

Package ‘simba’

April 19, 2009

Type Package

Version 0.2-5

Date 2007-08-16

Title A Collection of functions for similarity calculation of binary data

Author Gerald Jurasinski <terhorab@mac.com>

Maintainer Gerald Jurasinski <terhorab@mac.com>

Depends R (>= 2.5.0), stats, vegan, gplots, geoR, plotrix

Description Besides a function for the calculation of similarity measures with binary data (for instance presence/absence species data) the package contains some simple wrapper functions for reshaping species lists into matrices and vice versa and some other functions for further processing of similarity data (Mantel-like permutation procedures) as well as some other useful stuff.

License GPL-2

URL <http://www.r-project.org>

Repository CRAN

Date/Publication 2007-08-27 13:49:54

R topics documented:

simba-package	2
abis	3
ads.ternaries	5
aslopect	6
bcoov	7
boxes	9
com.sim	10
dfcor	12
diffmean	14

diffmich	15
diffslope	17
direct	20
dist.tmp	21
hexgrid	22
liste	23
makead	25
mama	27
occ.time	28
pcol	30
plot.mrpp	34
sim	35
sim.het	42
sim.pat	44
sim.relt	47
sim.tmp	49
sim.yo	50

Index	53
--------------	-----------

simba-package	<i>Calculate similarity measures for binary data</i>
---------------	------------------------------------------------------

Description

Besides a function for the calculation of similarity measures with binary data (for instance presence/absence species data) the package contains some simple wrapper functions for reshaping species lists into matrices and vice versa and some other functions for further processing of similarity data.

Details

Package:	simba
Type:	Package
Version:	0.2-5
Date:	2007-08-16
License:	GPL version 2

The functions in this package can be used to calculate similarities between species records (in binary format). Functions related to the correlation of similarity matrices and some other useful functions for the analysis of spatial patterns and their change in time are included as well.

Author(s)

Gerald Jurasinski <terhorab@mac.com>, with some help by Vroni Retzer <vroni.retzer@uni-bayreuth.de>

References

- Legendre, P. & Legendre, L. (1998) *Numerical Ecology*. – Elsevier.
- Wilson, M. V. & Shmida, A. (1984) Measuring beta-diversity with presence-absence data. – *Journal of Ecology* 72, 1055–1064.
- Gower, J.C. and Legendre, P. (1986) Metric and Euclidean properties of dissimilarity coefficients. *Journal of Classification*, 3, 5–48.
- Faith, D. P, Minchin, P. R. and Belbin, L. (1987). Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio* 69, 57–68.
- Krebs, C. J. (1999). *Ecological Methodology*. Addison Wesley Longman.
- Legendre, P, & Legendre, L. (1998) *Numerical Ecology*. 2nd English Edition. Elsevier.
- Mountford, M. D. (1962). An index of similarity and its application to classification problems. In: P.W.Murphy (ed.), *Progress in Soil Zoology*, 43–50. Butterworths.
- Wolda, H. (1981). Similarity indices, sample size and diversity. *Oecologia* 50, 296–302.

See Also

[vegdist](#), [dist.binary](#), [dsvdis](#), [dist](#), [reshape](#), [cor.test](#)

Examples

```
##see functions for examples
```

abis	<i>Vegetation and environmental data recorded in a Tundra ecosystem in Northern Sweden (Southeast of Abisko).</i>
------	-------------------------------------------------------------------------------------------------------------------

Description

Vegetation and related data from field sites in a Tundra Ecosystem in Northern Sweden. `abis.spec` holds abundance information of 158 species on 61 sites.

Usage

```
data(abis)
```

Format

`abis.env` contains some environmental variables from the Abisko field sites. These are:

- X Numeric: The x-coordinates of the field sites (utm). The sites are arranged in a regular equidistant grid. Each site is a hexagon plot with 5m radius.
- Y Numeric: The y-coordinates of the geographic position of the field sites.
- alt Integer: Height a.s.l. of the field sites.
- aspect Exposition of the sites. In degree from 0 to 180/-180. Counts negative in westward direction and positive in eastward direction.

slope	Numeric: the inclination of the sites in degree
pH	Numeric: pH value of soil samples taken on the sites (mixed samples from three
n.spec	Integer: Number of species on the site.
shannon	Numeric: Shannon index calculated from the species data.
even	Numeric: Evenness calculated from the species data.
simps	Numeric: Simpson diversity index for the site.
ephwtrs1	Numeric: Cover of ephemeral waters
ephwtrs2	Numeric: Cover of dried out ephemeral waters?
roughn	Numeric: Roughness index for the site. Obtained through calculating the ratio between the length of the straight lines dissecting the hexagonal plot and the lengths of lines along this dissections but exactly following the surface.
stones.cov	Numeric: Proportion of area of site covered by rocks which are covered with vegetation.
stones.uncov	Numeric: Proportion of area of site with uncovered, bare rocks.
bare.soil	Numeric: Open soil (most often sandy substrates).
water	Numeric: Proportion of area covered by permanent water bodies.
bog	Numeric: Proportion of area covered with boggy depressions.
lemming	Numeric: Counts of Lemming feces.
elk	Numeric: Counts of Elk feces.
ren	Numeric: Counts of Ren feces.
ripa	Numeric: Counts of Ripa feces.
fox	Numeric: Counts of Fox feces.
grubbings	Numeric: Amount of mole grubbings.
tread	Numeric: Amount of hoof tread.
betula	Numeric: Cover of <i>Betula</i> .
CEC	Numeric: Cation Exchange Capacity measured from the mixed soil sample taken in the field.
base.satur	Numeric: Base saturation.
heath	Binary: Does the site belong to the vegetation type 'heath'?
shrubs	Binary: Does the site belong to the vegetation type 'shrubby vegetation'?
protect	Binary: Does the site belong to the vegetation type 'protected snow heath'?

Source

Jurasinski, G. and Retzer, V. (in. prep.) Measuring multi-plot similarity with presence-absence data.
 Rettenmaier, N. 2004. Räumliche Muster der Biodiversität in der skandinavischen Tundra - Diploma thesis (unpublished), Department of Biogeography, University of Bayreuth, pp. 96.

Examples

```
data(abis)
```

ads.ternaries	<i>Artificial data-set for studying the mathematical behavior of asymmetric similarity coefficients</i>
---------------	---------------------------------------------------------------------------------------------------------

Description

Artificial data-set as utilized in Koleff et al. 2003, and Jurasinski 2007 to study the mathematical behavior of asymmetrical similarity indices. The corresponding values of all indices computable with `sim` are contained. See example!

Usage

```
data(ads.ternaries)
```

Format

ads.ternaries `data.frame` with the three matching components of asymmetric binary similarity measures (`a`, `b`, `c`) with all possible combinations of these components derived from a virtual data-set with 100 variables (species). These are the first three columns. The preceding columns contain the values of the similarity coefficients computable with `sim` according to the three matching components. This information can be used to study the mathematical behavior of the indices. See example

Source

Jurasinski, G. (2007) Spatio-Temporal Patterns of Biodiversity and their Drivers - Method Development and a Case Study from Northeastern Morocco. PhD-Thesis, Department of Biogeography, University of Bayreuth

Koleff, P., Gaston, K. J. & Lennon, J. J. (2003) Measuring beta diversity for presence-absence data. *Journal of Animal Ecology* **72**: 367-382.

Examples

```
data(ads.ternaries)
library(plotrix)
##take any index you want to study, see the help for sim() for available
##asymmetric indices or the names of the data.frame:
names(ads.ternaries)

##make a tmp from the index you want to study (we perform a transformation
##to obtain values between 0 and 1). if you want another index,
##just change the name in the next line:
tmp <- ads.ternaries$mounford
tmp <- (tmp-min(tmp))/max(tmp)
triax.plot(ads.ternaries[,c(2,3,1)], main="mounford",
col.symbols=grey(seq(0.1,1,0.1))[floor((tmp*100)/5)+1], pch=16)

##don't wonder: mounford is strange, just try another one:
```

```
##this time with rainbow-colors
tmp <- ads.ternaries$soerensen
tmp <- (tmp-min(tmp))/max(tmp)
triax.plot(ads.ternaries[,c(2,3,1)], main="sørensen",
col.symbols=rainbow(10)[floor((tmp*100)/10)+1], pch=16)

##and an interesting shape: routledge in greyscale...
tmp <- ads.ternaries$routledge
tmp <- (tmp-min(tmp))/max(tmp)
triax.plot(ads.ternaries[,c(2,3,1)], main="routledge",
col.symbols=grey(seq(0.1,1,0.1))[floor((tmp*100)/5)+1], pch=16)
```

aslopect

Calculate similarity of plots based on slope aspect and inclination

Description

Allows for the comparison of plots regarding the two variables slope aspect and slope inclination at once. To obtain a distance measure integrating aspect and inclination the model of a unit sphere is used and great-circle distances between virtual locations are calculated. For each plot a virtual location on the sphere is defined using the values for aspect as longitude and 90°-inclination as latitude. See `details` for more...

Usage

```
aslopect(asp, slo, names=rownames(asp), listout = FALSE)
```

Arguments

<code>asp</code>	Numeric vector with aspect values, given in degree. Expects values between 0° (North) and 180°/-180° (South). Eastward directions count positive, westward directions count negative.
<code>slo</code>	Numeric vector with slope inclination values. Expects values between 0° (flat) and 90° (vertical wall)
<code>names</code>	Plot names, defaults to the rownames of <code>asp</code> , but a separate vector can be specified. Its length has to match the length of <code>asp</code> and <code>slo</code> .
<code>listout</code>	Shall the results be given in list-format (<code>data.frame</code>). Defaults to <code>FALSE</code> which means that a matrix of class <code>dist</code> is returned

Details

To obtain a distance measure integrating aspect and inclination the model of a unit sphere is used and great-circle distances between virtual locations are calculated. For each plot a virtual location on the sphere is defined using the values for aspect as longitude and 90°-inclination as latitude. This means that as long as the inclination is low the virtual points are located in the pole region so that, regardless of aspect, plots with low inclination are rather close to each other regarding these

qualities. The idea behind is, that solar radiation, wind or other factors highly depending on aspect and inclination are not really differing between plots with different aspect as long as the slope is low. The longitude values on the unit sphere are derived from the values of slope aspect. The equator of the sphere is thought as the compass circle. The Prime Meridian of the virtual sphere is the great circle through North and South of the compass. As in geographic terms longitude counts positive in Eastern and negative in Western direction. With ϕ = latitude = 90° -inclination and λ = longitude = aspect the great-circle distance between A and B can be calculated with the following formula.

$$sim = \zeta = \arccos \left(\sin(\phi_A) \cdot \sin(\phi_B) + \cos(\phi_A) \cdot \cos(\phi_B) \cdot \cos(\lambda_B - \lambda_A) \right)$$

Value

Returns a `dist` object or a `data.frame` (depending if `listout = FALSE` or `TRUE`). As a unit sphere is used, the maximum distance between two inclination/aspect pairs is `perimeter/2` of the sphere which is by definition π . To scale the possible distances between 0 and 1 the results are divided by π . Thus, a great-circle distance of 1 is rather scarce in the real world, however, two vertical rock walls with opposite aspect would share it. If `listout = TRUE` a `data.frame` with the following variables returns.

NBX	one of the compared sampling units
NBY	the other part of the pair
x	The returned aslopect value

Author(s)

Gerald Jurasinski

Examples

```
data(abis)
## identify columns with slope and aspect data
names(abis.env)

## calculate aslopect
abis.aslop <- aslopect(abis.env[,4], abis.env[,5])
```

bcoov

Calculate Bray-Curtis distance for only one variable

Description

Calculates Bray-Curtis distance for only one variable: How dissimilar are the units regarding for instance pH?

Usage

```
bcoov(x, names, listout = FALSE)
```

Arguments

<code>x</code>	numeric vector with quantitative data from sampling-units, standardization may be applied before
<code>names</code>	where should the names be taken from, must be a vector of the same length as <code>x</code> and should give the names of the sampling-units
<code>listout</code>	Shall the result given in list-format (<code>data.frame</code>)? Defaults to FALSE

Details

It is just a simple way to calculate similarity based on only one variable. In the future there might be a possibility to choose from some functions. However, you could also use the difference in the data-values instead (which is the Euclidean distance in one dimensional space).

Value

Returns a `dist` object or a `data.frame` (in the case of `listout = TRUE`) with

<code>NBX</code>	one of the compared sampling units
<code>NBY</code>	the other part of the pair
<code>x</code>	The returned distance value. It is named like the tested variable

Author(s)

Gerald Jurasinski

See Also

[vegdist](#), [dist.quant](#), [dsvdis](#)

Examples

```
data(abis)
names(abis.env) ##take a look at the data
##calculate the similarity (Bray-Curtis) between the plots
##regarding pH
pH.dist <- bcoov(abis.env$pH, names=rownames(abis.env))

## directly give it as a list (data.frame)
pH.dist.ls <- bcoov(abis.env$pH, names=rownames(abis.env)
, listout=TRUE)
```

`boxes`*An adaption of `boxplot.n`.*

Description

Uses `boxplot` to produce a boxplot, which is then annotated with the number of observations in each group. Does allow for more flexibility compared to `boxplot.n`. Default size of the text is bigger and per default the numbers are not plotted directly on the x-axis and their distance from the axis can be changed.

Usage

```
boxes(..., top = FALSE, shrink = 1, textcolor = NULL, yadj = NULL)
```

Arguments

<code>...</code>	Parameters passed to <code>boxplot</code> .
<code>top</code>	Should the numbers of observations be printed below or above the boxes? Defaults to below (<code>top = FALSE</code>).
<code>shrink</code>	Parameter to scale the size of the numbers of observations. Above 1 increases size, below 1 until 0 decreases size.
<code>textcolor</code>	Color of the text. Defaults to <code>NULL</code> which then uses the actual plotting colour of the graphics device.
<code>yadj</code>	Can be used to adjust the vertical plotting positions of the numbers of observations. Defaults to <code>NULL</code> - no adjustment.

Author(s)

Gerald Jurasinski

See Also

[boxplot](#), [plot](#), [boxplot.n](#)

Examples

```
data(abis)

## see environmental data (see documentation on data for details)
abis.env

## calculate the difference in similarities for the three major
## vegetation types
## therefore created a vector from the data expressing belonging
## to the vegetation types:
tcs.sub <- rep(0, 61)
```

```

tcs.sub[abis.env[,29]==1] <- 1
tcs.sub[abis.env[,30]==1] <- 2
tcs.sub[abis.env[,31]==1] <- 3

## calculate similarity (Jaccard) between all pairs of plots
abis.jacc <- sim(abis.spec, method="jaccard")

## make boxplots regarding the similarities for each vegetation
## type, including the number of pairs contained in each box.
boxes(as.matrix(abis.jacc)~tcs.sub, shrink=1.2)

## prettier
boxes(as.matrix(abis.jacc)~tcs.sub, notch=TRUE, col="grey50",
shrink=1.2, ylim=c(-0.1, 0.9))

```

com.sim

Compare mean similarity between subsets of data

Description

Related to [mrpp](#). Are the differences in mean similarity between data subsets significant? Function takes the whole data-set (species matrix) and a subsetting vector and computes a specified similarity between all sampling units (rows). Then subsets are compared regarding their mean similarity. Statistical inference is obtained through permutation.

Usage

```

com.sim(veg, subs, simil = "soerensen", binary = TRUE,
        permutations = 1000, alpha = 0.05, bonfc = TRUE, ...)

```

Arguments

veg	Species matrix with columns = sites, rows = species. Deliver presence/absence data or abundance data. However, <code>binary</code> has to be set accordingly.
subs	Vector containing the subset definition. Same entries are understood to indicate belonging to the same subset (can be characters, factors or numerics). For each subset similarities/distances are calculated. Then all subsets are compared regarding mean and variance of the similarities/distances.
simil	Sets the coefficient to be used for calculating similarities/distances. If <code>binary = TRUE</code> , see sim , otherwise see vegdist for possible choices.
binary	Changes the function used for the calculation of similarity/distance. If binary species data is provided in <code>veg</code> keep the default (<code>binary = TRUE</code>). In this case sim is used to calculate the similarities. Set to <code>FALSE</code> when abundance or frequency data is provided. This calls vegdist to calculate the distances between sites in species similarity space.
permutations	Number of permutations performed to obtain the statistical inference. See Details.

alpha	Initial alpha level to test against. Defaults to 0.05.
bonfc	Shall Bonferroni correction be applied? Defaults to true.
...	Further arguments to functions.

Details

Entries of similarity/distance matrices are not independent. Therefore normal statistics might fail. One possibility is the application of permutation procedures. This means that the statistical distribution against which significance is tested is derived from the data.

Here it is implemented as follows: For each subset the similarities/distances between all sites (plots) are calculated with the specified coefficient. Then the resulting similarity/distance matrices are compared with `diffmean`. This is done for the comparison of each subset with each other subset. If specified (defaults to TRUE), Bonferroni correction is applied (to correct for multiple testing).

Depending on the number of subsets and the number of sites per subset it may take some seconds to be computed.

Value

Returns an object of class `cslist` containing the call to the function, the used method for similarity/distance calculation, a comparison matrix showing the connections between data-subsets (rows and columns connected with "*" are significantly different), the number of subsets involved, the number of permutations and a matrix giving information about the following components for each comparison between subsets:

X	Subset identifier for one of the compared subsets
Y	Subset identifier for the other compared subset
mean.x	Average distance/similarity for subset X.
mean.y	Average distance/similarity for subset Y.
diff	Difference in average distance/similarity for this comparison
sig	Significance of the difference in mean of the similarities.
sigs	Significance flag for the comparison ("*" means significant differences, "ns" means that the differences are not significant).
F	F-value for the Comparison.
sigF	Is F significant?
sigsF	Significance flag for F.

Author(s)

Gerald Jurasinski

See Also

[mrpp](#) for an anova like approach for comparing the differences of species data subsets.

Examples

```

data(abis)

## see environmental data (see documentation on data for details)
abis.env

## calculate the difference in similarities for the three major
## vegetation types
## therefore create a vector from the data expressing belonging
## to the vegetation types:
tcs.sub <- rep(0, 61)
tcs.sub[abis.env[,29]==1] <- 1
tcs.sub[abis.env[,30]==1] <- 2
tcs.sub[abis.env[,31]==1] <- 3

## calculate differences with Bray-Curtis as the distance measure
com.sim(abis.spec, tcs.sub, simil="bray", binary=FALSE)

## calculate differences with Soerensen as the similarity measure
com.sim(abis.spec, tcs.sub)

```

dfcor

Calculate permuted (Mantel) correlations between one and many variables

Description

The function uses [permcors2](#) to calculate permuted correlation on vectors. One vector is compared to various vectors of the same length. Useful e.g. if one variable has to be tested against various variables.

Usage

```
dfcor(ox, y, method = "pearson", permutations = 1000, ...)
```

Arguments

ox	Numeric vector. If it is a similarity matrix (i.e. a <code>dist</code> object, extract vector via <code>as.vector(x)</code> beforehand).
y	A <code>data.frame</code> containing numeric vectors to correlate <code>x</code> with. Number of rows has to equal the length of <code>x</code>
method	Method for correlation. Defaults to "pearson". See cor for other possibilities.
permutations	Number of permutations. Defaults to 1000, which gives reasonable results and allows to test against $\alpha = 0.001$.
...	Further arguments passed to internal functions (i.e. to cor).

Details

dfcor is a wrapper for permcor2, which is usually called as a part of pcol. Here, the numeric vector in x is compared to each column vector of y.

Value

A list with the following:

call	The function call
method	P-value obtained by testing the initial correlation against the permuted correlation values.
out	A table with statistics. See details below.
gesN	The included number of cases.
strata	The number of variables against which x was tested.
permutations	The number of permutations.
	The included printing method gives nice output (where information appears in a slightly different order) ending with the table of out. It is based on a data.frame with nrow = ncol(y) giving the statistics for the correlation between x and each column of y as follows.
corr	Correlation value (regarding to method).
sig	P-value obtained by testing the initial correlation against the permuted correlation values.
nop	Number of included pairs. The function tests for complete cases before calculation starts. Pairs containing NA's are not included.
miss	Number of missing pairs.

Note

Maybe pcol will get this functionality in future releases.

Author(s)

Gerald Jurasinski

References

Legendre, P, & Legendre, L. (1998) *Numerical Ecology*. 2nd English Edition. Elsevier.

See Also

For related functions of simba permcor, permcor2, mancor, pcol. Further see mantel of package vegan for a different implementation of permuted correlation on distance matrices.

diffmean

*Calculate the difference in Mean between two vectors***Description**

The function can be used to calculate the difference in mean between two vectors. Statistical inference is obtained through permutation. F-ratio is also calculated. For data which is not normally distributed or lacks independence. The plotting method plots the actual values of the difference in mean and F against an histogram of the results of the permuted runs.

Usage

```
diffmean(x, y, permutations = 1000)
## S3 method for class 'dmn':
plot(x, y, which=3, two=2, ...)
```

Arguments

<code>x</code>	Numeric vector. For the plotting method the <code>dmn</code> -object which should be printed (results from a <code>diffmean</code> operation). Plotting object in the plotting method.
<code>y</code>	Numeric vector. Plotting object in the plotting method, optional when <code>x</code> has appropriate structure
<code>permutations</code>	Number of permutations.
<code>which</code>	which histogram should be plotted? 1 triggers the histogram for difference in mean, 2 the one for F. It defaults to 3: both histograms are plotted. If it is changed from default, the next argument (<code>two</code>) is automatically set to 1!
<code>two</code>	Should the histograms be printed on a divided display? And how? Can only be set if <code>which</code> is set to 3. Defaults to 2, which means that the display is divided in two halves and the histogram-plots are plotted side by side. 3 causes histograms to be plotted one on top of the other. If <code>two = 1</code> , the display is NOT automatically divided. Might be useful if more than one <code>dmn</code> -objekt is to be plotted on one display. Otherwise the function overrides the actual display settings.
<code>...</code>	Further arguments to the plotting method.

Details

The two vectors do not need to share the same length but they should not be too different. Otherwise the function might give spurious results.

Value

Returns a list giving the function call, the difference in Mean, the mean of vector `x` and `y`, the mean of means, the F-value, the significance of the difference in Mean and the significance of F, as well as the number of permutations. The results of the permutation runs can be retrieved with `result$bootSM` (for the difference in mean) and `result$bootF` (for the F-values). There is a plot method for easily illustrating the test. The difference is plotted against an histogram displaying the distribution of the permuted values.

Author(s)

Gerald Jurasinski

See Also[diffslope](#), [diffmich](#)**Examples**

```

data(abis)

## create subsetting vector describing the belonging to different
## vegetationtypes
tcs.sub <- rep(0, 61)
tcs.sub[abis.env[,29]==1] <- 1
tcs.sub[abis.env[,30]==1] <- 2
tcs.sub[abis.env[,31]==1] <- 3

## check distribution
summary(as.factor(tcs.sub))

## compare vegetation types "shrubby vegetation" (shrub=2) and
## "protected by snowcover" (protect=3) regarding difference in
## similarities
abis2.soer <- sim(abis.spec[tcs.sub==2,])
abis3.soer <- sim(abis.spec[tcs.sub==3,])
abis.23cmp <- diffmean(abis2.soer, abis3.soer)

```

diffmich

Calculate the difference in parameters of a Michaelis-Menten kinetik fitted to (PAM) data

Description

The function can be used to calculate the difference in the two parameters of the Michaelis-Menten Kinetik $y = \frac{a*x}{b+x}$ between two datasets containing each two vectors. Through permutation it is possible to compute significance of the difference. [fitmich](#) is used to calculate the Michaelis-Menten fit to the data. With the corresponding plot method a plot of the actual difference in the parameters against a histogram of the permuted values can easily be achieved.

Usage

```

diffmich(x1, y1, x2, y2, permutations = 1000, a=3, b=0.5, trace=FALSE, ...)
fitmich(x, y, a=3, b=0.5)
## S3 method for class 'diffmich':
plot(x, y, which=3, two=2, ...)

```

Arguments

<code>x1</code>	Vector containing an independent variable, for instance PAR measurements.
<code>y1</code>	Vector containing a variable dependent on <code>x1</code> (for instance ETR measurements). Must have the same length as <code>x1</code> .
<code>x2</code>	Vector containing a second independent variable (for instance PAR measurements).
<code>y2</code>	Vector containing a variable dependent on <code>x2</code> (for instance ETR measurements). Must have the same length as <code>x2</code> .
<code>permutations</code>	Number of permutations.
<code>a</code>	start value for parameter a, defaults to 3, usually there is no change necessary, but if the function gets trapped in the first run, changing the parameters might solve the problem.
<code>b</code>	Start value for parameter b, defaults to 0.5.
<code>trace</code>	set to TRUE for displaying the progress of the calculation
<code>...</code>	Arguments to other functions (for instance to <code>lm</code> , which is used to calculate the regression lines)
<code>x</code>	Vector containing an independent variable, for instance PAR measurements. Function <code>fitmich</code> is usually called only internally by <code>diffmich</code> . Plotting object in the plot method.
<code>y</code>	Vector containing a variable dependent on <code>x1</code> (for instance ETR measurements). Must have the same length as <code>x1</code> . Plotting object in the plotting method, optional when <code>x</code> has appropriate structure
<code>which</code>	Which histogram should be plotted? 1 triggers the histogram for parameter a, 2 the one for parameter b. It defaults to 3: both histograms are plotted. If it is changed from default the next argument (<code>two</code>) is automatically set to 1!
<code>two</code>	Should the histograms be printed on a divided display? And how? Can only be set if <code>which</code> is set to 3. Defaults to 2, which means that the display is divided in two halves and the histogram-plots are plotted side by side. 3 causes histograms to be plotted one on top of the other. If <code>two</code> = 1, the display is NOT automatically divided. Might be useful if more than one <code>diffmich</code> -objekt is to be plotted on one display. Otherwise the function overrides the actual display settings.

Details

As the function was initially built to easily calculate the difference of parameters of the Michaelis-Menten Kinetik for PAM measurements, the independent vectors are meant to contain PAR values whereas the dependent vectors should represent ETR values. But you can use it for anything else which can be fitted with Michaelis-Menten. The vectors belonging together are formed into a `data.frame`. For each permutation run the rows are interchanged randomly between the two `data.frames` and the difference in the parameters is calculated and collected into a vector. The p-value is then computed as the ratio between the number of cases where the differences in Parameter exceed the difference in parameter of the initial configuration and the number of permutations.

As it uses a `for` loop it takes a while to calculate. So get a coffee while it is running, or set `trace` = TRUE to avoid boring moments ...

Value

Returns a `diffmich`-object with the function call, the difference in the two parameters and their significance. Furthermore the number of permutations. If you want to change the way `fitmich` is computed you can change the starting values. Per default it is calculated with starting values `a=3` and `b=0.5`. There's no change needed unless the function gets trapped.

Author(s)

Gerald Jurasinski

See Also

[nls](#), [sample](#)

diffslope

Calculate the difference in slope of two regression lines

Description

The function can be used to calculate the difference in slope between two datasets containing each two vectors. Follows an idea of Nekola & White (1999) for calculating the statistical inference of the difference in slope between two regression lines. `diffslope2` has the same purpose as `diffslope` but implementation is without for-loop. The plot method allows easy plotting of the actual difference in slope against the distribution of permuted values.

Usage

```
diffslope(x1, y1, x2, y2, permutations = 1000, resc.x = FALSE,
          resc.y = TRUE, trace=FALSE, ...)

diffslope2(x1, y1, x2, y2, permutations = 1000, resc.x = FALSE,
           resc.y = TRUE, ...)

## S3 method for class 'dsl':
plot(x, y, ...)
```

Arguments

<code>x1</code>	vector containing an independent variable (for instance distance between plots).
<code>y1</code>	vector containing a variable dependent on <code>x1</code> (for instance similarity between the plots. must have the same length as <code>x1</code>).
<code>x2</code>	vector containing a second independent variable (for instance distance between plots). can be the same as in <code>x1</code> .
<code>y2</code>	vector containing a variable dependent on <code>x2</code> (for instance similarity between the plots. must have the same length as <code>x2</code>).
<code>permutations</code>	number of permutations

<code>resc.x</code>	Shall the values of the independent variables be rescaled to a common mean?
<code>resc.y</code>	Shall the values of the dependent variables be rescaled to a common mean? Defaults to TRUE (Nekola & White 1999).
<code>trace</code>	Set to true if progress shall be printed with increasing numbers. Defaults to FALSE
<code>...</code>	Arguments to other functions (for instance to <code>lm</code> , which is used to calculate the regression lines).
<code>x</code>	<code>dsl</code> -object (given back by <code>diffslope</code>) which is to be plotted.
<code>y</code>	Plotting object, usually not necessary

Details

As the function was initially build to easily calculate the difference in slope between the regression lines of distance decay plots, the independent vectors are meant to contain distance values whereas the dependent vectors should represent similarity values. But you can use it for anything else, as you wish. The vectors belonging together are formed into a `data.frame`. For each permutation run the rows are interchanged randomly between the two `data.frames` and the difference in slope calculated thereafter is calculated and collected into a vector. The p-value is then computed as the ratio between the number of cases where the differences in slope exceed the difference in slope of the initial configuration and the number of permutations.

If the difference in slope returns negative, the slope (distance decay) of the second relationship is less pronounced, if it returns positive, the second relationship exhibits a stronger distance decay (slope) than the first. This holds for distance decay relationships. If `y` increases with `x`, it is vice versa.

As it uses a `for` loop, it takes a while to calculate. So get a coffee while it is running, or set `trace` to TRUE to avoid being bored ...

Value

Returns a list giving the function call, the difference in slope, the significance of this difference, and the number of permutations. If you want to change the way `lm` is computed you must send the arguments to `lm` via `...`. Per default it is calculated with the default arguments of `lm`.

Author(s)

Gerald Jurasinski

References

- Nekola, J. C. and White, P. S. (1999) The distance decay of similarity in biogeography and ecology. *Journal of Biogeography* 26: 867-878.
- Steinitz, O., Heller, J., Tsoar, A., Rotem, D. and Kadmon, R. (2005) Predicting Regional Patterns of Similarity in Species Composition for Conservation Planning. *Conservation Biology* 19: 1978-1988.
- Steinitz, O., Heller, J., Tsoar, A., Rotem, D. and Kadmon, R. (2006) Environment, dispersal and patterns of species similarity. *Journal of Biogeography* 33: 1044-1054.

See Also

[lm, sample](#)

Examples

```

data(abis)
names(abis.env) ##take a look at the data
pert.dist <- 1-vegdist(abis.env[,19:25])
##calculate the similarity (Bray-Curtis) between the plots
##regarding disturbance variables

soil.dist <- 1-vegdist(abis.env[,c(6,27:28)])
##calculate the similarity (Bray-Curtis) between the plots
##regarding soil parameters

coord.dist <- dist(abis.env[,1:2]) ##calculate geographical distance between plots

##transform all distance matrices into list format:
struc.dist.ls <- liste(pert.dist, entry="BC.struc")
soil.dist.ls <- liste(soil.dist, entry="BC.soil")
coord.dist.ls <- liste(coord.dist, entry="dist")

##create a data.frame containg plot information, geographical
##distance, similarity of soil parameters, and similarity of
##structural parameters:

df <- data.frame(coord.dist.ls, soil.dist.ls[,3], struc.dist.ls[,3])
names(df) ##see names

##give better names:
names(df)[4:5] <- c("soil", "struc")
attach(df)

##prepare graphics device:
par(mfrow=c(2,1))

##plot and compare distance decay (decrease of similarity with
##distance):
plot(dist, soil)
plot(dist, struc)
##remove problematic zero entries:
df <- subset(df, struc != 0)

##plot again, this time with regression lines (in red for better
##visability):
detach(df)
attach(df)
plot(dist, soil)
abline(lm(soil~dist), col="red4")
plot(dist, struc)
abline(lm(struc~dist), col="red4")
##is the slope significantly different?

```

```
res <- diffslope(dist, soil, dist, struc)
##go for a coffee, as it takes a while...
```

direct

Obtain Direction Classes from Geographic Coordinates

Description

The functions calculates direction classes from geographic coordinates (not lat/lon). All possible connections between these points are established and the direction of each link is calculated. This is followed by a designation of direction-classes.

Usage

```
direct(coord, listout=FALSE)
direct2(coord, listout=FALSE)
```

Arguments

coord	A <code>data.frame</code> containing coordinates. Should have the same number of points as the data for which the direction-classes are calculated.
listout	Logical value, indicating whether the result is given back in (<code>data.frame</code>)-format instead of returning a <code>dist</code> -object.

Value

Returns a matrix containing the direction-classes of the connections between the coordinates as a `dist`-object. If `listout = TRUE`, the result is given as a list (`data.frame`).

`direct` returns 4 directions (North-South, Northwest-Southeast, West-East, Northeast-Southwest).

`direct2` returns 6 directions.

Note

as with `mantel` it takes a while to calculate

Author(s)

Gerald Jurasinski

See Also

`mantel`, `cor.test`

Examples

```

data(abis)
dirclass <- direct(abis.env[,1:2])
dirclass
dirclass.ls <- direct(abis.env[,1:2], listout=TRUE)
dirclass.ls

```

dist.tmp	<i>Calculate the distance between the instances or variables of two similar data-sets.</i>
----------	--------------------------------------------------------------------------------------------

Description

The function calculates the distance between the instances or variables of two data-sets, preferably two recordings from the same places/variables at different time steps. Could for example be useful to calculate the dissimilarity of species records from the same set of plots at different recording campaigns.

Usage

```
dist.tmp(x, y, method = "euclidean", margin = 1, adjust=TRUE)
```

Arguments

x	data.frame with numeric columns giving the data at time one
y	data.frame with numeric columns giving the data at time two
method	Which distance measure should be employed? The following measurees are available: "manhattan", "euclidean", "bray", "canberra", "kulczynski", "gower", "jaccard". See details.
margin	Shall the distance between the instances (rows, margin = 2), or the distance between the variables (columns, margin = 1) be calculated?
adjust	The default adjust = TRUE takes care of your input data. However, instances and variable names have to be unique, as the matching relies on these names. If your data meets this requirement the function automatically detects which instances and variables can be found in both data-sets and calculates the distance based on these data.

Details

The function provides just an alternative to code your repeated measurements from the same plots, calculate the distances between all instances and sort out the distances between the same instances at different time steps. Here just the latter is calculated directly.

For details regarding the different distance coefficients, see [vegdist](#). The formulae behind were taken from there. However, not all choices have been implemented here.

Value

Returns a vector of distances. Length depends on `margin`. If you choose `margin = 1` it equals the number of instances (plots) in your two data-sets. If you choose `margin = 2` it equals the number of variables (species) in your data-sets.

Author(s)

Gerald Jurasinski (terhorab@mac.com)

References

See references in [vegdist](#)

See Also

[vegdist](#), [dist](#), [sim](#)

Examples

```
data(abis)

##just to see that the function takes care of matching instances
##and variables we compare abis.spec with a permuted version of
##itself and skip some instances and variables in the original matrix.
dist.tmp(abis.spec[5:50,3:120], apply(abis.spec, 2, sample), method="bray")
```

hexgrid

Produces the nodes of an equidistant grid.

Description

Given the coordinates of a starting point (left upper point of the grid), the function produces the nodes of an equidistant grid. Extent and distance between plots can be specified.

Usage

```
hexgrid(x, y, r = 100, nro = 10, nco = 20)
```

Arguments

<code>x</code>	x-value of the starting point. Defaults to 0.
<code>y</code>	y-value of the starting point. Defaults to 0.
<code>r</code>	Distance between nodes. Defaults to 100.
<code>nro</code>	Number of rows in the grid. Defaults to 10
<code>nco</code>	Number of columns in the grid. Defaults to 20. They are doubled (see the grid) compared to the rows. That's why <code>2*nro</code> produces a quadratic grid.

Details

If the overall shape of the grid is not square, the user has to delete by hand the superfluous units. Might get more flexible in future versions.

Value

Returns a data.frame giving informations on the produced point/unit/plot locations with the following columns:

ROW	Number of row in the grid to which the point/unit/plot belongs.
COL	Number of column in the grid to which the point/unit/plot belongs.
X	x-coordinate of the point.
Y	y-coordinate of the point.

Author(s)

Gerald Jurasinski

References

<http://homepage.mac.com/terhorab/gerald/downloads/whyhexaagons.pdf>

Examples

```
## produces a grid with r=400:
test.grd <- hexgrid(456000, 7356700, r=400)
## for plotting the following is recommended as it preserves
## real positions:
library(geoR)
points.geodata(coords=test.grd[,3:4], data=rnorm(nrow(test.grd)))
```

liste

Convert dist-objects and matrices to database format

Description

Transposes `dist` objects to database list format (where each line represents a similarity value calculated between two plots, so the list has three columns containing information on plot x, plot y and information on similarity/dissimilarity). This might be useful if further database processing is intended. If the given matrix is not of class `dist`, the whole matrix is converted. This might be handy if species matrices are to be converted to database format.

Usage

```
liste(obj, x="NBX", y="NBY", entry=NULL, factorize=TRUE,
      splist=FALSE)
```

Arguments

<code>obj</code>	A distance object as it is returned from <code>sim</code> , <code>dist</code> , <code>vegdist</code> or <code>dist.binary</code> or a similar matrix with <code>class="dist"</code> . If the matrix is not of class <code>dist</code> a <code>data.frame</code> is expected and all entries are converted. Species matrices are assumed to contain sites (or plots) in rows and species in columns.
<code>x</code>	How the second column should be named, standard is that it contains "NBX" the neighbour X. Obsolete when <code>splist = TRUE</code> .
<code>y</code>	How the second column should be named, standard is that it contains "NBY" the neighbour Y. Obsolete when <code>splist = TRUE</code> .
<code>entry</code>	How the third column should be named. If there is nothing given, it is named "we" (whatever). Obsolete when <code>splist = TRUE</code> .
<code>factorize</code>	If naming (first two) columns should be given as factors. Defaults to <code>TRUE</code> .
<code>splist</code>	Set to <code>TRUE</code> if a species matrix is transformed to a database list format. Automatically removes entries with zero occurrence of the species' and names columns correspondingly (see Value).

Details

Column `x` represents the column names of the input matrix. So if you want to reshape a species matrix, `x` will be the species names and `y` will be the plot names. If it is needed vice versa, you have to change column order. For convenience you can set `splist = TRUE` and you will get a species list in database format with columns representing the plot, species and occurrence information. Furthermore zero occurrences are already omitted.

Value

Returns a `data.frame` with three columns:

<code>Col1</code>	The column names of the input matrix. If it is a <code>dist</code> -object only the lower triangle is used. Named <code>plot</code> when a species matrix is transformed with <code>splist = TRUE</code> .
<code>Col2</code>	The row names of the input matrix. Named <code>spec</code> when a species matrix is transformed with <code>splist = TRUE</code> .
<code>Col3</code>	The respective matrix entries. Named <code>occ</code> when a species matrix is transformed with <code>splist = TRUE</code> .

Author(s)

Gerald Jurasinski

See Also

[reshape](#), [data.frame](#). It resembles functionality of `reshape` - which is more flexible (but also more complex).

Examples

```
data(abis)
## there are empty species entries:
sum(colSums(abis.spec)==0)
## remove empty species
abis.spec <- abis.spec[,colSums(abis.spec)!=0]
abis.spec.ls <- liste(abis.spec, splist=TRUE)
```

makead

Create artificial data set (species matrix).

Description

The functions allow for the automated creation of artificial data (species matrix). The user can choose between random organization or a gradient. The gradient can be defined via a gradient vector which allows for fine tuning of the gradient. `ads` has a different implementation and produces better results for gradients.

Usage

```
makead(nspec, nplots, avSR = NULL, anc = NULL, grad.v = NULL,
       cf = 0.2, puq = 0.01)
```

```
ads(nspec, nplots, avSR = NULL, anc = NULL, grad.v = NULL,
    reord = TRUE, cf = 0.2, puq = 0.01)
```

```
ads.hot(nspec, nplots, avSR = NULL, anc = NULL, grad.v = NULL,
        frac=0.5, reord=TRUE, cf=0.2, puq=0.01)
```

Arguments

<code>nspec</code>	Numbers of species you want to be in the data-set. Meaningless if <code>anc != NULL</code> .
<code>nplots</code>	Numbers of plots you want to be in the data-set. Meaningless if <code>anc != NULL</code> .
<code>avSR</code>	Average species richness. If <code>anc</code> is given, it is calculated from the data when the default is not changed. If <code>avSR != NULL</code> , the given value is taken instead. In the actual version not implemented in <code>ads</code> .
<code>anc</code>	If a model species matrix is available (either a real data-set, or another artificial data-set) on which creation should be based, give it here. Rows must be plots and columns be species. The first three parameters are then obtained from this set. However average species richness (<code>avSR</code>) can still be given by the user.
<code>grad.v</code>	A numeric vector describing the gradient, or - in case of <code>ads.hot</code> - the hotspot. Must have the same length as <code>nplots</code> (or <code>nrow(anc)</code> respectively). See details.

<code>cf</code>	Determines the probability of the species to occur on the plots. In other words, it changes the shape of the species accumulation curve. Set to NULL if no natural species accumulation should be applied (may sometimes increase the visibility of the gradient)
<code>puq</code>	Percentage of ubiquitous species. Set to NULL if the produced gradients seem to be unclear or if you don't want ubiquitous species to be in the data-set. Only used if a gradient vector is given (which is then not applied to the given percentage of species).
<code>reord</code>	Triggers reordering of the columns in the produced gradient matrix (see details). May considerably change the resulting matrix. Defaults to TRUE.
<code>frac</code>	Numeric between 0 and 1 giving the percentage of species which should occur on the hotspot-gradient only (see details).

Details

There are three different implementations to create an artificial species matrix. `makead` first applies the natural species accumulation curve, the gradient for each species is represented by a vector containing numerics between 0 and 1. Both matrices are added so that values between 0 and 2 result. Through an iteration procedure a breakvalue is defined above which all entries are converted to 1. Values below are converted to 0 resulting in a presence/absence matrix. However the random element seems to be too strong to get evident gradient representations.

Therefore `ads` is implemented. It works different. First, a gradient is applied. As with `makead` the gradient is always applied in two directions so that half of the species are more likely to occur on plots on one side of the gradient, whereas the others are more likely to occur on the other side of the gradient. Subsequently, species occurrence for all species will oscillate around `nplots/2`.

If `puq` is specified the given percentage of species is divided from the whole matrix before the gradient is applied. With the parameter `cf` a vector is produced representing natural quasi-natural occurrence of the species on the plots: Most species are rare and few species are very common. This is described by a power function $y = \frac{1}{x^cf}$ with `x` starting at 2 and gives a vector of length `nspec` representing the number of times each species is occurring.

These numbers are applied to the gradient matrix and from the species occurrences only as many as specified by the respective number are randomly sampled. In cases where the occurrence number given by the vector exceeds the occurrences resulting from the gradient matrix, the species in the gradient matrix is replaced by a new one for which occurrence is not following the gradient and represents the number of occurrences given by the vector. The idea behind this is, that also in nature a species occurring on more than about half of the plots will likely be independent from a specific gradient.

In both cases (`makead` and `ads`) a totally random species matrix (under consideration of natural species occurrence, see `cf`) is obtained by randomly shuffling these occurrences on the columns (species) of the "natural species occurrence" matrix.

Contrarily to the other two functions, `ads.hot` allows for the creation of an artificial data-set including a hotspot of species richness and composition. In this case, `frac` can be used to specify which proportion of the total number of species should only occur on the hotspot gradient. All other species occur randomly on the plots. However, with the hotspot-gradient (`grad.v`) you can influence the explicitness of the hotspot.

Value

Returns a species matrix with rows representing plots/sampling units and columns representing species.

Author(s)

Gerald Jurasinski, Vroni Retzer

Examples

```
## create a random data-set with 200 species on 60 plots
artda <- makead(200, 60, avSR=25)

## create a gradient running from North to South (therefore you
## need a spatially explicit model of your data which is obtained
## with hexgrid())
coor <- hexgrid(0, 4000, 200)
coor <- coor[order(coor$ROW),] #causes coordinates to be in order.
## then the gradient vektor can easily be generated from the ROW names
gradvek <- as.numeric(coor$ROW)
## check how many plots your array has
nrow(coor)
## create a data-set with 200 species
artda <- ads(200, 100, grad.v=gradvek)
## see the species frequency distribution curve
plot(sort(colSums(artda)))
```

mama

A (convenience) wrapper function to make matrix from a data.frame

Description

The function `mama` uses [reshape](#) to transpose species data given in database list format (where each line represents a species in a plot, so the list has three columns containing information on plot, species and information on occurrence) into a plot species matrix (where rows represent plots and columns represent species) for further use with other functions on vegetational data.

Usage

```
mama(dat)
```

Arguments

`dat` Species data in list format. The columns have to represent plot, species, occurrence information (presence/absence or abundances). Column names may differ but they must be in that order!

Details

You could reach the same result with [reshape](#). I was just always quite confused with this. That's why i decided to do this little wrapper for convenience. It needs quite a while to run though, but this is due to [reshape](#).

Value

Returns a [data.frame](#) which contains the presence/absence or abundance data of the species list. Rows represent plots, columns represent species. If you want to have it vice versa you have to use the function on a list with columns species, plots, occurrence information (in that order).

Author(s)

Gerald Jurasinski

See Also

[reshape](#), [data.frame](#)

Examples

```
data(abis)
abis.spcls <- liste(abis.spec, splist=TRUE)
## see the list, it like what you get from a database
## and return to matrix-format:
abis.test <- mama(abis.spcls)
```

occ.time

Track species occurrence

Description

Calculate the change in occurrence of species on plots in general or on specific plots. Allows for the tracking of temporal changes in species abundance throughout an area or the simple quantification and comparison of species occurrences between time steps.

Usage

```
occ.time(x, y, times = NULL, adjust = TRUE, gen.occ = FALSE,
perc = TRUE, nc.acc = FALSE, ...)
```

```
occ.tmp(x, y, adjust=TRUE, gen.occ=FALSE, perc=TRUE,
nc.acc=FALSE, ...)
```

Arguments

<code>x</code>	Species data in matrix or database-format representing species occurrence at time step one or throughout a time series. The latter means that you have a table with three columns where the columns represent <code>plots</code> , <code>species</code> and <code>occurrence</code> information (in this order!). These are typically exported from a database. When <code>times</code> are given and data represents more than one time steps it has to be in database format. Conversion is handled automatically - if three columns are in the table it is assumed to be in database format. When there are only three species give the data in database-format. If there are many plots and/or species, internal matrix conversion might be quite slow.
<code>y</code>	Species data in matrix or database-format representing species occurrence at time step two. Obsolete when <code>times</code> are given. Otherwise the same as for <code>x</code> applies.
<code>times</code>	A vector describing the timesteps which has to be coercible to a factor. If your data comes from a database and contains species records for different time-steps, just export the time information with the species data. If you have single matrices for each time step, you could reshape them to database format via liste and concatenate these or calculate each time step alone.
<code>adjust</code>	Do not change the default behaviour (TRUE) unless you know what you do. Would spare some calculation time if set to FALSE, when your species data do not need adjustment, which means that in both or all time steps, there are exactly the same species and the same plots. However in most cases it will be more convenient to rely on the function (see details).
<code>gen.occ</code>	Triggers if general occurrence is regarded or specific occurrence. The latter is default (<code>gen.occ=FALSE</code>) and it means that it is calculated on which exact plots a species is changing. When set to TRUE only general occurrence is regarded and it is calculated on how many plots a species occurs more or less then before. See details.
<code>perc</code>	If output shall be in percentage of species. Defaults to TRUE.
<code>nc.acc</code>	Per default, species which are not changing on a plot are counted as single species (also when they do not change on more than one plot). This can be changed when setting <code>nc.acc = TRUE</code> . Then each occurrence of species which has not changed is counted.
<code>...</code>	Further arguments to functions.

Details

If you compare species data among time steps there will be most likely different numbers of species (and often also different numbers of plots for which information is available). The function takes care of this and you can give any species matrices you want. If one plot is the same, it will calculate what changed on this plot. There will be an error message if no plot is shared. The function relies on plot and species names!! As in a database - they must be unique!!

With the resulting named vector or table it is easy to draw a barplot tracking the changing occurrence of plants. Walter et al. (2005) used such plots to illuminate the changing (and increasing) occurrence of plant species on Alpine summits due to climate change.

However they only considered change in general occurrence. We added the possibility to track changes on the specific plots. If a species is occurring on 5 plots at time one and on 4 plots at time two one can't be sure that this species changed occurrence on only one plot. If it occurs partly on different plots then before it will contribute to loss and gain respectively in this function when `gen.occ` is set to `FALSE`.

Value

Returns a named vector or a table (when multiple time steps are evaluated at once) with information on change in species occurrence. The names tell on how many plots a certain species has changed. The values tell how many species (or percent of species) exhibit this change. Additionally the number of matching plots and species for each comparison are given back.

Per default nice output is given. However, the table can be accessed for printing with `*$bac` (see Example for Details). A plotting method will be added in the near future.

Author(s)

Gerald Jurasinski (terhorab@mac.com) <http://homepage.mac.com/terhorab/gerald/>

References

Walther, G.-R., Beißner, S. & Burga, C. A. 2005. Trends in the upward shift of alpine plants. *Journal of Vegetation Science* **16**: 541-548.

See Also

to calculate similarity based on plant species occurrence between time steps see [sim.tmp](#)

pcol

Permuted Correlation (on strata)

Description

The function is a wrapper for several functions related to the permuted correlation between matrices or vectors: It calls `permcorm2` to calculate permuted correlation between vectors and `permcorm` to calculate permuted correlation on strata. This can be useful to obtain data-points for a multivariate Mantel correlogram. Two matrices or vectors and a variable dividing these vectors into strata (levels) are to be specified. If the last is not given only permuted correlation between the two matrices (`dist`-objects) is done. If the second matrix/vector is a subsetting object the correlation can be done on the first object for each of the strata with `mancorm`.

Usage

```
pcol(x, y, z = NULL, method="pearson", permutations=1000,
solo=FALSE, ...)
```

```
permcorm(x, y, subsetter, method="pearson", permutations=1000,
```

```

alpha=0.05, trace=FALSE, ...)

permc2(x, y, method="pearson", permutations=1000, subset=NULL,
complete=TRUE, ...)

mancor(dis, classes, width=NULL, method="pearson", permutations=1000,
alpha=0.05, trace=FALSE, ...)

## S3 method for class 'permc2':
plot(x, y, ...)

```

Arguments

x	Matrix (dist-objekt) or vector of numeric values (atomic) containing distances or similarities returned by <code>sim</code> , <code>vegdist</code> , <code>dist</code> , <code>dist.binary</code> or a similar matrix. Conversion is done automatically and triggered by <code>is.vector</code> . If it is not a vector it is assumed to be a dist-objekt or a similar matrix (with <code>nrow==ncol</code>). For the plotting method a <code>permc2</code> -object.
y	If <code>solo=FALSE</code> an object of the same qualities and dimensions (or length) as x to correlate to. See <code>Details</code> for the case <code>solo=TRUE</code> . For the plotting method an optional object. Not needed here.
z	Vector or dist-objekt of a variable which defines the strata, or levels for which the permuted correlation shall be calculated. Doesn't have to be a factor but it has to be convertible into factor. see <code>Details</code> for more.
subsetter	Vector of a variable which defines the strata, or levels for which the permuted correlation shall be calculated. Usually conversion from dist-object to vector is done in <code>pcol</code> . If <code>permc2</code> is run separately it has to be a vector (same holds for x, y, and z if the background functions are run separately). Doesn't have to be a factor but it has to be convertible into factor. see <code>Details</code> for more.
method	Method of correlation, as it is done by <code>cor.test</code> , see help there for details. Defaults to Pearson correlation coefficients. Other options are Kendall and Spearman rank correlations.
permutations	number of permutations, defaults to 1000 to get a significance level of $p = 0.001$.
solo	If <code>TRUE</code> y is assumed to be a subsetting matrix, dist-object or vector giving strata to correlate x with and calculate the data-points for a Multivariate Mantel Correlogram for one distance-matrix.
width	If <code>solo=TRUE</code> the subsetting object is assumed to contain classes already. If width is specified (defaults to <code>NULL</code>), the classes are defined inside the function and width gives the class width.
trace	Set to <code>TRUE</code> to follow the runs of the for-loops in functions <code>permc2</code> and <code>mancor</code> . See <code>details</code> for when it is appropriate to set via <code>pcol</code> .
complete	Should only complete cases be considered? Defaults to <code>TRUE</code> .
alpha	The initial alpha-level against which should be tested. Depending on sub-function it is internally changed via Bonferroni-correction if necessary.
subset	If Case is 1 (see <code>details</code>) a subset of cases from x and y can be defined for correlation.

<code>dis</code>	Same as <code>x</code> for <code>mancor</code> .
<code>classes</code>	If <code>mancor</code> shall be run, second item is a vector or <code>dist</code> -objekt of a variable which defines the strata, or levels for which the permuted correlation shall be calculated. Doesn't have to be a factor but it has to be convertible into factor. see <code>Details</code> for more.
<code>permc</code>	(In <code>plot.permcor</code>) an object returned by <code>permc2</code> is easily plotted with <code>plot.permcor</code> . The actual correlation value is plotted against an histogram of the distribution of the permuted values.
<code>...</code>	Arguments to other functions, for instance to <code>cor.test</code> regarding specifications of the test, however only the correlation value is taken from this function. but here you could change from pearson to kendall for instance.

Details

`pcol` is a wrapper for the other functions. Depending on the input and the setting of `solo` the following functions are invoked (They can also be run separately. In this case `x`, `y`, and `z` must be given as vectors).

1. If `x` and `y` are `dist`-objects, or vectors containing distance or similarity values and everything else is set to defaults a simple permuted correlation with `permc2` is run. This corresponds to a Mantel test. The two data-objects are correlated with `cor.test`, then the `y` is permuted and with `cor` the correlation is calculated again and written to a vector. This is repeated `permutation` times. Finally, the initial correlation value is compared to the permuted values. The number of times, the permuted values exceed the initial value is divided by the number of permutations to obtain a significance value. Thus, with 1000 permutations a minimum `p` of 0.001 can be tested.
2. If a `subsetter` is given in `z`, the permuted correlation is done for every stratum or level given by the subsetting object - this could e.g. be direction or distance classes flagging which plots share a similar distance and therefore fall into the same class. The resulting data-points can be used to plot a correlogram which allows for the analysis of non-stationarity in the relationships between `x` and `y`.
3. If `y` is itself a subsetting object (distance classes or so) you have to set `solo=TRUE`. Then the matrix or vector in `x` is correlated against this classes. This is handled by `mancor`. The parameter `width` allows for the calculation of classes inside the function. If for instance a distance-matrix with geographical distances is given, `width` specifies the width of the distance classes, they are computed and used to correlate the data in `x` against. This produces the data-points for a multivariate Mantel correlogram in the sense of Oden & Sokal (1986) (see also Legendre & Legendre 1998 for a comprehensive coverage of the subject).

Value

Returns different objects, depending on given arguments and triggers.

<code>case 1</code>	a <code>permc</code> -object with the following items is returned:
<code>call</code>	The call to the function.
<code>method</code>	The correlation method as used by <code>cor.test</code> .
<code>statistic</code>	The initial correlation value which is tested against the permuted values.

- signif The significance of the calculation.
- n The number of cases.
- permutations The number of permutations as specified by `permutations`.
- perms The result of the permuted runs. It is not printed by default but can be accessed via `result$perms`. The correlation value can be plotted against an histogram of the distribution of the permuted values to visualize the significance with the default plotting method.
- case 2 a `pclist`-object with the following items (in this case it might be nice to set `trace=TRUE` to display the progress of the calculation) is returned:
 - call The call to the function.
 - method The correlation method as used by `cor.test`.
 - gesN The total number of cases.
 - strata The number of strata (or levels) for which permutation has been done.
 - permutations The number of permutations as specified by `permutations`.
 - out A `data.frame` with 3 columns containing the result for each stratum in the rows: `statistic` contains the correlation values for the corresponding stratum, `sig` the obtained significance, and `nop` the number of cases found and used for permutation on this very level.
- case 3 a `pclist`-object with the same items as in Case 2 (in this case it might be as well interesting to set `trace=TRUE` to display the progress of the calculation) is returned.

Note

Depending on the background-function and the size of the matrices or vectors it may take a while to calculate. The slowest is `mancor` (case 3).

Author(s)

Gerald Jurasinski

References

- Legendre, P. & Legendre, L. (1998) *Numerical Ecology*. 2nd English Edition. Elsevier.
- Oden, N. L. & Sokal, R. R. (1986) Directional Autocorrelation: An Extension of Spatial Correlograms to Two Dimensions. *Systematic Zoology* **35**: 608-617.

See Also

[mantel](#) for a different implementation of Mantel tests, and [mantell](#), [cor.test](#)

Examples

```
data(abis)

## calculate soerensen of species data
```

```
abis.soer <- sim(abis.spec)
## calculate distance (Euclidean) regarding some disturbance
## variables (feces counts)
abis.pert <- dist(abis.env[,19:25])
## are compositional similarity and dissimilarity of disturbance related?
pcol(abis.soer, abis.pert)
## the relationship is significant, but not very strong
```

plot.mrpp

plot an mrpp-object

Description

There is no mrpp plotting function in the vegan package. For convenience it is provided here.

Usage

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

Arguments

x	a mrpp-object as computed with mrpp
y	Optional plotting object, not needed here.
...	Arguments to the plotting function

Value

Returns a histogram with the distribution of the permutation values plotted against delta: For explanations see [mrpp](#)

Author(s)

Gerald Jurasinski

See Also

[mrpp](#), [anosim](#)

sim *Calculate similarities for binary vegetation data*

Description

One of 56 (dis)similarity measures for binary data can be set to calculate (dis)similarities. The vegetational data can be in either database (list) or matrix format. Same holds for the output. Simultaneous calculation of geographical distances between plots and the virtual position of the calculated similarity values between the parental units can be achieved if a `data.frame` with coordinates is given.

Usage

```
sim(x, coord=NULL, method = "soer", dn=NULL, normalize = FALSE,
    listin = FALSE, listout = FALSE, ...)
```

Arguments

x	Vegetation data, either as matrix with rows = plots and columns = species (similarities are calculated between rows!), or as <code>data.frame</code> with first three columns representing plots, species and occurrence information respectively. All further columns are dumped before calculation. Occurrence is only considered as binary. If your list or matrix contains abundances or frequencies they are transformed automatically.
coord	A <code>data.frame</code> with two columns containing the coordinate values of the sampling units. If given, it triggers the simultaneous calculation of the geographical distances between the sampling units, the coordinates of virtual centre-points between all possible pairs of plots, and the geographical distances in either x- or y-direction. If <code>coord</code> is given, output is always in database format (no matrix).
method	Binary Similarity index (see Details for references and formulae), partial match to "soerensen", "jaccard", "ochiai", "mountford", "whittaker", "lande", "wilsonshmida", "cocogaston", "magurran", "harrison", "cody", "williams", "williams2", "harte", "simpson", "lennon", "weiher", "ruggiero", "lennon2", "rout1ledge", "rout2ledge", "rout3ledge", "sokal1", "dice", "kulczlinsky", "kulcz2insky", "mcconnagh", "manhattan", "simplematching", "margaleff", "pearson", "roger", "baroni", "dennis", "fossum", "gower", "legendre", "sokal2", "sokal3", "sokal4", "stiles", "yule", "michael", "hamann", "forbes", "chisquare", "peirce", "eyraud", "simpson2", "legendre2", "fager", "maarel", "lamont", "johnson", "sorgenfrei", "johnson2". See details.
dn	Neighbor definition. A geographic distance represented by a numeric or a two value vector defining a ring around each plot. Only takes effect when <code>coord != NULL</code> . If specified, the output does only contain similarities between neighboring plots. A plot is a neighbour to any given plot if it is within the range of the neighbor definition. See details.

<code>normalize</code>	Logical value indicating whether the values for <code>a</code> , <code>b</code> and <code>c</code> which are calculated in the process should be normalized to 100% (per row, which means per plot comparison). If <code>normalize = TRUE</code> an asymmetric index must be chosen (see details).
<code>listin</code>	if <code>x</code> is given in database (list) format this must be set to <code>TRUE</code> (there is no automatic detection of the format)
<code>listout</code>	If output is wanted in database format rather than as a <code>dist</code> -object set this to <code>TRUE</code> . Output is automatically given in database-format, when <code>coord</code> is specified.
<code>...</code>	Arguments to other functions

Details

All binary similarity indices are based on the variables `a`, `b` and `c` (or can be expressed as such). Some of them also use `d`. Where `a` is the number of species shared by two compared plots, `b` is the number of species found only in one of the compared plots, and `c` is the number of species only found in the other of the compared plots. `d` refers to species which are absent from both the compared plots but present in the whole dataset. Indices incorporating `d` are discussed critically by Legendre & Legendre (1998) and elsewhere. They are called symmetric and expose a "double zero" problem as they take species into account which are absent from both compared units. Absence of species from a sampling site might be due to various factors, it does not necessarily reflect differences in the environment. Hence, it is preferable to avoid drawing ecological conclusions from the absence of species at two sites (Legendre & Legendre 1998). The indices presented here come from various sources as indicated. Comparative reviews can be found in e.g. Huhta (1979), Wolda (1981), Janson & Vegelius (1981), Shi (1993), Koleff et al. (2003), Albatineh (2006)

The indices considerably differ in their behaviour. For classification purposes and in ecology, Jaccard and Sørensen have been found to give robust and meaningful results (e.g. Janson & Vegelius 1981, Shi 1993). For other purposes other indices might be better suited. However, you are invited to use (at least with the asymmetric indices) ternary plots as suggested by Koleff et al. 2003. The matching components `a`, `b`, and `c` can be displayed in a `ternary.plot` to evaluate the position of the plots in similarity space. When output is in database-format, the matching components are always given and `triax.plot` can be used to plot them into a triangle-plot. Koleff et al. (2003) used an artificial set of matching components including all possibilities of values that `a`, `b`, and `c` can take from 0 to 100 to display the mathematical behavior of indices. An artificial data-set with this properties - together with the values for the asymmetric indices included here - is part of this package (`ads.ternaries`) and can be used to study the behavior of the indices prior to analysis. See details and examples there.

If `coord` is given, the geographic distances between plots/sampling units are calculated automatically, which may be of value when the display or further analyses of distance decay (sensu Tobler 1970, Nekola & White 1999) is in focus. For convenience the `dn`-trigger can be used to tell the function to only return similarities calculated between neighboring plots. Similarities between neighboring plots in an equidistant array are not subjected to the problem of auto-correlation because all plots share the same distance (Juraskinski & Beierkuhnlein 2006). Therefore, any variation occurring in the data are most likely caused by environmental differences alone.

In the following formulae...

`a` = number of shared species

b = number of species only found on one of the compared units

c = number of species only found on the other of the compared units

d = number of species not found on the compared plots but in the dataset

$$N = a + b + c + d$$

with ($n_1 \leq n_2$)...

n_1 = number of species of the plot with fewer species ($a + b$) or ($a + c$)

n_2 = number of species of the plot with more species ($a + b$) or ($a + c$)

Computable asymmetric indices:

soerensen	$sim = \frac{2a}{2a+b+c}$	Soerensen (1948)
jaccard	$sim = \frac{a}{a+b+c}$	Jaccard (1912)
ochiai	$sim = \frac{a}{\sqrt{(a+b)(a+c)}}$	Ochiai (1957), Shi (1993)
mountford	$sim = \frac{2a}{(a(b+c)+2bc)}$	Mountford (1962), Shi (1993)
whittaker	$sim = \frac{a+b+c}{2a+b+c} - 1$	Whittaker (1960), Magurran (1988)
lande	$sim = \frac{b+c}{2}$	Lande (1996)
wilsonshmid	$sim = \frac{b+c}{2a+b+c}$	Wilson & Shmida (1984)
cocogaston	$sim = \frac{b+c}{a+b+c}$	Colwell & Coddington (1948), Gaston et al. (2001)
magurran	$sim = (2a + b + c)(1 - \frac{a}{a+b+c})$	Magurran (1988)
harrison	$sim = \frac{\min(b,c)}{\max(b,c)+a}$	Harrison et al. (1992), Koleff et al. (2003)
cody	$sim = 1 - \frac{a(2a+b+c)}{2(a+b)(a+c)}$	Cody (1993)
williams	$sim = \frac{\min(b,c)}{a+b+c}$	Williams (1996), Koleff et al. (2003)
williams2	$\frac{(bc)+1}{(a+b+c)^2-(a+b+c)}$	Williams (1996), Koleff et al. (2003)
harte	$1 - \frac{2a}{2a+b+c}$	Harte & Kinzig (1997), Koleff et al. (2003)
simpson	$\frac{\min(b,c)}{\min(b,c)+a}$	Simpson (1949), Koleff et al. (2003)
lennon	$\frac{2 b-c }{2a+b+c}$	Lennon et al. (2001), Koleff et al. (2003)
weiher	$sim = b + c$	Weiher & Boylen (1994)
ruggiero	$sim = \frac{a}{a+c}$	Ruggiero et al. (1998), Koleff et al. (2003)
lennon2	$sim = 1 - \left[\frac{\log\left(\frac{2a+b+c}{a+b+c}\right)}{\log 2} \right]$	Lennon et al. (2001), Koleff et al. (2003)
routledge	$sim = \frac{(a+b+c)^2}{(a+b+c)^2-2bc}$	Routledge (1977), Magurran (1988)
routledge2	<i>toolong, seebelow</i>	Routledge (1977), Wilson & Shmida (1984)
routledge3	$sim = e^{routledge} - 1$	Routledge (1977)
sokal1	$sim = \frac{a}{a+2(b+c)}$	Sokal & Sneath (1963)
dice	$sim = \frac{a}{\min((b+a),(c+a))}$	Association index of Dice (1945), Wolda (1981)
kulczlinsky	$sim = \frac{a}{b+c}$	Oosting (1956), Southwood (1978)
kulcz2insky	$sim = \frac{\frac{a}{2}(2a+b+c)}{(a+b)(a+c)}$	Oosting (1956), Southwood (1978)
mcconnagh	$sim = \frac{a^2-bc}{(a+b)(a+c)}$	Hubalek (1982)
simpson2	$sim = \frac{a}{a+b}$	Simpson (1960), Shi (1993)
legendre2	$sim = \frac{3a}{3a+b+c}$	Legendre & Legendre (1998)
fager	$sim = \frac{a}{\sqrt{n_1 n_2}} - \frac{1}{2*\sqrt{n_2}}$	Fager (1957), Shi (1993)
maarel	$sim = \frac{2a-(b+c)}{2a+b+c}$	van der Maarel (1969)
lamont	$sim = \frac{a}{2a+b+c}$	Lamont and Grant (1979)

johnson	$sim = \frac{a}{2b}$	Johnson (1971)
sorgenfrei	$sim = \frac{a^2}{(a+b)(a+c)}$	Sorgenfrei (1959)
johnson2	$sim = \frac{a}{a+b} + \frac{a}{a+c}$	Johnson (1967)

Computable symmetric indices (including unshared species):

manhattan	$sim = \frac{b+c}{a+b+c+d}$	Mean Manhattan, Legendre & Legendre (1998)
simplematching	$sim = \frac{a+d}{a+b+c+d}$	Sokal & Michener 1958
margaleff	$sim = \frac{a(a+b+c+d)}{(a+b)(a+c)}$	Clifford & Stevenson (1975)
pearson	$sim = \frac{ad-bc}{\sqrt{(a+b)(a+c)(d+b)(d+c)}}$	Phi of Pearson, Gower & Legendre (1986)
roger	$sim = \frac{a+d}{a+2(b+c)+d}$	Rogers & Tanimoto (1960), Gower & Legendre (1986)
baroni	$sim = \frac{\sqrt{ad+c}}{\sqrt{ad+a+b+c}}$	Baroni-Urbani & Buser (1976), Wolda (1981)
dennis	$sim = \frac{ad-bc}{\sqrt{(a+b+c+d)(a+b)(a+c)}}$	Holliday et al. (2002)
fossum	$sim = \frac{(a+b)(a+c)}{(a+b+c+d)(-\frac{a}{2})^2}$	Holliday et al. (2002)
gower	$sim = \frac{a-(b+c)+d}{a+b+c+d}$	Gower & Legendre (1986)
legendre	$sim = \frac{a}{a+b+c+d}$	Gower & Legendre (1986)
sokal2	$sim = \frac{ad}{\sqrt{(a+b)(a+c)(d+b)(d+c)}}$	Sokal & Sneath (1963)
sokal3	$sim = \frac{2a+2d}{(a+d+(a+b+c+d))}$	Sokal & Sneath (1963)
sokal4	$sim = \frac{a+d}{b+c}$	Sokal & Sneath (1963)
stiles	$sim = \frac{(a+b+c+d)(ad-bc - \frac{a+b+c+d}{2})^2}{(a+b)(a+c)(b+d)(c+d)}$	Stiles (1946)
yule	$sim = \frac{ad-bc}{ad+bc}$	Yule & Kendall (1973)
michael	$sim = \frac{4(ad-bc)}{(a+b)^2+(b+c)^2}$	Michael (1920), Shi (1993)
hamann	$sim = \frac{(a+d)-(b+c)}{N}$	Hamann (1961)
forbes	$sim = \frac{(aN-2n_2)}{(Nn_1-2n_2)}$	Forbes (1925), Shi (1993)
chisquare	$sim = \frac{2(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)}$	Yule & Kendall (1950)
peirce	$sim = \frac{(ad-bc)}{(a+c)(b+d)}$	Peirce (1984) in Shi (1993)
eyraud	$sim = \frac{a-(a+b)(a+c)}{(a+b)(a+c)(b+d)(c+d)}$	Eyraud (1936) in Shi (1993)

routledge formula (Routledge, 1977; Koleff et al. 2003):

$$\beta_{R2} = \log(2a+b+c) - \left(\frac{1}{2a+b+c} 2a \log 2 \right) - \left(\frac{1}{2a+b+c} ((a+b) \log(a+b) + (a+c) \log(a+c)) \right)$$

Value

If `listout = FALSE` a distance matrix of class `dist` is returned. If `listout = TRUE`, a `data.frame` is returned with 7 columns giving the names of the compared plots in the first two and the calculated similarity measure in the third column. The rest of the columns give the values for `a`, `b`, `c`, and `d` (in this order). Naming of the first three columns can be changed but defaults to `NBX` (one of the compared plots), `NBY` (the other one), `used_index` (the values of the calculated index). If `coord != NULL`, the following columns are given in addition and the columns `a:d` shift to the end of the `data.frame`.

`distance` Geographical distance between compared plots

X	For plotting purposes, the x-coordinate of the virtual position of the calculated similarity value in the center between the two compared plots
Y	For plotting purposes, the y-coordinate of the virtual position of the calculated similarity value in the center between the two compared plots
xdist	Geographical distance between compared plots, on the x-axis only
ydist	Geographical distance between compared plots, on the y-axis only

Note

In general, concepts of data-handling are taken from [vegdist](#) and the calculation of a, b, c and d is taken from [dist.binary](#). Thanks to Jari Oksanen for his `vegan` package. The indices were collected from the literature and are applicable in different fields of research.

Author(s)

Gerald Jurasinski (terhorab@mac.com)

References

- Albatineh, A. N., Niewiadomska-Bugaj, M. & Mihalko, D. (2006) On Similarity Indices and Correction for Chance Agreement. *Journal of Classification* **V23**: 301-313.
- Baroni-Urbani, C. & Buser, M. W. (1976) Similarity of Binary Data. *Systematic Zoology* **25**: 251-259.
- Clifford, H. T. & Stephenson, W. (1975) *An introduction to numerical classification*. Academic Press, New York, San Francisco, London.
- Cody, M. L. (1993) Bird diversity components within and between habitats in Australia. - In: Ricklefs, R. E. & Schluter, D. (eds.), *Species Diversity in Ecological Communities: historical and geographical perspectives*, pp. 147-158, University of Chicago Press, Chicago
- Colwell, R. K. & Coddington, J. A. (1994) Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **345**: 101-118.
- Dice, L. R. (1945) Measures of the amount of ecological association between species. *Ecology* **26**: 297-302.
- Fager, E. W. (1957) Determination and analysis of recurrent groups. *Ecology* **38**: 586-595.
- Faith, D. P., Minchin, P. R. & Belbin, L. (1987) Compositional dissimilarity as a robust measure of ecological distance. *Plant Ecology* **69**: 57-68.
- Gaston, K. J., Rodrigues, A. S. L., van Rensburg, B. J., Koleff, P. & Chown, S. L. (2001) Complementary representation and zones of ecological transition. *Ecology Letters* **4**: 4-9.
- Gower, J. C. & Legendre, P. (1986) Metric and Euclidean properties of dissimilarity coefficients. *Journal of Classification* **3**: 5-48.
- Hajdu, L. J. (1981) Graphical comparison of resemblance measures in phytosociology. *Plant Ecology* **V48**: 47-59.
- Harrison, S., Ross, S. J. & Lawton, J. H. (1992) Beta diversity on geographic gradients in Britain. *Journal of Animal Ecology* **61**: 151-158.

- Harte, J. & Kinzig, A. (1997) On the implications of species-area relationships for endemism, spatial turnover and food web patterns. *Oikos* **80**.
- Holliday, J. D., Hu, C.-Y. & Willett, P. (2002) Grouping of Coefficients for the Calculation of Inter-Molecular Similarity and Dissimilarity using 2D Fragment Bit-Strings. *Combinatorial Chemistry & High Throughput Screening* **5**: 155-166.
- Hubalek, Z. (1982) Coefficients of association and similarity, based on binary (presence-absence) data: An evaluation. *Biological Reviews of the Cambridge Philosophical Society* **57**: 669-689.
- Huhta, V. (1979) Evaluation of different similarity indices as measures of succession in arthropod communities of the forest floor after clear-cutting. *Oecologia* **V41**: 11-23.
- Jaccard, P. (1901) Etude comparative de la distribution florale d'une portion des Alpes et du Jura. *Bulletin de la Société Vaudoise des Sciences Naturelles* **37**: 547-579.
- Jaccard, P. (1912) The distribution of the flora of the alpine zone. *New Phytologist* **11**: 37-50.
- Johnson, J. G. (1971) A quantitative approach to faunal province analysis. *American Journal of Science* **270**: 257-280.
- Johnson, S. C. (1967) Hierarchical clustering schemes. *Psychometrika* **32**: 241-254.
- Jurasinski, G. & Beierkuhnlein, C. (2006) Spatial patterns of biodiversity - assessing vegetation using hexagonal grids. *Proceedings of the Royal Irish Academy - Biology and Environment* **106B**: 401-411.
- Jurasinski, G. & Beierkuhnlein, C. (submitted) Distance decay and non-stationarity in a semi-arid Mediterranean ecosystem. *Journal of Vegetation Science*.
- Koleff, P., Gaston, K. J. & Lennon, J. J. (2003) Measuring beta diversity for presence-absence data. *Journal of Animal Ecology* **72**: 367-382.
- Lamont, B. B. & Grant, K. J. (1979) A comparison of twenty-one measures of site dissimilarity. - In: Orlóci, L., Rao, C. R. & Stiteler, W. M. (eds.), *Multivariate Methods in Ecological Work*, pp. 101-126, Int. Coop. Publ. House, Fairland, MD
- Lande, R. (1996) Statistics and partitioning of species diversity and similarity along multiple communities. *Oikos* **76**: 25-39.
- Legendre, P. & Legendre, L. (1998) *Numerical Ecology*. Elsevier, Amsterdam.
- Lennon, J. J., Koleff, P., Greenwood, J. J. D. & Gaston, K. J. (2001) The geographical structure of British bird distributions: diversity, spatial turnover and scale. *J Anim Ecology* **70**: 966-979.
- Magurran, A. E. (1988) *Ecological Diversity and its Measurement*. Chapman & Hall, London.
- Mountford, M. D. (1962) An index of similarity and its application to classification problems. - In: Murphy, P. W. (ed.) *Progress in Soil Zoology*, pp. 43-50, Butterworths, Oosting, H. J. (1956) *The study of plant communities: an introduction to plant ecology*. W. H. Freeman, San Francisco.
- Rogers, D. J. & Tanimoto, T. T. (1960) A computer program for classifying plants. *Science* **132**: 1115-1118.
- Routledge, R. D. (1977) On Whittaker's components of diversity. *Ecology* **58**: 1120-1127.
- Ruggiero, A., Lawton, J. H. & Blackburn, T. M. (1998) The geographic ranges of mammalian species in South America: spatial patterns in environmental resistance and anisotropy. *Journal of Biogeography* **25**: 1093-1103.
- Shi, G. R. (1993) Multivariate data analysis in palaeoecology and palaeobiogeography—a review. *Palaeogeography, Palaeoclimatology, Palaeoecology* **105**: 199-234.

- Simpson, E. H. (1949) The measurement of diversity. *Nature* **163**: 688.
- Simpson, G. G. (1960) Notes on the measurement of faunal resemblance. *American Journal of Science* **258-A**: 300-311.
- Sokal, R. R. & Michener, C. D. (1958) A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin* **38**: 1409-1438.
- Sokal, R. R. & Sneath, P. H. A. (1963) *Principles of numerical taxonomy*. W. H. Freeman, San Francisco.
- Sørensen, T. (1948) A method of establishing groups of equal amplitude in plant sociology based on similarity of species content. *Biologiske Skrifter* **5**: 1-34.
- Sorgenfrei, T. (1959) Molluscan assemblages from the marine middle Miocene of South Jutland and their environments. *Danmark Geologiske Undersøgelse. Serie 2* **79**: 403-408.
- Southwood, T. S. (1978) *Ecological Methods*. Chapman and Hall, London.
- Weiher, E. & Boylen, C. W. (1994) Patterns and prediction of alpha and beta diversity of aquatic plants in Adirondack (New York) lakes. *Canadian Journal of Botany-Revue Canadienne De Botanique* **72**: 1797-1804.
- Whittaker, R. H. (1960) Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs* **30**: 279-338.
- Williams, P. H. (1996) Mapping variations in the strength and breadth of biogeographic transition zones using species turnover. *Proceedings of the Royal Society of London Series B-Biological Sciences* **263**: 579-588.
- Williams, P. H., Klerk, H. M. & Crowe, T. M. (1999) Interpreting biogeographical boundaries among Afrotropical birds: spatial patterns in richness gradients and species replacement. *J Biogeography* **26**: 459-474.
- Wilson, M. V. & Shmida, A. (1984) Measuring beta-diversity with presence-absence data. *Journal of Ecology* **72**: 1055-1064.
- Wolda, H. (1981) Similarity indices, sample size and diversity. *Oecologia* **50**: 296-302.
- Yule, G. U. & Kendall, M. G. (1973) *An introduction to the theory of statistics*. Griffin, London.

See Also

[vegdist](#), [dist.binary](#), [dsvdis](#), [dist](#)

Examples

```
data(abis)
##calculate jaccard similarity and output as dist-object
jacc.dist <- sim(abis.spec, method="jaccard")

##calculate Whittaker similarity (with prior normalisation) and
##output as data.frame
whitt.list <- sim(abis.spec, method="whittaker", normalize=TRUE,
listout=TRUE)

##calculate similarity from a database list after Harte & Kinzig (1997)
##and output as dist-object
abis.spec.ls <- liste(abis.spec, splist=TRUE)
```

```

hart.dist <- sim(abis.spec.ls, method="harte", listin=TRUE)

## calculate the geographic distances between sites simultaneously
## and return only similarities calculated between neighboring plots
abis.soer <- sim(abis.spec, coord=abis.env[,1:2], dn=100)

## in an equidistant array
## you can plot this nice between the original positions of the
## sites (with the size of the dots expressing number of species
## for the sites, and value of the Sørensen coefficient in between)
require(geoR)
points.geodata(coord=abis.env[,1:2], data=abis.env$n.spec,
cex.min=1, cex.max=5)
points.geodata(coord=abis.soer[,5:6], data=abis.soer$soerensen,
cex.min=1, cex.max=5, col="grey50", add=TRUE)

```

sim.het

Calculate a mean similarity to adjacent cells in an array

Description

Useful for systematic sampling grids for calculating the mean similarity and its standard deviation between a focal unit and its surrounding units. Surrounding can be specified by the user.

Usage

```
sim.het(mat, coord=NULL, dn, method="soerensen", test=TRUE,
permutations=100, ...)
```

```
het2nbs(mat, coord=NULL, dn, method="soerensen", ...)
```

Arguments

mat	Species matrix or <code>dist</code> object resulting from similarity calculation.
coord	A <code>data.frame</code> with two columns containing the coordinates of the plots for which species data or a <code>dist</code> matrix is given. Defaults to <code>NULL</code> . Then, mean similarity and standard deviation of similarities from each plot to all other plots is calculated
dn	A numeric or a two value vector. In the first case it gives the distance until which sampling units should be considered as neighbors. In the second the two values define a ring around each plot. Plots which fall into the ring are considered as neighbors. Not necessary and skipped when <code>coord=NULL</code>
method	Only necessary when a species matrix is given. Specifies the method for calculating the similarity between plots with <code>sim</code> . For more details see the documentation of <code>sim</code> .
test	Shall significance be calculated? Defaults to <code>TRUE</code> which means that it is tested if the observed pattern is significantly different from random

`permutations` Number of permutations for the significance test. Species are permuted on plots and the index is recalculated for each permuted species matrix. The initial values are then tested against the permuted values to obtain a significance level.

`...` Further arguments passed to `sim`.

Details

If similarities (e.g. of species composition) are calculated between plots in a regular or irregular array (whereas the latter might be problematic because of spatial auto-correlation) the resulting values have no position. However they could be associated to a point in the centre between the compared plots (for each compared pair). Sometimes it is wanted to associate a dissimilarity value to the focal unit which comprises the similarities/dissimilarities of a focal unit to all surrounding units. One possibility is the calculation of a mean similarity (Williams 1996, Lennon 2001, Gaston 2001) out of the single similarities between a focal unit and its surrounding units. This might be problematic as the single similarity values might result from different sets of species. To include the variability of the single similarities between a focal unit and its neighbours as a measure of spatial heterogeneity the variation (sd) in the similarity values is included. See `sim.pat` for the calculation of similarity between a focal unit and its neighbours whilst preserving species identity.

Significance is tested against random expectations with a permutation procedure. After calculating the values for `mean` and `sd` the species/similarity matrix is permuted and the values are calculated again. This is done `permutation` times. Then the initial values are tested against the obtained distribution. If the initial values are under the mean of the respective values among plots they are tested against the lower tail of the permuted distribution. If they exceed the mean, they are tested against the upper tail of the permuted distribution. If a value is significant, this means that it is significantly different from a random distribution of species and therefore might likely be caused by underlying environmental patterns.

Value

A `data.frame` with the following columns

<code>nnbs</code>	Number of neighbours of the respective plot. Results from the neighbour definition <code>dn</code> .
<code>m.sims</code>	Mean similarity between plot and its neighbours.
<code>sd.sims</code>	Standard deviation of the similarity values calculated between the plot and its neighbours.
<code>sim.test.mean</code>	The p-value of the significance test for the <code>mean</code> -values.
<code>sig.mean</code>	The significance decision tested against $\alpha = 0.05$. "*" indicates, that mean value is significantly different from random.
<code>sig.prefix.mean</code>	Marks against which tail significance of <code>mean</code> was tested. If the initial similarity value is lower than the mean of the permuted values the lower tail is tested (<code>sig.prefix = "-"</code>) and vice versa (<code>sig.prefix = "+"</code>).
<code>sim.test.sd</code>	The p-value of the significance test for the <code>sd</code> -values.
<code>sig.sd</code>	The significance decision tested against $\alpha = 0.05$. "*" indicates, that <code>sd</code> value is significantly different from random.

sig.prefix.sd

Marks against which tail significance of sd was tested. If the initial similarity value is lower than the mean of the permuted values the lower tail is tested (sig.prefix = "-") and vice versa (sig.prefix = "+").

Author(s)

Gerald Jurasinski (terhorab@mac.com); Vroni Retzer (vroni.retzer@uni.bayreuth.de)

References

Gaston, K. J., Rodrigues, A. S. L., van Rensburg, B. J., Koleff, P. & Chown, S. L. (2001) Complementary representation and zones of ecological transition. *Ecology Letters* **4**: 4-9.

Jurasinski, G. & Retzer, V. in prep. Measuring multi-plot similarity with presence-absence data. *Ecology*

Lennon, J. J., Koleff, P., Greenwood, J. J. D. & Gaston, K. J. (2001) The geographical structure of British bird distributions: diversity, spatial turnover and scale. *Journal of Animal Ecology* **70**: 966-979.

Williams, P. H. (1996) Mapping variations in the strength and breadth of biogeographic transition zones using species turnover. *Proceedings of the Royal Society of London Series B-Biological Sciences* **263**: 579-588.

See Also

See also [sim.pat](#) and [sim](#)

Examples

```
data(abis)
## calculate average similarity for the focal plots
abis.het <- sim.het(abis.spec, coord=abis.env[,1:2], dn=100)
```

sim.pat

Calculate pattern diversity

Description

The function `sim.pat` uses `sima` to calculate compositional similarity between a focal sampling unit and all units in the data-set and `simn` to calculate compositional similarity between a focal unit and its surrounding units. Surrounding can be specified as a radius or a ring. Significance can be tested.

Usage

```
sim.pat(veg, coord = NULL, dn, presabs = TRUE, test = TRUE,
        permutations = 100, ...)
```

```
sima(veg, presabs = TRUE, d.inc = FALSE, ...)
```

```
simn(veg, coord, dn, presabs = TRUE, d.inc = FALSE, ...)
```

Arguments

veg	A (species) matrix with rows representing plots/sampling units and columns representing species/variables. Up to now all values are transformed to presence/absence: all values > 0 are transformed to 1, all other values are 0.
coord	A two column <code>data.frame</code> defining the geographic positions of the sampling units. Should give (in this order) the x- and y-values in UTM or similar coordinates. Defaults to <code>NULL</code> . Triggers also if similarity to neighbours (when it is given) or to all units in the data-set (default) should be calculated.
dn	Neighbour definition. A positive numeric or a two value vector. In the first case it gives the distance from each sampling unit until which other sampling units should be seen as neighbours. In the second the two values define a ring around each plot. Plots which fall into the ring are considered as neighbors. See details. If <code>coord</code> is not given it becomes obsolete.
presabs	Triggers the calculation method. Should only the presence in the focal unit (<code>FALSE</code>) or also the absence in the focal unit (<code>TRUE</code>) be regarded?
d.inc	If species not occurring on the focal plot or its surroundings - but in the rest of the data-set should be considered, set it to <code>TRUE</code> . However, default behaviour is to do not. Note that when calling <code>sim.pat</code> without coordinates or <code>sima</code> this setting takes not effect unless there are non-occurring species in the matrix.
test	Should significance be tested? Defaults to <code>true</code> . See details.
permutations	If <code>test=TRUE</code> , how many permutations of the species matrix should be carried out? Defaults to 100. See details.
...	Further arguments to functions.

Details

The idea behind the function is the extension of available similarity measures by the possibility to calculate similarities between a focal unit and several other (surrounding) units at once whilst preserving species identities. In an equidistant grid one could also calculate the similarities between a focal unit and its six neighbours and then calculate a mean similarity value. However, this might be problematic as the values are based on the presence of different species. So the same similarity value can derive from a different set of species.

`sim.pat` is different as it respects species identities of all included units at once. It has two modes. If a coordinate file and a neighbour definition is given it calculates the similarity between each unit and its surrounding units. Neighbours are all units which fall into the specified radius or ring of radiuses (`dn`). On an equidistant grid the distance between grid units can simply be given if first hand neighbours shall be included, otherwise a ring has to be specified with a two value vector.

If there is no `coord`, `dn` is obsolete. Then the similarity from each unit to all other units in the data-set is calculated.

The significance of the obtained similarity value can be tested in both cases. In the neighbour-case it is tested against a randomly rearranged species matrix `veg`. For each permuted matrix, `simn` is calculated for each unit and stored. The initial value is then tested against the resulting random values. If the initial value is lower than the mean of the permuted values (for each unit), the function looks at the lower end. If it is higher, the upper tail is tested. The test direction is given in the results file. Thus, for each unit it is known, whether the pattern is a deviation from random, if it is lower or higher than random, and if it is significantly different from random. In the all-case it is virtually the same but as a rearrangement of species on plots would change nothing, the function tests against an artificially produced data-set. This is done with `ads` and the key parameters resemble the original species matrix `veg`.

Value

Returns a `data.frame` with the similarity values calculated between the focal unit and its neighbours (`sim2nbs`) or all other units in the data-set (`sim2all`). In case of the first, the two first columns of the `data.frame` give the number of species on the focal plot (`n.spec`) and the number of neighbours (`nbs`) to which similarity was calculated based on the neighbour definition (`dn`).

If `test = TRUE` the following columns are given in addition:

<code>sim.test</code>	The p-value of the significance test.
<code>sig</code>	The significance decision, tested against $\alpha = 0.05$. "*" means similarity value is significantly different from random.
<code>sig.prefix</code>	Marks against which tail significance was tested. If the initial similarity value is lower than the mean of the permuted values the lower tail is tested (<code>sig.prefix = "-"</code>) and vice versa (<code>sig.prefix = "+"</code>).

Author(s)

Gerald Jurasinski <terhorab@mac.com> <http://homepage.mac.com/terhorab/gerald/>,
Vroni Retzer <vroni.retzer@uni-bayreuth.de>

References

Jurasinski, G. & Retzer, V. in prep. Measuring multi-plot similarity with presence-absence data. *Ecology* in prep.

See Also

`sim` from this package and the following functions of other packages for the calculation of similarities between two sites: `vegdist`, `dist.binary`, `dsvdis`, `dist`.

 sim.relt

 Calculate additive and relative partitioning.

Description

Enables the calculation of three β -diversity measures closely related to each other: additive partitioning, multiplicative partitioning (Whittaker 1960), reverse multiplicative partitioning (Whittaker 1960, full change notion).

Usage

```
sim.relt(veg, coord = NULL, dn, method = "additive", test = TRUE,
         permutations = 100, ...)
```

```
sim.rel(veg, coord=NULL, dn, method="additive", ...)
```

Arguments

veg	A species matrix. The function takes only presence/absence into account. However the translation is done automatically by assuming that zero means no occurrence, whereas any value > 0 is transformed to 1
coord	A two column <code>data.frame</code> defining the geographic positions of the sampling units. Should give (in this order) the x- and y-values in UTM or similar coordinates. Defaults to NULL. Triggers also if the similarity measures should only incorporate a focal plot and its neighbours (moving window) or if all units in the data-set (default) should be considered.
dn	A numeric or a two value vector. In the first case it gives the distance until which sampling units should be considered as neighbors. In the second the two values define a ring around each plot. Plots which fall into the ring are considered as neighbors. If <code>coord</code> is not given it becomes obsolete.
method	Which similarity measure should be calculated. Give a partial match to "additive", "relative" or "reverse". See details.
test	Shall significance be tested? Defaults to true. Does only make sense when <code>coord</code> and <code>dn</code> is given. In general <code>sim.relt</code> calls <code>sim.rel</code> to calculate the values. So the latter will rarely be used by the user directly.
permutations	The number of permutations. Defaults to 100 which allows for a p-level of 0.01
...	Arguments to other functions.

Details

Three different diversity measures related to the concept of β -diversity can be calculated with this function. Besides the definition of β -diversity as the turnover in species composition (Whittaker 1960), Whittaker defined the *relation* between γ - and α -diversity as β -diversity where α is the mean species richness found on local plots or sampling units and γ -diversity is the species richness throughout a reference area (e.g. an investigation area). However, γ could also be defined for

smaller units e.g. for a focal plot and its neighbours, which is done in the function when `coord` and `dn` are set. β -diversity in terms of *multiplicative* diversity is calculated when `method` is set to "relative". "reverse" simply means that the relation is calculated vice versa ($\frac{\alpha}{\gamma}$) which results in numbers between 0 and 1 whereas the original formula gives results from 1 ($\bar{\alpha} = \gamma$) upwards. An upper end is not defined. However, increasing values indicate increasing heterogeneity in species composition.

The idea of "additive partitioning" (Lande 1996, Vellend 2001, Veech 2002) takes Whittakers approach further. Out of the criticism that β in Whittakers sense does not exhibit the same units (species numbers) as α - and γ -diversity here β -diversity results as the subtract of γ - and α -diversity. It expresses the average amount of diversity not found in a single, randomly-chosen sample. It is also rather calculated for a whole data-set. Here we apply it as well to a moving window of a focal plot and its neighbours if `coord` and `dn` are set.

The idea of γ -diversity might be questioned in general, as its quality is not different from α . Only the geographic extent is changed and often definition becomes problematic. Imagine a temporal study where different numbers of species are found throughout the years - Is γ then the overall species richness, or the species richness in one year? Furthermore β -diversity is not clearly defined. There are even more definitions to it then mentioned here (e.g. Qian et al. 2005) so it may be better to use 'differentiation-diversity' instead.

Significance is tested with a simple Monte-Carlo procedure. The initial value of the respective index is tested against a number of values which are calculated from a random reshuffling of the original species matrix. So the hypothesis tested is, that the observed pattern (for each focal plot) is different from random. This is meaningless when `coord` and `dn` are not set.

Value

The function returns either a named vector (if `coord` and `dn` are set) or a `data.frame` with the following items. The last three are only added if `test = TRUE`.

<code>nnbs</code>	Number of neighbours of the respective plot. Results from the neighbour definition <code>dn</code> . If this was not given it reflects the total number of plots in the data-set.
<code>n.spec</code>	Number of species on the focal plot. If <code>dn</code> and <code>coord</code> were not given it reflects the total number of species in the data-set
<code>value</code>	The calculated β -diversity. The name of the column changes respectively.
<code>sim.test</code>	The p-value of the significance test for the diversity values.
<code>sig</code>	The significance decision tested against $\alpha = 0.05$. "*" means that the respective β -value is significantly different from random.
<code>sig.prefix</code>	Marks against which tail significance of β -values was tested. If the initial value is lower than the mean of the permuted values the lower tail is tested (<code>sig.prefix = "-"</code>) and vice versa (<code>sig.prefix = "+"</code>).

Author(s)

Gerald Jurasinski (terhorab@mac.com)

References

- Jurasinski, G. & Retzer, V. in prep. Measuring multi-plot similarity with presence-absence data. *Ecology* in prep.
- Lande, R. (1996) Statistics and partitioning of species diversity and similarity along multiple communities. *Oikos* **76**: 25-39.
- Qian, H., Ricklefs, R. E. & White, P. S. (2005) Beta diversity of angiosperms in temperate floras of eastern Asia and eastern North America. *Ecology Letters* **8**: 15-22.
- Veech, J. A., Summerville, K. S., Crist, T. O. & Gering, J. C. (2002) The additive partitioning of species diversity: recent revival of an old idea. *Oikos* **99**: 3-9.
- Vellend, M. (2001) Do commonly used indices of beta-diversity measure species turnover? *Journal of Vegetation Science* **12**: 545-552.
- Whittaker, R. H. (1960) Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs* **30**: 279-338.

See Also

see also [sim.pat](#), [sim.het](#), [sim](#), and for quantitative similarity measures [vegdist](#), [dsvdis](#), [dist](#). More qualitative similarity indices can be calculated with [dist.binary](#)

sim.tmp

Calculate binary similarity in time

Description

The function applies one of 56 similarity measures for binary data to calculate compositional similarity of plots between time steps.

Usage

```
sim.tmp(x, y, method = "soer", normalize = FALSE, adjust = TRUE, ...)
```

Arguments

- | | |
|--------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| x | Vegetation data, either as matrix with rows = plots and columns = species, or as data.frame with first three columns representing plots, species and occurrence information respectively. All further columns are dumped before calculation. Occurrence is only considered as binary. If your list or matrix contains abundances or frequencies they are transformed automatically. |
| y | Same as x for time-step two. |
| method | One of 42 similarity measures for binary data. The function uses the same indices as sim . See details there. Per default <code>soerensen</code> similarity is calculated. |

normalize	Logical value indicating whether the values for a, b and c which are calculated in the process should be normalized to 100% (per row, which means per plot comparison). If normalize = TRUE an asymmetric index must be chosen (for details see sim).
adjust	Do not change the default behaviour (TRUE) unless you know what you do. Would spare some calculation time if set to FALSE, when your species data do not need adjustment, which means that in both or all time steps, there are exactly the same species and the same plots. However in most cases it will be more convenient to rely on the function (see details).
...	Other arguments to <code>sim</code>

Details

If you compare species data among time steps there will be most likely different numbers of species (and often also different numbers of plots for which information is available). The function takes care of this and you can give any species matrices you want. If one plot is the same, it will calculate what changed on this plot. There will be an error message if no plot is shared. The function relies on plot and species names!! As in a database - they must be unique!!

Value

Returns a named vector with the similarities for each site between time steps for each plot.

Author(s)

Gerald Jurasinski (terhorab@mac.com)

References

See references in [sim](#)

See Also

See Also as [sim](#) (where you can find a much more elaborate see also section as well).

sim.yo

Calculate a binary similarity index you define.

Description

Enables the calculation of any binary similarity index via the provision of a formula.

Usage

```
sim.yo(x, coord = NULL, method = "(2*a)/((2*a) + b + c)", dn = NULL,
normalize = FALSE, listin = FALSE, listout = FALSE, ...)
```

Arguments

<code>x</code>	Vegetation data, either as matrix with rows = plots and columns = species (similarities are calculated between rows!), or as <code>data.frame</code> with first three columns representing plots, species and occurrence information respectively. All further columns are dumped before calculation. Occurrence is only considered as binary. If your list or matrix contains abundances or frequencies they are transformed automatically.
<code>coord</code>	A <code>data.frame</code> with two columns containing the coordinate values of the sampling units. If given, it triggers the simultaneous calculation of the geographical distances between the sampling units, the coordinates of virtual centre-points between all possible pairs of plots, and the geographical distances in either x- or y-direction. If <code>coord</code> is given, output is always in database format (no matrix).
<code>method</code>	Give the formula for a binary similarity index. Defaults to the formula of Sørensen index. See <code>sim</code> for more examples and general explanations.
<code>dn</code>	Neighbor definition. A geographic distance represented by a numeric or a two value vector defining a ring around each plot. Only takes effect when <code>coord != NULL</code> . If specified, the output does only contain similarities between neighboring plots. A plot is a neighbour to any given plot if it is within the range of the neighbor definition. See details.
<code>normalize</code>	Logical value indicating whether the values for a, b and c which are calculated in the process should be normalized to 100% (per row, which means per plot comparison). If <code>normalize = TRUE</code> an asymmetric index must be chosen (see details in <code>sim</code>).
<code>listin</code>	if <code>x</code> is given in database (list) format this must be set to <code>TRUE</code> (there is no automatic detection of the format)
<code>listout</code>	If output is wanted in database format rather than as a <code>dist</code> -object set this to <code>TRUE</code> . Output is automatically given in database-format, when <code>coord</code> is specified.
<code>...</code>	Arguments to other functions

Details

Presumably this function will rarely be used because `sim` already allows for the calculation of a large variety of binary similarity coefficients. But just in case you found or thought of an alternative this function is provided. For details regarding similarity indices see `sim`. You have to give your formula in quotation marks like this: `"(2*a)/((2*a) + b + c)"`.

Value

If `listout = FALSE` a distance matrix of class `dist` is returned. If `listout = TRUE`, a `data.frame` is returned with 7 columns giving the names of the compared plots in the first two and the calculated similarity measure in the third column. The rest of the columns give the values for a, b, c, and d (in this order). Naming of the first three columns can be changed but defaults to NBX (one of the compared plots), NBY (the other one), `used index` (the values of the calculated index). If `coord != NULL`, the following columns are given in addition and the columns a:d shift to the end of the `data.frame`.

distance	Geographical distance between compared plots
X	For plotting purposes, the x-coordinate of the virtual position of the calculated similarity value in the center between the two compared plots
Y	For plotting purposes, the y-coordinate of the virtual position of the calculated similarity value in the center between the two compared plots
xdist	Geographical distance between compared plots, on the x-axis only
ydist	Geographical distance between compared plots, on the y-axis only

Note

In general, concepts of data-handling are taken from [vegdist](#) and the calculation of a, b, c and d is taken from [dist.binary](#). Thanks to Jari Oksanen for his `vegan` package and the idea to provide a custom build distance formula.

Author(s)

Gerald Jurasinski (terhorab@mac.com)

References

Sørensen, T. (1948) A method of establishing groups of equal amplitude in plant sociology based on similarity of species content. *Biologiske Skrifter* **5**: 1-34.

See Also

[vegdist](#), [dist.binary](#), [dsvdis](#), [dist](#) for other dissimilarity coefficients, and [sim](#) for a variety of formulae for similarity coefficients.

Examples

```
data(abis)

##calculate the default Sørensen index
abis.soer <- sim.yo(abis.spec)

##calculate a custom similarity index
abis.sim <- sim.yo(abis.spec, method="(2*a)/((a) + b + c)")
```

Index

- *Topic **datagen**
 - hexgrid, 22
 - makead, 24
- *Topic **datasets**
 - abis, 3
 - ads.ternaries, 4
- *Topic **hplot**
 - boxes, 8
- *Topic **htest**
 - pcol, 30
- *Topic **manip**
 - liste, 23
 - mama, 27
- *Topic **methods**
 - aslopect, 5
 - bcoov, 7
 - com.sim, 9
 - dfcor, 11
 - diffmean, 13
 - diffmich, 15
 - diffslope, 16
 - direct, 19
 - dist.tmp, 20
 - hexgrid, 22
 - occ.time, 28
 - pcol, 30
 - plot.mrpp, 33
 - sim, 34
 - sim.het, 41
 - sim.pat, 44
 - sim.relt, 46
 - sim.tmp, 48
 - sim.yo, 50
- *Topic **multivariate**
 - com.sim, 9
 - dfcor, 11
 - dist.tmp, 20
 - occ.time, 28
 - sim, 34
 - sim.het, 41
 - sim.pat, 44
 - sim.relt, 46
 - sim.tmp, 48
 - sim.yo, 50
- *Topic **package**
 - simba-package, 2
- *Topic **univar**
 - aslopect, 5
 - bcoov, 7
 - diffmean, 13
 - diffmich, 15
 - diffslope, 16
- abis, 3
- ads, 24, 45
- ads (*makead*), 24
- ads.ternaries, 4, 36
- anosim, 34
- aslopect, 5
- bcoov, 7
- boxes, 8
- boxplot, 8, 9
- boxplot.n, 8, 9
- com.sim, 9
- coord, 35, 50
- cor, 12, 32
- cor.test, 2, 20, 31–33
- data.frame, 16, 17, 19, 20, 24, 27, 34, 38, 43, 49–51
- dfcor, 11
- diffmean, 10, 13
- diffmich, 14, 15
- diffslope, 14, 16
- diffslope2 (*diffslope*), 16
- direct, 19
- direct2 (*direct*), 19

- dist, 2, 6, 19–21, 23, 30, 32, 35, 38, 41, 46, 48, 50, 51
- dist.binary, 2, 23, 30, 38, 41, 46, 48, 51
- dist.quant, 8
- dist.tmp, 20
- dsvdis, 2, 8, 41, 46, 48, 51
- fitmich, 15, 16
- fitmich (*diffmich*), 15
- het2nbs (*sim.het*), 41
- hexgrid, 22
- is.vector, 30
- liste, 23, 28
- lm, 15, 17, 18
- makead, 24
- mama, 27
- mancor, 13, 30
- mancor (*pcol*), 30
- mantel, 13, 20, 33
- mantell, 33
- mrpp, 9, 11, 34
- nls, 16
- occ.time, 28
- occ.tmp (*occ.time*), 28
- pcol, 12, 13, 30
- permcors, 13, 30
- permcors (*pcol*), 30
- permcors2, 11, 13, 30
- permcors2 (*pcol*), 30
- plot, 9
- plot.diffmich (*diffmich*), 15
- plot.dmn (*diffmean*), 13
- plot.dsl (*diffslope*), 16
- plot.mrpp, 33
- plot.permcors (*pcol*), 30
- reshape, 2, 24, 27
- sample, 16, 18
- sim, 4, 10, 21, 23, 30, 34, 42, 44, 46, 48–51
- sim.het, 41, 48
- sim.pat, 42, 44, 44, 48
- sim.rel (*sim.relt*), 46
- sim.relt, 46
- sim.tmp, 30, 48
- sim.yo, 50
- sima (*sim.pat*), 44
- simba (*simba-package*), 2
- simba-package, 2
- simn (*sim.pat*), 44
- triax.plot, 36
- vegdist, 2, 8, 10, 21, 23, 30, 38, 41, 46, 48, 51