <compositions>, 3
- aDist, 4
- adjust, 5
- adtest, 6, 8
- adtestWrapper, 7, 7, 28, 32
- aitchison359, 9
- aitchison360, 9
- aitchison395, 10
- alrEM, 11
- help, 17, 19
- ilr, 4, 8, 12, 15
- impCoda, 6, 12, 13, 16, 25, 29, 33
- impKNNa, 15, 15, 25, 29, 33
- invilr, 13, 17
- machineOperators, 18
- MAD, 31
- outCoDa, 18, 26, 29
- par, 24
- pcaCoDa, 20, 27, 30
- pfa, 21
- plot.acomp, 25
- plot.imp, 23
- plot.outCoDa, 25
- plot.pcaCoDa, 21, 26, 30
- print.adtestWrapper, 27
- print.imp, 28
- print.outCoDa, 29
- print.pcaCoDa, 21, 30

robCompositions
 (*robCompositions-package*),
 2
robCompositions-package, 2
robVariation, 31

scatterplot.matrix, 25
summary.adtestWrapper, 28, 32
summary.imp, 33

variation, 31