

Package ‘biotools’

July 2, 2014

Type Package

Title Tools for Biometry and Applied Statistics in Agricultural Science

Version 1.2

Date 2014-05-25

LazyLoad yes

LazyData yes

Author Anderson Rodrigo da Silva

Maintainer Anderson Rodrigo da Silva <anderson.agro@hotmail.com>

Depends R (>= 2.15), utils, rpanel, tkrplot, MASS, boot

Suggests soilphysics

Description This package contains tools designed to perform and work with cluster analysis, discriminant analysis and path analysis (standard and with collinearity), as well as some useful miscellaneous tools for dealing with sample size and optimum plot size calculations.

License GPL (>= 2)

NeedsCompilation no

Repository CRAN

Date/Publication 2014-05-26 01:09:46

R topics documented:

biotools-package	2
aer	3
boxM	4
confusionmatrix	5
coph.tocher	6
creategroups	7

D2.disc	8
D2.dist	9
distClust	10
findSubsample	11
fitplotsize	12
garlicdist	14
multcor.test	14
optimumplotsize	16
pathanalysis	17
peppercorr	18
raise.matrix	19
samplesize	19
singh	21
tocher	22

Index	24
--------------	-----------

biotools-package	<i>Tools for Biometry and Applied Statistics in Agricultural Science</i>
------------------	--

Description

This package contains tools designed to perform and work with cluster analysis, discriminant analysis and path analysis (standard and with collinearity), as well as some useful miscellaneous tools for dealing with sample size and optimum plot size calculations.

Details

Package: biotools
 Type: Package
 Version: 1.2
 Date: 2014-05-25
 License: GPL (>= 2)

Note

biotools is an ongoing project. Any and all criticism, comments and suggestions are welcomed.

Author(s)

Anderson Rodrigo da Silva

Maintainer: Anderson Rodrigo da Silva <anderson.agro@hotmail.com>

References

- Carvalho, S.P. (1995) *Metodos alternativos de estimacao de coeficientes de trilha e indices de selecao, sob multicolinearidade*. Ph.D. Thesis, Federal University of Vicosa (UFV), Vicosa, MG, Brazil.
- Cruz, C.D.; Ferreira, F.M.; Pessoni, L.A. (2011) *Biometria aplicada ao estudo da diversidade genetica*. Visconde do Rio Branco: Suprema.
- Lessman, K. J. & Atkins, R. E. (1963). Optimum plot size and relative efficiency of lattice designs for grain sorghum yield tests. *Crop Sci.*, 3:477-481.
- Mahalanobis, P. C. (1936) On the generalized distance in statistics. *Proceedings of The National Institute of Sciences of India*, 12:49-55.
- Manly, B.F.J. (2004) *Multivariate statistical methods: a primer*. CRC Press.
- Meier, V. D. & Lessman, K. J. (1971) Estimation of optimum field plot shape and size for testing yield in *Crambe abyssinica* Hochst. *Crop Sci.*, 11:648-650.
- Morrison, D.F. (1976) *Multivariate Statistical Methods*.
- Rao, R.C. *Advanced statistical methods in biometric research*. New York: John Wiley & Sons, 1952.
- Sharma, J.R. (2006) *Statistical and biometrical techniques in plant breeding*. Delhi: New Age International.
- Silva, A.R. & Dias, C.T.S. (2013) A cophenetic correlation coefficient for Tocher's method. *Pesquisa Agropecuaria Brasileira*, 48:589-596.
- Silva et al. (2013) Path analysis in multicollinearity for fruit traits of pepper. *Idesia*, 31:55-60.
- Singh, D. (1981) The relative importance of characters affecting genetic divergence. *Indian Journal Genetics & Plant Breeding*, 41:237-245.

aer

Apparent Error Rate

Description

A function to calculate the apparent error rate of two classification vectors, i.e., the proportion of observed cases incorrectly predicted. It can be useful for evaluating discriminant analysis or other classification systems.

$$aer = \frac{1}{n} \sum_{i=1}^n I(y_i \neq \hat{y}_i)$$

Usage

```
aer(obs, predict)
```

Arguments

obs a vector containing the observed classes.

predict a vector with the same length of obs containing the predicted classes.

Value

The apparent error rate, a number between 0 (no agreement) and 1 (thorough agreement).

Author(s)

Anderson Rodrigo da Silva <anderson.agro@hotmail.com>

See Also

[confusionmatrix](#), [lda](#)

Examples

```
data(iris)
da <- lda(Species ~ ., data = iris)
pred <- predict(da, dimen = 1)
aer(iris$Species, pred$class)

# End (not run)
```

boxM

Box's M-test

Description

It performs the Box's M-test for homogeneity of covariance matrices obtained from multivariate normal data according to one classification factor. The test is based on the chi-square approximation.

Usage

```
boxM(data, grouping)
```

Arguments

data	a numeric data.frame or matrix containing n observations of p variables; it is expected that $n > p$.
grouping	a vector of length n containing the class of each observation; it is usually a factor.

Value

A list with class "htest" containing the following components:

statistic	an approximated value of the chi-square distribution.
parameter	the degrees of freedom related of the test statistic in this case that it follows a Chi-square distribution.
p.value	the p-value of the test.
cov	a list containing the within covariance matrix for each level of grouping.

pooled	the pooled covariance matrix.
logDet	a vector containing the natural logarithm of each matrix in cov.
data.name	a character string giving the names of the data.
method	the character string "Box's M-test for Homogeneity of Covariance Matrices".

Author(s)

Anderson Rodrigo da Silva <anderson.agro@hotmail.com>

References

Morrison, D.F. (1976) *Multivariate Statistical Methods*.

Examples

```
data(iris)
boxM(iris[, -5], iris[, 5])

# End (not run)
```

confusionmatrix	<i>Confusion Matrix</i>
-----------------	-------------------------

Description

A function to compute the confusion matrix of two classification vectors. It can be useful for evaluating discriminant analysis or other classification systems.

Usage

```
confusionmatrix(obs, predict)
```

Arguments

obs	a vector containing the observed classes.
predict	a vector with the same length of obs containing the predicted classes.

Value

A square matrix containing the number of objects in each class, observed (rows) and predicted (columns). Diagonal elements refers to agreement of obs and predict.

Author(s)

Anderson Rodrigo da Silva <anderson.agro@hotmail.com>

See Also[aer](#), [lda](#)**Examples**

```
data(iris)
da <- lda(Species ~ ., data = iris)
pred <- predict(da, dimen = 1)
confusionmatrix(iris$Species, pred$class)

# End (not run)
```

coph.tocher

Cophenetic Distances for Tocher's Clustering

Description

It computes the cophenetic distance matrix for a Tocher's clustering.

Usage

```
coph.tocher(obj)
```

Arguments

obj an object of class [tocher](#).

Value

An object of class "dist".

Author(s)

Anderson Rodrigo da Silva <anderson.agro@hotmail.com>

References

Silva, A.R. & Dias, C.T.S. (2013) A cophenetic correlation coefficient for Tocher's method. *Pesquisa Agropecuaria Brasileira*, 48:589-596.

See Also

[tocher](#)

Examples

```
# Example 1
data(garlicdist)
garlic <- tocher(garlicdist)
coph <- coph.tocher(garlic)
cor(coph, garlicdist)

# Example 2
usa <- tocher(dist(USArrests))
cophUS <- coph.tocher(usa)
cor(cophUS, dist(USArrests))

# End (not run)
```

creategroups

Creating Homogeneous Groups

Description

A function to create homogeneous groups of named objects according to an objective function evaluated at a covariate. It can be useful to design experiments which contain a fixed covariate factor.

Usage

```
creategroups(x, ngroups, sizes, fun = mean, tol = 0.01, maxit = 200)
```

Arguments

x	a numeric vector of a covariate at which to evaluate the objective function.
ngroups	the number of groups to create.
sizes	a numeric vector of length equal to ngroups containing the group sizes.
fun	the objective function, i.e., to create groups with similar fun; default is mean.
tol	the tolerance level to define the groups as homogenous; see details.
maxit	the maximum number of iterations; default is 200.

Details

creategroups uses a tol value to evaluate the following statistic: $h = \sum_j^{ngroups} abs(t_{j+1} - t_j) / ngroups$, where $t_j = fun(group_j)$. If $h \leq tol$, the groups are considered homogeneous.

Value

A list of

covar	a character indicating the name of the covariate.
func	a character indicating the name of the objective function.
val.func	a numeric vector containing the values evaluated by func on each group.
niter	the number of iteration require to achieve convergence.
labels	a list containing the labels of the objects in each group.
groups	a list of named vectors containing the values for the groups

Author(s)

Anderson Rodrigo da Silva <anderson.agro@hotmail.com>

Examples

```
x <- rnorm(10, 1, 0.5)
names(x) <- letters[1:10]
creategroups(x, ngroups = 2, sizes = c(5, 5))
creategroups(x, ngroups = 3, sizes = c(3, 4, 3), tol = 0.05)

# End (not run)
```

D2.disc

Discriminant Analysis Based on Mahalanobis Distance

Description

A function to perform discriminant analysis based on the squared generalized Mahalanobis distance (D2) of the observations to the center of the groups.

Usage

```
D2.disc(data, grouping, pooled.cov = NULL)
```

Arguments

data	a numeric data.frame or matrix ($n \times p$).
grouping	a vector of length n containing the class of each observation (row) in data.
pooled.cov	a grouping-pooled covariance matrix ($p \times p$). If NULL (default), D2.disc will automatically compute a pooled covariance matrix.

Value

A list of

D2 a matrix containing the Mahalanobis distances between each row of data and the center of each class of grouping. In addition, the original and the predicted (lowest distance) class are displayed, as well as a character vector indicating where the misclassification has occurred.

means a matrix containing the vector of means of each class in grouping.

pooled the pooled covariance matrix.

confusion.matrix an object of class `confusionmatrix`.

Author(s)

Anderson Rodrigo da Silva <anderson.agro@hotmail.com>

References

Manly, B.F.J. (2004) *Multivariate statistical methods: a primer*. CRC Press. (p. 105-106).

Mahalanobis, P.C. (1936) On the generalized distance in statistics. *Proceedings of The National Institute of Sciences of India*, 12:49-55.

See Also

[D2.dist](#), [confusionmatrix](#), [lda](#)

Examples

```
data(iris)
D2.disc(iris[, -5], iris[, 5])

# End (not run)
```

D2.dist

Pairwise Squared Generalized Mahalanobis Distances

Description

Function to calculate the squared generalized Mahalanobis distance between all pairs of rows in a data frame with respect to a covariance matrix. The element of the i -th row and j -th column of the distance matrix is defined as

$$D_{ij}^2 = (\mathbf{x}_i - \mathbf{x}_j)' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \mathbf{x}_j)$$

Usage

```
D2.dist(data, cov, inverted = FALSE)
```

Arguments

data a data frame or matrix of data ($n \times p$).
 cov a variance-covariance matrix ($p \times p$).
 inverted logical. If FALSE (default), cov is supposed to be a variance-covariance matrix.

Value

An object of class "dist".

Author(s)

Anderson Rodrigo da Silva <anderson.agro@hotmail.com>

References

Mahalanobis, P. C. (1936) On the generalized distance in statistics. *Proceedings of The National Institute of Sciences of India*, 12:49-55.

See Also

[dist](#), [singh](#)

Examples

```
# Manly (2004, p.65-66)
x1 <- c(131.37, 132.37, 134.47, 135.50, 136.17)
x2 <- c(133.60, 132.70, 133.80, 132.30, 130.33)
x3 <- c(99.17, 99.07, 96.03, 94.53, 93.50)
x4 <- c(50.53, 50.23, 50.57, 51.97, 51.37)
x <- cbind(x1, x2, x3, x4)
Cov <- matrix(c(21.112,0.038,0.078,2.01, 0.038,23.486,5.2,2.844,
0.078,5.2,24.18,1.134, 2.01,2.844,1.134,10.154), 4, 4)
D2.dist(x, Cov)

# End (not run)
```

distClust

Cluster Distance Matrix

Description

Function to compute a matrix of average distances within and between clusters.

Usage

```
distClust(d, nobj.cluster, id.cluster)
```

Arguments

d	an object of class "dist" containing the distances between objects.
nobj.cluster	a numeric vector containing the numbers of objects per cluster.
id.cluster	a numeric vector for identification of the objects per cluster.

Value

A squared matrix containing distances within (diagonal) and between (off-diagonal) clusters.

Author(s)

Anderson Rodrigo da Silva <anderson.agro@hotmail.com>

See Also

[tocher](#), [dist](#)

findSubsample

Finding an Optimized Subsample

Description

It allows one to find an optimized (minimized or maximized) numeric subsample according to a statistic of interest. For example, it might be of interest to determine a subsample whose standard deviation is the lowest among all of those obtained from all possible subsamples of the same size.

Usage

```
findSubsample(x, size, fun = sd, minimize = TRUE, niter = 10000)
```

Arguments

x	a numeric vector.
size	an integer; the size of the subsample.
fun	an object of class function; the statistic at which to evaluate the subsample.
minimize	logical; if TRUE (default) findSubsample will find a subsample that minimizes stat.
niter	an integer indicating the number of iterations, i.e., the number of subsamples to be selected (without replacement) from the original sample, x. The larger is this number, the more optimized is the subsample to be found, but this also implies in time-consuming.

Value

A list of

dataname	a character.
niter	the number of iterations.
fun	the objective function.
stat	the achieved statistic for the optimized subsample.
criterion	a character indicating the type of optimization.
subsample	a numeric vector; the optimized subsample.
labels	a string containg the labels of the subsample values.

Author(s)

Anderson Rodrigo da Silva <anderson.agro@hotmail.com>

See Also

[sample](#), [creategroups](#)

Examples

```
# Example 1
y <- rnorm(40, 5, 2)
findSubsample(x = y, size = 6)

# Example 2
f <- function(x) diff(range(x)) # max(x) - min(x)
findSubsample(x = y, size = 6, fun = f, minimize = FALSE, niter = 20000)

# End (not run)
```

fitplotsize

Parameter Estimation of the Plot Size Model

Description

Function to estimate the parameters of the nonlinear Lessman & Atkins (1963) model for determining the optimum plot size as a function of the experimental coefficient of variation (CV) or as a function of the residual standard error.

$$CV = a * plotsize^{-b}.$$

It creates initial estimates of the parameters a and b by log-linearization and uses them to provide its least-squares estimates via [nls](#).

Usage

```
fitplotsize(plotsize, CV)
```

Arguments

plotsize	a numeric vector containing estimates of plot size.
CV	a numeric vector of experimental coefficient of variation or residual standard error.

Value

A [nls](#) output.

Side Effects

A summary table ([summary.nls](#)), if convergence is achieved.

Author(s)

Anderson Rodrigo da Silva <anderson.agro@hotmail.com>

References

Lessman, K. J. & Atkins, R. E. (1963) Optimum plot size and relative efficiency of lattice designs for grain sorghum yield tests. *Crop Sci.*, 3:477-481.

See Also

[optimumplotsize](#)

Examples

```
ps <- c(1, 2, 3, 4, 6, 8, 12)
cv <- c(35.6, 29, 27.1, 25.6, 24.4, 23.3, 21.6)
out <- fitplotsize(plotsize = ps, CV = cv)
predict(out) # fitted.values
plot(cv ~ ps)
curve(coef(out)[1] * x^(-coef(out)[2]), add = TRUE)

# End (not run)
```

`garlicdist`*Distances Between Garlic Cultivars*

Description

The data give the squared generalized Mahalanobis distances between 17 garlic cultivars. The data are taken from the article published by Silva & Dias (2013).

Usage

```
data(garlicdist)
```

Format

An object of class "dist" based on 17 objects.

Source

Silva, A.R. & Dias, C.T.S. (2013) A cophenetic correlation coefficient for Tocher's method. *Pesquisa Agropecuaria Brasileira*, 48:589-596.

Examples

```
data(garlicdist)
tocher(garlicdist)

# End (not run)
```

`multcor.test`*Pairwise Correlation t-Test*

Description

It performs multiple correlation t-tests from a correlation matrix based on the statistic:

$$t = r * \sqrt{df / (1 - r^2)}$$

where, in general, $df = n - 2$.

Usage

```
multcor.test(x, n = NULL, Df = NULL,
alternative = c("two.sided", "less", "greater"), adjust = "none")
```

Arguments

x	a correlation matrix.
n	the number of observations; if NULL (default), the argument Df must be passed.
Df	the number of degrees of freedom of the t statistic; if NULL (default), the argument n must be passed and, in this case, multcor.test considers $Df = n - 2$.
alternative	the alternative hypothesis. It must be one of "two.sided", "greater" or "less". You can specify just the initial letter. "greater" corresponds to positive association, "less" to negative association. The default is "two.sided".
adjust	The adjustment method for multiple tests. It must be one of "holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none" (default). For more information, see p.adjust .

Value

A list with class "multcor.test" containing the following components:

t.values	the t-value calculated for each correlation.
p.values	the p.value for each t-test, adjusted for multiple tests.
p.check	a matrix containing the p.values for each t-test (lower triangular) and a symbol indicating the significance level at which one can to reject the null hypothesis (upper triangular).
adjustemnt	a character indicating the p-value adjustment method.
df	the degrees of freedom of the tests.
alternative	a character indicating the type of alternative hypothesis.
data.name	a character string giving the name of the data.

Author(s)

Anderson Rodrigo da Silva <anderson.agro@hotmail.com>

See Also

[cor](#), [cor.test](#), [p.adjust](#)

Examples

```
data(peppercorr)
multcor.test(peppercorr, n = 20)

# End (not run)
```

optimumplotsize	<i>Maximum Curvature Point for Optimum Plot Size</i>
-----------------	--

Description

The Meier & Lessman (1971) method to determine the maximum curvature point for optimum plot size as a function of the experimental coefficient of variation.

Usage

```
optimumplotsize(a, b)
```

Arguments

a	a parameter estimate of the plot size model; see fitplotsize .
b	a parameter estimate of the plot size model; see fitplotsize .

Value

The (approximated) optimum plot size value.

Author(s)

Anderson Rodrigo da Silva <anderson.agro@hotmail.com>

References

Meier, V. D. & Lessman, K. J. (1971) Estimation of optimum field plot shape and size for testing yield in *Crambe abyssinica* Hochst. *Crop Sci.*, 11:648-650.

See Also

[fitplotsize](#)

Examples

```
ps <- c(1, 2, 3, 4, 6, 8, 12)
cv <- c(35.6, 29, 27.1, 25.6, 24.4, 23.3, 21.6)
out <- fitplotsize(plotsize = ps, CV = cv)
plot(cv ~ ps)
curve(coef(out)[1] * x^(-coef(out)[2]), add = TRUE)
optimumplotsize(a = coef(out)[1], b = coef(out)[2])

# End (not run)
```

pathanalysis

Path Analysis, Simple and Under Collinearity

Description

Function to perform the simple path analysis and the path analysis under collinearity (sometimes called *ridge path analysis*). It computes the direct (diagonal) and indirect (off-diagonal) effects of each explanatory variable over a response one.

Usage

```
pathanalysis(corMatrix, resp.col, collinearity = FALSE)
```

Arguments

corMatrix	a correlation matrix.
resp.col	an integer value indicating the column in corMatrix that corresponds to the response variable.
collinearity	logical; if TRUE, an external interactive display is used to pass a value, say k , at which to evaluate the system: $(\mathbf{X}'\mathbf{X} + \mathbf{I}k)\mathbf{B} = \mathbf{X}'\mathbf{Y}$, being $\mathbf{X}'\mathbf{X}$ the correlation matrix between explanatory variables, $\mathbf{X}'\mathbf{Y}$ the correlation vector between all explanatory variables and the response variable, \mathbf{B} is the vector of path coefficients and k is a value between 0 and 1; default is FALSE, i.e., $k = 0$.

Value

A list of

coef	a matrix containing the direct (diagonal) and indirect (off-diagonal) effects of each variable.
Rsq	the coefficient of determination.
ResidualEffect	the residual effect.
VIF	a vector containing the variance inflation factors.
CN	the condition number.

Side Effects

If `collinearity = TRUE`, an interactive graphic is displayed for dealing with collinearity.

Author(s)

Anderson Rodrigo da Silva <anderson.agro@hotmail.com>

References

Carvalho, S.P. (1995) *Metodos alternativos de estimacao de coeficientes de trilha e indices de selecao, sob multicolinearidade*. Ph.D. Thesis, Federal University of Vicosa (UFV), Vicosa, MG, Brazil.

Examples

```
data(peppercorr)
pathanalysis(peppercorr, 6, collinearity = TRUE)

# End (not run)
```

peppercorr

Correlations Between Pepper Variables

Description

The data give the correlations between 6 pepper variables. The data are taken from the article published by Silva et al. (2013).

Usage

```
data(peppercorr)
```

Format

An object of class "matrix".

Source

Silva et al. (2013) Path analysis in multicollinearity for fruit traits of pepper. *Idesia*, 31:55-60.

Examples

```
data(peppercorr)
print(peppercorr)

# End (not run)
```

raise.matrix	<i>Raising a Square Matrix to a Power</i>
--------------	---

Description

raise.matrix raises a square matrix to a power by using spectral decomposition.

Usage

```
raise.matrix(x, power = 1)
```

Arguments

x	a square matrix.
power	numeric; default is 1.

Value

An object of class "matrix".

Author(s)

Anderson Rodrigo da Silva <anderson.agro@hotmail.com>

See Also

[eigen](#), [svd](#)

Examples

```
m <- matrix(c(1, -2, -2, 4), 2, 2)
raise.matrix(m)
raise.matrix(m, 2)

# End (not run)
```

samplesize	<i>Minimum Sample Size</i>
------------	----------------------------

Description

Function to determine the minimum sample size for calculating a statistic based on its the confidence interval.

Usage

```
samplesize(x, fun, sizes = NULL, lcl = NULL, ucl = NULL,
nboot = 200, conf.level = 0.95, nrep = 500, graph = TRUE, ...)
```

Arguments

<code>x</code>	a numeric vector.
<code>fun</code>	an objective function at which to evaluate the sample size; see details.
<code>sizes</code>	a numeric vector containing sample sizes; if <code>NULL</code> (default), <code>samplesize</code> creates a vector ranging from 2 to $n-1$.
<code>lcl</code>	the lower confidence limit for the statistic defined in <code>fun</code> ; if <code>NULL</code> (default), <code>samplesize</code> estimates <code>lcl</code> based on bootstrap percentile interval.
<code>ucl</code>	the upper confidence limit for the statistic defined in <code>fun</code> ; if <code>NULL</code> (default), <code>samplesize</code> estimates <code>ucl</code> based on bootstrap percentile interval.
<code>nboot</code>	the number of bootstrap samples; it is used only if <code>lcl</code> or <code>ucl</code> is <code>NULL</code> .
<code>conf.level</code>	the confidence level for calculating the <code>lcl</code> and <code>ucl</code> ; it is used only if <code>lcl</code> or <code>ucl</code> is <code>NULL</code> .
<code>nrep</code>	the resampling (with replacement) number for each sample size in <code>sizes</code> ; default is 500.
<code>graph</code>	logical; default is <code>TRUE</code> .
<code>...</code>	further graphical arguments.

Details

If `ucl` or `lcl` is `NULL`, `fun` must be defined as in `boot`, i.e., the first argument passed will always be the original data and the second will be a vector of indices, frequencies or weights which define the bootstrap sample. By now, `samplesize` considers the second argument only as index.

Value

A list of	
<code>CI</code>	a vector containing the lower and the upper confidence limit for the statistic evaluated.
<code>pointsOut</code>	a data frame containing the sample sizes (in <code>sizes</code>), the number of points out of the CI (<code>n.out</code>) and the proportion of this number (<code>prop</code>).

Side Effects

If `graph = TRUE`, a graphic with the dispersion of the estimates for each sample size, as well as the graphic containing the number of points out of the confidence interval for the reference sample.

Author(s)

Anderson Rodrigo da Silva <anderson.agro@hotmail.com>

Examples

```

cv <- function(x, i) sd(x[i]) / mean(x[i]) # coefficient of variation
x = rnorm(20, 15, 2)
cv(x)
samplesize(x, cv)

par(mfrow = c(2, 1), cex = 0.8)
samplesize(x, cv, lcl = 0.05, ucl = 0.20)
abline(h = 0.05 * 500, col = "blue") # sample sizes with 5% (or less) out CI

# End (not run)

```

singh

*Importance of Variables According to the Singh (1981) Criterion***Description**

A function to calculate the Singh (1981) criterion for importance of variables based on the squared generalized Mahalanobis distance.

$$S_{.j} = \sum_{i=1}^{n-1} \sum_{i'>i}^n (x_{ij} - x_{i'j}) * (\mathbf{x}_i - \mathbf{x}_{i'})' * \Sigma_j^{-1}$$

Usage

```
singh(data, cov, inverted = FALSE, graph = TRUE, ...)
```

Arguments

data	a data frame or matrix of data ($n \times p$).
cov	a variance-covariance matrix ($p \times p$).
inverted	logical. If FALSE (default), cov is supposed to be a variance-covariance matrix.
graph	logical; if TRUE (default), a pie chart containing the importance proportions is plotted.
...	further graphical arguments.

Value

singh returns a matrix containing the Singh statistic, the importance proportion and the cumulative proportion of each variable (column) in data.

Side Effects

A pie chart containing the importance of the variables, if graph = TRUE.

Author(s)

Anderson Rodrigo da Silva <anderson.agro@hotmail.com>

References

Singh, D. (1981) The relative importance of characters affecting genetic divergence. *Indian Journal Genetics & Plant Breeding*, 41:237-245.

See Also

[D2.dist](#)

Examples

```
# Manly (2004, p.65-66)
x1 <- c(131.37, 132.37, 134.47, 135.50, 136.17)
x2 <- c(133.60, 132.70, 133.80, 132.30, 130.33)
x3 <- c(99.17, 99.07, 96.03, 94.53, 93.50)
x4 <- c(50.53, 50.23, 50.57, 51.97, 51.37)
x <- cbind(x1, x2, x3, x4)
Cov <- matrix(c(21.112,0.038,0.078,2.01, 0.038,23.486,5.2,2.844,
0.078,5.2,24.18,1.134, 2.01,2.844,1.134,10.154), 4, 4)
singh(x, Cov)

# End (not run)
```

tocher

Tocher's Clustering

Description

tocher performs the Tocher (Rao, 1952) optimization clustering from a distance matrix.

Usage

```
tocher(d)
```

Arguments

d an object of class "dist".

Value

A list of

clusters a list of length k (the number of clusters), containing the labels of the objects in d for each cluster.

distClust a matrix of distances within (diagonal) and between (off-diagonal) clusters.

d the input object.

Author(s)

Anderson Rodrigo da Silva <anderson.agro@hotmail.com>

References

- Cruz, C.D.; Ferreira, F.M.; Pessoni, L.A. (2011) *Biometria aplicada ao estudo da diversidade genetica*. Visconde do Rio Branco: Suprema.
- Rao, R.C. (1952) *Advanced statistical methods in biometric research*. New York: John Wiley & Sons.
- Sharma, J.R. (2006) *Statistical and biometrical techniques in plant breeding*. Delhi: New Age International.
- Silva, A.R. & Dias, C.T.S. (2013) A cophenetic correlation coefficient for Tocher's method. *Pesquisa Agropecuaria Brasileira*, 48:589-596.

See Also

[D2.dist](#), [coph.tocher](#), [distClust](#), [hclust](#)

Examples

```
# example 1
data(garlicdist)
(garlic <- tocher(garlicdist))
garlic$distClust # cluster distances

# example 2
data(USArrests)
(usa <- tocher(dist(USArrests)))
usa$distClust

# example 3
data(eurodist)
(euro <- tocher(eurodist))
euro$distClust

# End (not run)
```

Index

*Topic **datasets**
 garlicdist, [14](#)
 peppercorr, [18](#)
*Topic **package**
 biotools-package, [2](#)

aer, [3](#), [6](#)

biotools (biotools-package), [2](#)
biotools-package, [2](#)
boot, [20](#)
boxM, [4](#)

confusionmatrix, [4](#), [5](#), [9](#)
coph.tocher, [6](#), [23](#)
cor, [15](#)
cor.test, [15](#)
creategroups, [7](#), [12](#)

D2.disc, [8](#)
D2.dist, [9](#), [9](#), [22](#), [23](#)
dist, [10](#), [11](#)
distClust, [10](#), [23](#)

eigen, [19](#)

findSubsample, [11](#)
fitplotsize, [12](#), [16](#)

garlicdist, [14](#)

hclust, [23](#)

lda, [4](#), [6](#), [9](#)

multcor.test, [14](#)

nls, [12](#), [13](#)

optimumplotsize, [13](#), [16](#)

p.adjust, [15](#)

pathanalysis, [17](#)
peppercorr, [18](#)

raise.matrix, [19](#)

sample, [12](#)
samplesize, [19](#)
singh, [10](#), [21](#)
summary.nls, [13](#)
svd, [19](#)

tocher, [6](#), [11](#), [22](#)