**Package ‘vegan’**

May 17, 2018

**Title**  Community Ecology Package
**Version**  2.5-2
**Date**  2018-05-16
**Author**  Jari Oksanen, F. Guillaume Blanchet, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlinn, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, Eduard Szöcs, Helene Wagner

**Maintainer**  Jari Oksanen <jari.oksanen@oulu.fi>
**Depends**  permute (>= 0.9-0), lattice, R (>= 3.2.0)
**Suggests**  parallel, tcltk, knitr
**Imports**  MASS, cluster, mgcv

**VignetteBuilder**  utils, knitr
**Description**  Ordination methods, diversity analysis and other functions for community and vegetation ecologists.
**License**  GPL-2

**BugReports**  [https://github.com/vegandevs/vegan/issues](https://github.com/vegandevs/vegan/issues)
**URL**  [https://cran.r-project.org/](https://cran.r-project.org/), [https://github.com/vegandevs/vegan](https://github.com/vegandevs/vegan)
**NeedsCompilation**  yes
**Repository**  CRAN
**Date/Publication**  2018-05-17 09:05:08 UTC

---

**R topics documented:**

- vegan-package ...........................................  4
- add1.cca .............................................  5
- adipart ...............................................  7
- adonis ............................................... 11
- anosim ............................................... 14
- anova.cca .......................................... 16
- avgdist ........................................... 19
### R topics documented:

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCI</td>
<td>21</td>
</tr>
<tr>
<td>beals</td>
<td>22</td>
</tr>
<tr>
<td>betadisper</td>
<td>25</td>
</tr>
<tr>
<td>betadiver</td>
<td>30</td>
</tr>
<tr>
<td>bgdispersal</td>
<td>33</td>
</tr>
<tr>
<td>bioenv</td>
<td>34</td>
</tr>
<tr>
<td>biplot.rda</td>
<td>37</td>
</tr>
<tr>
<td>capscale</td>
<td>39</td>
</tr>
<tr>
<td>cascadeKM</td>
<td>42</td>
</tr>
<tr>
<td>cca</td>
<td>45</td>
</tr>
<tr>
<td>cca.object</td>
<td>49</td>
</tr>
<tr>
<td>CCorA</td>
<td>52</td>
</tr>
<tr>
<td>clamtest</td>
<td>55</td>
</tr>
<tr>
<td>commsim</td>
<td>57</td>
</tr>
<tr>
<td>contribdiv</td>
<td>64</td>
</tr>
<tr>
<td>decorana</td>
<td>66</td>
</tr>
<tr>
<td>decostand</td>
<td>69</td>
</tr>
<tr>
<td>designdist</td>
<td>72</td>
</tr>
<tr>
<td>deviance.cca</td>
<td>75</td>
</tr>
<tr>
<td>dispindmorisita</td>
<td>76</td>
</tr>
<tr>
<td>dispweight</td>
<td>78</td>
</tr>
<tr>
<td>distconnected</td>
<td>80</td>
</tr>
<tr>
<td>diversity</td>
<td>82</td>
</tr>
<tr>
<td>dune</td>
<td>84</td>
</tr>
<tr>
<td>dune.taxon</td>
<td>85</td>
</tr>
<tr>
<td>eigenvals</td>
<td>86</td>
</tr>
<tr>
<td>envfit</td>
<td>87</td>
</tr>
<tr>
<td>eventstar</td>
<td>91</td>
</tr>
<tr>
<td>fisherfit</td>
<td>93</td>
</tr>
<tr>
<td>goodness.cca</td>
<td>96</td>
</tr>
<tr>
<td>goodness.metaMDS</td>
<td>99</td>
</tr>
<tr>
<td>humpfit</td>
<td>100</td>
</tr>
<tr>
<td>indpower</td>
<td>103</td>
</tr>
<tr>
<td>influence.cca</td>
<td>105</td>
</tr>
<tr>
<td>isomap</td>
<td>107</td>
</tr>
<tr>
<td>kendall.global</td>
<td>110</td>
</tr>
<tr>
<td>linestack</td>
<td>112</td>
</tr>
<tr>
<td>make.cepnames</td>
<td>114</td>
</tr>
<tr>
<td>mantel</td>
<td>115</td>
</tr>
<tr>
<td>mantel.correlog</td>
<td>117</td>
</tr>
<tr>
<td>MDSrotate</td>
<td>120</td>
</tr>
<tr>
<td>metaMDS</td>
<td>121</td>
</tr>
<tr>
<td>mite</td>
<td>127</td>
</tr>
<tr>
<td>monoMDS</td>
<td>128</td>
</tr>
<tr>
<td>MOSTest</td>
<td>131</td>
</tr>
<tr>
<td>mrpp</td>
<td>134</td>
</tr>
<tr>
<td>mso</td>
<td>138</td>
</tr>
<tr>
<td>multipart</td>
<td>141</td>
</tr>
</tbody>
</table>
### R topics documented:

- `nestedtemp` .................................................. 143
- `nobs.adonis` ............................................... 146
- `nullmodel` .................................................. 147
- `ocecosimu` .................................................. 151
- `ordiarrows` .................................................. 155
- `ordiArrowTextXY` .......................................... 157
- `ordihull` .................................................... 159
- `ordilabel` ................................................... 163
- `ordiplot` ..................................................... 164
- `ordipointlabel` ............................................. 166
- `ordiresids` ................................................... 168
- `ordistep` ..................................................... 169
- `ordsurf` ....................................................... 172
- `orditkplot` ................................................... 177
- `orditorp` ..................................................... 180
- `ordixyplot` ................................................... 182
- `pcnm` ........................................................ 184
- `permat` ........................................................ 186
- `permustats` ................................................... 191
- `permutations` ................................................ 193
- `permute.test.betadisper` ................................ 197
- `plot.cca` ..................................................... 199
- `prc` ........................................................... 202
- `predict.cca` ................................................... 205
- `procrustes` ................................................... 209
- `pyrifos` ....................................................... 212
- `radfit` ........................................................ 213
- `rankindex` ..................................................... 217
- `rarefy` ........................................................ 219
- `raupcrick` ..................................................... 221
- `read.cep` ...................................................... 223
- `renyi` .......................................................... 225
- `reorder.hclust` ............................................. 227
- `RsquareAdj` ................................................... 229
- `scores` ........................................................ 230
- `screeplot.cca` ............................................... 231
- `simper` ........................................................ 234
- `simulate.rda` .................................................. 236
- `sipoo` .......................................................... 238
- `spantree` ...................................................... 238
- `specaccum` .................................................... 241
- `specpool` ...................................................... 246
- `spsscores` ..................................................... 249
- `SSarrhenius` .................................................. 250
- `stepacross` .................................................... 253
- `stressplot.wcmdscale` .................................... 255
- `taxondive` .................................................... 256
- `tolerance` ..................................................... 259
Description

The vegan package provides tools for descriptive community ecology. It has most basic functions of diversity analysis, community ordination and dissimilarity analysis. Most of its multivariate tools can be used for other data types as well.

Details

The functions in the vegan package contain tools for diversity analysis, ordination methods and tools for the analysis of dissimilarities. Together with the labdsv package, the vegan package provides most standard tools of descriptive community analysis. Package ade4 provides an alternative comprehensive package, and several other packages complement vegan and provide tools for deeper analysis in specific fields. Package BiodiversityR provides a GUI for a large subset of vegan functionality.

The vegan package is developed at GitHub (https://github.com/vegandevs/vegan/). GitHub provides up-to-date information and forums for bug reports.

Most important changes in vegan documents can be read with news(package="vegan") and vignettes can be browsed with browseVignettes("vegan"). The vignettes include a vegan FAQ, discussion on design decisions, short introduction to ordination and discussion on diversity methods. A tutorial of the package at http://cc.oulu.fi/~jarioksa/opetus/metodi/vegantutor.pdf provides a more thorough introduction to the package.

To see the preferable citation of the package, type citation("vegan").

Author(s)

The vegan development team is Jari Oksanen, F. Guillaume Blanchet, Roeland Kindt, Pierre Legendre, Peter R. Minchin, R. B. O’Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, Helene Wagner. Many other people have contributed to individual functions: see credits in function help pages.
Examples

### Example 1: Unconstrained ordination
## NMDS
data(varespec)
data(varechem)
ord <- metaMDS(varespec)
plot(ord, type = "t")
## Fit environmental variables
ef <- envfit(ord, varechem)
plot(ef, p.max = 0.05)
### Example 2: Constrained ordination (RDA)
## The example uses formula interface to define the model
data(dune)
data(dune.env)
## No constraints: PCA
mod0 <- rda(dune ~ 1, dune.env)
plot(mod0)
## All environmental variables: Full model
mod1 <- rda(dune ~ ., dune.env)
plot(mod1)
## Automatic selection of variables by permutation P-values
mod <- ordistep(mod0, scope=formula(mod1))
plot(mod)
## Permutation test for all variables
anova(mod)
## Permutation test of "type III" effects, or significance when a term
## is added to the model after all other terms
anova(mod, by = "margin")
## Plot only sample plots, use different symbols and draw SD ellipses
## for Management classes
plot(mod, display = "sites", type = "n")
with(dune.env, points(mod, disp = "si", pch = as.numeric(Management)))
with(dune.env, legend("topleft", levels(Management), pch = 1:4,
    title = "Management"))
with(dune.env, ordiellipse(mod, Management, label = TRUE))
## add fitted surface of diversity to the model
ordisurf(mod, diversity(dune), add = TRUE)
### Example 3: analysis of dissimilarites a.k.a. non-parametric
### permutational anova
adonis(dune ~ ., dune.env)
adonis(dune ~ Management + Moisture, dune.env)
Description

Compute all single terms that can be added to or dropped from a constrained ordination model.

Usage

```r
## S3 method for class 'cca'
add1(object, scope, test = c("none", "permutation"),
     permutations = how(nperm=199), ...)
## S3 method for class 'cca'
drop1(object, scope, test = c("none", "permutation"),
      permutations = how(nperm=199), ...)
```

Arguments

- **object**: A constrained ordination object from `cca`, `rda` or `capscale`.
- **scope**: A formula giving the terms to be considered for adding or dropping; see `add1` for details.
- **test**: Should a permutation test be added using `anova.cca`.
- **permutations**: a list of control values for the permutations as returned by the function `how`, or the number of permutations required, or a permutation matrix where each row gives the permuted indices.
- **...**: Other arguments passed to `add1.default`, `drop1.default`, and `anova.cca`.

Details

With argument `test = "none"` the functions will only call `add1.default` or `drop1.default`. With argument `test = "permutation"` the functions will add test results from `anova.cca`. Function `drop1.cca` will call `anova.cca` with argument `by = "margin"`. Function `add1.cca` will implement a test for single term additions that is not directly available in `anova.cca`.

Functions are used implicitly in `step`, `ordiR2step` and `ordistep`. The `deviance.cca` and `deviance.rda` used in `step` have no firm basis, and setting argument `test = "permutation"` may help in getting useful insight into validity of model building. Function `ordistep` calls alternately `drop1.cca` and `add1.cca` with argument `test = "permutation"` and selects variables by their permutation P-values. Meticulous use of `add1.cca` and `drop1.cca` will allow more judicious model building.

The default number of permutations is set to a low value, because permutation tests can take a long time. It should be sufficient to give a impression on the significances of the terms, but higher values of `permutations` should be used if P values really are important.

Value

Returns a similar object as `add1` and `drop1`.

Author(s)

Jari Oksanen
See Also

`addQ`, `dropQ` and `anova.cca` for basic methods. You probably need these functions with `step` and `ordistep`. Functions `deviance.cca` and `extractAIC.cca` are used to produce the other arguments than test results in the output. Functions `cca`, `rda` and `capscale` produce result objects for these functions.

Examples

```r
data(dune)
data(dune.env)
## Automatic model building based on AIC but with permutation tests
step(cca(dune ~ 1, dune.env), reformulate(names(dune.env)), test="perm")
## see ?ordistep to do the same, but based on permutation P-values
## Not run:
ordistep(cca(dune ~ 1, dune.env), reformulate(names(dune.env)))

## End(Not run)
## Manual model building
## -- define the maximal model for scope
mbig <- rda(dune ~ ., dune.env)
## -- define an empty model to start with
m0 <- rda(dune ~ 1, dune.env)
## -- manual selection and updating
addQ(m0, scope=formula(mbig), test="perm")
m0 <- update(m0, . ~ . + Management)
addQ(m0, scope=formula(mbig), test="perm")
m0 <- update(m0, . ~ . + Moisture)
## -- included variables still significant?
dropQ(m0, test="perm")
addQ(m0, scope=formula(mbig), test="perm")
```

**adipart**

*Additive Diversity Partitioning and Hierarchical Null Model Testing*

**Description**

In additive diversity partitioning, mean values of alpha diversity at lower levels of a sampling hierarchy are compared to the total diversity in the entire data set (gamma diversity). In hierarchical null model testing, a statistic returned by a function is evaluated according to a nested hierarchical sampling design (`hiersimu`).

**Usage**

```r
adipart(...)
## Default S3 method:
adipart(y, x, index=c("richness", "shannon", "simpson"),
weights=c("unif", "prop"), relative = FALSE, nsimul=99,
method = "r2dtable", ...)
```
Arguments

y  A community matrix.
x  A matrix with same number of rows as in y, columns coding the levels of sampling hierarchy. The number of groups within the hierarchy must decrease from left to right. If x is missing, function performs an overall decomposition into alpha, beta and gamma diversities.
formula  A two sided model formula in the form y ~ x, where y is the community data matrix with samples as rows and species as column. Right hand side (x) must be grouping variables referring to levels of sampling hierarchy, terms from right to left will be treated as nested (first column is the lowest, last is the highest level, at least two levels specified). Interaction terms are not allowed.
data  A data frame where to look for variables defined in the right hand side of formula. If missing, variables are looked in the global environment.
index  Character, the diversity index to be calculated (see Details).
weights  Character, "unif" for uniform weights, "prop" for weighting proportional to sample abundances to use in weighted averaging of individual alpha values within strata of a given level of the sampling hierarchy.
relative  Logical, if TRUE then alpha and beta diversity values are given relative to the value of gamma for function adipart.
nsimul  Number of permutations to use. If nsimul = 0, only the FUN argument is evaluated. It is thus possible to reuse the statistic values without a null model.
method  Null model method: either a name (character string) of a method defined in make.commsim or a commsim function. The default ”r2dtable” keeps row sums and column sums fixed. See oecosimu for Details and Examples.
FUN  A function to be used by hiersimu. This must be fully specified, because currently other arguments cannot be passed to this function via . . .
location  Character, identifies which function (mean or median) is to be used to calculate location of the samples.
drop.highest  Logical, to drop the highest level or not. When FUN evaluates only arrays with at least 2 dimensions, highest level should be dropped, or not selected at all.
... Other arguments passed to functions, e.g. base of logarithm for Shannon diversity, or method, thin or burnin arguments for \texttt{oecosimu}.

### Details

Additive diversity partitioning means that mean alpha and beta diversities add up to gamma diversity, thus beta diversity is measured in the same dimensions as alpha and gamma (Lande 1996). This additive procedure is then extended across multiple scales in a hierarchical sampling design with \( i = 1, 2, 3, \ldots, m \) levels of sampling (Crist et al. 2003). Samples in lower hierarchical levels are nested within higher level units, thus from \( i = 1 \) to \( i = m \) grain size is increasing under constant survey extent. At each level \( i \), \( \alpha_i \) denotes average diversity found within samples.

At the highest sampling level, the diversity components are calculated as

\[
\beta_m = \gamma - \alpha_m
\]

For each lower sampling level as

\[
\beta_i = \alpha_{i+1} - \alpha_i
\]

Then, the additive partition of diversity is

\[
\gamma = \alpha_1 + \sum_{i=1}^{m} \beta_i
\]

Average alpha components can be weighted uniformly (\texttt{weight="unif"}) to calculate it as simple average, or proportionally to sample abundances (\texttt{weight="prop"}) to calculate it as weighted average as follows

\[
\alpha_i = \sum_{j=1}^{n_i} D_{ij} w_{ij}
\]

where \( D_{ij} \) is the diversity index and \( w_{ij} \) is the weight calculated for the \( j \)th sample at the \( i \)th sampling level.

The implementation of additive diversity partitioning in \texttt{adipart} follows Crist et al. 2003. It is based on species richness (\( S \), not \( S - 1 \)), Shannon’s and Simpson’s diversity indices stated as the index argument.

The expected diversity components are calculated \texttt{nsimul} times by individual based randomisation of the community data matrix. This is done by the "\texttt{r2dtable}" method in \texttt{oecosimu} by default.

\texttt{hiersimu} works almost in the same way as \texttt{adipart}, but without comparing the actual statistic values returned by \texttt{FUN} to the highest possible value (cf. gamma diversity). This is so, because in most of the cases, it is difficult to ensure additive properties of the mean statistic values along the hierarchy.

### Value

An object of class "\texttt{adipart}" or "\texttt{hiersimu}" with same structure as \texttt{oecosimu} objects.

### Author(s)

Péter Sólymos, <solymos@ualberta.ca>
References


See Also

See `oecosimu` for permutation settings and calculating $p$-values. `multipart` for multiplicative diversity partitioning.

Examples

```r
## NOTE: 'nsimul' argument usually needs to be $\geq$ 99
## here much lower value is used for demonstration

data(mite)
data(mite.xy)
data(mite.env)

## Function to get equal area partitions of the mite data

cutter <- function (x, cut = seq(0, 10, by = 2.5)) {
  out <- rep(1, length(x))
  for (i in 2:(length(cut) - 1))
    out[which(x > cut[i] & x <= cut[(i + 1)])] <- i
  return(out)
}

## The hierarchy of sample aggregation

levs $\leftarrow$ with(mite.xy, data.frame(
  1 = 1:nrow(mite),
  2 = cutter(y, cut = seq(0, 10, by = 2.5)),
  3 = cutter(y, cut = seq(0, 10, by = 5)),
  4 = cutter(y, cut = seq(0, 10, by = 10)))

## Let's see in a map

par(mfrow=c(1,3))

plot(mite.xy, main="1", col=as.numeric(levs$1)+1, asp=1)
plot(mite.xy, main="2", col=as.numeric(levs$2)+1, asp=1)
plot(mite.xy, main="3", col=as.numeric(levs$3)+1, asp=1)
par(mfrow=c(1,1))

## Additive diversity partitioning

adipart(mite, index="richness", nsimul=19)
adipart(mite ~ ., levsm, index="richness", nsimul=19)

## Hierarchical null model testing

## diversity analysis (similar to adipart)
hiersimu(mite, FUN=diversity, relative=TRUE, nsimul=19)
hiersimu(mite ~ ., levsm, FUN=diversity, relative=TRUE, nsimul=19)

## Hierarchical testing with the Morisita index

morfun $\leftarrow$ function(x) dispindmorisita(x)$\mid$smst
hiersimu(mite ~ ., levsm, morfun, drop.highest=TRUE, nsimul=19)
```
Description

Analysis of variance using distance matrices — for partitioning distance matrices among sources of variation and fitting linear models (e.g., factors, polynomial regression) to distance matrices; uses a permutation test with pseudo-\(F\) ratios.

Usage

```r
adonis(formula, data, permutations = 999, method = "bray",
       sqrt.dist = FALSE, add = FALSE, by = "terms",
       parallel = getOption("mc.cores"), ...)
```

Arguments

- `formula`: Model formula. The LHS must be either a community data matrix or a dissimilarity matrix, e.g., from `vegdist` or `dist`. If the LHS is a data matrix, function `vegdist` will be used to find the dissimilarities. The RHS defines the independent variables. These can be continuous variables or factors, they can be transformed within the formula, and they can have interactions as in a typical `formula`. If a dissimilarity object is supplied, no species coefficients can be calculated in `adonis` (see Value below).
- `data`: the data frame for the independent variables.
- `permutations`: a list of control values for the permutations as returned by the function `how`, or the number of permutations required, or a permutation matrix where each row gives the permuted indices.
- `method`: the name of any method used in `vegdist` to calculate pairwise distances if the left hand side of the formula was a data frame or a matrix.
- `sqrt.dist`: Take square root of dissimilarities. This often euclidifies dissimilarities.
- `add`: Add a constant to the non-diagonal dissimilarities such that all eigenvalues are non-negative in the underlying Principal Co-ordinates Analysis (see `wcmdscale` for details). Choice "lingoes" (or TRUE) use the recommended method of Legendre & Anderson (1999: "method 1") and "cailliez" uses their "method 2".
- `by`: `by = "terms"` will assess significance for each term (sequentially from first to last), setting `by = "margin"` will assess the marginal effects of the terms (each marginal term analysed in a model with all other variables), and `by = NULL` will assess the overall significance of all terms together. The arguments is passed on to `anova.cca`.
- `strata`: groups (strata) within which to constrain permutations.
contr.unordered, contr.ordered

contrasts used for the design matrix (default in R is dummy or treatment contrasts for unordered factors).

parallel

Number of parallel processes or a predefined socket cluster. With parallel = 1 uses ordinary, non-parallel processing. The parallel processing is done with parallel package.

Details

adonis2 and adonis are functions for the analysis and partitioning sums of squares using dissimilarities. Function adonis is directly based on the algorithm of Anderson (2001) and performs a sequential test of terms. Function adonis2 is based on the principles of McArdle & Anderson (2001) and can perform sequential, marginal and overall tests. Function adonis2 also allows using additive constants or squareroot of dissimilarities to avoid negative eigenvalues. but both functions can handle semimetric indices (such as Bray-Curtis) that produce negative eigenvalues. Function adonis2 can be much slower than adonis, in particular with several terms. With the same random permutation, tests are identical in both functions, and the results are also identical to anova.cca of dbrda and capscale. With Euclidean distances, the tests are also identical to anova.cca of rda.

The functions partition sums of squares of a multivariate data set, and they are directly analogous to MANOVA (multivariate analysis of variance). McArdle and Anderson (2001) and Anderson (2001) refer to the method as “permutational manova” (formerly “nonparametric manova”). Further, as the inputs are linear predictors, and a response matrix of an arbitrary number of columns, they are a robust alternative to both parametric MANOVA and to ordination methods for describing how variation is attributed to different experimental treatments or uncontrolled covariates. Functions are also analogous to distance-based redundancy analysis in functions dbrda and capscale (Legendre and Anderson 1999). Functions provide an alternative to AMOVA (nested analysis of molecular variance, Excoffier, Smouse, and Quattro, 1992; amova in the ade4 package) for both crossed and nested factors.

Value

Function adonis2 returns an anova.cca result object with a new column for partial $R^2$: This is the proportion of sum of squares from the total, and in marginal models (by = "margin") the $R^2$ terms do not add up to 1.

Function adonis returns an object of class "adonis" with following components:

aov.tab Typical AOV table showing sources of variation, degrees of freedom, sequential sums of squares, mean squares, $F$ statistics, partial $R^2$ and $P$ values, based on $N$ permutations.

coefficients matrix of coefficients of the linear model, with rows representing sources of variation and columns representing species; each column represents a fit of a species abundance to the linear model. These are what you get when you fit one species to your predictors. These are NOT available if you supply the distance matrix in the formula, rather than the site x species matrix

coef.sites matrix of coefficients of the linear model, with rows representing sources of variation and columns representing sites; each column represents a fit of a sites
adonis

distances (from all other sites) to the linear model. These are what you get when you fit distances of one site to your predictors.

f.perms
an \( N \) by \( m \) matrix of the null \( F \) statistics for each source of variation based on \( N \) permutations of the data. The permutations can be inspected with `permustats` and its support functions.

model.matrix
The `model.matrix` for the right hand side of the formula.

terms
The `terms` component of the model.

Note
Anderson (2001, Fig. 4) warns that the method may confound location and dispersion effects: significant differences may be caused by different within-group variation (dispersion) instead of different mean values of the groups (see Warton et al. 2012 for a general analysis). However, it seems that `adonis` is less sensitive to dispersion effects than some of its alternatives (`anosim`, `mrpp`). Function `betadisper` is a sister function to `adonis` to study the differences in dispersion within the same geometric framework.

Author(s)
Martin Henry H. Stevens (`adonis`) and Jari Oksanen (`adonis2`).

References


See Also
`mrpp`, `anosim`, `mantel`, `varpart`.

Examples
```r
data(dune)
data(dune.env)
## default test by terms
adonis2(dune ~ Management*A1, data = dune.env)
## overall tests
adonis2(dune ~ Management*A1, data = dune.env, by = NULL)
```
### Example of use with strata, for nested (e.g., block) designs.

```r
dat <- expand.grid(rep=gl(2, 1), NO3=factor(c(0, 10)), field=gl(3, 1))
dat
Agropyron <- with(dat, as.numeric(field) + as.numeric(NO3)+2 + rnorm(12)/2
Schizachyrium <- with(dat, as.numeric(field) - as.numeric(NO3)+2 + rnorm(12)/2
total <- Agropyron + Schizachyrium
dotplot(total ~ NO3, dat, jitter.x=TRUE, groups=field,
    type=c("p", "a"), xlab="NO3", auto.key=list(columns=3, lines=TRUE))
```

Y <- data.frame(Agropyron, Schizachyrium)
mod <- metaMDS(Y, trace = FALSE)
plot(mod)
### Ellipsoid hulls show treatment
with(dat, ordiellipse(mod, field, kind = "ehull", label = TRUE))
### Spider shows fields
with(dat, ordispider(mod, field, lty=3, col="red"))

### Incorrect (no strata)
perm <- how(nperm = 199)
adonis2 (Y ~ NO3, data = dat, permutations = perm)

### Correct with strata
setBlocks(perm) <- with(dat, field)
adonis2(Y ~ NO3, data = dat, permutations = perm)
```

---

**anosim**

**Analysis of Similarities**

**Description**

Analysis of similarities (ANOSIM) provides a way to test statistically whether there is a significant difference between two or more groups of sampling units.

**Usage**

```r
anosim(x, grouping, permutations = 999, distance = "bray", strata = NULL,
    parallel = getOption("mc.cores"))
```

**Arguments**

- `x`: Data matrix or data frame in which rows are samples and columns are response variable(s), or a dissimilarity object or a symmetric square matrix of dissimilarities.
- `grouping`: Factor for grouping observations.
- `permutations`: a list of control values for the permutations as returned by the function `how`, or the number of permutations required, or a permutation matrix where each row gives the permuted indices.
**anosim**

**distance**
Choice of distance metric that measures the dissimilarity between two observations. See `vegdist` for options. This will be used if `x` was not a dissimilarity structure or a symmetric square matrix.

**strata**
An integer vector or factor specifying the strata for permutation. If supplied, observations are permuted only within the specified strata.

**parallel**
Number of parallel processes or a predefined socket cluster. With `parallel = 1` uses ordinary, non-parallel processing. The parallel processing is done with `parallel` package.

**Details**
Analysis of similarities (ANOSIM) provides a way to test statistically whether there is a significant difference between two or more groups of sampling units. Function `anosim` operates directly on a dissimilarity matrix. A suitable dissimilarity matrix is produced by functions `dist` or `vegdist`. The method is philosophically allied with NMDS ordination (`monomds`), in that it uses only the rank order of dissimilarity values.

If two groups of sampling units are really different in their species composition, then compositional dissimilarities between the groups ought to be greater than those within the groups. The `anosim` statistic $R$ is based on the difference of mean ranks between groups ($r_B$) and within groups ($r_W$):

$$R = (r_B - r_W)/(N(N - 1)/4)$$

The divisor is chosen so that $R$ will be in the interval $-1 ... +1$, value 0 indicating completely random grouping.

The statistical significance of observed $R$ is assessed by permuting the grouping vector to obtain the empirical distribution of $R$ under null-model. See `permutations` for additional details on permutation tests in Vegan. The distribution of simulated values can be inspected with the `permustats` function.

The function has `summary` and `plot` methods. These both show valuable information to assess the validity of the method: The function assumes that all ranked dissimilarities within groups have about equal median and range. The `plot` method uses `boxplot` with options `notch=TRUE` and `varwidth=TRUE`.

**Value**
The function returns a list of class "anosim" with following items:

- **call**
  Function call.

- **statistic**
  The value of ANOSIM statistic $R$

- **signif**
  Significance from permutation.

- **perm**
  Permutation values of $R$. The distribution of permutation values can be inspected with function `permustats`.

- **class.vec**
  Factor with value `Between` for dissimilarities between classes and class name for corresponding dissimilarity within class.

- **dis.rank**
  Rank of dissimilarity entry.

- **dissimilarity**
  The name of the dissimilarity index: the "method" entry of the `dist` object.

- **control**
  A list of control values for the permutations as returned by the function `how`.
Note

The anosim function can confound the differences between groups and dispersion within groups and the results can be difficult to interpret (cf. Warton et al. 2012). The function returns a lot of information to ease studying its performance. Most anosim models could be analysed with adonis2 which seems to be a more robust alternative.

Author(s)

Jari Oksanen, with a help from Peter R. Minchin.

References


See Also

mrpp for a similar function using original dissimilarities instead of their ranks. dist and vegdist for obtaining dissimilarities, and rank for ranking real values. For comparing dissimilarities against continuous variables, see mantel. Function adonis2 is a more robust alternative that should preferred.

Examples

data(dune)
data(dune.env)
dune.dist <- vegdist(dune)
dune.ano <- with(dune.env, anosim(dune.dist, Management))
summary(dune.ano)
plot(dune.ano)

Description

The function performs an ANOVA like permutation test for Constrained Correspondence Analysis (cca), Redundancy Analysis (rda) or distance-based Redundancy Analysis (dbRDA, capscale) to assess the significance of constraints.
Usage

## S3 method for class 'cca'
anova(object, ..., permutations = how(nperm=999),
   by = NULL, model = c("reduced", "direct", "full"),
   parallel =getOption("mc.cores"), strata = NULL,
   cutoff = 1, scope = NULL)

## S3 method for class 'cca'
permutest(x, permutations = how(nperm = 99),
   model = c("reduced", "direct", "full"), by = NULL, first = FALSE,
   strata = NULL, parallel = getOption("mc.cores"), ...)

Arguments

**object**
One or several result objects from `cca`, `rda`, `dbrda` or `capscale`. If there are several result objects, they are compared against each other in the order they were supplied. For a single object, a test specified in by or an overall test is given.

**x**
A single ordination result object.

**permutations**
a list of control values for the permutations as returned by the function `how`, or the number of permutations required, or a permutation matrix where each row gives the permuted indices.

**by**
Setting `by = "axis"` will assess significance for each constrained axis, and setting `by = "terms"` will assess significance for each term (sequentially from first to last), and setting `by = "margin"` will assess the marginal effects of the terms (each marginal term analysed in a model with all other variables). Function `permutest` accepts choices "terms" for sequential test of terms, and "onedf" for sequential test of one-degree-of-freedom contrasts.

**model**
Permutation model: `model="direct"` permutes community data, `model="reduced"` permutes residuals of the community data after Conditions (partial model), `model = "full"` permutes residuals after Conditions and Constraints.

**parallel**
Use parallel processing with the given number of cores.

**strata**
An integer vector or factor specifying the strata for permutation. If supplied, observations are permuted only within the specified strata. It is an error to use this when permutations are a matrix, or a `how` defines blocks. This is a legacy argument that will be deprecated in the future: use `permutations = how(..., blocks)` instead.

**cutoff**
Only effective with `by="axis"` where stops permutations after an axis exceeds the cutoff $p$-value.

**scope**
Only effective with `by="margin"` where it can be used to select the marginal terms for testing. The default is to test all marginal terms in `drop.scope`.

**first**
Analyse only significance of the first axis.

... Parameters passed to other functions. `anova.cca` passes all arguments to `permutest.cca`. In `anova` with `by = "axis"` you can use argument `cutoff` (defaults 1) which stops permutations after exceeding the given level.
Details

Functions `anova.cca` and `permutest.cca` implement ANOVA like permutation tests for the joint effect of constraints in `cca`, `rda`, `dbRDA` or `capscale`. Function `anova` is intended as a more user-friendly alternative to `permutest` (that is the real workhorse).

Function `anova` can analyse a sequence of constrained ordination models. The analysis is based on the differences in residual deviance in permutations of nested models.

The default test is for the sum of all constrained eigenvalues. Setting `first = TRUE` will perform a test for the first constrained eigenvalue. Argument `first` can be set either in `anova.cca` or in `permutest.cca`. It is also possible to perform significance tests for each axis or for each term (constraining variable) using argument `by` in `anova.cca`. Setting `by = "axis"` will perform separate significance tests for each constrained axis. All previous constrained axes will be used as conditions (“partialled out”) and a test for the first constrained eigenvalues is performed (Legendre et al. 2011). You can stop permutation tests after exceeding a given significance level with argument `cutoff` to speed up calculations in large models. Setting `by = "terms"` will perform separate significance test for each term (constraining variable). The terms are assessed sequentially from first to last, and the order of the terms will influence their significances. Setting `by = "margin"` will perform separate significance test for each marginal term in a model with all other terms. The marginal test also accepts a scope argument for the `drop.scope` which can be a character vector of term labels that are analysed, or a fitted model of lower scope. The marginal effects are also known as “Type III” effects, but the current function only evaluates marginal terms. It will, for instance, ignore main effects that are included in interaction terms. In calculating pseudo-\(F\), all terms are compared to the same residual of the full model.

Community data are permuted with choice `model="direct"`, and residuals after partial CCA/ RDA/ dbRDA with choice `model="reduced"` (default). If there is no partial CCA/ RDA/ dbRDA stage, `model="reduced"` simply permutes the data and is equivalent to `model="direct"`. The test statistic is “pseudo-\(F\)”, which is the ratio of constrained and unconstrained total Inertia (Chi-squares, variances or something similar), each divided by their respective degrees of freedom. If there are no conditions (“partial” terms), the sum of all eigenvalues remains constant, so that pseudo-\(F\) and eigenvalues would give equal results. In partial CCA/ RDA/ dbRDA, the effect of conditioning variables (“covariables”) is removed before permutation, and the total Chi-square is not fixed, and test based on pseudo-\(F\) would differ from the test based on plain eigenvalues.

Value

The function `anova.cca` calls `permutest.cca` and fills an `anova` table. Additional attributes are `random.seed` (the random seeds used), `control` (the permutation design, see `how`) and `F.perm` (the permuted test statistics).

Author(s)

Jari Oksanen

References


avgdist

See Also

anova.cca, cca, rda, capscale to get something to analyse. Function drop1.cca calls anova.cca with by = "margin", and add1.cca an analysis for single terms additions, which can be used in automatic or semiautomatic model building (see deviance.cca).

Examples

data(varespec, varechem)
mod <- cca(varespec ~ Al + P + K, varechem)
## overall test
anova(mod)
## tests for individual terms
anova(mod, by="term")
anova(mod, by="margin")
## test for adding all environmental variables
anova(mod, cca(varespec ~ ., varechem))

avgdist

Averaged Subsampled Dissimilarity Matrices

Description

The function computes the dissimilarity matrix of a dataset multiple times using vegdist while randomly subsampling the dataset each time. All of the subsampled iterations are then averaged (mean) to provide a distance matrix that represents the average of multiple subsampling iterations. This emulates the behavior of the distance matrix calculator within the Mothur microbial ecology toolkit.

Usage

avgdist(x, sample, distfun = vegdist, meanfun = mean, transf = NULL, iterations = 100, dmethod = "bray", ...)

Arguments

x Community data matrix.
sample The subsampling depth to be used in each iteration. Samples that do not meet this threshold will be removed from the analysis, and their identify returned to the user in stdout.
distfun The dissimilarity matrix function to be used. Default is the vegan vegdist
meanfun The calculation to use for the average (mean or median).
transf Option for transforming the count data before calculating the distance matrix. Any base transformation option can be used (e.g. sqrt)
iterations The number of random iterations to perform before averaging. Default is 100 iterations.
dmethod   Dissimilarity index to be used with the specified dissimilarity matrix function. Default is Bray-Curtis
...
Any additional arguments to add to the distance function or mean/median function specified.

Note
The function builds on the function **rrarefy** and and additional distance matrix function (e.g. **vegdist**) to add more meaningful representations of distances among randomly subsampled datasets by presenting the average of multiple random iterations. This function runs using the **vegdist**. This functionality has been utilized in the Mothur standalone microbial ecology toolkit here.

Author(s)
Geoffrey Hannigan.

See Also
This function utilizes the **vegdist** and **rrarefy** functions.

Examples
```r
# Import an example count dataset
data(BCI)
# Test the base functionality
mean.avg.dist <- avgdist(BCI, sample = 50, iterations = 10)
# Test the transformation function
mean.avg.dist.t <- avgdist(BCI, sample = 50, iterations = 10, transf = sqrt)
# Test the median functionality
median.avg.dist <- avgdist(BCI, sample = 50, iterations = 10, meanfun = median)
# Print the resulting tables
head(as.matrix(mean.avg.dist))
head(as.matrix(mean.avg.dist.t))
head(as.matrix(median.avg.dist))
# Run example to illustrate low variance of mean, median, and stdev results
# Mean and median std dev are around 0.05
sdd <- avgdist(BCI, sample = 50, iterations = 100, meanfun = sd)
sdd <- avgdist(BCI, sample = 450, iterations = 10)
# Print the result
depth.avg.dist
```
Barro Colorado Island Tree Counts

Description

Tree counts in 1-hectare plots in the Barro Colorado Island and associated site information.

Usage

data(BCI)
data(BCI.env)

Format

A data frame with 50 plots (rows) of 1 hectare with counts of trees on each plot with total of 225 species (columns). Full Latin names are used for tree species. The names were updated against http://www.theplantlist.org and Kress et al. (2009) which allows matching 207 of species against doi: 10.5061/dryad.63q27 (Zanne et al., 2014). The original species names are available as attribute original.names of BCI. See Examples for changed names.

For BCI.env, a data frame with 50 plots (rows) and nine site variables derived from Pyke et al. (2001) and Harms et al. (2001):

- **UTM.EW**: UTM coordinates (zone 17N) East-West.
- **UTM.NS**: UTM coordinates (zone 17N) North-South.
- **Precipitation**: Precipitation in mm per year.
- **Elevation**: Elevation in m above sea level.
- **Age.cat**: Forest age category.
- **Geology**: The Underlying geological formation.
- **Habitat**: Dominant habitat type based on the map of habitat types in 25 grid cells in each plot (Harms et al. 2001, excluding streamside habitat). The habitat types are Young forests (ca. 100 years), old forests on > 7 degree slopes (OldSlope), old forests under 152 m elevation (OldLow) and at higher elevation (OldHigh) and Swamp forests.
- **River**: "Yes" if there is streamside habitat in the plot.
- **EnvHet**: Environmental Heterogeneity assessed as the Simpson diversity of frequencies of Habitat types in 25 grid cells in the plot.

Details

Data give the numbers of trees at least 10 cm in diameter at breast height (1.3 m above the ground) in each one hectare square of forest. Within each one hectare square, all individuals of all species were tallied and are recorded in this table.

The data frame contains only the Barro Colorado Island subset of the original data.

The quadrats are located in a regular grid. See BCI.env for the coordinates.

A full description of the site information in BCI.env is given in Pyke et al. (2001) and Harms et al. (2001). *N.B.* Pyke et al. (2001) and Harms et al. (2001) give conflicting information about forest age categories and elevation.
Source

http://www.sciencemag.org/cgi/content/full/295/5555/666/DC1 for community data and References for environmental data.

References


See Also


Examples

data(BCI, BCI.env)
head(BCI.env)
## see changed species names
oldnames <- attr(BCI, "original.names")
taxa <- cbind("Old Names" = oldnames, "Current Names" = names(BCI))
noquote(taxa[taxa[,1] != taxa[,2], ])

beals Beals Smoothing and Degree of Absence

Description

Beals smoothing replaces each entry in the community data with a probability of a target species occurring in that particular site, based on the joint occurrences of the target species with the species that actually occur in the site. Swan’s (1970) degree of absence applies Beals smoothing to zero items so long that all zeros are replaced with smoothed values.
Usage

beals(x, species = NA, reference = x, type = 0, include = TRUE)
swan(x, maxit = Inf, type = 0)

Arguments

x Community data frame or matrix.
species Column index used to compute Beals function for a single species. The default
(NA) indicates that the function will be computed for all species.
reference Community data frame or matrix to be used to compute joint occurrences. By
default, x is used as reference to compute the joint occurrences.
type Numeric. Specifies if and how abundance values have to be used in function
beals. See details for more explanation.
include This logical flag indicates whether the target species has to be included when
computing the mean of the conditioned probabilities. The original Beals (1984)
definition is equivalent to include=TRUE, while the formulation of Münzbergová and Herben is equal to include=FALSE.

maxit Maximum number of iterations. The default Inf means that iterations are con-
tinued until there are no zeros or the number of zeros does not change. Probably
only maxit = 1 makes sense in addition to the default.

Details

Beals smoothing is the estimated probability $p_{ij}$ that species $j$ occurs at site $i$. It is defined as $p_{ij} = \frac{1}{S_i} \sum_k \frac{N_{jk}I_{ik}}{N_k}$, where $S_i$ is the number of species at site $i$, $N_{jk}$ is the number of joint occurrences of species $j$ and $k$, $N_k$ is the number of occurrences of species $k$, and $I$ is the incidence (0 or 1) of species (this last term is usually omitted from the equation, but it is necessary). As $N_{jk}$ can be interpreted as a mean of conditioned probability, the beals function can be interpreted as a mean
of conditioned probabilities (De Cáceres & Legendre 2008). The present function is generalized to
abundance values (De Cáceres & Legendre 2008).

The type argument specifies if and how abundance values have to be used. type = 0 presence/absence mode. type = 1 abundances in reference (or x) are used to compute conditioned probabilities. type = 2 abundances in x are used to compute weighted averages of conditioned probabilities. type = 3 abundances are used to compute both conditioned probabilities and weighted averages.

Beals smoothing was originally suggested as a method of data transformation to remove excessive
zeros (Beals 1984, McCune 1994). However, it is not a suitable method for this purpose since it does not maintain the information on species presences: a species may have a higher probability of occurrence at a site where it does not occur than at sites where it occurs. Moreover, it regularizes data too strongly. The method may be useful in identifying species that belong to the species pool (Ewald 2002) or to identify suitable unoccupied patches in metapopulation analysis
(Münzbergová & Herben 2004). In this case, the function should be called with include=FALSE
for cross-validation smoothing for species; argument species can be used if only one species is studied.

Swan (1970) suggested replacing zero values with degrees of absence of a species in a community
data matrix. Swan expressed the method in terms of a similarity matrix, but it is equivalent to
applying Beals smoothing to zero values, at each step shifting the smallest initially non-zero item to value one, and repeating this so many times that there are no zeros left in the data. This is actually very similar to extended dissimilarities (implemented in function \texttt{stepacross}), but very rarely used.

**Value**

The function returns a transformed data matrix or a vector if Beals smoothing is requested for a single species.

**Author(s)**

Miquel De Cáceres and Jari Oksanen

**References**


**See Also**

\texttt{decostand} for proper standardization methods, \texttt{specpool} for an attempt to assess the size of species pool. Function \texttt{indpower} assesses the power of each species to estimate the probabilities predicted by \texttt{beals}.

**Examples**

data(dune)

```r
## Default
x <- beals(dune)
## Remove target species
x <- beals(dune, include = FALSE)
## Smoothed values against presence or absence of species
pa <- decostand(dune, "pa")
boxplot(as.vector(x) ~ unlist(pa), xlab="Presence", ylab="Beals")
## Remove the bias of target species: Yields lower values.
beals(dune, type =3, include = FALSE)
## Uses abundance information.
## Vector with beals smoothing values corresponding to the first species
```
## betadisper

**Multivariate homogeneity of groups dispersions (variances)**

### Description

Implements Marti Anderson’s PERMDISP2 procedure for the analysis of multivariate homogeneity of group dispersions (variances). `betadisper` is a multivariate analogue of Levene’s test for homogeneity of variances. Non-euclidean distances between objects and group centroids are handled by reducing the original distances to principal coordinates. This procedure has latterly been used as a means of assessing beta diversity. There are `anova`, `scores`, `plot` and `boxplot` methods. `TukeyHSD`. `betadisper` creates a set of confidence intervals on the differences between the mean distance-to-centroid of the levels of the grouping factor with the specified family-wise probability of coverage. The intervals are based on the Studentized range statistic, Tukey’s ‘Honest Significant Difference’ method.

### Usage

```r
betadisper(d, group, type = c("median","centroid"), bias.adjust = FALSE, sqrt.dist = FALSE, add = FALSE)
```

## S3 method for class 'betadisper'

```r
anova(object, …)
```

## S3 method for class 'betadisper'

```r
scores(x, display = c("sites","centroids"), choices = c(1,2), …)
```

## S3 method for class 'betadisper'

```r
eigenvals(x, …)
```

## S3 method for class 'betadisper'

```r
plot(x, axes = c(1,2), cex = 0.7, pch = seq_len(ng), col = NULL, lty = "solid", lwd = 1, hull = TRUE, ellipse = FALSE, conf, segments = TRUE, seg.col = "grey", seg.lty = lty, seg.lwd = lwd, label = TRUE, label.cex = 1, ylab, xlab, main, sub, …)
```

## S3 method for class 'betadisper'

```r
boxplot(x, ylab = "Distance to centroid", …)
```

## S3 method for class 'betadisper'

```r
TukeyHSD(x, which = "group", ordered = FALSE, conf.level = 0.95, …)
```
**Arguments**

- **d**: a distance structure such as that returned by `dist`, `betadiver` or `vegdist`.
- **group**: vector describing the group structure, usually a factor or an object that can be coerced to a factor using `as.factor`. Can consist of a factor with a single level (i.e., one group).
- **type**: the type of analysis to perform. Use the spatial median or the group centroid? The spatial median is now the default.
- **bias.adjust**: logical: adjust for small sample bias in beta diversity estimates?
- **sqrt.dist**: Add a constant to the non-diagonal dissimilarities such that all eigenvalues are non-negative in the underlying Principal Co-ordinates Analysis (see `cmdscale` for details). Choice "lingoes" (or TRUE) use the recommended method of Legendre & Anderson (1999: "method 1") and "cailliez" uses their "method 2".
- **display**: character; partial match to access scores for "sites" or "species".
- **object, x**: an object of class "betadisper", the result of a call to `betadisper`.
- **choices, axes**: the principal coordinate axes wanted.
- **hull**: logical; should the convex hull for each group be plotted?
- **ellipse**: logical; should the standard deviation data ellipse for each group be plotted?
- **conf**: Expected fractions of data coverage for data ellipses, e.g. 0.95. The default is to draw a 1 standard deviation data ellipse, but if supplied, `conf` is multiplied with the corresponding value found from the Chi-squared distribution with 2df to provide the requested coverage (probability contour).
- **pch**: plot symbols for the groups, a vector of length equal to the number of groups.
- **col**: colors for the plot symbols for the groups, a vector of length equal to the number of groups.
- **lty, lwd**: linetype, linewidth for convex hulls and confidence ellipses.
- **segments**: logical; should segments joining points to their centroid be drawn?
- **seg.col**: colour to draw segments between points and their centroid. Can be a vector, in which case one colour per group.
- **seg.lty, seg.lwd**: linetype and line width for segments.
- **label**: logical; should the centroids by labelled with their respective factor label?
- **label.cex**: numeric; character expansion for centroid labels.
- **cex, ylab, xlab, main, sub**: graphical parameters. For details, see `plot.default`.
- **which**: A character vector listing terms in the fitted model for which the intervals should be calculated. Defaults to the grouping factor.
- **ordered**: logical; see `TukeyHSD`.
- **conf.level**: A numeric value between zero and one giving the family-wise confidence level to use.
- **...**: arguments, including graphical parameters (for `plot.betadisper` and `boxplot.betadisper`), passed to other methods.
Details

One measure of multivariate dispersion (variance) for a group of samples is to calculate the average distance of group members to the group centroid or spatial median (both referred to as ‘centroid’ from now on unless stated otherwise) in multivariate space. To test if the dispersions (variances) of one or more groups are different, the distances of group members to the group centroid are subject to ANOVA. This is a multivariate analogue of Levene’s test for homogeneity of variances if the distances between group members and group centroids is the Euclidean distance.

However, better measures of distance than the Euclidean distance are available for ecological data. These can be accommodated by reducing the distances produced using any dissimilarity coefficient to principal coordinates, which embeds them within a Euclidean space. The analysis then proceeds by calculating the Euclidean distances between group members and the group centroid on the basis of the principal coordinate axes rather than the original distances.

Non-metric dissimilarity coefficients can produce principal coordinate axes that have negative Eigenvalues. These correspond to the imaginary, non-metric part of the distance between objects. If negative Eigenvalues are produced, we must correct for these imaginary distances.

The distance to its centroid of a point is

\[ z_{ij}^c = \sqrt{\Delta^2(u_{ij}^+, c_i^+)} - \Delta^2(u_{ij}^-, c_i^-) \]

where \( \Delta^2 \) is the squared Euclidean distance between \( u_{ij} \), the principal coordinate for the \( j \)th point in the \( i \)th group, and \( c_i \), the coordinate of the centroid for the \( i \)th group. The super-scripted ‘+’ and ‘−’ indicate the real and imaginary parts respectively. This is equation (3) in Anderson (2006). If the imaginary part is greater in magnitude than the real part, then we would be taking the square root of a negative value, resulting in NaN. Function takes the absolute value of the real distance minus the imaginary distance, before computing the square root. This is in line with the behaviour of Marti Anderson’s PERMDISP2 programme.

To test if one or more groups is more variable than the others, ANOVA of the distances to group centroids can be performed and parametric theory used to interpret the significance of \( F \). An alternative is to use a permutation test. *permutestNbetadisper* permutes model residuals to generate a permutation distribution of \( F \) under the Null hypothesis of no difference in dispersion between groups.

Pairwise comparisons of group mean dispersions can also be performed using *permutestNbetadisper*. An alternative to the classical comparison of group dispersions, is to calculate Tukey’s Honest Significant Differences between groups, via *TukeyHSD*.betadisper. This is a simple wrapper to *TukeyHSD*. The user is directed to read the help file for *TukeyHSD* before using this function. In particular, note the statement about using the function with unbalanced designs.

The results of the analysis can be visualised using the *plot* and *boxplot* methods.

One additional use of these functions is in assessing beta diversity (Anderson *et al* 2006). Function *betadiver* provides some popular dissimilarity measures for this purpose.

As noted in passing by Anderson (2006) and in a related context by O’Neill (2000), estimates of dispersion around a central location (median or centroid) that is calculated from the same data will be biased downward. This bias matters most when comparing diversity among treatments with small, unequal numbers of samples. Setting *bias.adjust=TRUE* when using *betadisper* imposes a \( \sqrt{n/(n - 1)} \) correction (Stier *et al*. 2013).
Value

The anova method returns an object of class "anova" inheriting from class "data.frame".
The scores method returns a list with one or both of the components "sites" and "centroids".
The plot function invisibly returns an object of class "ordiplot", a plotting structure which can
be used by identify.ordiplot (to identify the points) or other functions in the ordiplot family.
The boxplot function invisibly returns a list whose components are documented in boxplot.
eigenvals.betadisper returns a named vector of eigenvalues.
TukeyHSD.betadisper returns a list. See TukeyHSD for further details.
betadisper returns a list of class "betadisper" with the following components:

- eig numeric; the eigenvalues of the principal coordinates analysis.
- vectors matrix; the eigenvectors of the principal coordinates analysis.
- distances numeric; the Euclidean distances in principal coordinate space between the sam-
ples and their respective group centroid.
- group factor; vector describing the group structure
- centroids matrix; the locations of the group centroids on the principal coordinates.
- call the matched function call.

Warning

Stewart Schultz noticed that the permutation test for type="centroid" had the wrong type I error
and was anti-conservative. As such, the default for type has been changed to "median", which uses
the spatial median as the group centroid. Tests suggest that the permutation test for this type of
analysis gives the correct error rates.

Note

If group consists of a single level or group, then the anova and permutest methods are not appro-
priate and if used on such data will stop with an error.
Missing values in either d or group will be removed prior to performing the analysis.

Author(s)

Gavin L. Simpson; bias correction by Adrian Stier and Ben Bolker.

References

Stier, A.C., Geange, S.W., Hanson, K.M., & Bolker, B.M. (2013) Predator density and timing of
See Also

`permutest.betadisper`, `anova.lm`, `scores`, `boxplot`, `TukeyHSD`. Further measure of beta diversity can be found in `betadiver`.

Examples

data(varespec)

```r
## Bray-Curtis distances between samples
dis <- vegdist(varespec)

## First 16 sites grazed, remaining 8 sites ungrazed
groups <- factor(c(rep(1,16), rep(2,8)), labels = c("grazed","ungrazed"))

## Calculate multivariate dispersions
mod <- betadisper(dis, groups)

## Perform test
anova(mod)

## Permutation test for F
permutest(mod, pairwise = TRUE, permutations = 99)

## Tukey's Honest Significant Differences
(mod.HSD <- TukeyHSD(mod))
plot(mod.HSD)

## Plot the groups and distances to centroids on the
## first two PCoA axes
plot(mod)

## with data ellipses instead of hulls
plot(mod, ellipse = TRUE, hull = FALSE) # 1 sd data ellipse
plot(mod, ellipse = TRUE, hull = FALSE, conf = 0.90) # 90% data ellipse

## can also specify which axes to plot, ordering respected
plot(mod, axes = c(3,1), seg.col = "forestgreen", seg.lty = "dashed")

## Draw a boxplot of the distances to centroid for each group
boxplot(mod)

## `scores` and `eigenvals` also work
scrs <- scores(mod)
str(scrs)
head(scores(mod, 1:4, display = "sites"))
# group centroids/medians
scores(mod, 1:4, display = "centroids")
# eigenvalues from the underlying principal coordinates analysis
eigenvals(mod)

## try out bias correction; compare with mod3
```
(mod3B <- betadisper(dis, groups, type = "median", bias.adjust=TRUE))
anova(mod3B)
permutest(mod3B, permutations = 99)

## should always work for a single group
group <- factor(rep("grazed", NROW(varespec)))
(tmp <- betadisper(dis, group, type = "median"))
(tmp <- betadisper(dis, group, type = "centroid"))

## simulate missing values in 'd' and 'group'
## using spatial medians
groups[c(2,20)] <- NA
dis[c(2, 20)] <- NA
mod2 <- betadisper(dis, groups) ## messages
mod2
permutest(mod2, permutations = 99)
anova(mod2)
plot(mod2)
boxplot(mod2)
plot(TukeyHSD(mod2))

## Using group centroids
mod3 <- betadisper(dis, groups, type = "centroid")
mod3
permutest(mod3, permutations = 99)
anova(mod3)
plot(mod3)
boxplot(mod3)
plot(TukeyHSD(mod3))

---

**betadiver**

### Indices of beta Diversity

**Description**

The function estimates any of the 24 indices of beta diversity reviewed by Koleff et al. (2003). Alternatively, it finds the co-occurrence frequencies for triangular plots (Koleff et al. 2003).

**Usage**

```r
betadiver(x, method = NA, order = FALSE, help = FALSE, ...)
## S3 method for class 'betadiver'
plot(x, ...)
## S3 method for class 'betadiver'
scores(x, triangular = TRUE, ...)
```
Arguments

x          Community data matrix, or the betadiver result for plot and scores functions.
method     The index of beta diversity as defined in Koleff et al. (2003), Table 1. You can use either the subscript of $\beta$ or the number of the index. See argument help below.
order      Order sites by increasing number of species. This will influence the configuration in the triangular plot and non-symmetric indices.
help       Show the numbers, subscript names and the defining equations of the indices and exit.
triangular Return scores suitable for triangular plotting of proportions. If FALSE, returns a 3-column matrix of raw counts.
...        Other arguments to functions.

Details

The most commonly used index of beta diversity is $\beta_w = S/\alpha - 1$, where $S$ is the total number of species, and $\alpha$ is the average number of species per site (Whittaker 1960). A drawback of this model is that $S$ increases with sample size, but the expectation of $\alpha$ remains constant, and so the beta diversity increases with sample size. A solution to this problem is to study the beta diversity of pairs of sites (Marion et al. 2017). If we denote the number of species shared between two sites as $a$ and the numbers of unique species (not shared) as $b$ and $c$, then $S = a + b + c$ and $\alpha = (2a + b + c)/2$ so that $\beta_w = (b + c)/(2a + b + c)$. This is the Sørensen dissimilarity as defined in vegan function vegdist with argument binary = TRUE. Many other indices are dissimilarity indices as well.

Function betadiver finds all indices reviewed by Koleff et al. (2003). All these indices could be found with function designdist, but the current function provides a conventional shortcut. The function only finds the indices. The proper analysis must be done with functions such as betadisper, adonis2 or mantel.

The indices are directly taken from Table 1 of Koleff et al. (2003), and they can be selected either by the index number or the subscript name used by Koleff et al. The numbers, names and defining equations can be seen using betadiver(help = TRUE). In all cases where there are two alternative forms, the one with the term $-1$ is used. There are several duplicate indices, and the number of distinct alternatives is much lower than 24 formally provided. The formulations used in functions differ occasionally from those in Koleff et al. (2003), but they are still mathematically equivalent. With method = NA, no index is calculated, but instead an object of class betadiver is returned. This is a list of elements $a$, $b$ and $c$. Function plot can be used to display the proportions of these elements in triangular plot as suggested by Koleff et al. (2003), and scores extracts the triangular coordinates or the raw scores. Function plot returns invisibly the triangular coordinates as an "ordiplot" object.

Value

With method = NA, the function returns an object of class "betadisper" with elements $a$, $b$, and $c$. If method is specified, the function returns a "dist" object which can be used in any function analysing dissimilarities. For beta diversity, particularly useful functions are betadisper to study the betadiversity in groups, adonis2 for any model, and mantel to compare beta diversities to other dissimilarities or distances (including geographical distances). Although betadiver returns a
"dist" object, some indices are similarities and cannot be used as such in place of dissimilarities, but that is a severe user error. Functions 10 ("j") and 11 ("sor") are two such similarity indices.

Warning

Some indices return similarities instead of dissimilarities.

Author(s)

Jari Oksanen

References


See Also

designdist can be used to implement all these functions, and also allows using notation with alpha and gamma diversities. vegdist has some canned alternatives. Functions betadisper, adonis2 and mantel can be used for analysing beta diversity objects. The returned dissimilarities can be used in any distance-based methods, such as metaMDS, capscale and dbda. Functions nestedbetasor and nestedbetajac implement decomposition beta diversity measures (Sørensen and Jaccard) into turnover and nestedness components following Baselga (2010).

Examples

```r
## Raw data and plotting
data(sipoo)
m <- betadiver(sipoo)
plot(m)
## The indices
betadiver(help=TRUE)
## The basic Whittaker index
d <- betadiver(sipoo, "w")
## This should be equal to Sorensen index (binary Bray-Curtis in ## vegan)
range(d - vegdist(sipoo, binary=TRUE))
```
Description

This function computes coefficients of dispersal direction between geographically connected areas, as defined by Legendre and Legendre (1984), and also described in Legendre and Legendre (2012, section 13.3.4).

Usage

`bgdispersal(mat, PAonly = FALSE, abc = FALSE)`

Arguments

- `mat` Data frame or matrix containing a community composition data table (species presence-absence or abundance data).
- `PAonly` FALSE if the four types of coefficients, DD1 to DD4, are requested; TRUE if DD1 and DD2 only are sought (see Details).
- `abc` If TRUE, return tables a, b and c used in DD1 and DD2.

Details

The signs of the DD coefficients indicate the direction of dispersal, provided that the asymmetry is significant. A positive sign indicates dispersal from the first (row in DD tables) to the second region (column); a negative sign indicates the opposite. A McNemar test of asymmetry is computed from the presence-absence data to test the hypothesis of a significant asymmetry between the two areas under comparison.

In the input data table, the rows are sites or areas, the columns are taxa. Most often, the taxa are species, but the coefficients can be computed from genera or families as well. DD1 and DD2 only are computed for presence-absence data. The four types of coefficients are computed for quantitative data, which are converted to presence-absence for the computation of DD1 and DD2. `PAonly = FALSE` indicates that the four types of coefficients are requested. `PAonly = TRUE` if DD1 and DD2 only are sought.

Value

Function `bgdispersal` returns a list containing the following matrices:

- **DD1**
  \[ DD1_{j,k} = \frac{(a(b-c))}{((a + b + c)^2)} \]

- **DD2**
  \[ DD2_{j,k} = \frac{(2a(b-c))}{((2a + b + c)(a + b + c))} \text{ where } a, b, \text{ and } c \text{ have the same meaning as in the computation of binary similarity coefficients.} \]

- **DD3**
  \[ DD3_{j,k} = \frac{W(A - B)}{(A + B - W)^2} \]

- **DD4**
  \[ DD4_{j,k} = 2\frac{W(A-B)}{(A+B)(A-B-W))} \text{ where } W = \text{sum}(\text{pmin(vector1, vector2)}), \text{ A = sum(vector1), B = sum(vector2)} \]
### McNemar

McNemar chi-square statistic of asymmetry (Sokal and Rohlf 1995): $2(b \log(b) + c \log(c) - (b + c) \log((b + c)/2))/q$, where $q = 1 + 1/(2(b + c))$ (Williams correction for continuity)

### prob. McNemar

probabilities associated with McNemar statistics, chi-square test. H0: no asymmetry in $(b - c)$.

#### Note

The function uses a more powerful alternative for the McNemar test than the classical formula. The classical formula was constructed in the spirit of Pearson’s Chi-square, but the formula in this function was constructed in the spirit of Wilks Chi-square or the $G$ statistic. Function `mcnemar.test` uses the classical formula. The new formula was introduced in `vegan` version 1.10-11, and the older implementations of `bgdispersal` used the classical formula.

### Author(s)

Pierre Legendre, Departement de Sciences Biologiques, Universite de Montreal

### References


### Examples

```r
mat <- matrix(c(32,15,14,10,70,30,100,4,10,30,25,0,18,0,40, 
0,0,20,0,0,4,0,0,0,0,0,30,20,0,0,0,0,0,0,25,74,42,1,45,89,5,16,16,20), 
4, 10, byrow=TRUE)
bgdispersal(mat)
```

### bioenv

*Best Subset of Environmental Variables with Maximum (Rank) Correlation with Community Dissimilarities*

#### Description

Function finds the best subset of environmental variables, so that the Euclidean distances of scaled environmental variables have the maximum (rank) correlation with community dissimilarities.
Usage

## Default S3 method:

```r
bioenv(comm, env, method = "spearman", index = "bray",
upto = ncol(env), trace = FALSE, partial = NULL,
metric = c("euclidean", "mahalanobis", "manhattan", "gower"),
parallel =getOption("mc.cores"), ...)
```

## S3 method for class 'formula'

```r
bioenv(formula, data, ...)
```

```r
bioenvdist(x, which = "best")
```

Arguments

- `comm`: Community data frame or a dissimilarity object or a square matrix that can be interpreted as dissimilarities.
- `env`: Data frame of continuous environmental variables.
- `method`: The correlation method used in `cor`.
- `index`: The dissimilarity index used for community data (`comm`) in `vegdist`. This is ignored if `comm` are dissimilarities.
- `upto`: Maximum number of parameters in studied subsets.
- `formula, data`: Model formula and data.
- `trace`: Trace the calculations
- `partial`: Dissimilarities partialled out when inspecting variables in `env`.
- `metric`: Metric used for distances of environmental distances. See Details.
- `parallel`: Number of parallel processes or a predefined socket cluster. With `parallel = 1` uses ordinary, non-parallel processing. The parallel processing is done with `parallel` package.
- `x`: bioenv result object.
- `which`: The number of the model for which the environmental distances are evaluated, or the "best" model.
- `...`: Other arguments passed to `cor`.

Details

The function calculates a community dissimilarity matrix using `vegdist`. Then it selects all possible subsets of environmental variables, `scales` the variables, and calculates Euclidean distances for this subset using `dist`. The function finds the correlation between community dissimilarities and environmental distances, and for each size of subsets, saves the best result. There are \(2^p - 1\) subsets of \(p\) variables, and an exhaustive search may take a very, very, very long time (parameter `upto` offers a partial relief).

The argument `metric` defines distances in the given set of environmental variables. With `metric = "euclidean"`, the variables are scaled to unit variance and Euclidean distances are calculated. With `metric = "mahalanobis"`, the Mahalanobis distances are calculated: in addition to scaling to unit variance, the matrix of the current set of environmental variables is also made orthogonal (uncorrelated). With `metric = "manhattan"`, the variables are scaled to unit range and Manhattan distances are calculated, so that the distances
are sums of differences of environmental variables. With metric = "gower", the Gower distances are calculated using function daisy. This allows also using factor variables, but with continuous variables the results are equal to metric = "manhattan".

The function can be called with a model formula where the LHS is the data matrix and RHS lists the environmental variables. The formula interface is practical in selecting or transforming environmental variables.

With argument partial you can perform “partial” analysis. The partializing item must be a dissimilarity object of class dist. The partial item can be used with any correlation method, but it is strictly correct only for Pearson.

Function bioenvdist recalculates the environmental distances used within the function. The default is to calculate distances for the best model, but the number of any model can be given.

Clarke & Ainsworth (1993) suggested this method to be used for selecting the best subset of environmental variables in interpreting results of nonmetric multidimensional scaling (NMDS). They recommended a parallel display of NMDS of community dissimilarities and NMDS of Euclidean distances from the best subset of scaled environmental variables. They warned against the use of Procrustes analysis, but to me this looks like a good way of comparing these two ordinations.

Clarke & Ainsworth wrote a computer program BIO-ENV giving the name to the current function. Presumably BIO-ENV was later incorporated in Clarke’s PRIMER software (available for Windows). In addition, Clarke & Ainsworth suggested a novel method of rank correlation which is not available in the current function.

Value

The function returns an object of class bioenv with a summary method.

Note

If you want to study the ‘significance’ of bioenv results, you can use function mantel or mantel.partial which use the same definition of correlation. However, bioenv standardizes environmental variables depending on the used metric, and you must do the same in mantel for comparable results (the standardized data are returned as item x in the result object). It is safest to use bioenvdist to extract the environmental distances that really were used within bioenv. NB., bioenv selects variables to maximize the Mantel correlation, and significance tests based on a priori selection of variables are biased.

Author(s)

Jari Oksanen

References

See Also

vegdist, dist, cor for underlying routines, monomDS and metaMDS for ordination, procrustes for Procrustes analysis, protest for an alternative, and rankindex for studying alternatives to the default Bray-Curtis index.

Examples

# The method is very slow for large number of possible subsets.
# Therefore only 6 variables in this example.

data(varespec)
data(varechem)
sol <- bioenv(wisconsin(varespec) ~ log(N) + P + K + Ca + pH + Al, varechem)
sol
summary(sol)

Description

Draws a PCA biplot with species scores indicated by biplot arrows

Usage

## S3 method for class 'rda'
biplot(x, choices = c(1, 2), scaling = "species",
       display = c("sites", "species"), type, xlim, ylim, col = c(1,2),
      const, correlation = FALSE, ...)

Arguments

x  A *rda* result object.
choices  Axes to show.
scaling  Scaling for species and site scores. Either species (2) or site (1) scores are scaled by eigenvalues, and the other set of scores is left unscaled, or with 3 both are scaled symmetrically by square root of eigenvalues. With negative scaling values in rda, species scores are divided by standard deviation of each species and multiplied with an equalizing constant. Unscaled raw scores stored in the result can be accessed with scaling = 0.

The type of scores can also be specified as one of "none", "sites", "species", or "symmetric", which correspond to the values 0, 1, 2, and 3 respectively. Argument correlation can be used in combination with these character descriptions to get the corresponding negative value.

correlation  logical; if scaling is a character description of the scaling type, correlation can be used to select correlation-like scores for PCA. See argument scaling for details.
Scores shown. These must some of the alternatives "species" for species scores, and/or "sites" for site scores.

Type of plot: partial match to text for text labels, points for points, and none for setting frames only. If omitted, text is selected for smaller data sets, and points for larger. Can be of length 2 (e.g. type = c("text", "points")), in which case the first element describes how species scores are handled, and the second how site scores are drawn.

The x and y limits (min, max) of the plot.

Colours used for sites and species (in this order). If only one colour is given, it is used for both.

General scaling constant for scores.rda.

Other parameters for plotting functions.

Produces a plot or biplot of the results of a call to rda. It is common for the "species" scores in a PCA to be drawn as biplot arrows that point in the direction of increasing values for that variable. The biplot.rda function provides a wrapper to plot.cca to allow the easy production of such a plot.

biplot.rda is only suitable for unconstrained models. If used on an ordination object with constraints, an error is issued.

If species scores are drawn using "text", the arrows are drawn from the origin to 0.85 * species score, whilst the labels are drawn at the species score. If the type used is "points", then no labels are drawn and therefore the arrows are drawn from the origin to the actual species score.

The plot function returns invisibly a plotting structure which can be used by identify.ordiplot to identify the points or other functions in the ordiplot family.

Gavin Simpson, based on plot.cca by Jari Oksanen.

plot.cca, rda for something to plot, ordiplot for an alternative plotting routine and more support functions, and text, points and arrows for the basic routines.

Example:

```r
data(dune)
mod <- rda(dune, scale = TRUE)
bplot(mod, scaling = "symmetric")

# different type for species and site scores
bplot(mod, scaling = "symmetric", type = c("text", "points"))
```
Partial Distance-based Redundancy Analysis

Description

Distance-based redundancy analysis (dbRDA) is an ordination method similar to Redundancy Analysis (rda), but it allows non-Euclidean dissimilarity indices, such as Manhattan or Bray-Curtis distance. Despite this non-Euclidean feature, the analysis is strictly linear and metric. If called with Euclidean distance, the results are identical to rda, but dbRDA will be less efficient. Functions capscale and dbrda are constrained versions of metric scaling, a.k.a. principal coordinates analysis, which are based on the Euclidean distance but can be used, and are more useful, with other dissimilarity measures. The functions can also perform unconstrained principal coordinates analysis, optionally using extended dissimilarities.

Usage

capscale(formula, data, distance = "euclidean", sqrt.dist = FALSE, comm = NULL, add = FALSE, dfun = vegdist, metaMDSdist = FALSE, na.action = na.fail, subset = NULL, ...)
dbrda(formula, data, distance = "euclidean", sqrt.dist = FALSE, add = FALSE, dfun = vegdist, metaMDSdist = FALSE, na.action = na.fail, subset = NULL, ...)

Arguments

- **formula**: Model formula. The function can be called only with the formula interface. Most usual features of formula hold, especially as defined in cca and rda. The LHS must be either a community data matrix or a dissimilarity matrix, e.g., from vegdist or dist. If the LHS is a data matrix, function vegdist or function given in dfun will be used to find the dissimilarities. The RHS defines the constraints. The constraints can be continuous variables or factors, they can be transformed within the formula, and they can have interactions as in a typical formula. The RHS can have a special term Condition that defines variables to be “partialled out” before constraints, just like in rda or cca. This allows the use of partial dbRDA.

- **data**: Data frame containing the variables on the right hand side of the model formula.

- **distance**: The name of the dissimilarity (or distance) index if the LHS of the formula is a data frame instead of dissimilarity matrix.

- **sqrt.dist**: Take square roots of dissimilarities. See section Details below.

- **comm**: Community data frame which will be used for finding species scores when the LHS of the formula was a dissimilarity matrix. This is not used if the LHS is a data frame. If this is not supplied, the “species scores” are unavailable when dissimilarities were supplied. N.B., this is only available in capscale: dbrda does not return species scores. Function sppscores can be used to add species scores if they are missing.
Add a constant to the non-diagonal dissimilarities such that all eigenvalues are non-negative in the underlying Principal Co-ordinates Analysis (see `wcmdscale` for details). "lingoes" (or TRUE) uses the recommended method of Legendre & Anderson (1999: "method 1") and "cailliez" uses their "method 2". The latter is the only one in `cmdscale`.

dfun

Distance or dissimilarity function used. Any function returning standard "dist" and taking the index name as the first argument can be used.

metaMDSdist

Use `metaMDSdist` similarly as in `metaMDS`. This means automatic data transformation and using extended flexible shortest path dissimilarities (function `stepacross`) when there are many dissimilarities based on no shared species.

na.action

Handling of missing values in constraints or conditions. The default (na.fail) is to stop with missing values. Choices na.omit and na.exclude delete rows with missing values, but differ in representation of results. With na.omit only non-missing site scores are shown, but na.exclude gives NA for scores of missing observations. Unlike in `rda`, no WA scores are available for missing constraints or conditions.

subset

Subset of data rows. This can be a logical vector which is TRUE for kept observations, or a logical expression which can contain variables in the working environment, data or species names of the community data (if given in the formula or as comm argument).

... Other parameters passed to underlying functions (e.g., `metaMDSdist`).

Details

Functions `capscale` and `dbrda` provide two alternative implementations of dbRDA. Function `capscale` is based on Legendre & Anderson (1999): the dissimilarity data are first ordinated using metric scaling, and the ordination results are analysed as `rda`. Function `dbrda` is based on McArdle & Anderson (2001) and directly decomposes dissimilarities. It does not use `rda` but a parallel implementation adapted for analysing dissimilarities and returns a subset of `rda` items. With Euclidean distances both results are identical to `rda`. Other dissimilarities may give negative eigenvalues associated with imaginary axes. Negative eigenvalues are handled differently: `capscale` ignores imaginary axes and analyses only real axes with positive eigenvalues, and `dbrda` directly analyses dissimilarities and can give negative eigenvalues in any component.

If the user supplied a community data frame instead of dissimilarities, the functions will find dissimilarities using `vegdist` or distance function given in `dfun` with specified distance. The functions will accept distance objects from `vegdist`, `dist`, or any other method producing compatible objects. The constraining variables can be continuous or factors or both, they can have interaction terms, or they can be transformed in the call. Moreover, there can be a special term `condition` just like in `rda` and `cca` so that “partial” analysis can be performed.

Function `dbrda` does not return species scores, and they can also be missing in `capscale`, but they can be added after the analysis using function `sppscores`.

Non-Euclidean dissimilarities can produce negative eigenvalues (Legendre & Anderson 1999, McArdle & Anderson 2001). If there are negative eigenvalues, the printed output of `capscale` will add a column with sums of positive eigenvalues and an item of sum of negative eigenvalues, and `dbrda` will add a column giving the number of real dimensions with positive eigenvalues. If negative eigenvalues are disturbing, functions let you to distort the dissimilarities so that only non-negative
eigenvalues will be produced with argument add = TRUE. Alternatively, with sqrt.dist = TRUE, square roots of dissimilarities will be used which may help in avoiding negative eigenvalues (Legendre & Anderson 1999).

The functions can be also used to perform ordinary metric scaling a.k.a. principal coordinates analysis by using a formula with only a constant on the left hand side, or $\text{comm} \sim 1$. With metaMDSdist = TRUE, the function can do automatic data standardization and use extended dissimilarities using function stepacross similarly as in non-metric multidimensional scaling with metaMDS.

**Value**

The functions return an object of class capscale or dbRDA which inherits from rda. See cca.object for description of the result object.

**Note**

The function capscale was originally developed as a variant of constrained analysis of proximities (Anderson & Willis 2003), but these developments made it similar to dbRDA. However, it discards the imaginary dimensions with negative eigenvalues and ordination and significance tests area only based on real dimensions and positive eigenvalues.

The inertia is named after the dissimilarity index as defined in the dissimilarity data, or as unknown distance if such information is missing. If the largest original dissimilarity was larger than 4, capscale handles input similarly as rda and bases its analysis on variance instead of sum of squares. Keyword mean is added to the inertia in these cases, e.g. with Euclidean and Manhattan distances. Inertia is based on squared index, and keyword squared is added to the name of distance, unless data were square root transformed (argument sqrt.dist=TRUE). If an additive constant was used with argument add, Lingoes or Cailliez adjusted is added to the the name of inertia, and the value of the constant is printed.

**Author(s)**

Jari Oksanen

**References**


cascadeKM

**See Also**

`rda`, `cca`, `plot.cca`, `anova.cca`, `vegdist`, `dist`, `cmdscale`, `wcmdscale` for underlying and related functions. Function `sppscores` can add species scores or replace existing species scores.

The function returns similar result object as `rda` (see `cca.object`). This section for `rda` gives a more complete list of functions that can be used to access and analyse dbRDA results.

**Examples**

```r
data(varespec)
data(varechem)
## Basic Analysis
cap <- capscale(varespec ~ N + P + K + Condition(A1), varechem,
    dist="bray")
cap
plot(cap)
anova(cap)
## Avoid negative eigenvalues with additive constant
capscale(varespec ~ N + P + K + Condition(A1), varechem,
    dist="bray", add = TRUE)
## Avoid negative eigenvalues by taking square roots of dissimilarities
capscale(varespec ~ N + P + K + Condition(A1), varechem,
    dist = "bray", sqrt.dist = TRUE)
## Principal coordinates analysis with extended dissimilarities
capscale(varespec ~ 1, dist="bray", metaMDS = TRUE)
## dblda
dbrda(varespec ~ N + P + K + Condition(A1), varechem,
    dist="bray")
## avoid negative eigenvalues also with Jaccard distances
dbrda(varespec ~ N + P + K + Condition(A1), varechem,
    dist = "jaccard")
```

---

**cascadeKM**  
*K*-means partitioning using a range of values of *K*

**Description**

This function is a wrapper for the `kmeans` function. It creates several partitions forming a cascade from a small to a large number of groups.

**Usage**

```r
cascadeKM(data, inf.gr, sup.gr, iter = 100, criterion = "calinski")
cIndexKM(y, x, index = "all")
```

## S3 method for class 'cascadeKM'

```r
plot(x, min.g, max.g, grpmts.plot = TRUE,
    sortg = FALSE, gridcol = NA, ...)
```
Arguments

- **data**: The data matrix. The objects (samples) are the rows.
- **inf.gr**: The number of groups for the partition with the smallest number of groups of the cascade (min).
- **sup.gr**: The number of groups for the partition with the largest number of groups of the cascade (max).
- **iter**: The number of random starting configurations for each value of \( K \).
- **criterion**: The criterion that will be used to select the best partition. The default value is "calinski", which refers to the Calinski-Harabasz (1974) criterion. The simple structure index ("ssi") is also available. Other indices are available in function `clustIndex` (package `cclust`). In our experience, the two indices that work best and are most likely to return their maximum value at or near the optimal number of clusters are "calinski" and "ssi".
- **y**: Object of class "kmeans" returned by a clustering algorithm such as `kmeans`
- **x**: Data matrix where columns correspond to variables and rows to observations, or the plotting object in `plot`
- **index**: The available indices are: "calinski" and "ssi". Type "all" to obtain both indices. Abbreviations of these names are also accepted.
- **min.g, max.g**: The minimum and maximum numbers of groups to be displayed.
- **grpmts.plot**: Show the plot (TRUE or FALSE).
- **sortg**: Sort the objects as a function of their group membership to produce a more easily interpretable graph. See Details. The original object names are kept; they are used as labels in the output table \( x \), although not in the graph. If there were no row names, sequential row numbers are used to keep track of the original order of the objects.
- **gridcol**: The colour of the grid lines in the plots. NA, which is the default value, removes the grid lines.
- **...**: Other parameters to the functions (ignored).

Details

The function creates several partitions forming a cascade from a small to a large number of groups formed by `kmeans`. Most of the work is performed by function `cIndex` which is based on the `clustIndex` function (package `cclust`). Some of the criteria were removed from this version because computation errors were generated when only one object was found in a group.

The default value is "calinski", which refers to the well-known Calinski-Harabasz (1974) criterion. The other available index is the simple structure index "ssi" (Dolnicar et al. 1999). In the case of groups of equal sizes, "calinski" is generally a good criterion to indicate the correct number of groups. Users should not take its indications literally when the groups are not equal in size. Type "all" to obtain both indices. The indices are defined as:

**calinski**: \[ \frac{SSB/(K - 1)}{SSW/(n - K)} \], where \( n \) is the number of data points and \( K \) is the number of clusters. \( SSW \) is the sum of squares within the clusters while \( SSB \) is the sum of squares among the clusters. This index is simply an \( F \) (ANOVA) statistic.
ssi: the “Simple Structure Index” multiplicatively combines several elements which influence the interpretability of a partitioning solution. The best partition is indicated by the highest SSI value.

In a simulation study, Milligan and Cooper (1985) found that the Calinski-Harabasz criterion recovered the correct number of groups the most often. We recommend this criterion because, if the groups are of equal sizes, the maximum value of "calinski" usually indicates the correct number of groups. Another available index is the simple structure index "ssi". Users should not take the indications of these indices literally when the groups are not equal in size and explore the groups corresponding to other values of $K$.

Function cascadeKM has a plot method. Two plots are produced. The graph on the left has the objects in abscissa and the number of groups in ordinate. The groups are represented by colours. The graph on the right shows the values of the criterion ("calinski" or "ssi") for determining the best partition. The highest value of the criterion is marked in red. Points marked in orange, if any, indicate partitions producing an increase in the criterion value as the number of groups increases; they may represent other interesting partitions.

If sortg=TRUE, the objects are reordered by the following procedure: (1) a simple matching distance matrix is computed among the objects, based on the table of K-means assignments to groups, from $K = \min g$ to $K = \max g$. (2) A principal coordinate analysis (PCoA, Gower 1966) is computed on the centred distance matrix. (3) The first principal coordinate is used as the new order of the objects in the graph. A simplified algorithm is used to compute the first principal coordinate only, using the iterative algorithm described in Legendre & Legendre (2012). The full distance matrix among objects is never computed; this avoids the problem of storing it when the number of objects is large. Distance values are computed as they are needed by the algorithm.

Value

Function cascadeKM returns an object of class cascadeKM with items:

- **partition**: Table with the partitions found for different numbers of groups $K$, from $K = \inf gr$ to $K = \sup gr$.
- **results**: Values of the criterion to select the best partition.
- **criterion**: The name of the criterion used.
- **size**: The number of objects found in each group, for all partitions (columns).

Function cIndex returns a vector with the index values. The maximum value of these indices is supposed to indicate the best partition. These indices work best with groups of equal sizes. When the groups are not of equal sizes, one should not put too much faith in the maximum of these indices, and also explore the groups corresponding to other values of $K$.

Author(s)

Marie-Helene Ouellette <Marie-Helene.Ouellette@umontreal.ca>, Sebastien Durand <Sebastien.Durand@umontreal.ca> and Pierre Legendre <Pierre.Legendre@umontreal.ca>. Edited for vegan by Jari Oksanen.

References


See Also

`kmeans`, `clustIndex`.

Examples

```r
# Partitioning a (10 x 10) data matrix of random numbers
mat <- matrix(runif(100),10,10)
res <- cascadeKM(mat, 2, 5, iter = 25, criterion = 'calinski')
toto <- plot(res)

# Partitioning an autocorrelated time series
vec <- sort(matrix(runif(30),30,1))
res <- cascadeKM(vec, 2, 5, iter = 25, criterion = 'calinski')
toto <- plot(res)

# Partitioning a large autocorrelated time series
# Note that we remove the grid lines
vec <- sort(matrix(runif(1000),1000,1))
res <- cascadeKM(vec, 2, 7, iter = 10, criterion = 'calinski')
toto <- plot(res, gridcol=NA)
```

---

**Description**

Function `cca` performs correspondence analysis, or optionally constrained correspondence analysis (a.k.a. canonical correspondence analysis), or optionally partial constrained correspondence analysis. Function `rda` performs redundancy analysis, or optionally principal components analysis. These are all very popular ordination techniques in community ecology.
Usage

## S3 method for class 'formula'
cca(formula, data, na.action = na.fail, subset = NULL, ...

## S3 method for class 'formula'
rda(formula, data, scale=FALSE, na.action = na.fail, subset = NULL, ...)

## Default S3 method:
cca(X, Y, Z, ...)

## Default S3 method:
rda(X, Y, Z, scale=FALSE, ...)

Arguments

- **formula**
  - Model formula, where the left hand side gives the community data matrix, right hand side gives the constraining variables, and conditioning variables can be given within a special function `Condition`.

- **data**
  - Data frame containing the variables on the right hand side of the model formula.

- **X**
  - Community data matrix.

- **Y**
  - Constraining matrix, typically of environmental variables. Can be missing. If this is a `data.frame`, it will be expanded to a `model.matrix` where factors are expanded to contrasts ("dummy variables"). It is better to use `formula` instead of this argument, and some further analyses only work when `formula` was used.

- **Z**
  - Conditioning matrix, the effect of which is removed ("partialled out") before next step. Can be missing. If this is a `data.frame`, it is expanded similarly as constraining matrix.

- **scale**
  - Scale species to unit variance (like correlations).

- **na.action**
  - Handling of missing values in constraints or conditions. The default (`na.fail`) is to stop with missing value. Choice `na.omit` removes all rows with missing values. Choice `na.exclude` keeps all observations but gives NA for results that cannot be calculated. The WA scores of rows may be found also for missing values in constraints. Missing values are never allowed in dependent community data.

- **subset**
  - Subset of data rows. This can be a logical vector which is TRUE for kept observations, or a logical expression which can contain variables in the working environment, `data` or species names of the community data.

- **...**
  - Other arguments for `print` or `plot` functions (ignored in other functions).

Details

Since their introduction (ter Braak 1986), constrained, or canonical, correspondence analysis and its spin-off, redundancy analysis, have been the most popular ordination methods in community ecology. Functions `cca` and `rda` are similar to popular proprietary software Canoco, although the implementation is completely different. The functions are based on Legendre & Legendre’s (2012) algorithm: in `cca` Chi-square transformed data matrix is subjected to weighted linear regression on constraining variables, and the fitted values are submitted to correspondence analysis performed via
singular value decomposition (svd). Function rda is similar, but uses ordinary, unweighted linear regression and unweighted SVD. Legendre & Legendre (2012), Table 11.5 (p. 650) give a skeleton of the RDA algorithm of vegan. The algorithm of CCA is similar, but involves standardization by row and column weights.

The functions can be called either with matrix-like entries for community data and constraints, or with formula interface. In general, the formula interface is preferred, because it allows a better control of the model and allows factor constraints. Some analyses of ordination results are only possible if model was fitted with formula (e.g., most cases of anova.cca, automatic model building).

In the following sections, X, Y and Z, although referred to as matrices, are more commonly data frames.

In the matrix interface, the community data matrix X must be given, but the other data matrices may be omitted, and the corresponding stage of analysis is skipped. If matrix Z is supplied, its effects are removed from the community matrix, and the residual matrix is submitted to the next stage. This is called partial correspondence or redundancy analysis. If matrix Y is supplied, it is used to constrain the ordination, resulting in constrained or canonical correspondence analysis, or redundancy analysis. Finally, the residual is submitted to ordinary correspondence analysis (or principal components analysis). If both matrices Z and Y are missing, the data matrix is analysed by ordinary correspondence analysis (or principal components analysis).

Instead of separate matrices, the model can be defined using a model formula. The left hand side must be the community data matrix (X). The right hand side defines the constraining model. The constraints can contain ordered or unordered factors, interactions among variables and functions of variables. The defined contrasts are honoured in factor variables. The constraints can also be matrices (but not data frames). The formula can include a special term Condition for conditioning variables (“covariables”) partialled out before analysis. So the following commands are equivalent: cca(X, Y, Z), cca(X ~ Y + Condition(Z)), where Y and Z refer to constraints and conditions matrices respectively.

Constrained correspondence analysis is indeed a constrained method: CCA does not try to display all variation in the data, but only the part that can be explained by the used constraints. Consequently, the results are strongly dependent on the set of constraints and their transformations or interactions among the constraints. The shotgun method is to use all environmental variables as constraints. However, such exploratory problems are better analysed with unconstrained methods such as correspondence analysis (decorana, corresp) or non-metric multidimensional scaling (metaMDS) and environmental interpretation after analysis (envfit, ordisurf). CCA is a good choice if the user has clear and strong a priori hypotheses on constraints and is not interested in the major structure in the data set.

CCA is able to correct the curve artefact commonly found in correspondence analysis by forcing the configuration into linear constraints. However, the curve artefact can be avoided only with a low number of constraints that do not have a curvilinear relation with each other. The curve can reappear even with two badly chosen constraints or a single factor. Although the formula interface makes it easy to include polynomial or interaction terms, such terms often produce curved artefacts (that are difficult to interpret), these should probably be avoided.

According to folklore, rda should be used with “short gradients” rather than cca. However, this is not based on research which finds methods based on Euclidean metric as uniformly weaker than those based on Chi-squared metric. However, standardized Euclidean distance may be an appropriate measures (see Hellinger standardization in decostand in particular).
Partial CCA (pCCA; or alternatively partial RDA) can be used to remove the effect of some conditioning or background or random variables or covariates before CCA proper. In fact, pCCA compares models $\text{cca}(X \sim Z)$ and $\text{cca}(X \sim Y + Z)$ and attributes their difference to the effect of $Y$ cleansed of the effect of $Z$. Some people have used the method for extracting “components of variance” in CCA. However, if the effect of variables together is stronger than sum of both separately, this can increase total Chi-square after partialling out some variation, and give negative “components of variance”. In general, such components of “variance” are not to be trusted due to interactions between two sets of variables.

The functions have summary and plot methods which are documented separately (see plot.cca, summary.cca).

**Value**

Function cca returns a huge object of class cca, which is described separately in cca.object. Function rda returns an object of class rda which inherits from class cca and is described in cca.object. The scaling used in rda scores is described in a separate vignette with this package.

**Author(s)**

The responsible author was Jari Oksanen, but the code borrows heavily from Dave Roberts (Montana State University, USA).

**References**

The original method was by ter Braak, but the current implementation follows Legendre and Legendre.


**See Also**

This help page describes two constrained ordination functions, cca and rda. A related method, distance-based redundancy analysis (dBRA) is described separately (capscale). All these functions return similar objects (described in cca.object). There are numerous support functions that can be used to access the result object. In the list below, functions of type cca will handle all three constrained ordination objects, and functions of rda only handle rda and capsacle results.

The main plotting functions are plot.cca for all methods, and biplot.rda for RDA and dBRA. However, generic vegan plotting functions can also handle the results. The scores can be accessed and scaled with scores.cca, and summarized with summary.cca. The eigenvalues can be accessed with eigenvals.cca and the regression coefficients for constraints with coef.cca. The eigenvalues can be plotted with screeplot.cca, and the (adjusted) $R^2$ can be found with RsquareAdj.rda.
The scores can be also calculated for new data sets with `predict.cca` which allows adding points to ordinations. The values of constraints can be inferred from ordination and community composition with `calibrate.cca`.

Diagnostic statistics can be found with `goodness.cca`, `inertcomp`, `spenvcor`, `intersetcor`, `tolerance.cca`, and `vif.cca`. Function `as.mlm.cca` refits the result object as a multiple `lm` object, and this allows finding influence statistics (`lm.influence`, `cooks.distance` etc.).

Permutation based significance for the overall model, single constraining variables or axes can be found with `anova.cca`. Automatic model building with `Rstep` function is possible with `deviance.cca`, `add1.cca` and `drop1.cca`. Functions `ordistep` and `ordiR2step` (for RDA) are special functions for constrained ordination. Randomized data sets can be generated with `simulate.cca`.

Separate methods based on constrained ordination model are principal response curves (`prc`) and variance partitioning between several components (`varpart`).

Design decisions are explained in `vignette` on “Design decisions” which can be accessed with `browseVignettes("vegan")`.

Package `ade4` provides alternative constrained ordination function `pcaiv`.

### Examples

```r
data(varespec)
data(varechem)
## Common but bad way: use all variables you happen to have in your
## environmental data matrix
vare.cca <- cca(varespec, varechem)
vare.cca
plot(vare.cca)
## Formula interface and a better model
vare.cca <- cca(varespec ~ Al + P*(K + Baesoil), data=varechem)
vare.cca
plot(vare.cca)
## Partialling out and negative components of variance
cca(varespec ~ Ca, varechem)
cca(varespec ~ Ca + Condition(pH), varechem)
## RDA
data(dune)
data(dune.env)
dune.Manure <- rda(dune - Manure, dune.env)
plot(dune.Manure)
```

---

### Description

Ordination methods `cca`, `rda`, `dbrda` and `capscale` return similar result objects. All these methods use the same internal function `ordConstrained`. They differ only in (1) initial transformation of the data and in defining inertia, (2) weighting, and (3) the use of rectangular rows × columns data or symmetric rows × rows dissimilarities: `rda` initializes data to give variance or correlations
as inertia, \texttt{cca} is based on double-standardized data to give Chi-square inertia and uses row and column weights, \texttt{capscale} maps the real part of dissimilarities to rectangular data and performs RDA, and \texttt{dbrda} performs an RDA-like analysis directly on symmetric dissimilarities.

Function \texttt{ordConstrained} returns the same result components for all these methods, and the calling function may add some more components to the final result. However, you should not access these result components directly (using \$): the internal structure is not regarded as stable application interface (API), but it can change at any release. If you access the results components directly, you take a risk of breakage at any \texttt{vegan} release. The \texttt{vegan} provides a wide set of accessor functions to those components, and these functions are updated when the result object changes. This documentation gives an overview of accessor functions to the \texttt{cca} result object.

**Usage**

```r
ordiYbar(x, model = c("CCA", "CA", "pCCA", "partial", "initial"))
## S3 method for class 'cca'
model.frame(formula, ...)
## S3 method for class 'cca'
model.matrix(object, ...)
## S3 method for class 'cca'
weights(object, display = "sites", ...)
```

**Arguments**

- \texttt{object, x, formula}
  A result object from \texttt{cca, rda, dbrda, or capscale}.
- \texttt{model}
  Show constrained ("CCA"), unconstrained ("CA") or conditioned "partial" ("pCCA") results. In \texttt{ordiYbar} the value can also be "initial" for the internal working input data, and "partial" for the internal working input data after removing the partial effects.
- \texttt{display}
  Display either "sites" or "species".
- \texttt{...}
  Other arguments passed to the the function.

**Details**

The internal ("working") form of the dependent (community) data can be accessed with function \texttt{ordiYbar}. The form depends on the ordination method: for instance, in \texttt{cca} the data are weighted and Chi-square transformed, and in \texttt{dbrda} they are Gower-centred dissimilarities. The input data in the original ("response") form can be accessed with \texttt{fitted.cca} and \texttt{residuals.cca}. Function \texttt{predict.cca} can return either working or response data, and also their lower-rank approximations.

The model matrix of independent data ("Constraints" and "Conditions") can be extracted with \texttt{model.matrix}. In partial analysis, the function returns a list of design matrices called \texttt{Conditions} and \texttt{Constraints}. If either component was missing, a single matrix is returned. The redundant (aliased) terms do not appear in the model matrix. These terms can be found with \texttt{alias.cca}. Function \texttt{model.frame} tries to reconstruct the data frame from which the model matrices were derived. This is only possible if the original model was fitted with \texttt{formula} and \texttt{data} arguments, and still fails if the data are unavailable.
The number of observations can be accessed with `nobs.cca`, and the residual degrees of freedom with `df.residual.cca`. The information on observations with missing values can be accessed with `na.action`. The terms and formula of the fitted model can be accessed with `formula` and `terms`.

The weights used in `cca` can be accessed with `weights`. In unweighted methods (`rda`) all weights are equal.

The ordination results are saved in separate components for partial terms, constraints and residual unconstrained ordination. There is no guarantee that these components will have the same internal names as currently, and you should be cautious when developing scripts and functions that directly access these components.

The constrained ordination algorithm is based on QR decomposition of constraints and conditions (environmental data), and the QR component is saved separately for partial and constrained components. The QR decomposition of constraints can be accessed with `qr.cca`. This will also include the residual effects of partial terms (Conditions), and it should be used together with `ordiYbar(x, "partial")`. The environmental data are first centred in `rda` or weighted and centred in `cca`. The QR decomposition is used in many functions that access `cca` results, and it can be used to find many items that are not directly stored in the object. For examples, see `coef.cca`, `coef.rda`, `vif.cca`, `permute.cca`, `predict.cca`, `predict.rda`, `calibrate.cca`. See `qr` for other possible uses of this component. For instance, the rank of the constraints can be found from the QR decomposition.

The eigenvalues of the solution can be accessed with `eigenvals.cca`. Eigenvalues are not evaluated for partial component, and they will only be available for constrained and residual components.

The ordination scores are internally stored as (weighted) orthonormal scores matrices. These results can be accessed with `scores.cca` and `scores.rda` functions. The ordination scores are scaled when accessed with `scores` functions, but internal (weighted) orthonormal scores can be accessed by setting `scaling = FALSE`. Unconstrained residual component has species and site scores, and constrained component has also fitted site scores or linear combination scores for sites and biplot scores and centroids for constraint variables. The biplot scores correspond to the `model.matrix`, and centroids are calculated for factor variables when they were used. The scores can be selected by defining the axes, and there is no direct way of accessing all scores of a certain component. The number of dimensions can be assessed from `eigenvals`. In addition, some other types can be derived from the results although not saved in the results. For instance, regression scores and model coefficients can be accessed with `scores` and `coef` functions. Partial component will have no scores.

Distance-based methods (`dbrda`, `capscale`) can have negative eigenvalues and associated imaginary axis scores. There is no way of accessing these imaginary scores. In addition, species scores are initially missing in `dbrda` and they are accessory and found after analysis in `capscale` (and may be misleading). Function `sppscores` can be used to add species scores or replace them with more meaningful ones.

**Note**

Saving of “working” dependent (community) data changed in `vegan` version 2.5-0, and you should use `ordiYbar` function instead of direct access, or your scripts and functions will fail (ordiYbar has been available since `vegan` version 2.4-3, and it works both with the old and current result objects).

The `model.matrix` returns the unweighted model matrix also for `cca`. Prior to `vegan` version 2.5-0 it returned the weighted model matrix.
Canonical Correlation Analysis

Description

Canonical correlation analysis, following Brian McArdle's unpublished graduate course notes, plus improvements to allow the calculations in the case of very sparse and collinear matrices, and permutation test of Pillai's trace statistic.

Usage

CCorA(Y, X, stand.Y=FALSE, stand.X=FALSE, permutations = 0, ...)

## S3 method for class 'CCorA'
biplot(x, plot.type="ov", xlabs, plot.axes = 1:2, int=0.5,
   col.Y="red", col.X="blue", cex=c(0.7,0.9), ...)

Arguments

Y Left matrix (object class: matrix or data.frame).
X Right matrix (object class: matrix or data.frame).
stand.Y Logical; should Y be standardized?
stand.X Logical; should X be standardized?
permutations a list of control values for the permutations as returned by the function how. or the number of permutations required, or a permutation matrix where each row gives the permuted indices.
CCorA

x  CCoaR result object.

plot.type   A character string indicating which of the following plots should be produced: "objects", "variables", "ov" (separate graphs for objects and variables), or "biplots". Any unambiguous subset containing the first letters of these names can be used instead of the full names.

xlabs  Row labels. The default is to use row names, NULL uses row numbers instead, and NA suppresses plotting row names completely.

plot.axes  A vector with 2 values containing the order numbers of the canonical axes to be plotted. Default: first two axes.

int  Radius of the inner circles plotted as visual references in the plots of the variables. Default: int=0.5. With int=0, no inner circle is plotted.

col.Y  Color used for objects and variables in the first data table (Y) plots. In biplots, the objects are in black.

col.X  Color used for objects and variables in the second data table (X) plots.

cex  A vector with 2 values containing the size reduction factors for the object and variable names, respectively, in the plots. Default values: cex=c(0.7, 0.9).

...  Other arguments passed to these functions. The function biplot.CCorA passes graphical arguments to biplot and biplot.default. CCorA currently ignores extra arguments.

Details

Canonical correlation analysis (Hotelling 1936) seeks linear combinations of the variables of y that are maximally correlated to linear combinations of the variables of x. The analysis estimates the relationships and displays them in graphs. Pillai's trace statistic is computed and tested parametrically (F-test); a permutation test is also available.

Algorithmic note – The blunt approach would be to read the two matrices, compute the covariance matrices, then the matrix SQR EJE invHsRRI EJE tHsQQI. Its trace is Pillai’s trace statistic. This approach may fail, however, when there is heavy multicollinearity in very sparse data matrices. The safe approach is to replace all data matrices by their PCA object scores.

The function can produce different types of plots depending on the option chosen: "objects" produces two plots of the objects, one in the space of Y, the second in the space of X; "variables" produces two plots of the variables, one of the variables of Y in the space of Y, the second of the variables of X in the space of X; "ov" produces four plots, two of the objects and two of the variables; "biplots" produces two biplots, one for the first matrix (Y) and one for second matrix (X) solutions. For biplots, the function passes all arguments to biplot.default; consult its help page for configuring biplots.

Value

Function CCorA returns a list containing the following elements:

Pillai  Pillai's trace statistic = sum of the canonical eigenvalues.

Eigenvalues  Canonical eigenvalues. They are the squares of the canonical correlations.

CanCorr  Canonical correlations.
Mat. ranks  Ranks of matrices Y and X.
RDA.Rsquares  Bimultivariate redundancy coefficients (R-squares) of RDAs of Y|X and X|Y.
RDA.adj.Rsq  RDA.Rsquares adjusted for n and the number of explanatory variables.
nperm  Number of permutations.
p.Pillai  Parametric probability value associated with Pillai’s trace.
p.perm  Permutational probability associated with Pillai’s trace.
Cy  Object scores in Y biplot.
Cx  Object scores in X biplot.
corr.Y.Cy  Scores of Y variables in Y biplot, computed as cor(Y,Cy).
corr.X.Cx  Scores of X variables in X biplot, computed as cor(X,Cx).
corr.Y.Cx  cor(Y,Cy) available for plotting variables Y in space of X manually.
corr.X.Cy  cor(X,Cx) available for plotting variables X in space of Y manually.
control  A list of control values for the permutations as returned by the function how.
call  Call to the CCorA function.

Author(s)

Pierre Legendre, Departement de Sciences Biologiques, Universite de Montreal. Implemented in vegan with the help of Jari Oksanen.

References


Examples

# Example using two mite groups. The mite data are available in vegan
data(mite)
# Two mite species associations (Legendre 2005, Fig. 4)
group.1 <- c(1,2,4,8,10:15,17,19:22,24,26:30)
group.2 <- c(3,9,16,18,23,25,31:35)
# Separate Hellinger transformations of the two groups of species
mite.hel.1 <- decostand(mite[,group.1], "hel")
mite.hel.2 <- decostand(mite[,group.2], "hel")
rownames(mite.hel.1) = paste("S",1:nrow(mite),sep="")
rownames(mite.hel.2) = paste("S",1:nrow(mite),sep="")
out <- CCorA(mite.hel.1, mite.hel.2)
out
biplot(out, "ob")  # Two plots of objects
biplot(out, "v", cex=c(0.7,0.6))  # Two plots of variables
biplot(out, "ov", cex=c(0.7,0.6))  # Four plots (2 for objects, 2 for variables)
biplot(out, "b", cex=c(0.7,0.6))  # Two biplots
biplot(out, xlab = NA, plot.axes = c(3,5))  # Plot axes 3, 5. No object names
biplot(out, plot.type="biplots", xlab = NULL)  # Replace object names by numbers
# Example using random numbers. No significant relationship is expected

```r
mat1 <- matrix(rnorm(60), 20, 3)
mat2 <- matrix(rnorm(100), 20, 5)
out2 = CCorA(mat1, mat2, permutations=99)
out2
biplot(out2, "b")
```

## clamtest

### Multinomial Species Classification Method (CLAM)

#### Description

The CLAM statistical approach for classifying generalists and specialists in two distinct habitats is described in Chazdon et al. (2011).

#### Usage

```r
clamtest(comm, groups, coverage.limit = 10, specialization = 2/3,
  npoints = 20, alpha = .05/20)
## S3 method for class 'clamtest'
summary(object, ...)
## S3 method for class 'clamtest'
plot(x, xlab, ylab, main,  pch = 21:24, col.points = 1:4,
   col.lines = 2:4, lty = 1:3, position = "bottomright", ...)
```

#### Arguments

- **comm**: Community matrix, consisting of counts.
- **groups**: A vector identifying the two habitats. Must have exactly two unique values or levels. Habitat IDs in the grouping vector must match corresponding rows in the community matrix `comm`.
- **coverage.limit**: Integer, the sample coverage based correction is applied to rare species with counts below this limit. Sample coverage is calculated separately for the two habitats. Sample relative abundances are used for species with higher than or equal to `coverage.limit` total counts per habitat.
- **specialization**: Numeric, specialization threshold value between 0 and 1. The value of 2/3 represents ‘supermajority’ rule, while a value of 1/2 represents a ‘simple majority’ rule to assign shared species as habitat specialists.
- **npoints**: Integer, number of points used to determine the boundary lines in the plots.
- **alpha**: Numeric, nominal significance level for individual tests. The default value reduces the conventional limit of 0.05 to account for overdispersion and multiple testing for several species simultaneously. However, the is no firm reason for exactly this limit.
- **x, object**: Fitted model object of class "clamtest".
- **xlab, ylab**: Labels for the plot axes.
main

Main title of the plot.

pch, col.points

Symbols and colors used in plotting species groups.

lty, col.lines

Line types and colors for boundary lines in plot to separate species groups.

position

Position of figure legend, see legend for specification details. Legend not shown if position = NULL.

... Additional arguments passed to methods.

Details

The method uses a multinomial model based on estimated species relative abundance in two habitats (A, B). It minimizes bias due to differences in sampling intensities between two habitat types as well as bias due to insufficient sampling within each habitat. The method permits a robust statistical classification of habitat specialists and generalists, without excluding rare species a priori (Chazdon et al. 2011). Based on a user-defined specialization threshold, the model classifies species into one of four groups: (1) generalists; (2) habitat A specialists; (3) habitat B specialists; and (4) too rare to classify with confidence.

Value

A data frame (with class attribute "clamtest"), with columns:

- Species: species name (column names from comm),
- Total_{A}: total count in habitat A,
- Total_{B}: total count in habitat B,
- Classes: species classification, a factor with levels Generalist, Specialist_{A}, Specialist_{B}, and Too_rare.

_{A} and _{B} are placeholders for habitat names/labels found in the data.

The summary method returns descriptive statistics of the results. The plot method returns values invisibly and produces a bivariate scatterplot of species total abundances in the two habitats. Symbols and boundary lines are shown for species groups.

Note

The code was tested against standalone CLAM software provided on the website of Anne Chao (http://chao.stat.nthu.edu.tw/wordpress/); minor inconsistencies were found, especially for finding the threshold for 'too rare' species. These inconsistencies are probably due to numerical differences between the two implementation. The current R implementation uses root finding for iso-lines instead of iterative search.

The original method (Chazdon et al. 2011) has two major problems:

1. It assumes that the error distribution is multinomial. This is a justified choice if individuals are freely distributed, and there is no over-dispersion or clustering of individuals. In most ecological data, the variance is much higher than multinomial assumption, and therefore test statistic are too optimistic.
2. The original authors suggest that multiple testing adjustment for multiple testing should be based on the number of points (npoints) used to draw the critical lines on the plot, whereas the adjustment should be based on the number of tests (i.e., tested species). The function uses the same numerical values as the original paper, but there is no automatic connection between npoints and alpha arguments, but you must work out the adjustment yourself.

Author(s)

Peter Solymos <solymos@ualberta.ca>

References


Examples

data(mite)
data(mite.env)
sol <- with(mite.env, clamtest(mite, Shrub="None", alpha=0.005))
summary(sol)
head(sol)
plot(sol)

commsim

Create an Object for Null Model Algorithms

Description

The commsim function can be used to feed Null Model algorithms into nullmodel analysis. The make.commsim function returns various predefined algorithm types (see Details). These functions represent low level interface for community null model infrastructure in vegan with the intent of extensibility, and less emphasis on direct use by users.

Usage

commsim(method, fun, binary, isSeq, mode)
make.commsim(method)
## S3 method for class 'commsim'
print(x, ...)

Arguments

method Character, name of the algorithm.
fun A function. For possible formal arguments of this function see Details.
binary Logical, if the algorithm applies to presence-absence or count matrices.
isSeq Logical, if the algorithm is sequential (needs burnin and thinning) or not.
mode Character, storage mode of the community matrix, either "integer" or "double".
x An object of class commsim.
... Additional arguments.

Details

The function fun must return an array of dim(nr, nc, n), and must take some of the following arguments:

• x: input matrix,
• n: number of permuted matrices in output,
• nr: number of rows,
• nc: number of columns,
• rs: vector of row sums,
• cs: vector of column sums,
• rf: vector of row frequencies (non-zero cells),
• cf: vector of column frequencies (non-zero cells),
• s: total sum of x,
• fill: matrix fill (non-zero cells),
• thin: thinning value for sequential algorithms,
• ...: additional arguments.

You can define your own null model, but several null model algorithm are pre-defined and can be called by their name. The predefined algorithms are described in detail in the following chapters. The binary null models produce matrices of zeros (absences) and ones (presences) also when input matrix is quantitative. There are two types of quantitative data: Counts are integers with a natural unit so that individuals can be shuffled, but abundances can have real (floating point) values and do not have a natural subunit for shuffling. All quantitative models can handle counts, but only some are able to handle real values. Some of the null models are sequential so that the next matrix is derived from the current one. This makes models dependent from previous models, and usually you must thin these matrices and study the sequences for stability: see oecosimu for details and instructions.

See Examples for structural constraints imposed by each algorithm and defining your own null model.

Value

An object of class commsim with elements corresponding to the arguments (method, binary, isSeq, mode, fun).

If the input of make.commsim is a commsim object, it is returned without further evaluation. If this is not the case, the character method argument is matched against predefined algorithm names. An error message is issued if none such is found. If the method argument is missing, the function returns names of all currently available null model algorithms as a character vector.
Binary null models

All binary null models preserve fill: number of presences or conversely the number of absences. The classic models may also preserve column (species) frequencies ($c\theta$) or row frequencies or species richness of each site ($r\theta$) and take into account commonness and rarity of species ($r_1$, $r_2$). Algorithms swap, tswap, curveball, quasiswap and backtracking preserve both row and column frequencies. Three first ones are sequential but the two latter are non-sequential and produce independent matrices. Basic algorithms are reviewed by Wright et al. (1998).

- "r0": non-sequential algorithm for binary matrices that only preserves the number of presences (fill).
- "r\theta", "r\theta_old": non-sequential algorithm for binary matrices that preserves the site (row) frequencies. Methods "r\theta" and "r\theta_old" implement the same method, but use different random number sequences; use "r\theta_old" if you want to reproduce results in vegan 2.0-0 or older using commsimulator (now deprecated).
- "r1": non-sequential algorithm for binary matrices that preserves the site (row) frequencies, but uses column marginal frequencies as probabilities of selecting species.
- "r2": non-sequential algorithm for binary matrices that preserves the site (row) frequencies, and uses squared column marginal frequencies as as probabilities of selecting species.
- "c\theta": non-sequential algorithm for binary matrices that preserves species frequencies (Jonsson 2001).
- "swap": sequential algorithm for binary matrices that changes the matrix structure, but does not influence marginal sums (Gotelli & Entsminger 2003). This inspects $2 \times 2$ submatrices so long that a swap can be done.
- "tswap": sequential algorithm for binary matrices. Same as the "swap" algorithm, but it tries a fixed number of times and performs zero to many swaps at one step (according to the thin argument in the call). This approach was suggested by Miklós & Podani (2004) because they found that ordinary swap may lead to biased sequences, since some columns or rows are more easily swapped.
- "curveball": sequential method for binary matrices that implements the ‘Curveball’ algorithm of Strona et al. (2014). The algorithm selects two random rows and finds the set of unique species that occur only in one of these rows. The algorithm distributes the set of unique species to rows preserving the original row frequencies. Zero to several species are swapped in one step, and usually the matrix is perturbed more strongly than in other sequential methods.
- "quasiswap": non-sequential algorithm for binary matrices that implements a method where matrix is first filled honouring row and column totals, but with integers that may be larger than one. Then the method inspects random $2 \times 2$ matrices and performs a quasiswap on them. In addition to ordinary swaps, quasiswap can reduce numbers above one to ones preserving marginal totals (Miklós & Podani 2004). The method is non-sequential, but it accepts thin argument: the convergence is checked at every thin steps. This allows performing several ordinary swaps in addition to fill changing swaps which helps in reducing or removing the bias.
- "greedyqswap": A greedy variant of quasiswap. In greedy step, one element of the $2 \times 2$ matrix is taken from > 1 elements. The greedy steps are biased, but the method can be thinned, and only the first of thin steps is greedy. Even modest thinning (say thin = 20) removes or reduces the bias, and thin = 100 (1% greedy steps) looks completely safe and
still speeds up simulation. The code is experimental and it is provided here for further scrutiny, and should be tested for bias before use.

• "backtracking": non-sequential algorithm for binary matrices that implements a filling method with constraints both for row and column frequencies (Gotelli & Entsminger 2001). The matrix is first filled randomly, but typically row and column sums are reached before all incidences are filled in. After this begins "backtracking", where some of the incidences are removed, and filling is started again, and this backtracking is done so many times that all incidences will be filled into matrix. The results may be biased and should be inspected carefully before use.

Quantitative Models for Counts with Fixed Marginal Sums

These models shuffle individuals of counts and keep marginal sums fixed, but marginal frequencies are not preserved. Algorithm r2dtable uses standard R function r2dtable also used for simulated P-values in chisq.test. Algorithm quasiswap_count uses the same, but preserves the original fill. Typically this means increasing numbers of zero cells and the result is zero-inflated with respect to r2dtable.

• "r2dtable": non-sequential algorithm for count matrices. This algorithm keeps matrix sum and row/column sums constant. Based on r2dtable.

• "quasiswap_count": non-sequential algorithm for count matrices. This algorithm is similar as Carsten Dormann’s swap_web function in the package bipartite. First, a random matrix is generated by the r2dtable function preserving row and column sums. Then the original matrix fill is reconstructed by sequential steps to increase or decrease matrix fill in the random matrix. These steps are based on swapping $2 \times 2$ submatrices (see "swap_count" algorithm for details) to maintain row and column totals.

Quantitative Swap Models

Quantitative swap models are similar to binary swap, but they swap the largest permissible value. The models in this section all maintain the fill and perform a quantitative swap only if this can be done without changing the fill. Single step of swap often changes the matrix very little. In particular, if cell counts are variable, high values change very slowly. Checking the chain stability and independence is even more crucial than in binary swap, and very strong thinning is often needed. These models should never be used without inspecting their properties for the current data. These null models can also be defined using permatswap function.

• "swap_count": sequential algorithm for count matrices. This algorithm find $2 \times 2$ submatrices that can be swapped leaving column and row totals and fill unchanged. The algorithm finds the largest value in the submatrix that can be swapped ($d$). Swap means that the values in diagonal or antidiagonal positions are decreased by $d$, while remaining cells are increased by $d$. A swap is made only if fill does not change.

• "abuswap_r": sequential algorithm for count or nonnegative real valued matrices with fixed row frequencies (see also permatswap). The algorithm is similar to swap_count, but uses different swap value for each row of the $2 \times 2$ submatrix. Each step changes the the corresponding column sums, but honours matrix fill, row sums, and row/column frequencies (Hardy 2008; randomization scheme 2x).
• "abuswap_c": sequential algorithm for count or nonnegative real valued matrices with fixed column frequencies (see also permatswap). The algorithm is similar as the previous one, but operates on columns. Each step changes the the corresponding row sums, but honours matrix fill, column sums, and row/column frequencies (Hardy 2008; randomization scheme 3x).

Quantitative Swap and Shuffle Models

Quantitative Swap and Shuffle methods (swsh methods) preserve fill and column and row frequencies, and also either row or column sums. The methods first perform a binary quasiswap and then shuffle original quantitative data to non-zero cells. The samp methods shuffle original non-zero cell values and can be used also with non-integer data. The both methods redistribute individuals randomly among non-zero cells and can only be used with integer data. The shuffling is either free over the whole matrix, or within rows (r methods) or within columns (c methods). Shuffling within a row preserves row sums, and shuffling within a column preserves column sums. These models can also be defined with permatswap.

• "swsh_samp": non-sequential algorithm for quantitative data (either integer counts or non-integer values). Original non-zero values values are shuffled.
• "swsh_both": non-sequential algorithm for count data. Individuals are shuffled freely over non-zero cells.
• "swsh_samp_r": non-sequential algorithm for quantitative data. Non-zero values (samples) are shuffled separately for each row.
• "swsh_samp_c": non-sequential algorithm for quantitative data. Non-zero values (samples) are shuffled separately for each column.
• "swsh_both_r": non-sequential algorithm for count matrices. Individuals are shuffled freely for non-zero values within each row.
• "swsh_both_c": non-sequential algorithm for count matrices. Individuals are shuffled freely for non-zero values with each column.

Quantitative Shuffle Methods

Quantitative shuffle methods are generalizations of binary models r00, r0 and c0. The _ind methods shuffle individuals so that the grand sum, row sum or column sums are preserved. These methods are similar as r2dtable but with still slacker constraints on marginal sums. The _samp and _both methods first apply the corresponding binary model with similar restriction on marginal frequencies and then distribute quantitative values over non-zero cells. The _samp models shuffle original cell values and can therefore handle also non-count real values. The _both models shuffle individuals among non-zero values. The shuffling is over the whole matrix in r00_, and within row in r0_ and within column in c0_ in all cases.

• "r00_ind": non-sequential algorithm for count matrices. This algorithm preserves grand sum and individuals are shuffled among cells of the matrix.
• "r0_ind": non-sequential algorithm for count matrices. This algorithm preserves row sums and individuals are shuffled among cells of each row of the matrix.
• "c0_ind": non-sequential algorithm for count matrices. This algorithm preserves column sums and individuals are shuffled among cells of each column of the matrix.
• "r00_samp": non-sequential algorithm for count or nonnegative real valued (mode = "double") matrices. This algorithm preserves grand sum and cells of the matrix are shuffled.
• "r0_samp": non-sequential algorithm for count or nonnegative real valued (mode = "double") matrices. This algorithm preserves row sums and cells within each row are shuffled.
• "c0_samp": non-sequential algorithm for count or nonnegative real valued (mode = "double") matrices. This algorithm preserves column sums constant and cells within each column are shuffled.
• "r00_both": non-sequential algorithm for count matrices. This algorithm preserves grand sum and cells and individuals among cells of the matrix are shuffled.
• "r0_both": non-sequential algorithm for count matrices. This algorithm preserves grand sum and cells and individuals among cells of each row are shuffled.
• "c0_both": non-sequential algorithm for count matrices. This algorithm preserves grand sum and cells and individuals among cells of each column are shuffled.

Author(s)
Jari Oksanen and Peter Solymos

References

See Also
See `permatfull`, `permatswap` for alternative specification of quantitative null models. Function `oecosimu` gives a higher-level interface for applying null models in hypothesis testing and analysis of models. Function `nullmodel` and `simulate.nullmodel` are used to generate arrays of simulated null model matrices.
Examples

```r
## write the r00 algorithm
f <- function(x, n, ...) {
  array(replicate(n, sample(x)), c(dim(x), n))
}
(cs <- commsim("r00", fun=f, binary=TRUE, isSeq=FALSE, mode="integer"))

## retrieving the sequential swap algorithm
(cs <- make.commsim("swap"))

## feeding a commsim object as argument
make.commsim(cs)

## making the missing c1 model using r1 as a template
## non-sequential algorithm for binary matrices
## that preserves the species (column) frequencies, but uses row marginal frequencies
## as probabilities of selecting sites
f <- function (x, n, nr, nc, rs, cs, ...) {
  out <- array(0L, c(nr, nc, n))
  J <- seq_len(nc)
  storage.mode(rs) <- "double"
  for (k in seq_len(n))
    for (j in J)
      out[sample.int(nr, cs[j], prob = rs), j, k] <- 1L
  out
}
cs <- make.commsim("r1")
cs$method <- "c1"
cs$fun <- f

## structural constraints
diagfun <- function(x, y) {
  c(sum = sum(y) == sum(x),
    fill = sum(y > 0) == sum(x > 0),
    rowSums = all(rowSums(y) == rowSums(x)),
    colSums = all(colSums(y) == colSums(x)),
    rowFreq = all(rowSums(y > 0) == rowSums(x > 0)),
    colFreq = all(colSums(y > 0) == colSums(x > 0)))
}
evalfun <- function(meth, x, n) {
  m <- nullmodel(x, meth)
  y <- simulate(m, nsim=n)
  out <- rowMeans(sapply(1:dim(y)[3],
    function(i) diagfun(attr(y, "data"), y[, , i])))
  z <- as.numeric(c(attr(y, "binary"), attr(y, "isSeq"),
    attr(y, "mode") == "double"))
  names(z) <- c("binary", "isSeq", "double")
  c(z, out)
}
x <- matrix(rbinom(10*12, 1, 0.5)*rpois(10*12, 3), 12, 10)
algos <- make.commsim()
```
The contribution diversity approach is based on the differentiation of within-unit and among-unit diversity by using additive diversity partitioning and unit distinctiveness.

**Usage**

```r
contribdiv(comm, index = c("richness", "simpson"),
            relative = FALSE, scaled = TRUE, drop.zero = FALSE)
```

## S3 method for class 'contribdiv'

- `plot(x, sub, xlab, ylab, ylim, col, ...)`

**Arguments**

- `comm` The community data matrix with samples as rows and species as column.
- `index` Character, the diversity index to be calculated.
- `relative` Logical, if TRUE then contribution diversity values are expressed as their signed deviation from their mean. See details.
- `scaled` Logical, if TRUE then relative contribution diversity values are scaled by the sum of gamma values (if `index = "richness"`) or by sum of gamma values times the number of rows in `comm` (if `index = "simpson"`). See details.
- `drop.zero` Logical, should empty rows dropped from the result? If empty rows are not dropped, their corresponding results will be NAs.
- `x` An object of class "contribdiv".
- `sub, xlab, ylab, ylim, col` Graphical arguments passed to `plot`.
- `...` Other arguments passed to `plot`.

**Details**

This approach was proposed by Lu et al. (2007). Additive diversity partitioning (see [[`adipart`]] for more references) deals with the relation of mean alpha and the total (gamma) diversity. Although alpha diversity values often vary considerably. Thus, contributions of the sites to the total diversity are uneven. This site specific contribution is measured by contribution diversity components. A unit that has e.g. many unique species will contribute more to the higher level (gamma) diversity than another unit with the same number of species, but all of which common.

Distinctiveness of species $j$ can be defined as the number of sites where it occurs ($n_j$), or the sum of its relative frequencies ($p_j$). Relative frequencies are computed sitewise and $\sum_i p_{ij}$ at site $i$ sum up to 1.
The contribution of site $i$ to the total diversity is given by $\alpha_i = \sum_j (1/n_{ij})$ when dealing with richness and $\alpha_i = \sum (p_{ij} \times (1 - p_{ij}))$ for the Simpson index.

The unit distinctiveness of site $i$ is the average of the species distinctiveness, averaging only those species which occur at site $i$. For species richness: $\alpha_i = \text{mean}(n_i)$ (in the paper, the second equation contains a typo, $n$ is without index). For the Simpson index: $\alpha_i = \text{mean}(n_i)$.

The Lu et al. (2007) gives an in-depth description of the different indices.

**Value**

An object of class "contribdiv" inheriting from data frame.

Returned values are alpha, beta and gamma components for each sites (rows) of the community matrix. The "diff.coef" attribute gives the differentiation coefficient (see Examples).

**Author(s)**

Péter Sólymos, <solymos@ualberta.ca>

**References**


**See Also**

*adipart, diversity*

**Examples**

```r
## Artificial example given in
## Table 2 in Lu et al. 2007
x <- matrix(c(
  1/3, 1/3, 1/3, 0, 0, 0,
  0, 0, 1/3, 1/3, 1/3, 0,
  0, 0, 0, 1/3, 1/3, 1/3),
  3, 6, byrow = TRUE,
  dimnames = list(LETTERS[1:3], letters[1:6]))

x

## Compare results with Table 2
contribdiv(x, "richness")
contribdiv(x, "simpson")

## Relative contribution (C values), compare with Table 2
(cd1 <- contribdiv(x, "richness", relative = TRUE, scaled = FALSE))
(cd2 <- contribdiv(x, "simpson", relative = TRUE, scaled = FALSE))

## Differentiation coefficients
attr(cd1, "diff.coef") # D_ST
attr(cd2, "diff.coef") # D_DT

## BCI data set
data(BCI)

opar <- par(mfrow=c(2,2))
plot(contribdiv(BCI, "richness"), main = "Absolute")
plot(contribdiv(BCI, "richness", relative = TRUE), main = "Relative")
```
Decorana: Detrended Correspondence Analysis and Basic Reciprocal Averaging

Description

Performs detrended correspondence analysis and basic reciprocal averaging or orthogonal correspondence analysis.

Usage

```r
decorana(veg, iweigh=0, iresc=4, ira=0, mk=26, short=0,
    before=NULL, after=NULL)
```

## S3 method for class 'decorana'
```r
plot(x, choices=c(1,2), origin=TRUE,
    display=c("both","sites","species","none"),
    cex = 0.8, cols = c(1,2), type, xlim, ylim, ...)
```

## S3 method for class 'decorana'
```r
text(x, display = c("sites", "species"), labels,
    choices = 1:2, origin = TRUE, select, ...)
```

## S3 method for class 'decorana'
```r
points(x, display = c("sites", "species"),
    choices=1:2, origin = TRUE, select, ...)
```

## S3 method for class 'decorana'
```r
summary(object, digits=3, origin=TRUE,
    display=c("both", "species","sites","none"), ...)
```

## S3 method for class 'summary.decorana'
```r
print(x, head = NA, tail = head, ...)
```

## S3 method for class 'decorana'
```r
downweight(veg, fraction = 5)
```

## S3 method for class 'decorana'
```r
scores(x, display=c("sites","species"), choices=1:4,
    origin=TRUE, ...)
```

Arguments

- **veg**: Community data, a matrix-like object.
- **iweigh**: Downweighting of rare species (0: no).
decorana 67

iresc Number of rescaling cycles (0: no rescaling).
ira Type of analysis (0: detrended, 1: basic reciprocal averaging).
mk Number of segments in rescaling.
short Shortest gradient to be rescaled.
before Hill’s piecewise transformation: values before transformation.
after Hill’s piecewise transformation: values after transformation – these must correspond to values in before.
x, object A decorana result object.
choices Axes shown.
origin Use true origin even in detrended correspondence analysis.
display Display only sites, only species, both or neither.
cex Plot character size.
cols Colours used for sites and species.
type Type of plots, partial match to "text", "points" or "none".
labels Optional text to be used instead of row names.
select Items to be displayed. This can either be a logical vector which is TRUE for displayed items or a vector of indices of displayed items.
xlim, ylim the x and y limits (min,max) of the plot.
digits Number of digits in summary output.
head, tail Number of rows printed from the head and tail of species and site scores. Default NA prints all.
fraction Abundance fraction where downweighting begins.
... Other arguments for plot function.

Details

In late 1970s, correspondence analysis became the method of choice for ordination in vegetation science, since it seemed better able to cope with non-linear species responses than principal components analysis. However, even correspondence analysis can produce an arc-shaped configuration of a single gradient. Mark Hill developed detrended correspondence analysis to correct two assumed ‘faults’ in correspondence analysis: curvature of straight gradients and packing of sites at the ends of the gradient.

The curvature is removed by replacing the orthogonalization of axes with detrending. In orthogonalization successive axes are made non-correlated, but detrending should remove all systematic dependence between axes. Detrending is performed using a five-segment smoothing window with weights (1,2,3,2,1) on mk segments — which indeed is more robust than the suggested alternative of detrending by polynomials. The packing of sites at the ends of the gradient is undone by rescaling the axes after extraction. After rescaling, the axis is supposed to be scaled by ‘SD’ units, so that the average width of Gaussian species responses is supposed to be one over whole axis. Other innovations were the piecewise linear transformation of species abundances and downweighting of rare species which were regarded to have an unduly high influence on ordination axes.

It seems that detrending actually works by twisting the ordination space, so that the results look non-curved in two-dimensional projections (‘lolly paper effect’). As a result, the points usually
have an easily recognized triangular or diamond shaped pattern, obviously an artefact of detrending. Rescaling works differently than commonly presented, too. decorana does not use, or even evaluate, the widths of species responses. Instead, it tries to equalize the weighted variance of species scores on axis segments (parameter \(mk\) has only a small effect, since decorana finds the segment number from the current estimate of axis length). This equalizes response widths only for the idealized species packing model, where all species initially have unit width responses and equally spaced modes.

The summary method prints the ordination scores, possible prior weights used in downweighting, and the marginal totals after applying these weights. The plot method plots species and site scores. Classical decorana scaled the axes so that smallest site score was 0 (and smallest species score was negative), but summary, plot and scores use the true origin, unless origin = FALSE.

In addition to proper eigenvalues, the function also reports ‘decorana values’ in detrended analysis. These ‘decorana values’ are the values that the legacy code of decorana returns as eigenvalues. They are estimated internally during iteration, and it seems that detrending interferes the estimation so that these values are generally too low and have unclear interpretation. Moreover, ‘decorana values’ are estimated before rescaling which will change the eigenvalues. The proper eigenvalues are estimated after extraction of the axes and they are the ratio of biased weighted variances of site and species scores even in detrended and rescaled solutions. The ‘decorana values’ are provided only for the compatibility with legacy software, and they should not be used.

**Value**

`decorana` returns an object of class "decorana", which has print, summary and plot methods.

**Note**

`decorana` uses the central numerical engine of the original Fortran code (which is in the public domain), or about 1/3 of the original program. I have tried to implement the original behaviour, although a great part of preparatory steps were written in R language, and may differ somewhat from the original code. However, well-known bugs are corrected and strict criteria used (Oksanen & Minchin 1997).

Please note that there really is no need for piecewise transformation or even downweighting within `decorana`, since there are more powerful and extensive alternatives in R, but these options are included for compliance with the original software. If a different fraction of abundance is needed in downweighting, function `downweight` must be applied before `decorana`. Function `downweight` indeed can be applied prior to correspondence analysis, and so it can be used together with `cca`, too.

The function finds only four axes: this is not easily changed.

**Author(s)**

Mark O. Hill wrote the original Fortran code, the R port was by Jari Oksanen.

**References**


See Also

For unconstrained ordination, non-metric multidimensional scaling in `monoMDS` may be more robust (see also `metaMDS`). Constrained (or ‘canonical’) correspondence analysis can be made with `cca`. Orthogonal correspondence analysis can be made with `corresp`, or with `decorana` or `cca`, but the scaling of results vary (and the one in `decorana` corresponds to `scaling = "sites"` and `hill = TRUE` in `cca`). See `predict.decorana` for adding new points to an ordination.

Examples

data(varespec)
vare.dca <- decorana(varespec)
vare.dca
summary(vare.dca)
plot(vare.dca)

### the detrending rationale:
gaussresp <- function(x,u) exp(-(x-u)^2/2)
x <- seq(0,6,length=15) # The gradient
u <- seq(-2,8,len=23) # The optima
pack <- outer(x,u,gaussresp)
matplot(x, pack, type="l", main="Species packing")
opar <- par(mfrow=c(2,2))
plot(scores(prcomp(pack)), asp=1, type="b", main="PCA")
plot(scores(decorana(pack, ira=1)), asp=1, type="b", main="CA")
plot(scores(decorana(pack)), asp=1, type="b", main="DCA")
plot(scores(cca(pack ~ x), dis="sites"), asp=1, type="b", main="CCA")

### Let's add some noise:
noisy <- (0.5 + runif(length(pack))) * pack
par(mfrow=c(2,1))
matplot(x, pack, type="l", main="Ideal model")
matplot(x, noisy, type="l", main="Noisy model")
par(mfrow=c(2,2))
plot(scores(prcomp(noisy)), type="b", main="PCA", asp=1)
plot(scores(decorana(noisy, ira=1)), type="b", main="CA", asp=1)
plot(scores(decorana(noisy)), type="b", main="DCA", asp=1)
plot(scores(cca(noisy ~ x), dis="sites"), asp=1, type="b", main="CCA")
par(opar)

decostand

Description

The function provides some popular (and effective) standardization methods for community ecologists.
Usage

decostand(x, method, MARGIN, range.global, logbase = 2, na.rm = FALSE, ...)
  wisconsin(x)

Arguments

  x            Community data, a matrix-like object.
  method       Standardization method. See Details for available options.
  MARGIN       Margin, if default is not acceptable. 1 = rows, and 2 = columns of x.
  range.global Matrix from which the range is found in method = "range". This allows using
                 same ranges across subsets of data. The dimensions of MARGIN must match with
                 x.
  logbase      The logarithm base used in method = "log".
  na.rm        Ignore missing values in row or column standardizations.
  ...          Other arguments to the function (ignored).

Details

The function offers following standardization methods for community data:

  • total: divide by margin total (default MARGIN = 1).
  • max: divide by margin maximum (default MARGIN = 2).
  • freq: divide by margin maximum and multiply by the number of non-zero items, so that the
           average of non-zero entries is one (Oksanen 1983; default MARGIN = 2).
  • normalize: make margin sum of squares equal to one (default MARGIN = 1).
  • range: standardize values into range 0 . . . 1 (default MARGIN = 2). If all values are constant,
            they will be transformed to 0.
  • rank, rrank: rank replaces abundance values by their increasing ranks leaving zeros un-
                changed, and rrank is similar but uses relative ranks with maximum 1 (default MARGIN = 1).
    Average ranks are used for tied values.
  • standardize: scale x to zero mean and unit variance (default MARGIN = 2).
  • pa: scale x to presence/absence scale (0/1).
  • chi.square: divide by row sums and square root of column sums, and adjust for square
                root of matrix total (Legendre & Gallagher 2001). When used with the Euclidean distance,
                the distances should be similar to the Chi-square distance used in correspondence analysis.
                However, the results from cmdscale would still differ, since CA is a weighted ordination
                method (default MARGIN = 1).
  • hellinger: square root of method = "total" (Legendre & Gallagher 2001).
  • log: logarithmic transformation as suggested by Anderson et al. (2006): \( \log_b(x) + 1 \) for
           \( x > 0 \), where \( b \) is the base of the logarithm; zeros are left as zeros. Higher bases give less
           weight to quantities and more to presences, and logbase = Inf gives the presence/absence
           scaling. Please note this is not \( \log(x + 1) \). Anderson et al. (2006) suggested this for their
           (strongly) modified Gower distance (implemented as method = "altGower" in vegdist),
           but the standardization can be used independently of distance indices.
Standardization, as contrasted to transformation, means that the entries are transformed relative to other entries.

All methods have a default margin. MARGIN=1 means rows (sites in a normal data set) and MARGIN=2 means columns (species in a normal data set).

Command Wisconsin is a shortcut to common Wisconsin double standardization where species (MARGIN=2) are first standardized by maxima (max) and then sites (MARGIN=1) by site totals (tot).

Most standardization methods will give nonsense results with negative data entries that normally should not occur in the community data. If there are empty sites or species (or constant with method = "range"), many standardization will change these into NaN.

Value

Returns the standardized data frame, and adds an attribute "decostand" giving the name of applied standardization "method".

Note

Common transformations can be made with standard R functions.

Author(s)

Jari Oksanen, Etienne Laliberté (method = "log").

References


Examples

data(varespec)
sptrans <- decostand(varespec, "max")
apply(sptrans, 2, max)
sptrans <- wisconsin(varespec)

### Chi-square: PCA similar but not identical to CA.
### Use wcmdscale for weighted analysis and identical results.
sptrans <- decostand(varespec, "chi.square")
plot(procrustes(rda(sptrans)), cca(varespec)))
**Description**

Function `designdist` lets you define your own dissimilarities using terms for shared and total quantities, number of rows and number of columns. The shared and total quantities can be binary, quadratic or minimum terms. In binary terms, the shared component is number of shared species, and totals are numbers of species on sites. The quadratic terms are cross-products and sums of squares, and minimum terms are sums of parallel minima and row totals. Function `chaodist` lets you define your own dissimilarities using terms that are supposed to take into account the “unseen species” (see Chao et al., 2005 and Details in `vegdist`).

**Usage**

```r
designdist(x, method = "(A+B-2*J)/(A+B)",
           terms = c("binary", "quadratic", "minimum"),
           abcd = FALSE, alphagamma = FALSE, name)
chaodist(x, method = "1 - 2*U*V/(U+V)", name)
```

**Arguments**

- **x**  
  Input data.

- **method**  
  Equation for your dissimilarities. This can use terms $J$ for shared quantity, $A$ and $B$ for totals, $N$ for the number of rows (sites) and $P$ for the number of columns (species) or in `chaodist` it can use terms $U$ and $V$. The equation can also contain any $R$ functions that accepts vector arguments and returns vectors of the same length.

- **terms**  
  How shared and total components are found. For vectors $x$ and $y$ the "quadratic" terms are $J = \text{sum}(x*y)$, $A = \text{sum}(x^2)$, $B = \text{sum}(y^2)$, and "minimum" terms are $J = \text{sum}(\text{min}(x,y))$, $A = \text{sum}(x)$ and $B = \text{sum}(y)$, and "binary" terms are either of these after transforming data into binary form (shared number of species, and number of species for each row).

- **abcd**  
  Use 2x2 contingency table notation for binary data: $a$ is the number of shared species, $b$ and $c$ are the numbers of species occurring only one of the sites but not in both, and $d$ is the number of species that occur on neither of the sites.

- **alphagamma**  
  Use beta diversity notation with terms $alpha$ for average alpha diversity for compared sites, $gamma$ for diversity in pooled sites, and $delta$ for the absolute value of difference of average alpha and alpha diversities of compared sites. Terms $A$ and $B$ refer to alpha diversities of compared sites.

- **name**  
  The name you want to use for your index. The default is to combine the method equation and `terms` argument.
Details

Most popular dissimilarity measures in ecology can be expressed with the help of terms \( j \), \( a \) and \( b \), and some also involve matrix dimensions \( N \) and \( P \). Some examples you can define in `designdist` are:

- \( A + B - 2J \) \quad "quadratic" \quad squared Euclidean
- \( A + B - 2J \) \quad "minimum" \quad Manhattan
- \( (A + B - 2J)/(A + B) \) \quad "binary" \quad Bray-Curtis
- \( (A + B - 2J)/(A + B - J) \) \quad "binary" \quad Jaccard
- \( (A + B - 2J)/(A + B - J) \) \quad "minimum" \quad Ružička
- \( 1/J/sqrt(A*B) \) \quad "binary" \quad Ochiai
- \( 1/J/sqrt(A*B) \) \quad "quadratic" \quad cosine complement
- \( 1-\text{phyper}(J-1, A, P-A, B) \) \quad "binary" \quad Raup-Crick (but see `raupcrick`)

The function `designdist` can implement most dissimilarity indices in `vegdist` or elsewhere, and it can also be used to implement many other indices, amongst them, most of those described in Legendre & Legendre (2012). It can also be used to implement all indices of beta diversity described in Koleff et al. (2003), but there also is a specific function `betadiver` for the purpose.

If you want to implement binary dissimilarities based on the 2x2 contingency table notation, you can set `abcd = TRUE`. In this notation \( a = J, b = A-J, c = B-J, d = P-A-B+J \). This notation is often used instead of the more tangible default notation for reasons that are opaque to me.

With `alphagamma = TRUE` it is possible to use beta diversity notation with terms `alpha` for average alpha diversity and `gamma` for gamma diversity in two compared sites. The terms are calculated as `alpha = (A+B)/2`, `gamma = A+B-J` and `delta = abs(A-B)/2`. Terms `A` and `B` are also available and give the alpha diversities of the individual compared sites. The beta diversity terms may make sense only for binary terms (so that diversities are expressed in numbers of species), but they are calculated for quadratic and minimum terms as well (with a warning).

Function `chaodist` is similar to `designdist`, but uses terms `u` and `v` of Chao et al. (2005). These terms are supposed to take into account the effects of unseen species. Both `u` and `v` are scaled to range 0...1. They take the place of `A` and `B` and the product `U*V` is used in the place of `J` of `designdist`. Function `chaodist` can implement any commonly used Chao et al. (2005) style dissimilarity:

- \( 1 - 2*U*V/(U+V) \) \quad Sørensen type
- \( 1 - U*V/(U+V-U*V) \) \quad Jaccard type
- \( 1 - sqrt(U*V) \) \quad Ochiai type
- \( (pmin(U,V) - U*V)/pmin(U,V) \) \quad Simpson type

Function `vegdist` implements Jaccard-type Chao distance, and its documentation contains more complete discussion on the calculation of the terms.

Value

`designdist` returns an object of class `dist`. 
**Note**

designdist does not use compiled code, but it is based on vectorized R code. The designdist function can be much faster than vegdist, although the latter uses compiled code. However, designdist cannot skip missing values and uses much more memory during calculations.

The use of sum terms can be numerically unstable. In particularly, when these terms are large, the precision may be lost. The risk is large when the number of columns is high, and particularly large with quadratic terms. For precise calculations it is better to use functions like dist and vegdist which are more robust against numerical problems.

**Author(s)**

Jari Oksanen

**References**


**See Also**

vegdist, betadiver, dist, raupcrick.

**Examples**

data(BCI)
## Four ways of calculating the same Sørensen dissimilarity
d0 <- vegdist(BCI, "bray", binary = TRUE)
d1 <- designdist(BCI, "(A+B-2*I)/(A+B)"")
d2 <- designdist(BCI, "(b+c)/(2*a+b+c)", abcd = TRUE)
d3 <- designdist(BCI, "gamma/alpha - 1", alphagamma = TRUE)
## Arrhenius dissimilarity: the value of z in the species-area model
## S = c*A^z when combining two sites of equal areas, where S is the
## number of species, A is the area, and c and z are model parameters.
## The A below is not the area (which cancels out), but number of
## species in one of the sites, as defined in designdist().
dis <- designdist(BCI, "(log(A+B-1)-log(A+B)+log(2))/log(2)"")
## This can be used in clustering or ordination...
ordiplot(cmdscale(dis))
## ... or in analysing beta diversity (without gradients)
summary(dis)
The functions extract statistics that resemble deviance and AIC from the result of constrained correspondence analysis `cca` or redundancy analysis `rda`. These functions are rarely needed directly, but they are called by `step` in automatic model building. Actually, `cca` and `rda` do not have AIC and these functions are certainly wrong.

### Usage

```r
## S3 method for class 'cca'
deviance(object, ...)

## S3 method for class 'cca'
extractAIC(fit, scale = 0, k = 2, ...)
```

### Arguments

- `object`: the result of a constrained ordination (`cca` or `rda`).
- `fit`: fitted model from constrained ordination.
- `scale`: optional numeric specifying the scale parameter of the model, see `scale` in `step`.
- `k`: numeric specifying the "weight" of the equivalent degrees of freedom (=:edf) part in the AIC formula.
- `...`: further arguments.

### Details

The functions find statistics that resemble deviance and AIC in constrained ordination. Actually, constrained ordination methods do not have a log-Likelihood, which means that they cannot have AIC and deviance. Therefore you should not use these functions, and if you use them, you should not trust them. If you use these functions, it remains as your responsibility to check the adequacy of the result.

The deviance of `cca` is equal to the Chi-square of the residual data matrix after fitting the constraints. The deviance of `rda` is defined as the residual sum of squares. The deviance function of `rda` is also used for `capscale`. Function `extractAIC` mimics `extractAIC.lm` in translating deviance to AIC.

There is little need to call these functions directly. However, they are called implicitly in `step` function used in automatic selection of constraining variables. You should check the resulting model with some other criteria, because the statistics used here are unfounded. In particular, the penalty `k` is not properly defined, and the default `k = 2` is not justified theoretically. If you have only continuous covariates, the `step` function will base the model building on magnitude of eigenvalues, and the value of `k` only influences the stopping point (but the variables with the highest eigenvalues...
are not necessarily the most significant in permutation tests in \texttt{anova.cca}). If you also have multiclass factors, the value of \( k \) will have a capricious effect in model building. The \texttt{step} function will pass arguments to \texttt{add1.cca} and \texttt{drop1.cca}, and setting \texttt{test = "permutation"} will provide permutation tests of each deletion and addition which can help in judging the validity of the model building.

**Value**

The deviance functions return “deviance”, and \texttt{extractAIC} returns effective degrees of freedom and “AIC”.

**Note**

These functions are unfounded and untested and they should not be used directly or implicitly. Moreover, usual caveats in using \texttt{step} are very valid.

**Author(s)**

Jari Oksanen

**References**


**See Also**

\texttt{cca, rda, anova.cca, step, extractAIC, add1.cca, drop1.cca}.

**Examples**

```r
# The deviance of correspondence analysis equals Chi-square
data(dune)
data(dune.env)
chisq.test(dune)
deviance(cca(dune))

# Stepwise selection (forward from an empty model "dune ~ 1")
ord <- cca(dune ~ ., dune.env)
step(cca(dune ~ 1, dune.env), scope = formula(ord))
```

---

\texttt{dispindmorisita} \hspace{1cm} \textit{Morisita index of intraspecific aggregation}

**Description**

Calculates the Morisita index of dispersion, standardized index values, and the so called clumpedness and uniform indices.
Usage

dispindmorisita(x, unique.rm = FALSE, crit = 0.05, na.rm = FALSE)

Arguments

x community data matrix, with sites (samples) as rows and species as columns.
unique.rm logical, if TRUE, unique species (occurring in only one sample) are removed from the result.
crit two-sided p-value used to calculate critical Chi-squared values.
a.rm logical. Should missing values (including NaN) be omitted from the calculations?

Details

The Morisita index of dispersion is defined as (Morisita 1959, 1962):

$$Imor = n \times \frac{(\sum(x_i^2) - \sum(x_i))}{(\sum(x_i)^2 - \sum(x_i))}$$

where $x_i$ is the count of individuals in sample $i$, and $n$ is the number of samples ($i = 1, 2, \ldots, n$). $Imor$ has values from 0 to $n$. In uniform (hyperdispersed) patterns its value falls between 0 and 1, in clumped patterns it falls between 1 and $n$. For increasing sample sizes (i.e. joining neighbouring quadrats), $Imor$ goes to $n$ as the quadrat size approaches clump size. For random patterns, $Imor = 1$ and counts in the samples follow Poisson frequency distribution.

The deviation from random expectation (null hypothesis) can be tested using critical values of the Chi-squared distribution with $n - 1$ degrees of freedom. Confidence intervals around 1 can be calculated by the clumped $Mclu$ and uniform $Muni$ indices (Hairson et al. 1971, Krebs 1999) (Chi2Lower and Chi2Upper refers to e.g. 0.025 and 0.975 quantile values of the Chi-squared distribution with $n - 1$ degrees of freedom, respectively, for crit = 0.05):

$$Mclu = (\text{Chi2Lower} - n + \sum(x_i)) / (\sum(x_i) - 1)$$
$$Muni = (\text{Chi2Upper} - n + \sum(x_i)) / (\sum(x_i) - 1)$$

Smith-Gill (1975) proposed scaling of Morisita index from [0, n] interval into [-1, 1], and setting up -0.5 and 0.5 values as confidence limits around random distribution with rescaled value 0. To rescale the Morisita index, one of the following four equations apply to calculate the standardized index $Imst$:

(a) $Imor \geq Mclu > 1: Imst = 0.5 + 0.5 (Imor - Mclu) / (n - Mclu)$,
(b) $Mclu > Imor \geq 1: Imst = 0.5 (Imor - 1) / (Mclu - 1)$,
(c) $1 > Imor > Muni: Imst = -0.5 (Imor - 1) / (Muni - 1)$,
(d) $1 > Muni > Imor: Imst = -0.5 + 0.5 (Imor - Muni) / Muni$.

Value

Returns a data frame with as many rows as the number of columns in the input data, and with four columns. Columns are: $imor$ the unstandardized Morisita index, $mclu$ the clumpedness index, $muni$ the uniform index, $imst$ the standardized Morisita index, $pchisq$ the Chi-squared based probability for the null hypothesis of random expectation.
Note
A common error found in several papers is that when standardizing as in the case (b), the denominator is given as $\mu_n - 1$. This results in a hiatus in the $[0, 0.5]$ interval of the standardized index. The root of this typo is the book of Krebs (1999), see the Errata for the book (Page 217, http://www.zoology.ubc.ca/~krebs/downloads/errs2nd_printing.pdf).

Author(s)
Péter Sólymos, <solymos@ualberta.ca>

References

Examples

```r
data(dune)
x <- dispindmorisita(dune)
x
y <- dispindmorisita(dune, unique.rm = TRUE)
y
dim(x) ## with unique species
dim(y) ## unique species removed
```

dispweight

*Dispersion-based weighting of species counts*

Description
Transform abundance data downweighting species that are overdispersed to the Poisson error.

Usage

dispweight(comm, groups, nsimul = 999, nullmodel = "c0_ind", plimit = 0.05)
gdispweight(formula, data, plimit = 0.05)
## S3 method for class 'dispweight'
summary(object, ...)
dispweight

Arguments

- **comm**: Community data matrix.
- **groups**: Factor describing the group structure. If missing, all sites are regarded as belonging to one group. NA values are not allowed.
- **nsimul**: Number of simulations.
- **nullmodel**: The nullmodel used in commsim within groups. The default follows Clarke et al. (2006).
- **plimit**: Downweight species if their p-value is at or below this limit.
- **formula, data**: Formula where the left-hand side is the community data frame and right-hand side gives the explanatory variables. The explanatory variables are found in the data frame given in data or in the parent frame.
- **object**: Result object from dispweight or gdispweight.
- **...**: Other parameters passed to functions.

Details

The dispersion index \( (D) \) is calculated as ratio between variance and expected value for each species. If the species abundances follow Poisson distribution, expected dispersion is \( E(D) = 1 \), and if \( D > 1 \), the species is overdispersed. The inverse \( 1/D \) can be used to downweight species abundances. Species are only downweighted when overdispersion is judged to be statistically significant (Clarke et al. 2006).

Function dispweight implements the original procedure of Clarke et al. (2006). Only one factor can be used to group the sites and to find the species means. The significance of overdispersion is assessed freely distributing individuals of each species within factor levels. This is achieved by using nullmodel "c0_ind" (which accords to Clarke et al. 2006), but other nullmodels can be used, though they may not be meaningful (see commsim for alternatives). If a species is absent in some factor level, the whole level is ignored in calculation of overdispersion, and the number of degrees of freedom can vary among species. The reduced number of degrees of freedom is used as a divisor for overdispersion \( D \), and such species have higher dispersion and hence lower weights in transformation.

Function gdispweight is a generalized parametric version of dispweight. The function is based on glm with quasipoisson error family. Any glm model can be used, including several factors or continuous covariates. Function gdispweight uses the same test statistic as dispweight (Pearson Chi-square), but it does not ignore factor levels where species is absent, and the number of degrees of freedom is equal for all species. Therefore transformation weights can be higher than in dispweight. The gdispweight function evaluates the significance of overdispersion parametrically from Chi-square distribution (pchisq).

Functions dispweight and gdispweight transform data, but they add information on overdispersion and weights as attributes of the result. The summary can be used to extract and print that information.

Value

Function returns transformed data with the following new attributes:

- **D**: Dispersion statistic.
distconnected

**df** Degrees of freedom for each species.

**p** $\ p$-value of the Dispersion statistic $D$.

**weights** weights applied to community data.

**nsimul** Number of simulations used to assess the $p$-value, or NA when simulations were not performed.

**nullmodel** The name of `commsim` null model, or NA when simulations were not performed.

**Author(s)**

Eduard Szöcs <eduardszoesc@gmail.com> wrote the original dispweight, Jari Oksanen significantly modified the code, provided support functions and developed gdispweight.

**References**


**Examples**

```r
data(mite, mite.env)
## dispweight and its summary
mite.dw <- with(mite.env, dispweight(mite, Shrub, nsimul = 99))
summary(mite.dw)
## generalized dispersion weighting
mite.dw <- gdispweight(mite ~ Shrub + WatrCont, data = mite.env)
rda(mite.dw ~ Shrub + WatrCont, data = mite.env)
```

---

distconnected

**Connectedness of Dissimilarities**

**Description**

Function `distconnected` finds groups that are connected disregarding dissimilarities that are at or above a threshold or NA. The function can be used to find groups that can be ordinated together or transformed by `stepacross`. Function `no.shared` returns a logical dissimilarity object, where TRUE means that sites have no species in common. This is a minimal structure for `distconnected` or can be used to set missing values to dissimilarities.

**Usage**

```r
distconnected(dis, too long = 1, trace = TRUE)
no.shared(x)
```
Arguments

- **dis**: Dissimilarity data inheriting from class `dist` or an object, such as a matrix, that can be converted to a dissimilarity matrix. Functions `vegdist` and `dist` are some functions producing suitable dissimilarity data.
- **toolong**: Shortest dissimilarity regarded as NA. The function uses a fuzz factor, so that dissimilarities close to the limit will be made NA, too. If `toolong = 0` (or negative), no dissimilarity is regarded as too long.
- **trace**: Summarize results of `distconnected`.
- **x**: Community data.

Details

Data sets are disconnected if they have sample plots or groups of sample plots which share no species with other sites or groups of sites. Such data sets cannot be sensibly ordinated by any unconstrained method because these subsets cannot be related to each other. For instance, correspondence analysis will polarize these subsets with eigenvalue 1. Neither can such dissimilarities be transformed with `stepacross`, because there is no path between all points, and result will contain NAs. Function `distconnected` will find such subsets in dissimilarity matrices. The function will return a grouping vector that can be used for sub-setting the data. If data are connected, the result vector will be all 1s. The connectedness between two points can be defined either by a threshold `toolong` or using input dissimilarities with NAs.

Function `noNshared` returns a `dist` structure having value TRUE when two sites have nothing in common, and value FALSE when they have at least one shared species. This is a minimal structure that can be analysed with `distconnected`. The function can be used to select dissimilarities with no shared species in indices which do not have a fixed upper limit.

Function `distconnected` uses depth-first search (Sedgewick 1990).

Value

Function `distconnected` returns a vector for observations using integers to identify connected groups. If the data are connected, values will be all 1. Function `noNshared` returns an object of class `dist`.

Author(s)

Jari Oksanen

References


See Also

`vegdist` or `dist` for getting dissimilarities, `stepacross` for a case where you may need `distconnected`, and for connecting points `spantree`. 
Examples

```r
## There are no disconnected data in vegan, and the following uses an
## extremely low threshold limit for connectedness. This is for
## illustration only, and not a recommended practice.
data(dune)
dis <- vegdist(dune)
gr <- distconnected(dis, toolong=0.4)
# Make sites with no shared species as NA in Manhattan dissimilarities
dis <- vegdist(dune, "manhattan")
is.na(dis) <- no.shared(dune)
```

---

### Ecological Diversity Indices

#### Description

Shannon, Simpson, and Fisher diversity indices and species richness.

#### Usage

```r
diversity(x, index = "shannon", MARGIN = 1, base = exp(1))
fisher.alpha(x, MARGIN = 1, ...)
specnumber(x, groups, MARGIN = 1)
```

#### Arguments

- `x`: Community data, a matrix-like object or a vector.
- `index`: Diversity index, one of "shannon", "simpson" or "invsimpson".
- `MARGIN`: Margin for which the index is computed.
- `base`: The logarithm base used in shannon.
- `groups`: A grouping factor: if given, finds the total number of species in each group.
- `...`: Parameters passed to the function.

#### Details

Shannon or Shannon–Weaver (or Shannon–Wiener) index is defined as $H' = -\sum p_i \log_b p_i$, where $p_i$ is the proportional abundance of species $i$ and $b$ is the base of the logarithm. It is most popular to use natural logarithms, but some argue for base $b = 2$ (which makes sense, but no real difference).

Both variants of Simpson’s index are based on $D = \sum p_i^2$. Choice simpson returns $1 - D$ and invsimpson returns $1/D$.

`fisher.alpha` estimates the $\alpha$ parameter of Fisher’s logarithmic series (see `fisherfit`). The estimation is possible only for genuine counts of individuals.

Function `specnumber` finds the number of species. With `MARGIN = 2`, it finds frequencies of species. If `groups` is given, finds the total number of species in each group (see example on finding one kind of beta diversity with this option).
Better stories can be told about Simpson’s index than about Shannon’s index, and still grander narratives about rarefaction (Hurlbert 1971). However, these indices are all very closely related (Hill 1973), and there is no reason to despise one more than others (but if you are a graduate student, don’t drag me in, but obey your Professor’s orders). In particular, the exponent of the Shannon index is linearly related to inverse Simpson (Hill 1973) although the former may be more sensitive to rare species. Moreover, inverse Simpson is asymptotically equal to rarefied species richness in sample of two individuals, and Fisher’s $\alpha$ is very similar to inverse Simpson.

**Value**

A vector of diversity indices or numbers of species.

**Author(s)**

Jari Oksanen and Bob O’Hara (fisher.alpha).

**References**


**See Also**

These functions calculate only some basic indices, but many others can be derived with them (see Examples). Facilities related to diversity are discussed in a vegan vignette that can be read with browseVignettes("vegan"). Functions renyi and tsallis estimate a series of generalized diversity indices. Function rarefy finds estimated number of species for given sample size. Beta diversity can be estimated with betadiver. Diversities can be partitioned with adipart and multipart.

**Examples**

```r
data(BCI)
H <- diversity(BCI)
simp <- diversity(BCI, "simpson")
invsimp <- diversity(BCI, "inv")
## Unbiased Simpson (Hurlbert 1971, eq. 5) with rarefy:
unbias.simp <- rarefy(BCI, 2) - 1
## Fisher alpha
alpha <- fisher.alpha(BCI)
## Plot all
pairs(cbind(H, simp, invsimp, unbias.simp, alpha), pch="+", col="blue")
## Species richness (S) and Pielou’s evenness (J):
S <- specnumber(BCI) ## rowSums(BCI > 0) does the same...
J <- H/log(S)
## beta diversity defined as gamma/alpha - 1:
data(dune)
data(dune.env)
alpha <- with(dune.env, tapply(specnumber(dune), Management, mean))
```
gamma <- with(dune.env, specnumber(dune, Management))
gamma/alpha - 1

dune

Vegetation and Environment in Dutch Dune Meadows.

Description

The dune meadow vegetation data, dune, has cover class values of 30 species on 20 sites. The corresponding environmental data frame dune.env has following entries:

Usage

data(dune)
data(dune.env)

Format

dune is a data frame of observations of 30 species at 20 sites. The species names are abbreviated to 4+4 letters (see make.cepnames). The following names are changed from the original source (Jongman et al. 1987): Leontodon autumnalis to Scorzonoides, and Potentilla palustris to Comarum.
dune.env is a data frame of 20 observations on the following 5 variables:


Moisture: an ordered factor with levels: 1 < 2 < 4 < 5.

Management: a factor with levels: BF (Biological farming), HF (Hobby farming), NM (Nature Conservation Management), and SF (Standard Farming).

Use: an ordered factor of land-use with levels: Hayfield < Haypastu < Pasture.

Manure: an ordered factor with levels: 0 < 1 < 2 < 3 < 4.

Source


Examples

data(dune)
data(dune.env)
**Description**

Classification table of the species in the `dune` data set.

**Usage**

```r
data(dune.taxon)
data(dune.phylodis)
```

**Format**

dune.taxon is data frame with 30 species (rows) classified into five taxonomic levels (columns).
dune.phylodis is a dist object of estimated coalescence ages extracted from doi: 10.5061/dryad.63q27 (Zanne et al. 2014) using tools in packages ape and phylobase.

**Details**

The families and orders are based on APG IV (2016) in vascular plants and on Hill et al. (2006) in mosses. The higher levels (superorder and subclass) are based on Chase & Reveal (2009). Chase & Reveal (2009) treat Angiosperms and mosses as subclasses of class Equisetopsida (land plants), but bryologists have traditionally used much more inflated levels which are adjusted here to match Angiosperm classification.

**References**


**See Also**

Functions `taxondive`, `treedive`, and `treedist` use these data sets.
Examples

data(dune.taxon)
data(dune.phylodis)

\texttt{eigenvals} \hspace{1cm} \textit{Extract Eigenvalues from an Ordination Object}

Description

Function extracts eigenvalues from an object that has them. Many multivariate methods return such objects.

Usage

\begin{verbatim}
eigenvals(x, ...)  ## S3 method for class 'cca'
eigenvals(x, model = c("all", "unconstrained", "constrained"),
           constrained = NULL, ...)
## S3 method for class 'eigenvals'
summary(object, ...)
\end{verbatim}

Arguments

\begin{itemize}
\item \texttt{x} \hspace{1cm} An object from which to extract eigenvalues.
\item \texttt{object} \hspace{1cm} An \texttt{eigenvals} result object.
\item \texttt{model} \hspace{1cm} Which eigenvalues to return for objects that inherit from class "cca" only.
\item \texttt{constrained} \hspace{1cm} Return only constrained eigenvalues. Deprecated as of vegan 2.5-0. Use \texttt{model} instead.
\item \texttt{...} \hspace{1cm} Other arguments to the functions (usually ignored)
\end{itemize}

Details

This is a generic function that has methods for \texttt{cca}, \texttt{wcmdscale}, \texttt{pcnm}, \texttt{prcomp}, \texttt{princomp}, \texttt{dudi} (of \texttt{ade4}), and \texttt{pca} and \texttt{pco} (of \texttt{labdsv}) result objects. The default method also extracts eigenvalues if the result looks like being from \texttt{eigen} or \texttt{svd}. Functions \texttt{prcomp} and \texttt{princomp} contain square roots of eigenvalues that all called standard deviations, but \texttt{eigenvals} function returns their squares. Function \texttt{svd} contains singular values, but function \texttt{eigenvals} returns their squares. For constrained ordination methods \texttt{cca}, \texttt{rda} and \texttt{capscale} the function returns the both constrained and unconstrained eigenvalues concatenated in one vector, but the partial component will be ignored. However, with argument \texttt{constrained = TRUE} only constrained eigenvalues are returned.

The \texttt{summary} of \texttt{eigenvals} result returns eigenvalues, proportion explained and cumulative proportion explained. The result object can have some negative eigenvalues (\texttt{wcmdscale}, \texttt{capscale}, \texttt{pcnm}) which correspond to imaginary axes of Euclidean mapping of non-Euclidean distances (Gower 1985). In these cases, the sum of absolute values of eigenvalues is used in calculating the proportions explained, and real axes (corresponding to positive eigenvalues) will only explain a part of total variation (Mardia et al. 1979, Gower 1985).
**Value**

An object of class "eigenvals", which is a vector of eigenvalues.

The summary method returns an object of class "summary.eigenvals", which is a matrix.

**Author(s)**

Jari Oksanen.

**References**


**See Also**

eigen, svd, prcomp, princomp, cca, rda, capscale, wcmdscale, cca.object.

**Examples**

data(varespec)
data(varechem)
mod <- cca(varespec ~ A1 + P + K, varechem)
ev <- eigenvals(mod)
ev
summary(ev)

## choose which eigenvalues to return
eigenvals(mod, model = "unconstrained")

**Description**

The function fits environmental vectors or factors onto an ordination. The projections of points onto vectors have maximum correlation with corresponding environmental variables, and the factors show the averages of factor levels.

**Usage**

## Default S3 method:
envfit(ord, env, permutations = 999, strata = NULL,
       choices=c(1,2), display = "sites", w = weights(ord, display),
       na.rm = FALSE, ...)
## S3 method for class 'formula'
envfit(formula, data, ...)  
## S3 method for class 'envfit'
plot(x, choices = c(1,2), labels, arrow.mul, at = c(0,0),
    axis = FALSE, p.max = NULL, col = "blue", bg = add = TRUE, ...)
## S3 method for class 'envfit'
scores(x, display, choices, ...)
vectorfit(X, P, permutations = 0, strata = NULL, w, ...)
factorfit(X, P, permutations = 0, strata = NULL, w, ...)

Arguments

data frame, matrix or vector of environmental variables. The variables can be
of mixed type (factors, continuous variables) in data frames.

X Matrix or data frame of ordination scores.

P Data frame, matrix or vector of environmental variable(s). These must be con-
tinuous for vectorfit and factors for factorfit.

permutations a list of control values for the permutations as returned by the function how, or
the number of permutations required, or a permutation matrix where each row
gives the permuted indices. Set permutations = 0 to skip permutations.

formula, data Model formula and data.

na.rm Remove points with missing values in ordination scores or environmental vari-
ables. The operation is casewise: the whole row of data is removed if there is a
missing value and na.rm = TRUE.

x A result object from envfit. For ordiArrowMul and ordiArrowTextXY this
must be a two-column matrix (or matrix-like object) containing the coordinates
of arrow heads on the two plot axes, and other methods extract such a structure
from the envfit results.

choices Axes to plotted.

labels Change plotting labels. The argument should be a list with elements vectors
and factors which give the new plotting labels. If either of these elements is
omitted, the default labels will be used. If there is only one type of elements
(only vectors or only factors), the labels can be given as vector. The default
labels can be displayed with labels command.

arrow.mul Multiplier for vector lengths. The arrows are automatically scaled similarly as
in plot.cca if this is not given and add = TRUE.

at The origin of fitted arrows in the plot. If you plot arrows in other places then
origin, you probably have to specify arrow.mul.

axis Plot axis showing the scaling of fitted arrows.

p.max Maximum estimated P value for displayed variables. You must calculate P
values with setting permutations to use this option.

col Colour in plotting.

bg Background colour for labels. If bg is set, the labels are displayed with ordilabel
instead of text. See Examples for using semitransparent background.
add

Results added to an existing ordination plot.

strata

An integer vector or factor specifying the strata for permutation. If supplied, observations are permuted only within the specified strata.

display

In fitting functions these are ordinary site scores or linear combination scores ("1c") in constrained ordination (cca, rda, capscale). In scores function they are either "vectors" or "factors" (with synonyms "bp" or "cn", resp.).

w

Weights used in fitting (concerns mainly cca and decorana results which have nonconstant weights).

... Parameters passed to scores.

Details

Function envfit finds vectors or factor averages of environmental variables. Function plot.envfit adds these in an ordination diagram. If x is a data.frame, envfit uses factorfit for factor variables and vectorfit for other variables. If x is a matrix or a vector, envfit uses only vectorfit. Alternatively, the model can be defined a simplified model formula, where the left hand side must be an ordination result object or a matrix of ordination scores, and right hand side lists the environmental variables. The formula interface can be used for easier selection and/or transformation of environmental variables. Only the main effects will be analysed even if interaction terms were defined in the formula.

The ordination results are extracted with scores and all extra arguments are passed to the scores. The fitted models only apply to the results defined when extracting the scores when using envfit. For instance, scaling in constrained ordination (see scores.rda, scores.cca) must be set in the same way in envfit and in the plot or the ordination results (see Examples).

The printed output of continuous variables (vectors) gives the direction cosines which are the coordinates of the heads of unit length vectors. In plot these are scaled by their correlation (square root of the column r2) so that “weak” predictors have shorter arrows than “strong” predictors. You can see the scaled relative lengths using command scores. The plotted (and scaled) arrows are further adjusted to the current graph using a constant multiplier: this will keep the relative r2-scaled lengths of the arrows but tries to fill the current plot. You can see the multiplier using ordiArrowMul(result_of_envfit), and set it with the argument arrow.mul.

Functions vectorfit and factorfit can be called directly. Function vectorfit finds directions in the ordination space towards which the environmental vectors change most rapidly and to which they have maximal correlations with the ordination configuration. Function factorfit finds averages of ordination scores for factor levels. Function factorfit treats ordered and unordered factors similarly.

If permutations > 0, the significance of fitted vectors or factors is assessed using permutation of environmental variables. The goodness of fit statistic is squared correlation coefficient (r2). For factors this is defined as r2 = 1 − ss_w/ss_t, where ss_w and ss_t are within-group and total sums of squares. See permutations for additional details on permutation tests in Vegan.

User can supply a vector of prior weights w. If the ordination object has weights, these will be used. In practise this means that the row totals are used as weights with cca or decorana results. If you do not like this, but want to give equal weights to all sites, you should set w = NULL. The fitted vectors are similar to biplot arrows in constrained ordination only when fitted to LC scores (display = "1c") and you set scaling = "species" (see scores.cca). The weighted fitting gives similar results to biplot arrows and class centroids in cca.
The lengths of arrows for fitted vectors are automatically adjusted for the physical size of the plot, and the arrow lengths cannot be compared across plots. For similar scaling of arrows, you must explicitly set the arrowNmul argument in the plot command; see ordiArrowMul and ordiArrowTextXY. The results can be accessed with scores.envfit function which returns either the fitted vectors scaled by correlation coefficient or the centroids of the fitted environmental variables.

Value

Functions vectorfit and factorfit return lists of classes vectorfit and factorfit which have a print method. The result object have the following items:

- **arrows**: Arrow endpoints from vectorfit. The arrows are scaled to unit length.
- **centroids**: Class centroids from factorfit.
- **r**: Goodness of fit statistic: Squared correlation coefficient
- **permutations**: Number of permutations.
- **control**: A list of control values for the permutations as returned by the function how.
- **pvals**: Empirical P-values for each variable.

Function envfit returns a list of class envfit with results of vectorfit and envfit as items. Function plot.envfit scales the vectors by correlation.

Note

Fitted vectors have become the method of choice in displaying environmental variables in ordination. Indeed, they are the optimal way of presenting environmental variables in Constrained Correspondence Analysis cca, since there they are the linear constraints. In unconstrained ordination the relation between external variables and ordination configuration may be less linear, and therefore other methods than arrows may be more useful. The simplest is to adjust the plotting symbol sizes (cex, symbols) by environmental variables. Fancier methods involve smoothing and regression methods that abound in R, and ordisurf provides a wrapper for some.

Author(s)

Jari Oksanen. The permutation test derives from the code suggested by Michael Scroggie.

See Also

A better alternative to vectors may be ordisurf.

Examples

data(varespec, varechem)
library(MASS)
ord <- metaMDS(varespec)
(fit <- envfit(ord, varechem, perm = 999))
scores(fit, "vectors")
plot(ord)
plot(fit)
eventstar

Scale Parameter at the Minimum of the Tsallis Evenness Profile

Description

The function eventstar finds the minimum \( q^* \) of the evenness profile based on the Tsallis entropy. This scale factor of the entropy represents a specific weighting of species relative frequencies that leads to minimum evenness of the community (Mendes et al. 2008).

Usage

`eventstar(x, qmax = 5)`

Arguments

- **x**: A community matrix or a numeric vector.
- **qmax**: Maximum scale parameter of the Tsallis entropy to be used in finding the minimum of Tsallis based evenness in the range \( c(0, q_{max}) \).
Details

The function `eventstar` finds a characteristic value of the scale parameter $q$ of the Tsallis entropy corresponding to minimum of the evenness (equitability) profile based on Tsallis entropy. This value was proposed by Mendes et al. (2008) as $q^\ast$.

The $q^\ast$ index represents the scale parameter of the one parameter Tsallis diversity family that leads to the greatest deviation from the maximum equitability given the relative abundance vector of a community.

The value of $q^\ast$ is found by identifying the minimum of the evenness profile over scaling factor $q$ by one-dimensional minimization. Because evenness profile is known to be a convex function, it is guaranteed that underlying optimize function will find a unique solution if it is in the range $c(\emptyset, q_{\text{max}})$.

The scale parameter value $q^\ast$ is used to find corresponding values of diversity ($H_{q^\ast}$), evenness ($H_{q^\ast}(\text{max})$), and numbers equivalent ($D_{q^\ast}$). For calculation details, see `tsallis` and Examples below.

Mendes et al. (2008) advocated the use of $q^\ast$ and corresponding diversity, evenness, and Hill numbers, because it is a unique value representing the diversity profile, and is positively associated with rare species in the community, thus it is a potentially useful indicator of certain relative abundance distributions of the communities.

Value

A data frame with columns:

- `qstar` scale parameter value $q^\ast$ corresponding to minimum value of Tsallis based evenness profile.
- `estar` Value of evenness based on normalized Tsallis entropy at $q^\ast$.
- `hstar` Value of Tsallis entropy at $q^\ast$.
- `dstar` Value of Tsallis entropy at $q^\ast$ converted to numbers equivalents (also called as Hill numbers, effective number of species, ‘true’ diversity; cf. Jost 2007).

See `tsallis` for calculation details.

Note

Values for $q^\ast$ found by Mendes et al. (2008) ranged from 0.56 and 1.12 presenting low variability, so an interval between 0 and 5 should safely encompass the possibly expected $q^\ast$ values in practice, but profiling the evenness and changing the value of the qmax argument is advised if output values near the range limits are found.

Author(s)

Eduardo Ribeiro Cunha <edurcunha@gmail.com> and Heloisa Beatriz Antoniazi Evangelista <helobeatriz@gmail.com>, with technical input of Péter Sólymos.
References


See Also

Tsallis entropy: `tsallis`

Examples

data(BCI)
(x <- eventstar(BCI[1:5,]))
## profiling
y <- as.numeric(BCI[10,])
(z <- eventstar(y))
q <- seq(0, z, 0.05)
Eprof <- tsallis(y, scales=q, norm=TRUE)
Hprof <- tsallis(y, scales=q)
Dprof <- tsallis(y, scales=q, hill=TRUE)
opar <- par(mfrow=c(3,1))
plot(q, Eprof, type="l", main="Evenness")
abline(v=z$qstar, h=tsallis(y, scales=z$qstar, norm=TRUE), col=2)
plot(q, Hprof, type="l", main="Diversity")
abline(v=z$qstar, h=tsallis(y, scales=z$qstar), col=2)
plot(q, Dprof, type="l", main="Effective number of species")
abline(v=z$qstar, h=tsallis(y, scales=z$qstar, hill=TRUE), col=2)
par(opar)
Arguments

x Community data vector for fitting functions or their result object for plot functions.
tiesplit Split frequencies 1, 2, 4, 8 etc between adjacent octaves.
truncate Truncation point for log-Normal model, in log2 units. Default value −1 corresponds to the left border of zero Octave. The choice strongly influences the fitting results.
xlab, ylab Labels for x and y axes.
bar.col Colour of data bars.
line.col Colour of fitted line.
lwd Width of fitted line.
kind Kind of plot to drawn: "bar" is similar bar plot as in plot.fisherfit, "hiplot" draws vertical lines as with plot(..., type="h"), and "points" and "lines" are obvious.
add Add to an existing plot.
xadjust Adjustment of horizontal positions in octaves.
... Other parameters passed to functions. Ignored in prestonfit and tiesplit passed to as.preston in prestondistr.

Details

In Fisher's logarithmic series the expected number of species $f$ with $n$ observed individuals is $f_n = \alpha x^n / n$ (Fisher et al. 1943). The estimation is possible only for genuine counts of individuals. The parameter $\alpha$ is used as a diversity index, and $\alpha$ and its standard error can be estimated with a separate function fisher.alpha. The parameter $x$ is taken as a nuisance parameter which is not estimated separately but taken to be $n/(n + \alpha)$. Helper function as.fisher transforms abundance data into Fisher frequency table.

Preston (1948) was not satisfied with Fisher's model which seemed to imply infinite species richness, and postulated that rare species is a diminishing class and most species are in the middle of frequency scale. This was achieved by collapsing higher frequency classes into wider and wider "octaves" of doubling class limits: 1, 2, 3–4, 5–8, 9–16 etc. occurrences. It seems that Preston
regarded frequencies 1, 2, 4, etc. as “tied” between octaves (Williamson & Gaston 2005). This means that only half of the species with frequency 1 are shown in the lowest octave, and the rest are transferred to the second octave. Half of the species from the second octave are transferred to the higher one as well, but this is usually not as large a number of species. This practise makes data look more lognormal by reducing the usually high lowest octaves. This can be achieved by setting argument tiesplit = TRUE. With tiesplit = FALSE the frequencies are not split, but all ones are in the lowest octave, all twos in the second, etc. Williamson & Gaston (2005) discuss alternative definitions in detail, and they should be consulted for a critical review of log-Normal model.

Any logseries data will look like lognormal when plotted in Preston’s way. The expected frequency \( f \) at abundance octave \( o \) is defined by

\[
f_o = S_0 \exp\left(-\left(\log_2(o) - \mu\right)^2/2\sigma^2\right),
\]

where \( \mu \) is the location of the mode and \( \sigma \) the width, both in \( \log_2 \) scale, and \( S_0 \) is the expected number of species at mode. The lognormal model is usually truncated on the left so that some rare species are not observed. Function prestonfit fits the truncated lognormal model as a second degree log-polynomial to the octave pooled data using Poisson (when tiesplit = FALSE) or quasi-Poisson (when tiesplit = TRUE) error. Function prestondistr fits left-truncated Normal distribution to \( \log_2 \) transformed non-pooled observations with direct maximization of log-likelihood. Function prestondistr is modelled after function fitdistr which can be used for alternative distribution models.

The functions have common print, plot and lines methods. The lines function adds the fitted curve to the octave range with line segments showing the location of the mode and the width (sd) of the response. Function as.preston transforms abundance data to octaves. Argument tiesplit will not influence the fit in prestondistr, but it will influence the barplot of the octaves. The total extrapolated richness from a fitted Preston model can be found with function veiledspec. The function accepts results both from prestonfit and from prestondistr. If veiledspec is called with a species count vector, it will internally use prestonfit. Function specpool provides alternative ways of estimating the number of unseen species. In fact, Preston’s lognormal model seems to be truncated at both ends, and this may be the main reason why its result differ from lognormal models fitted in Rank–Abundance diagrams with functions rad.lognormal.

Value

The function prestonfit returns an object with fitted coefficients, and with observed (freq) and fitted (fitted) frequencies, and a string describing the fitting method. Function prestondistr omits the entry fitted. The function fisherfit returns the result of nlm, where item estimate is \( \alpha \). The result object is amended with the nuisance parameter and item fisher for the observed data from as.fisher.

Author(s)

Bob O’Hara and Jari Oksanen.

References


**See Also**

diversity, fisher.alpha, radfit, specpool. Function fitdistr of MASS package was used as the model for prestondistr. Function density can be used for smoothed non-parametric estimation of responses, and qqplot is an alternative, traditional and more effective way of studying concordance of observed abundances to any distribution model.

**Examples**

data(BCI)
mod <- fisherfit(BCI[5,])
mod
# prestonfit seems to need large samples
mod.oct <- prestonfit(colSums(BCI))
mod.11 <- prestondistr(colSums(BCI))
mod.oct
mod.11
plot(mod.oct)
lines(mod.11, line.col="blue3") # Different
## Smoothed density
den <- density(log2(colSums(BCI)))
lines(den$x, ncol(BCI)*den$y, lwd=2) # Fairly similar to mod.oct
## Extrapolated richness
veiledspec(mod.oct)
veiledspec(mod.11)

---


**Description**

Functions goodness and inertcomp can be used to assess the goodness of fit for individual sites or species. Function vif.cca and alias.cca can be used to analyse linear dependencies among constraints and conditions. In addition, there are some other diagnostic tools (see ‘Details’).

**Usage**

```r
# S3 method for class 'cca'
goodness(object, choices, display = c("species", "sites"),
         model = c("CCA", "CA"), summarize = FALSE, addprevious = FALSE, ...)
inertcomp(object, display = c("species", "sites"),
         unity = FALSE, proportional = FALSE)
spenvcor(object)
intersetcor(object)
vif.cca(object)
```
goodness.cca

## S3 method for class 'cca'
alias(object, names.only = FALSE, ...)

### Arguments

- **object**: A result object from `cca`, `rda`, `dbrda` or `capscale`.
- **display**: Display "species" or "sites". Species are not available in `dbrda` and `capscale`.
- **choices**: Axes shown. Default is to show all axes of the "model".
- **model**: Show constrained ("CCA") or unconstrained ("CA") results.
- **summarize**: Show only the accumulated total.
- **addprevious**: Add the variation explained by previous components when statistic="explained". For `model` = "CCA" add conditioned (partialled out) variation, and for `model` = "CA" add both conditioned and constrained variation. This will give cumulative explanation. The argument has no effect when statistic="distance", but this will always show the residual distance after current axis and all previous components.
- **unity**: Scale inertia components to unit sum (sum of all items is 1).
- **proportional**: Give the inertia components as proportional for the corresponding total of the item (sum of each row is 1). This option takes precedence over unity.
- **names.only**: Return only names of aliased variable(s) instead of defining equations.
- **...**: Other parameters to the functions.

### Details

Function `goodness` gives cumulative proportion of inertia accounted by species up to chosen axes. The proportions can be assessed either by species or by sites depending on the argument `display`, but species are not available in distance-based `dbrda`. The function is not implemented for `capscale`.

Function `inertcomp` decomposes the inertia into partial, constrained and unconstrained components for each site or species. Legendre & De Cáceres (2012) called these inertia components as local contributions to beta-diversity (LCBD) and species contributions to beta-diversity (SCBD), and they give these as relative contributions summing up to unity (argument `unity` = `TRUE`). For this interpretation, appropriate dissimilarity measures should be used in `dbrda` or appropriate standardization in `rda` (Legendre & De Cáceres 2012). The function is not implemented for `capscale`.

Function `spenvcor` finds the so-called “species – environment correlation” or (weighted) correlation of weighted average scores and linear combination scores. This is a bad measure of goodness of ordination, because it is sensitive to extreme scores (like correlations are), and very sensitive to overfitting or using too many constraints. Better models often have poorer correlations. Function `ordispider` can show the same graphically.

Function `intersetcor` finds the so-called “interset correlation” or (weighted) correlation of weighted averages scores and linear combination scores. The defined contrasts are used for factor variables. This is a bad measure since it is a correlation. Further, it focuses on correlations between single contrasts and single axes instead of looking at the multivariate relationship. Fitted vectors (`envfit`) provide a better alternative. Biplot scores (see `scores.cca`) are a multivariate alternative for (weighted) correlation between linear combination scores and constraints.
Function `vif.cca` gives the variance inflation factors for each constraint or contrast in factor constraints. In partial ordination, conditioning variables are analysed together with constraints. Variance inflation is a diagnostic tool to identify useless constraints. A common rule is that values over 10 indicate redundant constraints. If later constraints are complete linear combinations of conditions or previous constraints, they will be completely removed from the estimation, and no biplot scores or centroids are calculated for these aliased constraints. A note will be printed with default output if there are aliased constraints. Function `alias` will give the linear coefficients defining the aliased constraints, or only their names with argument `names.only = TRUE`.

**Value**

The functions return matrices or vectors as is appropriate.

**Note**

It is a common practise to use goodness statistics to remove species from ordination plots, but this may not be a good idea, as the total inertia is not a meaningful concept in `cca`, in particular for rare species.

**Author(s)**

Jari Oksanen. The `vif.cca` relies heavily on the code by W. N. Venables. `alias.cca` is a simplified version of `alias.lm`.

**References**


**See Also**

`cca`, `rda`, `dbrda`, `capscale`, `vif`.

**Examples**

```r
data(dune)
data(dune.env)
mod <- cca(dune ~ A1 + Management + Condition(Moisture), data=dune.env)
goodness(mod, addprevious = TRUE)
goodness(mod, addprevious = TRUE, summ = TRUE)
# Inertia components
inertcomp(mod, prop = TRUE)
inertcomp(mod)
# vif.cca
vif.cca(mod)
# Aliased constraints
mod <- cca(dune ~ ., dune.env)
```
goodness.metaMDS 99

goodness.metaMDS  Goodness of Fit and Shepard Plot for Nonmetric Multidimensional Scaling

description

Function goodness.metaMDS finds goodness of fit measure for points in nonmetric multidimensional scaling, and function stressplot makes a Shepard diagram.

usage

## S3 method for class 'metaMDS'

goodness(object, dis, ...)

## Default S3 method:

stressplot(object, dis, pch, p.col = "blue", l.col = "red",
           lwd = 2, ...)

arguments

object A result object from metaMDS, monoMDS or isoMDS.
dis Dissimilarities. This should not be used with metaMDS or monoMDS, but must be used with isoMDS.
pch Plotting character for points. Default is dependent on the number of points.
p.col, l.col Point and line colours.
lwd Line width. For monoMDS the default is lwd = 1 if more than two lines are drawn, and lwd = 2 otherwise.
... Other parameters to functions, e.g. graphical parameters.

details

Function goodness.metaMDS finds a goodness of fit statistic for observations (points). This is defined so that sum of squared values is equal to squared stress. Large values indicate poor fit. The absolute values of the goodness statistic depend on the definition of the stress: isoMDS expresses stress in percents, and therefore its goodness values are 100 times higher than those of monoMDS which expresses the stress as a proportion.

Function stressplot draws a Shepard diagram which is a plot of ordination distances and monotone or linear fit line against original dissimilarities. In addition, it displays two correlation-like statistics on the goodness of fit in the graph. The nonmetric fit is based on stress $\tau$ and defined
as $R^2 = 1 - S^2$. The “linear fit” is the squared correlation between fitted values and ordination distances. For `monoMDS`, the “linear fit” and $R^2$ from “stress type 2” are equal.

Both functions can be used with `metaMDS`, `monoMDS` and `isoMDS`. The original dissimilarities should not be given for `monoMDS` or `metaMDS` results (the latter tries to reconstruct the dissimilarities using `metaMDSredist` if `isoMDS` was used as its engine). With `isoMDS` the dissimilarities must be given. In either case, the functions inspect that dissimilarities are consistent with current ordination, and refuse to analyse inconsistent dissimilarities. Function `goodness.metaMDS` is generic in `vegan`, but you must spell its name completely with `isoMDS` which has no class.

Value

Function `goodness` returns a vector of values. Function `stressplot` returns invisibly an object with items for original dissimilarities, ordination distances and fitted values.

Author(s)

Jari Oksanen.

See Also

`metaMDS`, `monoMDS`, `isoMDS`, `Shepard`. Similar diagrams for eigenvector ordinations can be drawn with `stressplot.wcmdscale`, `stressplot.cca`, `stressplot.rda` and `stressplot.capscale`.

Examples

```r
data(vespa)
mod <- metaMDS(vespa)
stressplot(mod)
gof <- goodness(mod)
gof
plot(mod, display = "sites", type = "n")
points(mod, display = "sites", cex = 2*gof/mean(gof))
```

---

**humpfit**

No-interaction Model for Hump-backed Species Richness vs. Biomass

Description

Function `humpfit` fits a no-interaction model for species richness vs. biomass data (Oksanen 1996). This is a null model that produces a hump-backed response as an artifact of plant size and density.

Usage

```r
humpfit(mass, spno, family = poisson, start)
## S3 method for class 'humpfit'
summary(object, ...)
## S3 method for class 'humpfit'
predict(object, newdata = NULL, ...)
```
## S3 method for class 'humpfit'

plot(x, xlab = "Biomass", ylab = "Species Richness", lwd = 2,
     l.col = "blue", p.col = 1, type = "b", ...)

## S3 method for class 'humpfit'

points(x, ...)

## S3 method for class 'humpfit'

lines(x, segments = 101, ...)

## S3 method for class 'humpfit'

profile(fitted, parm = 1:3, alpha = 0.01, maxsteps = 20, del = zmax/5, ...)

### Arguments

- **mass**: Biomass.
- **spno**: Species richness.
- **start**: Vector of starting values for all three parameters.
- **family**: Family of error distribution. Any family can be used, but the link function is always Fisher’s diversity model, and other link functions are silently ignored.
- **x**, **object**, **fitted**: Result object of humpfit
- **newdata**: Values of mass used in predict. The original data values are used if missing.
- **xlab**, **ylab**: Axis labels in plot
- **lwd**: Line width
- **l.col**, **p.col**: Line and point colour in plot
- **type**: Type of plot: "p" for observed points, "l" for fitted lines, "b" for both, and "n" for only setting axes.
- **segments**: Number of segments used for fitted lines.
- **parm**: Profiled parameters.
- **alpha**, **maxsteps**, **del**: Parameters for profiling range and density.
- **...**: Other parameters to functions.

### Details

The no-interaction model assumes that the humped species richness pattern along biomass gradient is an artifact of plant size and density (Oksanen 1996). For low-biomass sites, it assumes that plants have a fixed size, and biomass increases with increasing number of plants. When the sites becomes crowded, the number of plants and species richness reaches the maximum. Higher biomass is reached by increasing the plant size, and then the number of plants and species richness will decrease. At biomasses below the hump, plant number and biomass are linearly related, and above the hump, plant number is proportional to inverse squared biomass. The number of plants is related to the number of species by the relationship (link function) from Fisher’s log-series (Fisher et al. 1943).

The parameters of the model are:

1. **hump**: the location of the hump on the biomass gradient.
2. scale: an arbitrary multiplier to translate the biomass into virtual number of plants.
3. alpha: Fisher’s \( \alpha \) to translate the virtual number of plants into number of species.

The parameters scale and alpha are intermingled and this function should not be used for estimating Fisher’s \( \alpha \). Probably the only meaningful and interesting parameter is the location of the hump.

Function may be very difficult to fit and easily gets trapped into local solutions, or fails with non-Poisson families, and function profile should be used to inspect the fitted models. If you have loaded package MASS, you can use functions plot.profile, pairs.profile for graphical inspection of the profiles, and confint.profile.glm for the profile based confidence intervals.

The original model intended to show that there is no need to speculate about “competition” and “stress” (Al-Mufti et al. 1977), but humped response can be produced as an artifact of using fixed plot size for varying plant sizes and densities.

**Value**

The function returns an object of class "humpfit" inheriting from class "glm". The result object has specific summary, predict, plot, points and lines methods. In addition, it can be accessed by the following methods for glm objects: AIC, extractAIC, deviance, coef, residuals.glm (except type = "partial"), fitted, and perhaps some others. In addition, function ellipse.glm (package ellipse) can be used to draw approximate confidence ellipses for pairs of parameters, if the normal assumptions look appropriate.

**Note**

The function is a replacement for the original GLIM4 function at the archive of Journal of Ecology. There the function was represented as a mixed glm with one non-linear parameter (hump) and a special one-parameter link function from Fisher’s log-series. The current function directly applies non-linear maximum likelihood fitting using function nlm. Some expected problems with the current approach are:

- The function is discontinuous at hump and may be difficult to optimize in some cases (the lines will always join, but the derivative jumps).
- The function does not try very hard to find sensible starting values and can fail. The user may supply starting values in argument start if fitting fails.
- The estimation is unconstrained, but both scale and alpha should always be positive. Perhaps they should be fitted as logarithmic. Fitting Gamma family models might become easier, too.

**Author(s)**

Jari Oksanen

**References**


See Also

`fisherfit.glm, profile.glm, confint.glm`.

Examples

```r
## Data approximated from Al-Mufti et al. (1977)

mass <- c(140, 230, 310, 310, 400, 510, 610, 670, 860, 900, 1050, 1160, 1900, 2480)
spno <- c(1, 4, 3, 9, 18, 30, 20, 14, 3, 2, 3, 2, 5, 2)
sol <- humpfit(mass, spno)
summary(sol) # Almost infinite alpha...
plot(sol)
# confint is in MASS, and implicitly calls profile.humpfit.
# Parameter 3 (alpha) is too extreme for profile and confint, and we
# must use only "hump" and "scale".
library(MASS)
plot(profile(sol, parm=1:2))
confint(sol, parm=c(1,2))
```

---

**indpower**

*Indicator Power of Species*

Description

Indicator power calculation of Halme et al. (2009) or the congruence between indicator and target species.

Usage

```r
indpower(x, type = 0)
```

Arguments

- **x**: Community data frame or matrix.
- **type**: The type of statistic to be returned. See Details for explanation.
Details

Halme et al. (2009) described an index of indicator power defined as \( IP_I = \sqrt{a \times b} \), where \( a = S/O_I \) and \( b = 1 - (O_T - S)/(N - O_I) \). \( N \) is the number of sites, \( S \) is the number of shared occurrences of the indicator (\( I \)) and the target (\( T \)) species. \( O_I \) and \( O_T \) are number of occurrences of the indicator and target species. The \texttt{type} argument in the function call enables to choose which statistic to return. \texttt{type = 0} returns \( IP_I \), \texttt{type = 1} returns \( a \), \texttt{type = 2} returns \( b \). Total indicator power (TIP) of an indicator species is the column mean (without its own value, see examples). Halme et al. (2009) explain how to calculate confidence intervals for these statistics, see Examples.

Value

A matrix with indicator species as rows and target species as columns (this is indicated by the first letters of the row/column names).

Author(s)

Peter Solymos

References


See Also

\texttt{indval} (package \texttt{labdsv}) for the indicator species analysis of Dufrêne & Legendre. Function \texttt{beals} estimates individual cell probabilities of species occurrences.

Examples

data(dune)
## IP values
ip <- indpower(dune)
## and TIP values
diag(ip) <- NA
(TIP <- rowMeans(ip, na.rm=TRUE))

## p value calculation for a species
## from Halme et al. 2009
## i is ID for the species
i <- 1
fun <- function(x, i) indpower(x)[i,-i]
## 'c0' randomizes species occurrences
os <- oecosimu(dune, fun, "c0", i=i, nsimul=99)
## get z values from oecosimu output
z <- os$oecosimu$z
## p-value
(p <- sum(z) / sqrt(length(z))))
## 'heterogeneity' measure
(chi2 <- sum((z - mean(z))^2))
pchisq(chi2, df=length(z)-1)
## Halme et al.'s suggested output

```r
c <- TIP[TIP[,1],
significance=p,
heterogeneity=chi2,
minIP=min(fun(dune, i=i)),
varIP=sd(fun(dune, i=i)^2)
```

### Description

This set of function extracts influence statistics and some other linear model statistics directly from a constrained ordination result object from `cca`, `rda`, `capscale` or `dbrda`. The constraints are linear model functions and these support functions return identical results as the corresponding linear models (`lm`), and you can use their documentation. The main functions for normal usage are leverage values (`hatvalues`), standardized residuals (`rstandard`), studentized or leave-one-out residuals (`rstudent`), and Cook's distance (`cooks.distance`). In addition, `vcov` returns the variance-covariance matrix of coefficients, and its diagonal values the variances of coefficients. Other functions are mainly support functions for these, but they can be used directly.

### Usage

```r
## S3 method for class 'cca'
hatvalues(model, ...)
## S3 method for class 'cca'
rstandard(model, type = c("response", "canoco"), ...)
## S3 method for class 'cca'
rstudent(model, type = c("response", "canoco"), ...)
## S3 method for class 'cca'
cooks.distance(model, type = c("response", "canoco"), ...)

## S3 method for class 'cca'
sigma(object, type = c("response", "canoco"), ...)
## S3 method for class 'cca'
vcov(object, type = "canoco", ...)
## S3 method for class 'cca'
SSD(object, type = "canoco", ...)

## S3 method for class 'cca'
qr(x, ...)
## S3 method for class 'cca'
```

### `influence.cca`

**Linear Model Diagnostics for Constrained Ordination**

**Description**

This set of function extracts influence statistics and some other linear model statistics directly from a constrained ordination result object from `cca`, `rda`, `capscale` or `dbrda`. The constraints are linear model functions and these support functions return identical results as the corresponding linear models (`lm`), and you can use their documentation. The main functions for normal usage are leverage values (`hatvalues`), standardized residuals (`rstandard`), studentized or leave-one-out residuals (`rstudent`), and Cook's distance (`cooks.distance`). In addition, `vcov` returns the variance-covariance matrix of coefficients, and its diagonal values the variances of coefficients. Other functions are mainly support functions for these, but they can be used directly.

### Usage

```r
## S3 method for class 'cca'
hatvalues(model, ...)
## S3 method for class 'cca'
rstandard(model, type = c("response", "canoco"), ...)
## S3 method for class 'cca'
rstudent(model, type = c("response", "canoco"), ...)
## S3 method for class 'cca'
cooks.distance(model, type = c("response", "canoco"), ...)

## S3 method for class 'cca'
sigma(object, type = c("response", "canoco"), ...)
## S3 method for class 'cca'
vcov(object, type = "canoco", ...)
## S3 method for class 'cca'
SSD(object, type = "canoco", ...)

## S3 method for class 'cca'
qr(x, ...)
## S3 method for class 'cca'
```
Arguments

model, object, x
A constrained ordination result object.

type
Type of statistics used for extracting raw residuals and residual standard deviation (sigma). Either "response" for species data or difference of WA and LC scores for "canoco".

... Other arguments to functions (ignored).

Details

The vegan algorithm for constrained ordination uses linear model (or weighted linear model in cca) to find the fitted values of dependent community data, and constrained ordination is based on this fitted response (Legendre & Legendre 2012). The hatvalues give the leverage values of these constraints, and the leverage is independent on the response data. Other influence statistics (rstandard, rstudent, cooks.distance) are based on leverage, and on the raw residuals and residual standard deviation (sigma). With type = "response" the raw residuals are given by the unconstrained component of the constrained ordination, and influence statistics are a matrix with dimensions no. of observations times no. of species. For cca the statistics are the same as obtained from the lm model using Chi-square standardized species data (see decostand) as dependent variable, and row sums of community data as weights, and for rda the lm model uses non-modified community data and no weights.

The algorithm in the CANOCO software constraints the results during iteration by performing a linear regression of weighted averages (WA) scores on constraints and taking the fitted values of this regression as linear combination (LC) scores (ter Braak 1984). The WA scores are directly found from species scores, but LC scores are linear combinations of constraints in the regression. With type = "canoco" the raw residuals are the differences of WA and LC scores, and the residual standard deviation (sigma) is taken to be the axis sum of squared WA scores minus one. These quantities have no relationship to residual component of ordination, but they rather are methodological artefacts of an algorithm that is not used in vegan. The result is a matrix with dimensions no. of observations times no. of constrained axes.

Function vcov returns the matrix of variances and covariances of regression coefficients. The diagonal values of this matrix are the variances, and their square roots give the standard errors of regression coefficients. The function is based on SSD that extracts the sum of squares and crossproducts of residuals. The residuals are defined similarly as in influence measures and with each type they have similar properties and limitations, and define the dimensions of the result matrix.

Note

Function as.mlm casts an ordination object to a multiple linear model of class "mlm" (see lm), and similar statistics can be derived from that modified object as with this set of functions. However, there are some problems in the R implementation of the further analysis of multiple linear model objects. When the results differ, the current set of functions is more probable to be correct. The use of as.mlm objects should be avoided.

Author(s)

Jari Oksanen
References


See Also

Corresponding lm methods and as.mlm.cca. Function ordiresids provides lattice graphics for residuals.

Examples

data(varespec, varechem)
mod <- cca(varespec ~ Al + P + K, varechem)
## leverage
hatvalues(mod)
plot(hatvalues(mod), type = "h")
## ordination plot with leverages: points with high leverage have
## similar LC and WA scores
plot(mod, type = "n")
ordispidr(mod) # segment from LC to WA scores
points(mod, dis="si", cex=5*hatvalues(mod), pch=21, bg=2) # WA scores
text(mod, dis="bp", col=4)

# deviation and influence
head(rstandard(mod))
head(cooks.distance(mod))

## Influence measures from lm
y <- decostand(varespec, "chi.square") # needed in cca
y1 <- with(y, Cladstel) # take one species for lm
lmod1 <- lm(y1 ~ Al + P + K, varechem, weights = rowSums(varespec))
## numerically identical within 2e-15
range(cooks.distance(lmod1) - cooks.distance(mod)[, "Cladstel"])

## t-values of regression coefficients based on type = "canoco"
## residuals
coef(mod)
coef(mod)/sqrt(diag(vcov(mod, type = "canoco")))

Description

The function performs isometric feature mapping which consists of three simple steps: (1) retain only some of the shortest dissimilarities among objects, (2) estimate all dissimilarities as shortest path distances, and (3) perform metric scaling (Tenenbaum et al. 2000).
Usage

isomap(dist, ndim=10, ...)  
isomap(dist, epsilon, k, path = "shortest", fragmentedOK = FALSE, ...)  
## S3 method for class 'isomap'  
summary(object, axes = 40, ...)  
## S3 method for class 'isomap'  
plot(x, net = TRUE, n.col = "gray", type = "points", ...)  

Arguments

dist  Dissimilarities.  
dim  Number of axes in metric scaling (argument k in cmdscale).  
epsilon  Shortest dissimilarity retained.  
k  Number of shortest dissimilarities retained for a point. If both epsilon and k are given, epsilon will be used.  
path  Method used in stepacross to estimate the shortest path, with alternatives "shortest" and "extended".  
fragmentedOK  What to do if dissimilarity matrix is fragmented. If TRUE, analyse the largest connected group, otherwise stop with error.  
x, object  An isomap result object.  
axes  Number of axes displayed.  
net  Draw the net of retained dissimilarities.  
n.col  Colour of drawn net segments. This can also be a vector that is recycled for points, and the colour of the net segment is a mixture of joined points.  
type  Plot observations either as "points", "text" or use "none" to plot no observations. The "text" will use ordilabel if net = TRUE and ordiplot if net = FALSE, and pass extra arguments to these functions.  
...  Other parameters passed to functions.  

Details

The function isomap first calls function isomapdist for dissimilarity transformation, and then performs metric scaling for the result. All arguments to isomap are passed to isomapdist. The functions are separate so that the isomapdist transformation could be easily used with other functions than simple linear mapping of cmdscale.

Function isomapdist retains either dissimilarities equal or shorter to epsilon, or if epsilon is not given, at least k shortest dissimilarities for a point. Then a complete dissimilarity matrix is reconstructed using stepacross using either flexible shortest paths or extended dissimilarities (for details, see stepacross).

De’ath (1999) actually published essentially the same method before Tenenbaum et al. (2000), and De’ath’s function is available in function xdist in non-CRAN package mvpart. The differences are that isomap introduced the k criterion, whereas De’ath only used epsilon criterion. In practice, De’ath also retains higher proportion of dissimilarities than typical isomap.

The plot function uses internally ordiplot, except that it adds text over net using ordilabel. The plot function passes extra arguments to these functions. In addition, vegan3d package has function rgl.isomap to make dynamic 3D plots that can be rotated on the screen.
isomap

Value

Function isomapdist returns a dissimilarity object similar to dist. Function isomap returns an object of class isomap with plot and summary methods. The plot function returns invisibly an object of class ordiplot. Function scores can extract the ordination scores.

Note

Tenenbaum et al. (2000) justify isomap as a tool of unfolding a manifold (e.g. a ‘Swiss Roll’). Even with a manifold structure, the sampling must be even and dense so that dissimilarities along a manifold are shorter than across the folds. If data do not have such a manifold structure, the results are very sensitive to parameter values.

Author(s)

Jari Oksanen

References


See Also

The underlying functions that do the proper work are stepacross, distconnected and cmdscale. Function metaMDS may trigger stepacross transformation, but usually only for longest dissimilarities. The plot method of vegan minimum spanning tree function (spantree) has even more extreme way of isomapping things.

Examples

```r
## The following examples also overlay minimum spanning tree to
## the graphics in red.
op <- par(mar=c(4,4,1,1)+0.2, mfrow=c(2,2))
data(BCI)
dis <- vegdist(BCI)
tr <- spantree(dis)
pl <- ordiplot(cmdscale(dis), main="cmdscale")
lines(tr, pl, col="red")
ord <- isomap(dis, k=3)
ord
pl <- plot(ord, main="isomap k=3")
lines(tr, pl, col="red")
pl <- plot(isomap(dis, k=5), main="isomap k=5")
lines(tr, pl, col="red")
pl <- plot(isomap(dis, epsilon=0.45), main="isomap epsilon=0.45")
lines(tr, pl, col="red")
par(op)
## colour points and web by the dominant species
```
dom <- apply(BCI, 1, which.max)
## need nine colours, but default palette has only eight
op <- palette(c(palette("default"), "sienna"))
plot(ord, pch = 16, col = dom, n.col = dom)
palette(op)

kendall.global  Kendall coefficient of concordance

Description

Function kendall.global computes and tests the coefficient of concordance among several judges (variables, species) through a permutation test.

Function kendall.post carries out \textit{a posteriori} tests of the contributions of individual judges (variables, species) to the overall concordance of their group through permutation tests.

If several groups of judges are identified in the data table, coefficients of concordance (kendall.global) or a posteriori tests (kendall.post) will be computed for each group separately. Use in ecology: to identify significant species associations.

Usage

kendall.global(Y, group, nperm = 999, mult = "holm")
kendall.post(Y, group, nperm = 999, mult = "holm")

Arguments

- **Y**: Data file (data frame or matrix) containing quantitative or semiquantitative data. Rows are objects and columns are judges (variables). In community ecology, that table is often a site-by-species table.
- **group**: A vector defining how judges should be divided into groups. See example below. If groups are not explicitly defined, all judges in the data file will be considered as forming a single group.
- **nperm**: Number of permutations to be performed. Default is 999.
- **mult**: Correct P-values for multiple testing using the alternatives described in \texttt{p.adjust} and in addition "sidak" (see Details). The Bonferroni correction is overly conservative; it is not recommended. It is included to allow comparisons with the other methods.

Details

- **Y** must contain quantitative data. They will be transformed to ranks within each column before computation of the coefficient of concordance.
- The search for species associations described in Legendre (2005) proceeds in 3 steps:
  - (1) Correlation analysis of the species. A possible method is to compute Ward’s agglomerative clustering of a matrix of correlations among the species. In detail: (1.1) compute a Pearson or
Spearman correlation matrix (correl.matrix) among the species: (1.2) turn it into a distance matrix: mat.D = as.dist(1-correl.matrix); (1.3) carry out Ward’s hierarchical clustering of that matrix using hclust: clust.ward = hclust(mat.D, "ward"); (1.4) plot the dendrogram: plot(clust.ward, hang=-1); (1.5) cut the dendrogram in two groups, retrieve the vector of species membership: group.2 = cutree(clust.ward, k=2). (1.6) After steps 2 and 3 below, you may have to come back and try divisions of the species into k = 3, 4, 5, ... groups.

(2) Compute global tests of significance of the 2 (or more) groups using the function kendall.global and the vector defining the groups. Groups that are not globally significant must be refined or abandoned.

(3) Compute a posteriori tests of the contribution of individual species to the concordance of their group using the function kendall.post and the vector defining the groups. If some species have negative values for "Spearman.mean", this means that these species clearly do not belong to the group, hence that group is too inclusive. Go back to (1.5) and cut the dendrogram more finely. The left and right groups can be cut separately, independently of the levels along the dendrogram; write your own vector of group membership if cutree does not produce the desired groups.

The corrections used for multiple testing are applied to the list of P-values (P); they take into account the number of tests (k) carried out simultaneously (number of groups in kendall.global, or number of species in kendall.post). The corrections are performed using function p.adjust; see that function for the description of the correction methods. In addition, there is Šidák correction which defined as $P_{corr} = 1 - (1 - P)^k$.

**Value**

A table containing the following information in rows. The columns correspond to the groups of "judges" defined in vector "group". When function Kendall.post is used, there are as many tables as the number of predefined groups.

- **W**: Kendall’s coefficient of concordance, W.
- **F**: F statistic. $F = W^*(m-1)/(1-W)$ where m is the number of judges.
- **Prob.F**: Probability associated with the F statistic, computed from the F distribution with $nu1 = n-1-(2/m)$ and $nu2 = nu1*(m-1)$; n is the number of objects.
- **Corrected prob.F**: Probabilities associated with F, corrected using the method selected in parameter mult. Shown only if there are more than one group.
- **Chi2**: Friedman’s chi-square statistic (Friedman 1937) used in the permutation test of W.
- **Prob.perm**: Permutational probabilities, uncorrected.
- **Corrected prob.perm**: Permutational probabilities corrected using the method selected in parameter mult. Shown only if there are more than one group.
- **Spearman.mean**: Mean of the Spearman correlations between the judge under test and all the other judges in the same group.
- **W.per.species**: Contribution of the judge under test to the overall concordance statistic for that group.
Author(s)
F. Guillaume Blanchet, University of Alberta, and Pierre Legendre, Université de Montréal

References


See Also
cor, friedman.test, hclust, cutree, kmeans, cascadeKM, indval

Examples

data(mite)
mite.hel <- decostand(mite, "hel")

data(mite)
mite.hel <- decostand(mite, "hel")

# Reproduce the results shown in Table 2 of Legendre (2005), a single group
mite.small <- mite.hel[c(4,9,14,22,31,34,45,53,61,69),c(13:15,23)]
kendall.global(mite.small, nperm=49)
kendall.post(mite.small, mult="holm", nperm=49)

# Reproduce the results shown in Tables 3 and 4 of Legendre (2005), 2 groups
group <- c(1,1,2,1,1,1,2,1,1,1,2,2,2,2,2,2,1,1,1,2,2,2,2,2,2,2,2,2)
kendall.global(mite.hel, group=group, nperm=49)
kendall.post(mite.hel, group=group, mult="holm", nperm=49)

# NOTE: 'nperm' argument usually needs to be larger than 49.
# It was set to this low value for demonstration purposes.

Description
Function linestack plots vertical one-dimensional plots for numeric vectors. The plots are always labelled, but the labels are moved vertically to avoid overwriting.
Usage

linestack(x, labels, cex = 0.8, side = "right", hoff = 2, air = 1.1,
at = 0, add = FALSE, axis = FALSE, ...)

Arguments

- **x**: Numeric vector to be plotted.
- **labels**: Labels used instead of default (names of x). May be expressions to be drawn with `plotmath`.
- **cex**: Size of the labels.
- **side**: Put labels to the "right" or "left" of the axis.
- **hoff**: Distance from the vertical axis to the label in units of the width of letter “m”.
- **air**: Multiplier to string height to leave empty space between labels.
- **at**: Position of plot in horizontal axis.
- **add**: Add to an existing plot.
- **axis**: Add axis to the plot.
- **...**: Other graphical parameters to labels.

Value

The function returns invisibly the shifted positions of labels in user coordinates.

Note

The function always draws labelled diagrams. If you want to have unlabelled diagrams, you can use, e.g., `plot`, `stripchart` or `rug`.

Author(s)

Jari Oksanen with modifications by Gavin L. Simpson

Examples

```r
## First DCA axis
data(dune)
ord <- decorana(dune)
linestack(scores(ord, choices=1, display="sp")
linestack(scores(ord, choices=1, display="si"), side="left", add=TRUE)
title(main="DCA axis 1")

## Expressions as labels
N <- 10  # Number of sites
df <- data.frame(Ca = rlnorm(N, 2), NO3 = rlnorm(N, 4), 
                 SO4 = rlnorm(N, 10), K = rlnorm(N, 3))
ord <- rda(df, scale = TRUE)
### vector of expressions for labels
labs <- expression(Ca^2+phantom()),
```
make.cepnames

Abbreviates a Botanical or Zoological Latin Name into an Eight-character Name

Description

A standard CEP name has four first letters of the generic name and four first letters of the specific epithet of a Latin name. The last epithet, that may be a subspecific name, is used in the current function. If the name has only one component, it is abbreviated to eight characters (see `abbreviate`). The returned names are made unique with function `make.unique` which adds numbers to the end of CEP names if needed.

Usage

`make.cepnames(names, seconditem = FALSE)`

Arguments

- `names` The names to be formatted into CEP names.
- `seconditem` Take always the second item of the original name to the abbreviated name instead of the last original item (default).

Details

Cornell Ecology Programs (CEP) used eight-letter abbreviations for species and site names. In species, the names were formed by taking four first letters of the generic name and four first letters of the specific or subspecific epithet. The current function first makes valid R names using `make.names`, and then splits these into elements. The CEP name is made by taking the four first letters of the first element, and four first letters of the last (default) or the second element (with `seconditem = TRUE`). If there was only one name element, it is abbreviated to eight letters. Finally, the names are made unique which may add numbers to duplicated names.

The CEP names were originally used, because old FORTRAN IV did not have CHARACTER data type, but text had to be stored in numerical variables, which in popular computers could hold four characters. In modern times, there is no reason for this limitation, but ecologists are used to these names, and they may be practical to avoid congestion in ordination plots.

Value

Function returns CEP names.
Note

The function is simpleminded and rigid. You must write a better one if you need.

Author(s)

Jari Oksanen

See Also

make.names, strsplit, substring, paste, abbreviate.

Examples

```r
make.cepnames(c("Aa maderoi", "Poa sp.", "Cladina rangiferina", "Cladonia cornuta", "Cladonia cornuta var. groenlandica", "Cladonia rangiformis", "Bryoerythrophyllum"))
data(BCI)
colnames(BCI) <- make.cepnames(colnames(BCI))
```

```
mantel(xdis, ydis, method="pearson", permutations=999, strata = NULL, na.rm = FALSE, parallel = getOption("mc.cores"))
mantel.partial(xdis, ydis, zdis, method = "pearson", permutations = 999, strata = NULL, na.rm = FALSE, parallel = getOption("mc.cores"))
```

Arguments

- **xdis, ydis, zdis**
  Dissimilarity matrices or a dist objects.
- **method**
  Correlation method, as accepted by `cor`: "pearson", "spearman" or "kendall".
- **permutations**
  a list of control values for the permutations as returned by the function `how`, or the number of permutations required, or a permutation matrix where each row gives the permuted indices.
- **strata**
  An integer vector or factor specifying the strata for permutation. If supplied, observations are permuted only within the specified strata.
**na.rm** Remove missing values in calculation of Mantel correlation. Use this option with care: Permutation tests can be biased, in particular if two matrices had missing values in matching positions.

**parallel** Number of parallel processes or a predefined socket cluster. With `parallel = 1` uses ordinary, non-parallel processing. The parallel processing is done with `parallel` package.

### Details

Mantel statistic is simply a correlation between entries of two dissimilarity matrices (some use cross products, but these are linearly related). However, the significance cannot be directly assessed, because there are $N(N - 1)/2$ entries for just $N$ observations. Mantel developed asymptotic test, but here we use permutations of $N$ rows and columns of dissimilarity matrix. See `permutations` for additional details on permutation tests in Vegan.

Partial Mantel statistic uses partial correlation conditioned on the third matrix. Only the first matrix is permuted so that the correlation structure between second and first matrices is kept constant. Although `mantel.partial` silently accepts other methods than "pearson", partial correlations will probably be wrong with other methods.

The function uses `cor`, which should accept alternatives `pearson` for product moment correlations and `spearman` or `kendall` for rank correlations.

### Value

The function returns a list of class `mantel` with following components:

- **Call** Function call.
- **method** Correlation method used, as returned by `cor.test`.
- **statistic** The Mantel statistic.
- **signif** Empirical significance level from permutations.
- **perm** A vector of permuted values. The distribution of permuted values can be inspected with `permustats` function.
- **permutations** Number of permutations.
- **control** A list of control values for the permutations as returned by the function `how`.

### Note

Legendre & Legendre (2012, Box 10.4) warn against using partial Mantel correlations.

### Author(s)

Jari Oksanen

### References

The test is due to Mantel, of course, but the current implementation is based on Legendre and Legendre.

See Also
cor for correlation coefficients, protest (Procrustes test) for an alternative with ordination
diagrams, anosim and mrpp for comparing dissimilarities against classification. For dissimilarity
matrices, see vegdist or dist. See bioenv for selecting environmental variables.

Examples

```r
# Is vegetation related to environment?
data(varespec)
data(varechem)

veg.dist <- vegdist(varespec)  # Bray-Curtis
env.dist <- vegdist(scale(varechem), "euclid")
mantel(veg.dist, env.dist)
mantel(veg.dist, env.dist, method="spear")
```

---

mantel.correlog  Mantel Correlogram

Description

Function mantel.correlog computes a multivariate Mantel correlogram. Proposed by Sokal (1986)
and Oden and Sokal (1986), the method is also described in Legendre and Legendre (2012, pp. 819–
821).

Usage

```r
mantel.correlog(D.eco, D.geo=NULL, XY=NULL, n.class=0, break.pts=NULL,
cutoff=TRUE, r.type="pearson", n.perm=999, mult="holm", progressive=TRUE)
```

Arguments

- **D.eco**  An ecological distance matrix, with class either dist or matrix.
- **D.geo**  A geographic distance matrix, with class either dist or matrix. Provide either
  D.geo or XY. Default: D.geo=NULL.
- **XY**  A file of Cartesian geographic coordinates of the points. Default: XY=NULL.
- **n.class**  Number of classes. If n.class=0, the Sturges equation will be used unless break
  points are provided.
- **break.pts**  Vector containing the break points of the distance distribution. Provide (n.class+1)
  breakpoints, that is, a list with a beginning and an ending point. Default: break.pts=NULL.
- **cutoff**  For the second half of the distance classes, cutoff = TRUE limits the correlogram to the distance classes that include all points. If cutoff = FALSE, the correlogram includes all distance classes.
mantel.correlog

r.type

Type of correlation in calculation of the Mantel statistic. Default: r.type="pearson". Other choices are r.type="spearman" and r.type="kendall", as in functions cor and mantel.

nperm

Number of permutations for the tests of significance. Default: nperm=999. For large data files, permutation tests are rather slow.

mult

Correct P-values for multiple testing. The correction methods are "holm" (default), "hochberg", "sidak", and other methods available in the p.adjust function: "bonferroni" (best known, but not recommended because it is overly conservative), "hommel", "BH", "BY", "fdr", and "none".

progressive

Default: progressive=TRUE for progressive correction of multiple-testing, as described in Legendre and Legendre (1998, p. 721). Test of the first distance class: no correction; second distance class: correct for 2 simultaneous tests; distance class k: correct for k simultaneous tests. progressive=FALSE: correct all tests for n.class simultaneous tests.

x

Output of mantel.correlog.

alpha

Significance level for the points drawn with black symbols in the correlogram. Default: alpha=0.05.

... Other parameters passed from other functions.

Details

A correlogram is a graph in which spatial correlation values are plotted, on the ordinate, as a function of the geographic distance classes among the study sites along the abscissa. In a Mantel correlogram, a Mantel correlation (Mantel 1967) is computed between a multivariate (e.g. multi-species) distance matrix of the user's choice and a design matrix representing each of the geographic distance classes in turn. The Mantel statistic is tested through a permutational Mantel test performed by vegan's mantel function.

When a correction for multiple testing is applied, more permutations are necessary than in the no-correction case, to obtain significant p-values in the higher correlogram classes.

The print.mantel.correlog function prints out the correlogram. See examples.

Value

mantel.res A table with the distance classes as rows and the class indices, number of distances per class, Mantel statistics (computed using Pearson's r, Spearman's r, or Kendall's tau), and p-values as columns. A positive Mantel statistic indicates positive spatial correlation. An additional column with p-values corrected for multiple testing is added unless mult="none".

n.class The number of distance classes.

break.pts The break points provided by the user or computed by the program.

mult The name of the correction for multiple testing. No correction: mult="none".

progressive A logical (TRUE, FALSE) value indicating whether or not a progressive correction for multiple testing was requested.

n.tests The number of distance classes for which Mantel tests have been computed and tested for significance.

call The function call.
Author(s)

Pierre Legendre, Université de Montréal

References


Examples

```r
# Mite data available in "vegan"
data(mite)
data(mite.xy)
mite.hel <- decostand(mite, "hellinger")

# Detrend the species data by regression on the site coordinates
mite.hel.resid <- resid(lm(as.matrix(mite.hel) ~ ., data=mite.xy))

# Compute the detrended species distance matrix
mite.hel.D <- dist(mite.hel.resid)

# Compute Mantel correlogram with cutoff, Pearson statistic
mite.correlog <- mantel.correlog(mite.hel.D, XY=mite.xy, nperm=49)
summary(mite.correlog)
mite.correlog
# or: print(mite.correlog)
# or: print.mantel.correlog(mite.correlog)
plot(mite.correlog)

# Compute Mantel correlogram without cutoff, Spearman statistic
mite.correlog2 <- mantel.correlog(mite.hel.D, XY=mite.xy, cutoff=FALSE, r.type="spearman", nperm=49)
summary(mite.correlog2)
mite.correlog2
plot(mite.correlog2)

# NOTE: 'nperm' argument usually needs to be larger than 49.
# It was set to this low value for demonstration purposes.
```
MDSrotate

Rotate First MDS Dimension Parallel to an External Variable

Description

Function rotates a multidimensional scaling result so that its first dimension is parallel to an external (environmental variable). The function can handle the results from metaMDS or monoMDS functions.

Usage

MDSrotate(object, vec, na.rm = FALSE, ...)

Arguments

- **object**: A result object from metaMDS or monoMDS.
- **vec**: An environmental variable or a matrix of such variables. The number of variables must be lower than the number of dimensions, and the solution is rotated to these variables in the order they appear in the matrix. Alternatively vec can be a factor, and the solution is rotated to optimal separation of factor levels using lda.
- **na.rm**: Remove missing values from the continuous variable vec.
- **...**: Other arguments (ignored).

Details

The orientation and rotation are undefined in multidimensional scaling. Functions metaMDS and metaMDS can rotate their solutions to principal components so that the dispersion of the points is highest on the first dimension. Sometimes a different rotation is more intuitive, and MDSrotate allows rotation of the result so that the first axis is parallel to a given external variable or two first variables are completely in a two-dimensional plane etc. If several external variables are supplied, they are applied in the order they are in the matrix. First axis is rotated to the first supplied variable, and the second axis to the second variable. Because variables are usually correlated, the second variable is not usually aligned with the second axis, but it is uncorrelated to later dimensions. There must be at least one free dimension: the number of external variables must be lower than the number of dimensions, and all used environmental variables are uncorrelated with that free dimension.

Alternatively the method can rotate to discriminate the levels of a factor using linear discriminant analysis (lda). This is hardly meaningful for two-dimensional solutions, since all rotations in two dimensions have the same separation of cluster levels. However, the function can be useful in finding a two-dimensional projection of clusters from more than two dimensions. The last dimension will always show the residual variation, and for k dimensions, only k − 1 discrimination vectors are used.

Value

Function returns the original ordination result, but with rotated scores (both site and species if available), and the pc attribute of scores set to FALSE.
**Note**

Rotation to a factor variable is an experimental feature and may be removed. The discriminant analysis weights dimensions by their discriminating power, but `MDSrotate` performs a rigid rotation. Therefore the solution may not be optimal.

**Author(s)**

Jari Oksanen

**See Also**

`metaMDS`, `monoMDS`.

**Examples**

```r
data(varespec)
data(varechem)
mod <- monoMDS(vegdist(varespec))
mod <- with(varechem, MDSrotate(mod, pH))
plot(mod)
ef <- envfit(mod ~ pH, varechem, permutations = 0)
plot(eff)
ordisurf(mod ~ pH, varechem, knots = 1, add = TRUE)
```

---

**Description**

Function `metaMDS` performs Nonmetric Multidimensional Scaling (NMDS), and tries to find a stable solution using several random starts. In addition, it standardizes the scaling in the result, so that the configurations are easier to interpret, and adds species scores to the site ordination. The `metaMDS` function does not provide actual NMDS, but it calls another function for the purpose. Currently `monoMDS` is the default choice, and it is also possible to call the `isoMDS` (`MASS` package).

**Usage**

```r
metaMDS(comm, distance = "bray", k = 2, try = 20, trymax = 20,
   engine = c("monoMDS", "isoMDS"), autotransform =TRUE,
   noshare = (engine == "isoMDS"), wascores = TRUE, expand = TRUE,
   trace = 1, plot = FALSE, previous.best, ...)
```

```r
# S3 method for class 'metaMDS'
plot(x, display = c("sites", "species"), choices = c(1, 2),
    type = "p", shrink = FALSE, ...)
# S3 method for class 'metaMDS'
points(x, display = c("sites", "species"),
    choices = c(1,2), shrink = FALSE, select, ...)
```
## metaMDS

S3 method for class 'metaMDS'

```r
text(x, display = c("sites", "species"), labels,
   choices = c(1,2), shrink = FALSE, select, ...)
```

S3 method for class 'metaMDS'

```r
scores(x, display = c("sites", "species"), shrink = FALSE,
   choices, ...)
```

```r
metaMDSdist(comm, distance = "bray", autotransform = TRUE,
   noshare = TRUE, trace = 1, commname, zerodist = "ignore",
   distfun = vegdist, ...)
```

```r
metaMDSiter(dist, k = 2, try = 20, trymax = 20, trace = 1, plot = FALSE,
   previous.best, engine = "monoMDS", maxit = 200,
   parallel = getOption("mc.cores"), ...)
```

```r
initMDS(x, k=2)
```

```r
postMDS(X, dist, pc=TRUE, center=TRUE, halfchange, threshold=0.8,
   nthreshold=10, plot=FALSE, ...)
```

```r
metaMDSredist(object, ...)
```

### Arguments

- **comm**
  - Community data. Alternatively, dissimilarities either as a `dist` structure or as a symmetric square matrix. In the latter case all other stages are skipped except random starts and centring and pc rotation of axes.

- **distance**
  - Dissimilarity index used in `vegdist`.

- **k**
  - Number of dimensions. NB., the number of points $n$ should be $n > 2k + 1$, and preferably higher in non-metric MDS.

- **try, trymax**
  - Minimum and maximum numbers of random starts in search of stable solution. After try has been reached, the iteration will stop when two convergent solutions were found or trymax was reached.

- **engine**
  - The function used for MDS. The default is to use the `monoMDS` function in `vegan`, but for backward compatibility it is also possible to use `isoMDS` of `MASS`.

- **autotransform**
  - Use simple heuristics for possible data transformation of typical community data (see below). If you do not have community data, you should probably set autotransform = FALSE.

- **noshare**
  - Triggering of calculation step-across or extended dissimilarities with function `stepacross`. The argument can be logical or a numerical value greater than zero and less than one. If TRUE, extended dissimilarities are used always when there are no shared species between some sites, if FALSE, they are never used. If noshare is a numerical value, stepacross is used when the proportion of site pairs with no shared species exceeds noshare. The number of pairs with no shared species is found with `noNshared` function, and noshare has no effect if input data were dissimilarities instead of community data.

- **wascores**
  - Calculate species scores using function `wascores`.

- **expand**
  - Expand weighted averages of species in `wascores`.

- **trace**
  - Trace the function; trace = 2 or higher will be more voluminous.

- **plot**
  - Graphical tracing: plot interim results. You may want to set `par(ask = TRUE)` with this option.
metaMDS

previous.best  Start searches from a previous solution.

x          metaMDS result (or a dissimilarity structure for initMDS).

choices      Axes shown.

type         Plot type: "p" for points, "t" for text, and "n" for axes only.

display      Display "sites" or "species".

shrink       Shrink back species scores if they were expanded originally.

labels       Optional test to be used instead of row names.

select       Items to be displayed. This can either be a logical vector which is TRUE for
                  displayed items or a vector of indices of displayed items.

X           Configuration from multidimensional scaling.

commname     The name of comm: should not be given if the function is called directly.

zerodist     Handling of zero dissimilarities: either "fail" or "add" a small positive value,
                  or "ignore". monomDS accepts zero dissimilarities and the default is zerodist = "ignore",
                  but with isomDS you may need to set zerodist = "add".

distfun      Dissimilarity function. Any function returning a dist object and accepting argu-
                  ment method can be used (but some extra arguments may cause name conflicts).

maxit        Maximum number of iterations in the single NMDS run; passed to the engine
                  function monomDS or isomDS.

parallel     Number of parallel processes or a predefined socket cluster. If you use pre-
                  defined socket clusters (say, clus), you must issue clusterEvalQ(clus, library(vegan))
                  to make available internal vegan functions. With parallel = 1 uses ordinary, non-parallel
                  processing. The parallel processing is done with parallel package.

dist         Dissimilarity matrix used in multidimensional scaling.

pc           Rotate to principal components.

center       Centre the configuration.

halfchange   Scale axes to half-change units. This defaults TRUE when dissimilarities were
                  evaluated within metaMDS and the dissimilarity index has an upper limit of 1. If
                  FALSE, the ordination dissimilarities are scaled to the same range as the input
                  dissimilarities.

threshold    Largest dissimilarity used in half-change scaling.

nthreshold   Minimum number of points in half-change scaling.

object       A result object from metaMDS.

...          Other parameters passed to functions. Function metaMDS passes all arguments to
                  its component functions metaMDSdist, metaMDSiter, postMDS, and to distfun
                  and engine.

Details

Non-metric Multidimensional Scaling (NMDS) is commonly regarded as the most robust uncon-
strained ordination method in community ecology (Minchin 1987). Function metaMDS is a wrapper
function that calls several other functions to combine Minchin’s (1987) recommendations into one
command. The complete steps in metaMDS are:
1. **Transformation:** If the data values are larger than common abundance class scales, the function performs a Wisconsin double standardization (*wisconsin*). If the values look very large, the function also performs *sqrt* transformation. Both of these standardizations are generally found to improve the results. However, the limits are completely arbitrary (at present, data maximum 50 triggers *sqrt* and > 9 triggers *wisconsin*). If you want to have a full control of the analysis, you should set `autotransform = FALSE` and standardize and transform data independently. The `autotransform` is intended for community data, and for other data types, you should set `autotransform = FALSE`. This step is performed using `metaMDSdist`.

2. **Choice of dissimilarity:** For a good result, you should use dissimilarity indices that have a good rank order relation to ordering sites along gradients (Faith et al. 1987). The default is Bray-Curtis dissimilarity, because it often is the test winner. However, any other dissimilarity index in `vegdist` can be used. Function `rankindex` can be used for finding the test winner for your data and gradients. The default choice may be bad if you analyse other than community data, and you should probably select an appropriate index using argument `distance`. This step is performed using `metaMDSdist`.

3. **Step-across dissimilarities:** Ordination may be very difficult if a large proportion of sites have no shared species. In this case, the results may be improved with `stepacross` dissimilarities, or flexible shortest paths among all sites. The default NMDS engine is `monoMDS` which is able to break tied values at the maximum dissimilarity, and this option is sufficient to handle cases with no shared species, and therefore the default is not to use `stepacross` with `monoMDS`. Function `isoMDS` does not handle tied values adequately, and therefore the default is to use `stepacross` always when there are sites with no shared species with engine = “isoMDS”. The `stepacross` is triggered by option `noshare`. If you do not like manipulation of original distances, you should set `noshare = FALSE`. This step is skipped if input data were dissimilarities instead of community data. This step is performed using `metaMDSdist`.

4. **NMDS with random starts:** NMDS easily gets trapped into local optima, and you must start NMDS several times from random starts to be confident that you have found the global solution. The strategy in `metaMDS` is to first run NMDS starting with the metric scaling (`cmdscale` which usually finds a good solution but often close to a local optimum), or use the previous best solution if supplied, and take its solution as the standard (Run 0). Then `metaMDS` starts NMDS from several random starts (minimum number is given by `try` and maximum number by `trymax`). These random starts are generated by `initMDS`. If a solution is better (has a lower stress) than the previous standard, it is taken as the new standard. If the solution is better or close to a standard, `metaMDS` compares two solutions using Procrustes analysis (function `procrustes` with option `symmetric = TRUE`). If the solutions are very similar in their Procrustes rmse and the largest residual is very small, the solutions are regarded as convergent and the better one is taken as the new standard. The conditions are stringent, and you may have found good and relatively stable solutions although the function is not yet satisfied. Setting `trace = TRUE` will monitor the final stresses, and `plot = TRUE` will display Procrustes overlay plots from each comparison. This step is performed using `metaMDSiter`. This is the only step performed if input data (comm) were dissimilarities.

5. **Scaling of the results:** `metaMDS` will run `postMDS` for the final result. Function `postMDS` provides the following ways of “fixing” the indeterminacy of scaling and orientation of axes in NMDS: Centring moves the origin to the average of the axes; Principal components rotate the configuration so that the variance of points is maximized on first dimension (with function `MDSrotate` you can alternatively rotate the configuration so that the first axis is parallel to an environmental variable); Half-change scaling scales the configuration so that one unit means halving of community similarity from replicate similarity. Half-change scaling is based on
closer dissimilarities where the relation between ordination distance and community dissimilarity is rather linear (the limit is set by argument `threshold`). If there are enough points below this threshold (controlled by the parameter `nthreshold`), dissimilarities are regressed on distances. The intercept of this regression is taken as the replicate dissimilarity, and half-change is the distance where similarity halves according to linear regression. Obviously the method is applicable only for dissimilarity indices scaled to 0...1, such as Kulczynski, Bray-Curtis and Canberra indices. If half-change scaling is not used, the ordination is scaled to the same range as the original dissimilarities.

6. Species scores: Function adds the species scores to the final solution as weighted averages using function `wascores` with given value of parameter `expand`. The expansion of weighted averages can be undone with `shrink = TRUE` in `plot` or `scores` functions, and the calculation of species scores can be suppressed with `wascores = FALSE`.

Value

Function `metaMDS` returns an object of class `metaMDS`. The final site ordination is stored in the item `points`, and species ordination in the item `species`, and the stress in item `stress` (NB, the scaling of the stress depends on the engine: `isomDS` uses percents, and `monomDS` proportions in the range 0...1). The other items store the information on the steps taken and the items returned by the engine function. The object has `print`, `plot`, `points` and `text` methods. Functions `metaMDSdist` and `metaMDSredist` return `vegdist` objects. Function `initMDS` returns a random configuration which is intended to be used within `isomDS` only. Functions `metaMDSiter` and `postMDS` returns the result of NMDS with updated configuration.

Convergence Problems

The function tries hard to find two convergent solutions, but it may fail. With default engine = "`monoMDS`" the function will tabulate the stopping criteria used, so that you can see which criterion should be made more stringent. The criteria can be given as arguments to `metaMDS` and their current values are described in `monoMDS`. In particular, if you reach the maximum number of iterations, you should increase the value of `maxit`. You may ask for a larger number of random starts without losing the old ones giving the previous solution in argument `previous.best`.

In addition to too slack convergence criteria and too low number of random starts, wrong number of dimensions (argument `k`) is the most common reason for not finding convergent solutions. NMDS is usually run with a low number dimensions (k=2 or k=3), and for complex data increasing `k` by one may help. If you run NMDS with much higher number of dimensions (say, k=10 or more), you should reconsider what you are doing and drastically reduce `k`. For very heterogeneous data sets with partial disjunctions, it may help to set `stepacross`, but for most data sets the default `weakties = TRUE` is sufficient.

Please note that you can give all arguments of other `metaMDS*` functions and NMDS engine (default `monoMDS`) in your `metaMDS` command, and you should check documentation of these functions for details.

Warning

`metaMDS` uses `monoMDS` as its NMDS engine from `vegan` version 2.0-0, when it replaced the `isoMDS` function. You can set argument `engine` to select the old engine.
Note

Function `metaMDS` is a simple wrapper for an NMDS engine (either `monoMDS` or `isoMDS`) and some support functions (`metaMDSdist`, `stepacross`, `metaMDSiter`, `initMDS`, `postMDS`, `wascores`). You can call these support functions separately for better control of results. Data transformation, dissimilarities and possible `stepacross` are made in function `metaMDSdist` which returns a dissimilarity result. Iterative search (with starting values from `initMDS` with `monoMDS`) is made in `metaMDSiter`. Processing of result configuration is done in `postMDS`, and species scores added by `wascores`. If you want to be more certain of reaching a global solution, you can compare results from several independent runs. You can also continue analysis from previous results or from your own configuration. Function may not save the used dissimilarity matrix (`monoMDS` does), but `metaMDSredist` tries to reconstruct the used dissimilarities with original data transformation and possible `stepacross`.

The `metaMDS` function was designed to be used with community data. If you have other type of data, you should probably set some arguments to non-default values: probably at least `wascores`, `autotransform` and `noshare` should be FALSE. If you have negative data entries, `metaMDS` will set the previous to FALSE with a warning.

Author(s)

Jari Oksanen

References


See Also

`monoMDS` (and `isoMDS`), `decostand`, `wisconsin`, `vegdist`, `rankindex`, `stepacross`, `procrustes`, `wascores`, `MDSrotate`, `ordiplot`.

Examples

```
## The recommended way of running NMDS (Minchin 1987)
##
data(dune)
# Global NMDS using monoMDS
sol <- metaMDS(dune)
sol
plot(sol, type="t")
## Start from previous best solution
sol <- metaMDS(dune, previous.best = sol)
## Local NMDS and stress 2 of monoMDS
sol2 <- metaMDS(dune, model = "local", stress=2)
sol2
## Use Arrhenius exponent 'z' as a binary dissimilarity measure
sol <- metaMDS(dune, distfun = betadiver, distance = "z")
sol
```
Oribatid Mite Data with Explanatory Variables

Description
Oribatid mite data. 70 soil cores collected by Daniel Borcard in 1989. See Borcard et al. (1992, 1994) for details.

Usage
```r
data(mite)
data(mite.env)
data(mite.pcnm)
data(mite.xy)
```

Format
There are three linked data sets: mite that contains the data on 35 species of Oribatid mites, mite.env that contains environmental data in the same sampling sites, mite.xy that contains geographic coordinates, and mite.pcnm that contains 22 PCNM base functions (columns) computed from the geographic coordinates of the 70 sampling sites (Borcard & Legendre 2002). The whole sampling area was 2.5 m x 10 m in size.

The fields in the environmental data are:

- **SubsDens**: Substrate density (g/L)
- **WatrCont**: Water content of the substrate (g/L)
- **Substrate**: Substrate type, factor with levels Sphagn1, Sphagn2, Sphagn3, Sphagn Litter, Barepeat, Interface
- **Shrub**: Shrub density, an ordered factor with levels 1 < 2 < 3
- **Topo**: Microtopography, a factor with levels Blanket and Hummock

Source
Pierre Legendre

References

Examples
```r
data(mite)
```
Description

Function implements Kruskal’s (1964a,b) non-metric multidimensional scaling (NMDS) using monotone regression and primary (“weak”) treatment of ties. In addition to traditional global NMDS, the function implements local NMDS, linear and hybrid multidimensional scaling.

Usage

```r
monoMDS(dist, y, k = 2, model = c("global", "local", "linear", "hybrid"),
threshold = 0.8, maxit = 200, weakties = TRUE, stress = 1,
scaling = TRUE, pc = TRUE, smin = 1e-4, sfgrmin = 1e-7,
sr=0.099999, ...)
```

```
## S3 method for class 'monoMDS'
scores(x, choices = NA, ...)
## S3 method for class 'monoMDS'
plot(x, choices = c(1,2), type = "t", ...)
## S3 method for class 'monoMDS'
points(x, choices = c(1,2), select, ...)
## S3 method for class 'monoMDS'
text(x, labels, choices = c(1,2), select, ...)
```

Arguments

dist Input dissimilarities.
y Starting configuration. A random configuration will be generated if this is missing.
k Number of dimensions. NB., the number of points n should be n > 2k + 1, and preferably higher in non-metric MDS.
model MDS model: "global" is normal non-metric MDS with a monotone regression, "local" is non-metric MDS with separate regressions for each point, "linear" uses linear regression, and "hybrid" uses linear regression for dissimilarities below a threshold in addition to monotone regression. See Details.
threshold Dissimilarity below which linear regression is used alternately with monotone regression.
maxit Maximum number of iterations.
weakties Use primary or weak tie treatment, where equal observed dissimilarities are allowed to have different fitted values. if FALSE, then secondary (strong) tie treatment is used, and tied values are not broken.
stress Use stress type 1 or 2 (see Details).
scaling Scale final scores to unit root mean squares.
monoMDS

pc
Rotate final scores to principal components.
smin, sfgrmin, sratmax
Convergence criteria: iterations stop when stress drops below smin, scale factor of the gradient drops below sfgrmin, or stress ratio between two iterations goes over sratmax (but is still < 1).
x
A monoMDS result.
choices
Dimensions returned or plotted. The default NA returns all dimensions.
type
The type of the plot: "t" for text, "p" for points, and "n" for none.
select
Items to be displayed. This can either be a logical vector which is TRUE for displayed items or a vector of indices of displayed items.
labels
Labels to be used used instead of row names.

Details
There are several versions of non-metric multidimensional scaling in R, but monoMDS offers the following unique combination of features:

- "Weak" treatment of ties (Kruskal 1964a,b), where tied dissimilarities can be broken in monotone regression. This is especially important for cases where compared sites share no species and dissimilarities are tied to their maximum value of one. Breaking ties allows these points to be at different distances and can help in recovering very long coenoclines (gradients). Function smacofSym (smacof package) also has adequate tie treatment.
- Handles missing values in a meaningful way.
- Offers "local" and "hybrid" scaling in addition to usual "global" NMDS (see below).
- Uses fast compiled code (isomds of the MASS package also uses compiled code).

Function monoMDS uses Kruskal’s (1964b) original monotone regression to minimize the stress. There are two alternatives of stress: Kruskal’s (1964a,b) original or “stress 1” and an alternative version or “stress 2” (Sibson 1972). Both of these stresses can be expressed with a general formula

\[ s^2 = \frac{\sum(d - \hat{d})^2}{\sum(d - d_0)^2} \]

where \(d\) are distances among points in ordination configuration, \(\hat{d}\) are the fitted ordination distances, and \(d_0\) are the ordination distances under null model. For “stress 1” \(d_0 = 0\), and for “stress 2” \(d_0 = \bar{d}\) or mean distances. “Stress 2” can be expressed as \(s^2 = 1 - R^2\), where \(R^2\) is squared correlation between fitted values and ordination distances, and so related to the “linear fit” of stressplot.

Function monoMDS can fit several alternative NMDS variants that can be selected with argument model. The default model = "global" fits global NMDS, or Kruskal’s (1964a,b) original NMDS similar to isoMDS (MASS) or smacofSym (smacof). Alternative model = "local" fits local NMDS where independent monotone regression is used for each point (Sibson 1972). Alternative model = "linear" fits a linear MDS. This fits a linear regression instead of monotone, and is not identical to metric scaling or principal coordinates analysis (cmdscale) that performs an eigenvector decomposition of dissimilarities (Gower 1966). Alternative model = "hybrid" implements
hybrid MDS that uses monotone regression for all points and linear regression for dissimilarities below or at a threshold dissimilarity in alternating steps (Faith et al. 1987). Function stressplot can be used to display the kind of regression in each model.

Scaling, orientation and direction of the axes is arbitrary. However, the function always centres the axes, and the default scaling is to scale the configuration of unit root mean square and to rotate the axes (argument pc) to principal components so that the first dimension shows the major variation. It is possible to rotate the solution so that the first axis is parallel to a given environmental variable using function MDSrotate.

Value

Returns an object of class "monoMDS". The final scores are returned in item points (function scores extracts these results), and the stress in item stress. In addition, there is a large number of other items (but these may change without notice in the future releases).

Convergence Criteria

NMDS is iterative, and the function stops when any of its convergence criteria is met. There is actually no criterion of assured convergence, and any solution can be a local optimum. You should compare several random starts (or use monoMDS via metaMDS) to assess if the solutions is likely a global optimum.

The stopping criteria are:

maxit: Maximum number of iterations. Reaching this criterion means that solutions was almost certainly not found, and maxit should be increased.

smin: Minimum stress. If stress is nearly zero, the fit is almost perfect. Usually this means that data set is too small for the requested analysis, and there may be several different solutions that are almost as perfect. You should reduce the number of dimensions (k), get more data (more observations) or use some other method, such as metric scaling (cmdscale, wcmdscale).

s ratmax: Change in stress. Values close to one mean almost unchanged stress. This may mean a solution, but it can also signal stranding on suboptimal solution with flat stress surface.

sfgrmin: Minimum scale factor. Values close to zero mean almost unchanged configuration. This may mean a solution, but will also happen in local optima.

Note

This is the default NMDS function used in metaMDS. Function metaMDS adds support functions so that NMDS can be run like recommended by Minchin (1987).

Author(s)

Peter R. Michin (Fortran core) and Jari Oksanen (R interface).

References


See Also

`metaMDS` for the `vegan` way of running NMDS, and `isoMDS` and `smacofSym` for some alternative implementations of NMDS.

Examples

```r
data(dune)
dis <- vegdist(dune)
m <- monoMDS(dis, model = "loc")
m
plot(m)
```

**Description**

Mitchell-Olds & Shaw test concerns the location of the highest (hump) or lowest (pit) value of a quadratic curve at given points. Typically, it is used to study whether the quadratic hump or pit is located within a studied interval. The current test is generalized so that it applies generalized linear models (`glm`) with link function instead of simple quadratic curve. The test was popularized in ecology for the analysis of humped species richness patterns (Mittelbach et al. 2001), but it is more general. With logarithmic link function, the quadratic response defines the Gaussian response model of ecological gradients (ter Braak & Looman 1986), and the test can be used for inspecting the location of Gaussian optimum within a given range of the gradient. It can also be used to replace Tokeshi’s test of “bimodal” species frequency distribution.

**Usage**

```r
MOStest(x, y, interval, ...)
## S3 method for class 'MOStest'
plot(x, which = c(1,2,3,6), ...)
fieller.MOStest(object, level = 0.95)
## S3 method for class 'MOStest'
```
profile(fitted, alpha = 0.01, maxsteps = 10, del = zmax/5, ...)
## S3 method for class 'MOStest'
confint(object, parm = 1, level = 0.95, ...)

Arguments

- **x**  The independent variable or plotting object in plot.
- **y**  The dependent variable.
- **interval**  The two points at which the test statistic is evaluated. If missing, the extremes of x are used.
- **which**  Subset of plots produced. Values which=1 and 2 define plots specific to MOStest (see Details), and larger values select graphs of plot.lm (minus 2).
- **object, fitted**  A result object from MOStest.
- **level**  The confidence level required.
- **alpha**  Maximum significance level allowed.
- **maxsteps**  Maximum number of steps in the profile.
- **del**  A step length parameter for the profile (see code).
- **parm**  Ignored.
- **...**  Other variables passed to functions. Function MOStest passes these to glm so that these can include family. The other functions pass these to underlying graphical functions.

Details

The function fits a quadratic curve $\mu = b_0 + b_1 x + b_2 x^2$ with given family and link function. If $b_2 < 0$, this defines a unimodal curve with highest point at $u = -b_1/(2b_2)$ (ter Braak & Looman 1986). If $b_2 > 0$, the parabola has a minimum at $u$ and the response is sometimes called "bimodal". The null hypothesis is that the extreme point $u$ is located within the interval given by points $p_1$ and $p_2$. If the extreme point $u$ is exactly at $p_1$, then $b_1 = 0$ on shifted axis $x - p_1$. In the test, origin of $x$ is shifted to the values $p_1$ and $p_2$, and the test statistic is based on the differences of deviances between the original model and model where the origin is forced to the given location using the standard anova.glm function (Oksanen et al. 2001). Mitchell-Olds & Shaw (1987) used the first degree coefficient with its significance as estimated by the summary.glm function. This give identical results with Normal error, but for other error distributions it is preferable to use the test based on differences in deviances in fitted models.

The test is often presented as a general test for the location of the hump, but it really is dependent on the quadratic fitted curve. If the hump is of different form than quadratic, the test may be insignificant.

Because of strong assumptions in the test, you should use the support functions to inspect the fit. Function plot(..., which=1) displays the data points, fitted quadratic model, and its approximate 95% confidence intervals (2 times SE). Function plot with which = 2 displays the approximate confidence interval of the polynomial coefficients, together with two lines indicating the combinations of the coefficients that produce the evaluated points of $x$. Moreover, the cross-hair shows the approximate confidence intervals for the polynomial coefficients ignoring their correlations. Higher values of which produce corresponding graphs from plot.lm. That is, you must add 2 to the value of which in plot.lm.
Function `fieller.MOSTest` approximates the confidence limits of the location of the extreme point (hump or pit) using Fieller's theorem following ter Braak & Looman (1986). The test is based on quasideviance except if the `family` is poisson or binomial. Function `profile` evaluates the profile deviance of the fitted model, and `confint` finds the profile based confidence limits following Oksanen et al. (2001).

The test is typically used in assessing the significance of diversity hump against productivity gradient (Mittelbach et al. 2001). It also can be used for the location of the pit (deepest points) instead of the Tokeshi test. Further, it can be used to test the location of the the Gaussian optimum in ecological gradient analysis (ter Braak & Looman 1986, Oksanen et al. 2001).

**Value**

The function is based on `glm`, and it returns the result of object of `glm` amended with the result of the test. The new items in the `MOSTest` are:

- `ishump` TRUE if the response is a hump.
- `isBracketed` TRUE if the hump or the pit is bracketed by the evaluated points.
- `hump` Sorted vector of location of the hump or the pit and the points where the test was evaluated.
- `coefficients` Table of test statistics and their significances.

**Note**

Function `fieller.MOSTest` is based on package `optgrad` in the Ecological Archives (https://figshare.com/articles/Full_Archive/3521975) accompanying Oksanen et al. (2001). The Ecological Archive package `optgrad` also contains profile deviance method for the location of the hump or pit, but the current implementation of `profile` and `confint` rather follow the example of `profile.glm` and `confint.glm` in the `MASS` package.

**Author(s)**

Jari Oksanen

**References**


**See Also**

The no-interaction model can be fitted with `humpfit`.
Examples

```r
## The Al-Mufti data analysed in humpfit():
mass <- c(140, 230, 310, 310, 400, 510, 610, 670, 860, 900, 1050, 1160, 1900, 2480)
spno <- c(1, 4, 3, 9, 18, 30, 20, 14, 3, 2, 3, 2, 5, 2)
mod <- MOSTest(mass, spno)
## Insignificant
mod
## ... but inadequate shape of the curve
op <- par(mfrow=c(2,2), mar=c(4,4,1,1)+1)
plot(mod)
## Looks rather like log-link with Poisson error and logarithmic biomass
mod <- MOSTest(log(mass), spno, family=quasipoisson)
mod
plot(mod)
par(op)
## Confidence Limits
fieller.MOSTest(mod)
confint(mod)
plot(profile(mod))
```

mrpp

Multi Response Permutation Procedure and Mean Dissimilarity Matrix

Description

Multiple Response Permutation Procedure (MRPP) provides a test of whether there is a significant difference between two or more groups of sampling units. Function `meandist` finds the mean within and between block dissimilarities.

Usage

```r
mrpp(dat, grouping, permutations = 999, distance = "euclidean",
     weight.type = 1, strata = NULL, parallel = getOption("mc.cores"))
meandist(dist, grouping, ...)
## S3 method for class 'meandist'
summary(object, ...)
## S3 method for class 'meandist'
plot(x, kind = c("dendrogram", "histogram"), cluster = "average",
      ylim, axes = TRUE, ...)
```

Arguments

dat data matrix or data frame in which rows are samples and columns are response variable(s), or a dissimilarity object or a symmetric square matrix of dissimilarities.
grouping Factor or numeric index for grouping observations.
permutations  a list of control values for the permutations as returned by the function how, or the number of permutations required, or a permutation matrix where each row gives the permuted indices. These are used to assess the significance of the MRPP statistic, delta.

distance  Choice of distance metric that measures the dissimilarity between two observations. See vegdist for options. This will be used if dat was not a dissimilarity structure of a symmetric square matrix.

weight.type  choice of group weights. See Details below for options.

strata  An integer vector or factor specifying the strata for permutation. If supplied, observations are permuted only within the specified strata.

parallel  Number of parallel processes or a predefined socket cluster. With parallel = 1 uses ordinary, non-parallel processing. The parallel processing is done with parallel package.

dist  A dist object of dissimilarities, such as produced by functions dist, vegdist or designdist.

object, x  A meandist result object.

kind  Draw a dendrogram or a histogram; see Details.

cluster  A clustering method for the hclust function for kind = "dendrogram". Any hclust method can be used, but perhaps only "average" and "single" make sense.

ylim  Limits for vertical axes (optional).

axes  Draw scale for the vertical axis.

...  Further arguments passed to functions.

Details

Multiple Response Permutation Procedure (MRPP) provides a test of whether there is a significant difference between two or more groups of sampling units. This difference may be one of location (differences in mean) or one of spread (differences in within-group distance; cf. Warton et al. 2012). Function mrpp operates on a data frame matrix where rows are observations and responses data matrix. The response(s) may be uni- or multivariate. The method is philosophically and mathematically allied with analysis of variance, in that it compares dissimilarities within and among groups. If two groups of sampling units are really different (e.g. in their species composition), then average of the within-group compositional dissimilarities ought to be less than the average of the dissimilarities between two random collection of sampling units drawn from the entire population.

The mrpp statistic δ is the overall weighted mean of within-group means of the pairwise dissimilarities among sampling units. The choice of group weights is currently not clear. The mrpp function offers three choices: (1) group size (n), (2) a degrees-of-freedom analogue (n – 1), and (3) a weight that is the number of unique distances calculated among n sampling units (n(n – 1)/2).

The mrpp algorithm first calculates all pairwise distances in the entire dataset, then calculates δ. It then permutes the sampling units and their associated pairwise distances, and recalculates δ based on the permuted data. It repeats the permutation step permutations times. The significance test is the fraction of permuted deltas that are less than the observed delta, with a small sample correction. The function also calculates the change-corrected within-group agreement A = 1 – δ/E(δ), where E(δ) is the expected δ assessed as the average of dissimilarities.
If the first argument dat can be interpreted as dissimilarities, they will be used directly. In other cases the function treats dat as observations, and uses \texttt{vegdist} to find the dissimilarities. The default distance is Euclidean as in the traditional use of the method, but other dissimilarities in \texttt{vegdist} also are available.

Function \texttt{meandist} calculates a matrix of mean within-cluster dissimilarities (diagonal) and between-cluster dissimilarities (off-diagonal elements), and an attribute \texttt{n} of grouping counts. Function \texttt{summary} finds the within-class, between-class and overall means of these dissimilarities, and the MRPP statistics with all \texttt{weightNtype} options and the Classification Strength, CS (Van Sickle and Hughes, 2000). CS is defined for dissimilarities as $\bar{B} - \bar{W}$, where $\bar{B}$ is the mean between cluster dissimilarity and $\bar{W}$ is the mean within cluster dissimilarity with \texttt{weightNtype = 1}. The function does not perform significance tests for these statistics, but you must use \texttt{mrpp} with appropriate \texttt{weightNtype}. There is currently no significance test for CS, but \texttt{mrpp} with \texttt{weightNtype = 1} gives the correct test for $\bar{W}$ and a good approximation for CS. Function \texttt{plot} draws a dendrogram or a histogram of the result matrix based on the within-group and between group dissimilarities. The dendrogram is found with the method given in the \texttt{cluster} argument using function \texttt{hclust}. The terminal segments hang to within-cluster dissimilarity. If some of the clusters are more heterogeneous than the combined class, the leaf segment are reversed. The histograms are based on dissimilarities, but are otherwise similar to those of Van Sickle and Hughes (2000): horizontal line is drawn at the level of mean between-cluster dissimilarity and vertical lines connect within-cluster dissimilarities to this line.

\textbf{Value}

The function returns a list of class mrpp with following items:

- \texttt{call} Function call.
- \texttt{delta} The overall weighted mean of group mean distances.
- \texttt{E.delta} expected delta, under the null hypothesis of no group structure. This is the mean of original dissimilarities.
- \texttt{CS} Classification strength (Van Sickle and Hughes, 2000). Currently not implemented and always \texttt{NA}.
- \texttt{n} Number of observations in each class.
- \texttt{classdelta} Mean dissimilarities within classes. The overall $\bar{\delta}$ is the weighted average of these values with given \texttt{weightNtype}.
- \texttt{pvalue} Significance of the test.
- \texttt{A} A chance-corrected estimate of the proportion of the distances explained by group identity; a value analogous to a coefficient of determination in a linear model.
- \texttt{distance} Choice of distance metric used; the "method" entry of the dist object.
- \texttt{weightNtype} The choice of group weights used.
- \texttt{boot.deltas} The vector of "permuted deltas," the deltas calculated from each of the permuted datasets. The distribution of this item can be inspected with \texttt{permustats} function.
- \texttt{permutations} The number of permutations used.
- \texttt{control} A list of control values for the permutations as returned by the function \texttt{how}.
Note

This difference may be one of location (differences in mean) or one of spread (differences in within-group distance). That is, it may find a significant difference between two groups simply because one of those groups has a greater dissimilarities among its sampling units. Most mrpp models can be analysed with adonis2 which seems not suffer from the same problems as mrpp and is a more robust alternative.

Author(s)

M. Henry H. Stevens <HStevens@muohio.edu> and Jari Oksanen.

References


See Also

anosim for a similar test based on ranks, and mantel for comparing dissimilarities against continuous variables, and vegdist for obtaining dissimilarities, adonis2 is a more robust alternative in most cases.

Examples

data(dune)
data(dune.env)
dune.mrpp <- with(dune.env, mrpp(dune, Management))
dune.mrpp

# Save and change plotting parameters
def.par <- par(no.readonly = TRUE)
layout(matrix(c(1:2, nr=1)))

plot(dune.ord <- metaMDS(dune), type="text", display="sites" )
with(dune.env, ordihull(dune.ord, Management))

with(dune.mrpp, {
    fig.dist <- hist(boot.deltas, xlim=range(c(delta.boot.deltas)),
    main="Test of Differences Among Groups")
    abline(v=delta);
    text(delta, 2*mean(fig.dist$counts), adj = -0.5,
    expression(bold(delta)), cex=1.5 )
})
par(def.par)
## meandist
dune.md <- with(dune.env, meandist(vegdist(dune), Management))
dune.md
summary(dune.md)
plot(dune.md)
plot(dune.md, kind="histogram")

### mso

**Functions for performing and displaying a spatial partitioning of cca or rda results**

#### Description

The function `mso` adds an attribute `vario` to an object of class "cca" that describes the spatial partitioning of the `cca` object and performs an optional permutation test for the spatial independence of residuals. The function `plot.mso` creates a diagnostic plot of the spatial partitioning of the "cca" object.

#### Usage

```r
mso(object.cca, object.xy, grain = 1, round.up = FALSE, permutations = 0)
msoplot(x, alpha = 0.05, explained = FALSE, ylim = NULL, legend = "topleft", ...)
```

#### Arguments

- `object.cca`: An object of class `cca`, created by the `cca` or `rda` function.
- `object.xy`: A vector, matrix or data frame with the spatial coordinates of the data represented by `object.cca`. The number of rows must match the number of observations (as given by `nobs`) in `cca.object`. Alternatively, interpoint distances can be supplied as a `dist` object.
- `grain`: Interval size for distance classes.
- `round.up`: Determines the choice of breaks. If false, distances are rounded to the nearest multiple of grain. If true, distances are rounded to the upper multiple of grain.
- `permutations`: a list of control values for the permutations as returned by the function `how`, or the number of permutations required, or a permutation matrix where each row gives the permuted indices.
- `x`: A result object of `mso`.
- `alpha`: Significance level for the two-sided permutation test of the Mantel statistic for spatial independence of residual inertia and for the point-wise envelope of the variogram of the total variance. A Bonferroni-type correction can be achieved by dividing the overall significance value (e.g. 0.05) by the number of distance classes.
- `explained`: If false, suppresses the plotting of the variogram of explained variance.
- `ylim`: Limits for y-axis.
legend

The x and y co-ordinates to be used to position the legend. They can be specified by keyword or in any way which is accepted by legend.

...

Other arguments passed to functions.

Details

The Mantel test is an adaptation of the function mantel of the vegan package to the parallel testing of several distance classes. It compares the mean inertia in each distance class to the pooled mean inertia of all other distance classes.

If there are explanatory variables (RDA, CCA, pRDA, pCCA) and a significance test for residual autocorrelation was performed when running the function mso, the function plot.mso will print an estimate of how much the autocorrelation (based on significant distance classes) causes the global error variance of the regression analysis to be underestimated.

Value

The function mso returns an amended cca or rda object with the additional attributes grain, H, H.test and vario.

grain

The grain attribute defines the interval size of the distance classes.

H

H is an object of class ’dist’ and contains the geographic distances between observations.

H.test

H.test contains a set of dummy variables that describe which pairs of observations (rows = elements of object $H$) fall in which distance class (columns).

vario

The vario attribute is a data frame that contains some or all of the following components for the rda case (cca case in brackets):

H Distance class as multiples of grain.

Dist Average distance of pairs of observations in distance class H.

n Number of unique pairs of observations in distance class H.

All Empirical (chi-square) variogram of total variance (inertia).

Sum Sum of empirical (chi-square) variograms of explained and residual variance (inertia).

CA Empirical (chi-square) variogram of residual variance (inertia).

CCA Empirical (chi-square) variogram of explained variance (inertia).

pCCA Empirical (chi-square) variogram of conditioned variance (inertia).

se Standard error of the empirical (chi-square) variogram of total variance (inertia).

CA.signif P-value of permutation test for spatial independence of residual variance (inertia).

Note

The function is based on the code published in the Ecological Archives E085-006 (doi: 10.1890/020738).
Author(s)

The responsible author was Helene Wagner.

References


See Also

Function `cca` and `rda`, `cca.object`.

Examples

```r
## Reconstruct worked example of Wagner (submitted):
X <- matrix(c(1, 2, 3, 2, 1, 0), 3, 2)
Y <- c(3, -1, -2)
tmat <- c(1:3)
## Canonical correspondence analysis (cca):
Example.cca <- cca(X, Y)
Example.cca <- mso(Example.cca, tmat)
msoplot(Example.cca)
Example.cca$vario

## Correspondence analysis (ca):
Example.ca <- mso(cca(X), tmat)
msoplot(Example.ca)

## Unconstrained ordination with test for autocorrelation
## using oribatid mite data set as in Wagner (2004)
data(mite)
data(mite.env)
data(mite.xy)

mite.cca <- cca(log(mite + 1))
mite.cca <- mso(mite.cca, mite.xy, grain = 1, permutations = 99)
msoplot(mite.cca)
mite.cca

## Constrained ordination with test for residual autocorrelation
## and scale-invariance of species-environment relationships
mite.cca <- cca(log(mite + 1) ~ Subsdens + WatrCont + Substrate + Shrub + Topo, mite.env)
mite.cca <- mso(mite.cca, mite.xy, permutations = 99)
msoplot(mite.cca)
mite.cca
```
### multipart

**Description**

In multiplicative diversity partitioning, mean values of alpha diversity at lower levels of a sampling hierarchy are compared to the total diversity in the entire data set or the pooled samples (gamma diversity).

**Usage**

```r
multipart(...)
```

### Default S3 method:

```r
multipart(y, x, index=c("renyi", "tsallis"), scales = 1,
    global = FALSE, relative = FALSE, nsimul=99, method = "r2table", ...)
```

### S3 method for class 'formula'

```r
multipart(formula, data, index=c("renyi", "tsallis"), scales = 1,
    global = FALSE, relative = FALSE, nsimul=99, method = "r2table", ...)
```

**Arguments**

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>y</code></td>
<td>A community matrix.</td>
</tr>
<tr>
<td><code>x</code></td>
<td>A matrix with same number of rows as in <code>y</code>, columns coding the levels of sampling hierarchy. The number of groups within the hierarchy must decrease from left to right. If <code>x</code> is missing, two levels are assumed: each row is a group in the first level, and all rows are in the same group in the second level.</td>
</tr>
<tr>
<td><code>formula</code></td>
<td>A two sided model formula in the form <code>y ~ x</code>, where <code>y</code> is the community data matrix with samples as rows and species as column. Right hand side (<code>x</code>) must be grouping variables referring to levels of sampling hierarchy, terms from right to left will be treated as nested (first column is the lowest, last is the highest level, at least two levels specified). Interaction terms are not allowed.</td>
</tr>
<tr>
<td><code>data</code></td>
<td>A data frame where to look for variables defined in the right hand side of <code>formula</code>. If missing, variables are looked in the global environment.</td>
</tr>
<tr>
<td><code>index</code></td>
<td>Character, the entropy index to be calculated (see Details).</td>
</tr>
<tr>
<td><code>relative</code></td>
<td>Logical, if TRUE then beta diversity is standardized by its maximum (see Details).</td>
</tr>
<tr>
<td><code>scales</code></td>
<td>Numeric, of length 1, the order of the generalized diversity index to be used.</td>
</tr>
<tr>
<td><code>global</code></td>
<td>Logical, indicates the calculation of beta diversity values, see Details.</td>
</tr>
<tr>
<td><code>nsimul</code></td>
<td>Number of permutations to use. If <code>nsimul = 0</code>, only the FUN argument is evaluated. It is thus possible to reuse the statistic values without a null model.</td>
</tr>
<tr>
<td><code>method</code></td>
<td>Null model method: either a name (character string) of a method defined in <code>make.commsim</code> or a <code>commsim</code> function. The default &quot;r2table&quot; keeps row sums and column sums fixed. See <code>oecosimu</code> for Details and Examples.</td>
</tr>
<tr>
<td><code>...</code></td>
<td>Other arguments passed to <code>oecosimu</code>, i.e. method, thin or burnin.</td>
</tr>
</tbody>
</table>
Details

Multiplicative diversity partitioning is based on Whittaker’s (1972) ideas, that has recently been
generalised to one parametric diversity families (i.e. Rényi and Tsallis) by Jost (2006, 2007). Jost
recommends to use the numbers equivalents (Hill numbers), instead of pure diversities, and proofs,
that this satisfies the multiplicative partitioning requirements.

The current implementation of multipart calculates Hill numbers based on the functions renyi
and tsallis (provided as index argument). If values for more than one scales are desired, it
should be done in separate runs, because it adds extra dimensionality to the implementation, which
has not been resolved efficiently.

Alpha diversities are then the averages of these Hill numbers for each hierarchy levels, the global
gamma diversity is the alpha value calculated for the highest hierarchy level. When global = TRUE,
beta is calculated relative to the global gamma value:

$$\beta_i = \gamma/\alpha_i$$

when global = FALSE, beta is calculated relative to local gamma values (local gamma means the
diversity calculated for a particular cluster based on the pooled abundance vector):

$$\beta_{ij} = \alpha_{(i+1)j}/\text{mean}(\alpha_{ij})$$

where $j$ is a particular cluster at hierarchy level $i$. Then beta diversity value for level $i$ is the mean
of the beta values of the clusters at that level, $\beta_i = \text{mean}(\beta_{ij})$.

If relative = TRUE, the respective beta diversity values are standardized by their maximum possible
values (mean($\beta_{ij}$)/$\beta_{max,ij}$) given as $\beta_{max,ij} = n_j$ (the number of lower level units in a given
cluster $j$).

The expected diversity components are calculated $\text{nsimul}$ times by individual based randomization
of the community data matrix. This is done by the "r2dtable" method in oecosimu by default.

Value

An object of class "multipart" with same structure as "oecosimu" objects.

Author(s)

Péter Sólymos, <solymos@ualberta.ca>

References

2427–2439.

See Also

See adipart for additive diversity partitioning, hiersimu for hierarchical null model testing and
oecosimu for permutation settings and calculating $p$-values.
Examples

## NOTE: 'nsimul' argument usually needs to be >= 99
## here much lower value is used for demonstration

data(mite)
data(mite.xy)
data(mite.env)

## Function to get equal area partitions of the mite data
cutter <- function(x, cut = seq(0, 10, by = 2.5)) {
  out <- rep(1, length(x))
  for (i in 2:(length(cut) - 1))
    out[which(x > cut[i] & x <= cut[(i + 1)])] <- i
  return(out)
}

## The hierarchy of sample aggregation
levsm <- with(mite.xy, data.frame(
  l1=nrow(mite),
  l2=cutter(y, cut = seq(0, 10, by = 2.5)),
  l3=cutter(y, cut = seq(0, 10, by = 5)),
  l4=cutter(y, cut = seq(0, 10, by = 10))))

## Multiplicative diversity partitioning
multipart(mite, levsm, index="renyi", scales=1, nsimul=19)
multipart(mite ~ ., levsm, index="renyi", scales=1, nsimul=19)
multipart(mite ~ ., levsm, index="renyi", scales=1, nsimul=19, relative=TRUE)
multipart(mite ~ ., levsm, index="renyi", scales=1, nsimul=19, global=TRUE)

Description

Patches or local communities are regarded as nested if they all could be subsets of the same community. In general, species poor communities should be subsets of species rich communities, and rare species should only occur in species rich communities.

Usage

nestedchecker(comm)
nestedn0(comm)
nesteddisc(comm, niter = 200)
nestedtemp(comm, ...)
nestednodf(comm, order = TRUE, weighted = FALSE, wbinary = FALSE)
nestedbetasor(comm)
nestedbetajac(comm)

## S3 method for class 'nestedtemp'
plot(x, kind = c("temperature", "incidence"),
  col=rev(heat.colors(100)), names = FALSE, ...)

## S3 method for class 'nestednodf'
plot(x, col = "red", names = FALSE, ...)
Arguments

- `comm`: Community data.
- `niter`: Number of iterations to reorder tied columns.
- `x`: Result object for a plot.
- `col`: Colour scheme for matrix temperatures.
- `kind`: The kind of plot produced.
- `names`: Label columns and rows in the plot using names in `comm`. If it is a logical vector of length 2, row and column labels are returned accordingly.
- `order`: Order rows and columns by frequencies.
- `weighted`: Use species abundances as weights of interactions.
- `wbinary`: Modify original method so that binary data give the same result in weighted and unweighted analysis.
- ... Other arguments to functions.

Details

The nestedness functions evaluate alternative indices of nestedness. The functions are intended to be used together with Null model communities and used as an argument in `oecosimu` to analyse the non-randomness of results.

Function `nestedchecker` gives the number of checkerboard units, or 2x2 submatrices where both species occur once but on different sites (Stone & Roberts 1990).

Function `nestednP` implements nestedness measure N0 which is the number of absences from the sites which are richer than the most pauperate site species occurs (Patterson & Atmar 1986).

Function `nesteddisc` implements discrepancy index which is the number of ones that should be shifted to fill a row with ones in a table arranged by species frequencies (Brualdi & Sanderson 1999). The original definition arranges species (columns) by their frequencies, but did not have any method of handling tied frequencies. The `nesteddisc` function tries to order tied columns to minimize the discrepancy statistic but this is rather slow, and with a large number of tied columns there is no guarantee that the best ordering was found (argument `niter` gives the maximum number of tried orders). In that case a warning of tied columns will be issued.

Function `nestedtemp` finds the matrix temperature which is defined as the sum of “surprises” in arranged matrix. In arranged unsurprising matrix all species within proportion given by matrix fill are in the upper left corner of the matrix, and the surprise of the absence or presences is the diagonal distance from the fill line (Atmar & Patterson 1993). Function tries to pack species and sites to a low temperature (Rodríguez-Gironés & Santamaria 2006), but this is an iterative procedure, and the temperatures usually vary among runs. Function `nestedtemp` also has a `plot` method which can display either incidences or temperatures of the surprises. Matrix temperature was rather vaguely described (Atmar & Patterson 1993), but Rodríguez-Gironés & Santamaria (2006) are more explicit and their description is used here. However, the results probably differ from other implementations, and users should be cautious in interpreting the results. The details of calculations are explained in the `vignette` `Design decisions and implementation` that you can read using functions `browseVignettes`. Function `nestedness` in the `bipartite` package is a direct port of the BINMATNEST programme of Rodríguez-Gironés & Santamaria (2006).
Function `nestednodf` implements a nestedness metric based on overlap and decreasing fill (Almeida-Neto et al., 2008). Two basic properties are required for a matrix to have the maximum degree of nestedness according to this metric: (1) complete overlap of 1’s from right to left columns and from down to up rows, and (2) decreasing marginal totals between all pairs of columns and all pairs of rows. The nestedness statistic is evaluated separately for columns (N columns) for rows (N rows) and combined for the whole matrix (NODF). If you set `order = FALSE`, the statistic is evaluated with the current matrix ordering allowing tests of other meaningful hypothesis of matrix structure than default ordering by row and column totals (breaking ties by total abundances when `weighted = TRUE`) (see Almeida-Neto et al. 2008). With `weighted = TRUE`, the function finds the weighted version of the index (Almeida-Neto & Ulrich, 2011). However, this requires quantitative null models for adequate testing. Almeida-Neto & Ulrich (2011) say that you have positive nestedness if values in the first row/column are higher than in the second. With this condition, weighted analysis of binary data will always give zero nestedness. With argument `wbinary = TRUE`, equality of rows/columns also indicates nestedness, and binary data will give identical results in weighted and unweighted analysis. However, this can also influence the results of weighted analysis so that the results may differ from Almeida-Neto & Ulrich (2011).

Functions `nestedbetasor` and `nestedbetajac` find multiple-site dissimilarities and decompose these into components of turnover and nestedness following Baselga (2012); the pairwise dissimilarities can be found with `designdist`. This can be seen as a decomposition of beta diversity (see `betadiver`). Function `nestedbetasor` uses Sørensen dissimilarity and the turnover component is Simpson dissimilarity (Baselga 2012), and `nestedbetajac` uses analogous methods with the Jaccard index. The functions return a vector of three items: turnover, nestedness and their sum which is the multiple Sørensen or Jaccard dissimilarity. The last one is the total beta diversity (Baselga 2012). The functions will treat data as presence/absence (binary) and they can be used with binary `nullmodel`. The overall dissimilarity is constant in all `nullmodels` that fix species (column) frequencies ("c0"), and all components are constant if row columns are also fixed (e.g., model "quasiswap"), and the functions are not meaningful with these null models.

Value

The result returned by a nestedness function contains an item called `statistic`, but the other components differ among functions. The functions are constructed so that they can be handled by `oecosimu`.

Author(s)

Jari Oksanen and Gustavo Carvalho (`nestednodf`).

References


**See Also**

In general, the functions should be used with `oecosimu` which generates Null model communities to assess the non-randomness of nestedness patterns.

**Examples**

```r
data(sipoo)
## Matrix temperature
out <- nestedtemp(sipoo)
out
plot(out)
plot(out, kind="incid")
## Use oecosimu to assess the non-randomness of checker board units
nestedchecker(sipoo)
oecosimu(sipoo, nestedchecker, "quasiswap")
## Another Null model and standardized checkerboard score
oecosimu(sipoo, nestedchecker, "r00", statistic = "C.score")
```

### nobs.adonis

**Extract the Number of Observations from a vegan Fit.**

**Description**

Extract the number of ‘observations’ from a vegan model fit.

**Usage**

```r
## S3 method for class 'adonis'
nobs(object, ...)
```

**Arguments**

- **object**: A fitted model object.
- **...**: Further arguments to be passed to methods.
Details

Function nobs is generic in R, and vegan provides methods for objects from adonis, betadisper, cca and other related methods, CCorA, decorana, isomap, metaMDS, pcnm, procrustes, radfit, varpart and wcmdscale.

Value

A single number, normally an integer, giving the number of observations.

Author(s)

Jari Oksanen

nullmodel

Null Model and Simulation

Description

The nullmodel function creates an object, which can serve as a basis for Null Model simulation via the simulate method. The update method updates the nullmodel object without sampling (effective for sequential algorithms). smbind binds together multiple simmat objects.

Usage

nullmodel(x, method)
## S3 method for class 'nullmodel'
print(x, ...)
## S3 method for class 'nullmodel'
simulate(object, nsim = 1, seed = NULL,
         burnin = 0, thin = 1, ...)
## S3 method for class 'nullmodel'
update(object, nsim = 1, seed = NULL, ...)
## S3 method for class 'simmat'
print(x, ...)
smbind(object, ..., MARGIN, strict = TRUE)

Arguments

x
A community matrix. For the print method, it is an object to be printed.

method
Character, specifying one of the null model algorithms listed on the help page of commsim. It can be a user supplied object of class commsim.

object
An object of class nullmodel returned by the function nullmodel. In case of smbind it is a simmat object as returned by the update or simulate methods.

nsim
Positive integer, the number of simulated matrices to return. For the update method, it is the number of burnin steps made for sequential algorithms to update the status of the input model object.
seed

An object specifying if and how the random number generator should be initialized ("seeded"). Either NULL or an integer that will be used in a call to set.seed before simulating the matrices. If set, the value is saved as the "seed" attribute of the returned value. The default, NULL will not change the random generator state, and return $\text{Random.seed}$ as the "seed" attribute, see Value.

burnin

Nonnegative integer, specifying the number of steps discarded before starting simulation. Active only for sequential null model algorithms. Ignored for non-sequential null model algorithms.

thin

Positive integer, number of simulation steps made between each returned matrix. Active only for sequential null model algorithms. Ignored for non-sequential null model algorithms.

MARGIN

Integer, indicating the dimension over which multiple simmat objects are to be bound together by smbind. $1$: matrices are stacked (row bound), $2$: matrices are column bound, $3$: iterations are combined. Needs to be of length $1$. The other dimensions are expected to match across the objects.

strict

Logical, if consistency of the time series attributes ("start", "end", "thin", and number of simulated matrices) of simmat objects are strictly enforced when binding multiple objects together using smbind. Applies only to input objects based on sequential null model algorithms.

Details

The purpose of the nullmodel function is to create an object, where all necessary statistics of the input matrix are calculated only once. This information is reused, but not recalculated in each step of the simulation process done by the simulate method.

The simulate method carries out the simulation, the simulated matrices are stored in an array. For sequential algorithms, the method updates the state of the input nullmodel object. Therefore, it is possible to do diagnostic tests on the returned simmat object, and make further simulations, or use increased thinning value if desired.

The update method makes burnin steps in case of sequential algorithms to update the status of the input model without any attempt to return matrices. For non-sequential algorithms the method does nothing.

update is the preferred way of making burnin iterations without sampling. Alternatively, burnin can be done via the simulate method. For convergence diagnostics, it is recommended to use the simulate method without burnin. The input nullmodel object is updated, so further samples can be simulated if desired without having to start the process all over again. See Examples.

The smbind function can be used to combine multiple simmat objects. This comes handy when null model simulations are stratified by sites (MARGIN = 1) or by species (MARGIN = 2), or in the case when multiple objects are returned by identical/consistent settings e.g. during parallel computations (MARGIN = 3). Sanity checks are made to ensure that combining multiple objects is sensible, but it is the user's responsibility to check independence of the simulated matrices and the null distribution has converged in case of sequential null model algorithms. The strict = FALSE setting can relax checks regarding start, end, and thinning values for sequential null models.
nullmodel

Value

The function `nullmodel` returns an object of class `nullmodel`. It is a set of objects sharing the same environment:

- `data`: original matrix in integer mode.
- `nrow`: number of rows.
- `ncol`: number of columns.
- `rowSums`: row sums.
- `colSums`: column sums.
- `rowFreq`: row frequencies (number of nonzero cells).
- `colFreq`: column frequencies (number of nonzero cells).
- `totalSum`: total sum.
- `fill`: number of nonzero cells in the matrix.
- `commsim`: the `commsim` object as a result of the `method` argument.
- `state`: current state of the permutations, a matrix similar to the original. It is NULL for non-sequential algorithms.
- `iter`: current number of iterations for sequential algorithms. It is NULL for non-sequential algorithms.

The `simulate` method returns an object of class `simmat`. It is an array of simulated matrices (third dimension corresponding to `nsim` argument).

The `update` method returns the current state (last updated matrix) invisibly, and update the input object for sequential algorithms. For non sequential algorithms, it returns NULL.

The `smbind` function returns an object of class `simmat`.

Note

Care must be taken when the input matrix only contains a single row or column. Such input is invalid for swapping and hypergeometric distribution (calling `r2dtable`) based algorithms. This also applies to cases when the input is stratified into subsets.

Author(s)

Jari Oksanen and Peter Solymos

See Also

`commsim`, `make.commsim`, `permatfull`, `permatswap`

Examples

```r
data(mite)
x <- as.matrix(mite)[1:12, 21:30]

## non-sequential nullmodel
(nm <- nullmodel(x, "r00"))
```
(sm <- simulate(nm, nsim=10))

## sequential nullmodel
(nm <- nullmodel(x, "swap"))
(sm1 <- simulate(nm, nsim=10, thin=5))
(sm2 <- simulate(nm, nsim=10, thin=5))

## sequential nullmodel with burnin and extra updating
(nm <- nullmodel(x, "swap"))
(sm1 <- simulate(nm, burnin=10, nsim=10, thin=5))
(sm2 <- simulate(nm, nsim=10, thin=5))

## sequential nullmodel with separate initial burnin
(nm <- nullmodel(x, "swap"))
nm <- update(nm, nsim=10)
(sm2 <- simulate(nm, nsim=10, thin=5))

## combining multiple simmat objects
## stratification
nm1 <- nullmodel(x[1:6,], "r00")
sm1 <- simulate(nm1, nsim=10)
nm2 <- nullmodel(x[7:12,], "r00")
sm2 <- simulate(nm2, nsim=10)

smbind(sm1, sm2, MARGIN=1)

## binding subsequent samples from sequential algorithms
## start, end, thin retained
nm <- nullmodel(x, "swap")
nm <- update(nm, nsim=10)
sm1 <- simulate(nm, nsim=10, thin=5)
sm2 <- simulate(nm, nsim=20, thin=5)
sm3 <- simulate(nm, nsim=10, thin=5)
smbind(sm3, sm2, sm1, MARGIN=3)

## 'replicate' based usage which is similar to the output
## of 'parLapply' or 'mclapply' in the 'parallel' package
## start, end, thin are set, also noting number of chains
smfun <- function(x, burnin, nsim, thin) {
  nm <- nullmodel(x, "swap")
  nm <- update(nm, nsim=burnin)
  simulate(nm, nsim=nsim, thin=thin)
}
smlist <- replicate(3, smfun(x, burnin=50, nsim=10, thin=5), simplify=FALSE)
smbind(smlist, MARGIN=3)  # Number of permuted matrices = 30

## Not run:
## parallel null model calculations
library(parallel)

if (.Platform$OS.type == "unix") {
## forking on Unix systems
smlist <- mclapply(1:3, function(i) smfun(x, burnin=50, nsim=10, thin=5))
## oecosimu

Evaluate Statistics with Null Models of Biological Communities

### Description

Function evaluates a statistic or a vector of statistics in community and evaluates its significance in a series of simulated random communities. The approach has been used traditionally for the analysis of nestedness, but the function is more general and can be used with any statistics evaluated with simulated communities. Function oecosimu collects and evaluates the statistics. The Null model communities are described in `makeNcommsim` and `permatfull/permatswap`, the definition of Null models in `nullmodel`, and nestedness statistics in `nestednodf` (which describes several alternative statistics, including nestedness temperature, $N_0$, checker board units, nestedness discrepancy and NODF).

### Usage

```r
oecosimu(comm, nestfun, method, nsimul = 99, burnin = 0, thin = 1,
  statistic = "statistic", alternative = c("two.sided", "less", "greater"),
  batchsize = NA, parallel = getOption("mc.cores"), ...)
```

### Arguments

- **comm**: Community data, or a Null model object generated by `nullmodel` or an object of class `simmat` (array of permuted matrices from `simulate.nullmodel`). If `comm` is a community data, null model simulation method must be specified. If `comm` is a `nullmodel`, the simulation method is ignored, and if `comm` is a `simmat` object, all other arguments are ignored except `nestfun`, statistic and alternative.

- **nestfun**: Function analysed. Some nestedness functions are provided in `vegan` (see `nestedtemp`), but any function can be used if it accepts the community as the first argument, and returns either a plain number or a vector or the result in list item with the name defined in argument statistic. See Examples for defining your own functions.
method Null model method: either a name (character string) of a method defined in make.commsim or a commsim function. This argument is ignored if comm is a nullmodel or a simmat object. See Details and Examples.

nsimul Number of simulated null communities (ignored if comm is a simmat object).

burnin Number of null communities discarded before proper analysis in sequential methods (such as "tswap") (ignored with non-sequential methods or when comm is a simmat object).

thin Number of discarded null communities between two evaluations of nestedness statistic in sequential methods (ignored with non-sequential methods or when comm is a simmat object).

statistic The name of the statistic returned by nestfun.

alternative a character string specifying the alternative hypothesis, must be one of "two.sided" (default), "greater" or "less". Please note that the p-value of two-sided test is approximately two times higher than in the corresponding one-sided test ("greater" or "less" depending on the sign of the difference).

batchsize Size in Megabytes of largest simulation object. If a larger structure would be produced, the analysis is broken internally into batches. With default NA the analysis is not broken into batches. See Details.

parallel Number of parallel processes or a predefined socket cluster. With parallel = 1 uses ordinary, non-parallel processing. The parallel processing is done with parallel package. If you define a nestfun in Windows that needs other R packages than vegan or permute, you must set up a socket cluster before the call.

x An oecosimu result object.

Details

Function oecosimu is a wrapper that evaluates a statistic using function given by nestfun, and then simulates a series of null models based on nullmodel, and evaluates the statistic on these null models. The vegan packages contains some nestedness functions that are described separately (nestedchecker, nesteddisc, nestedn0, nestedtemp, nestednodf), but many other functions can be used as long as they are meaningful with simulated communities. An applicable function must return either the statistic as a plain number or a vector, or as a list element "statistic" (like chisq.test), or in an item whose name is given in the argument statistic. The statistic can be a single number (like typical for a nestedness index), or it can be a vector. The vector indices can be used to analyse site (row) or species (column) properties, see treedive for an example. Raup-Crick index (raupcrick) gives an example of using a dissimilarities.

The Null model type can be given as a name (quoted character string) that is used to define a Null model in make.commsim. These include all binary models described by Wright et al. (1998), Jonsson (2001), Gotelli & Entsminger (2003), Miklós & Podani (2004), and some others. There are several quantitative Null models, such those discussed by Hardy (2008), and several that are unpublished (see make.commsim, permatfull, permatswap for discussion). The user can also define her own commsim function (see Examples).

Function works by first defining a nullmodel with given commsim, and then generating a series of simulated communities with simulate.nullmodel. A shortcut can be used for any of these stages and the input can be
1. Community data (comm), Null model function (nestfun) and the number of simulations (nsimul).
2. A nullmodel object and the number of simulations, and argument method is ignored.
3. A three-dimensional array of simulated communities generated with `simulate.nullmodel`, and arguments method and nsimul are ignored.

The last case allows analysing several statistics with the same simulations.

The function first generates simulations with given nullmodel and then analyses these using the nestfun. With large data sets and/or large number of simulations, the generated objects can be very large, and if the memory is exhausted, the analysis can become very slow and the system can become unresponsive. The simulation will be broken into several smaller batches if the simulated nullmodel objective will be above the set batchsize to avoid memory problems (see `object.size` for estimating the size of the current data set). The parallel processing still increases the memory needs. The parallel processing is only used for evaluating nestfun. The main load may be in simulation of the nullmodel, and parallel argument does not help there.

Function `as.ts` transforms the simulated results of sequential methods into a time series or a `ts` object. This allows using analytic tools for time series in studying the sequences (see examples). Function `as.mcmc` transforms the simulated results of sequential methods into an `mcmc` object of the `coda` package. The `coda` package provides functions for the analysis of stationarity, adequacy of sample size, autocorrelation, need of burn-in and much more for sequential methods, and summary of the results. Please consult the documentation of the `coda` package.

Function `permustats` provides support to the standard `density`, `densityplot`, `qqnorm` and `qqmath` functions for the simulated values.

Value

Function `oecosimu` returns an object of class "oecosimu". The result object has items statistic and oecosimu. The statistic contains the complete object returned by nestfun for the original data. The `oecosimu` component contains the following items:

- **statistic**: Observed values of the statistic.
- **simulated**: Simulated values of the statistic.
- **means**: Mean values of the statistic from simulations.
- **z**: Standardized effect sizes (SES, a.k.a. the z-values) of the observed statistic based on simulations.
- **pval**: The P-values of the statistic based on simulations.
- **alternative**: The type of testing as given in argument alternative.
- **method**: The method used in nullmodel.
- **isSeq**: TRUE if method was sequential.

Note

If you wonder about the name of `oecosimu`, look at journal names in the References (and more in `nestedtemp`).

The internal structure of the function was radically changed in `vegan 2.2-0` with introduction of `commsim` and `nullmodel` and depreciation of `commsimulator`. However, the results and the basic user interface remain the same (except that `method = "r8_old"` must be used to reproduce the old results of "method = r8").
Author(s)

Jari Oksanen and Peter Solymos

References


See Also

Function `oecosimu` currently defines null models with `commsim` and generates the simulated null model communities with `nullmodel` and `simulate.nullmodel`. For other applications of `oecosimu`, see `treedive` and `raupcrick`.

Function `rndtaxa` (labdsv package) randomizes a community table. See also `nestedtemp` (that also discusses other nestedness functions) and `treedive` for another application.

Examples

```r
## Use the first eigenvalue of correspondence analysis as an index
## of structure: a model for making your own functions.
data(sipoo)
## Traditional nestedness statistics (number of checkerboard units)
oecosimu(sipoo, nested checker, "r0")
## sequential model, one-sided test, a vector statistic
out <- oecosimu(sipoo, decorana, "swap", burnin=100, thin=10,
                statistic="evals", alt = "greater")
out
## Inspect the swap sequence as a time series object
plot(as.ts(out))
lag.plot(as.ts(out))
acf(as.ts(out))
## Density plot
densityplot(permustats(out), as.table = TRUE, layout = c(1,4))
## Use quantitative null models to compare
## mean Bray-Curtis dissimilarities
data(dune)
meandist <- function(x) mean(vegdist(x, "bray"))
mbc1 <- oecosimu(dune, meandist, "r2dtable")
mbc1
```
# ordiarrows

Add Arrows and Line Segments to Ordination Diagrams

## Description

Functions to add arrows, line segments, regular grids of points. The ordination diagrams can be produced by vegan `plot.cca`, `plot.decorana` or `ordiplot`.

## Usage

ordiarrows(ord, groups, levels, replicates, order.by, display = "sites",
           col = 1, show.groups, startmark, label = FALSE, length = 0.1, ...)
ordisegments(ord, groups, levels, replicates, order.by, display = "sites",
              col = 1, show.groups, label = FALSE, ...)
ordigrid(ord, levels, replicates, display = "sites", lty = c(1,1),
         col = c(1,1), lwd = c(1,1), ...)

## Arguments

- `ord` An ordination object or an ordiplot object.
- `groups` Factor giving the groups for which the graphical item is drawn.
levels, replicates
Alternatively, regular groups can be defined with arguments levels and replicates, where levels gives the number of groups, and replicates the number of successive items at the same group.

order.by
Order points by increasing order of this variable within groups. Reverse sign of the variable for decreasing ordering.

display
Item to displayed.

show.groups
Show only given groups. This can be a vector, or TRUE if you want to show items for which condition is TRUE. This argument makes it possible to use different colours and line types for groups. The default is to show all groups.

label
Label the groups by their names. In ordiellipse, ordihull and ordispider the the group name is in the centroid of the object, in ordiarrows in the start of the arrow, and in ordisegments at both ends. ordiellipse and ordihull use standard text, and others use ordilabel.

startmark
plotting character used to mark the first item. The default is to use no mark, and for instance, startmark = 1 will draw a circle. For other plotting characters, see pch in points.

col
Colour of lines, label borders and startmark in ordiarrows and ordisegments. This can be a vector recycled for groups. In ordigrid it can be a vector of length 2 used for levels and replicates.

length
Length of edges of the arrow head (in inches).

lty, lwd
Line type, line width used for levels and replicates in ordigrid.

...
Parameters passed to graphical functions such as lines, segments, arrows, or to scores to select axes and scaling etc.

Details
Function ordiarrows draws arrows and ordisegments draws line segments between successive items in the groups. Function ordigrid draws line segments both within the groups and for the corresponding items among the groups.

Note
These functions add graphical items to ordination graph: You must draw a graph first.

Author(s)
Jari Oksanen

See Also
The functions pass parameters to basic graphical functions, and you may wish to change the default values in arrows, lines and segments. You can pass parameters to scores as well.
Examples

```r
example(pyrifos)
mod <- rda(pyrifos)
plot(mod, type = "n")
## Annual succession by ditches, colour by dose
ordiarrows(mod, ditch, label = TRUE, col = as.numeric(dose))
legend("topright", levels(dose), lty=1, col=1:5, title="Dose")
## Show only control and highest Pyrifos treatment
plot(mod, type = "n")
ordiarrows(mod, ditch, label = TRUE, show = TRUE, show.groups = c("2", "3", "5", "11"))
ordiarrows(mod, ditch, label = TRUE, show = TRUE, show.groups = c("6", "9"),
            col = 2)
legend("topright", c("Control", "Pyrifos 44"), lty = 1, col = c(1,2))
```

Description

Support functions to assist with drawing of vectors (arrows) on ordination plots. `ordiArrowMul` finds the multiplier for the coordinates of the head of the vector such that they occupy fill proportion of the plot region. `ordiArrowTextXY` finds coordinates for the locations of labels to be drawn just beyond the head of the vector.

Usage

```r
ordiArrowTextXY(x, labels, display, choices = c(1,2),
                 rescale = TRUE, fill = 0.75, ...)
ordiArrowMul(x, at = c(0,0), fill = 0.75,
             display, choices = c(1,2), ...)```

Arguments

- **x**: An R object, from which `scores` can determine suitable ordination scores or an object created by `envfit`, or a two-column matrix of coordinates of arrow heads on the two plot axes.
- **labels**: Change plotting labels. A character vector of labels for which label coordinates are sought. If not supplied, these will be determined from the row names of x, or `scores(x, ...)` if required. If either of these are not defined, suitable labels will be generated.
- **display**: a character string known to `scores` or one of its methods which indicates the type of scores to extract. In fitting functions these are ordinary site scores or linear combination scores ("lc") in constrained ordination (`cca`, `rda`, `capscale`). If x was created by envfit then display can not be set by the user and takes the value "vectors". Ignored if x is a matrix.
- **choices**: Axes to be plotted.
`ordiArrowTextXY`  

`rescale`  logical: should the coordinates in or extracted from `x` be rescaled to fill `fill` proportion of the plot region? The default is to always rescale the coordinates as this is usually desired for objects `x` from which coordinates are retrieved. If supplying `x` a 2-column matrix that has already been rescaled, then set this to `FALSE`.

`fill` numeric: the proportion of the plot to fill by the span of the arrows.

`at` The origin of fitted arrows in the plot. If you plot arrows in other places than origin, you probably have to specify `arrow.mul`.

... Parameters passed to `scores` and `strwidth` and `strheight`.

**Details**

`ordiArrowMul` finds a multiplier to scale a bunch of arrows to fill an ordination plot, and `ordiArrowTextXY` finds the coordinates for labels of these arrows. NB., `ordiArrowTextXY` does not draw labels; it simply returns coordinates at which the labels should be drawn for use with another function, such as `text`.

**Value**

For `ordiArrowTextXY`, a 2-column matrix of coordinates for the label centres in the coordinate system of the currently active plotting device.

For `ordiArrowMul`, a length-1 vector containing the scaling factor.

**Author(s)**

Jari Oksanen, with modifications by Gavin L. Simpson

**Examples**

```r
## Scale arrows by hand to fill 80% of the plot
## Biplot arrows by hand
data(varespec, varechem)
ord <- cca(varespec ~ Al + P + K, varechem)
plot(ord, display = c("species","sites"))

## biplot scores
bip <- scores(ord, choices = 1:2, display = "bp")

## scaling factor for arrows to fill 80% of plot
(mul <- ordiArrowMul(bip, fill = 0.8))
bip.scl <- bip * mul          # Scale the biplot scores
labs <- rownames(bip)         # Arrow labels

## calculate coordinate of labels for arrows
(bip.lab <- ordiArrowTextXY(bip.scl, rescale = FALSE, labels = labs))

## draw arrows and text labels
arrows(0, 0, bip.scl[,1], bip.scl[,2], length = 0.1)
text(bip.lab, labels = labs)
```
## Handling of ordination objects directly

```r
mul2 <- ordiArrowMul(ord, display = "bp", fill = 0.8)
stopifnot(all.equal(mul, mul2))
```

### ordihull

#### Display Groups or Factor Levels in Ordination Diagrams

**Description**

Functions to add convex hulls, “spider” graphs, ellipses or cluster dendrogram to ordination diagrams. The ordination diagrams can be produced by vegan `plot.cca`, `plot.decorana` or `ordiplot`.

**Usage**

```r
ordihull(ord, groups, display = "sites", draw = c("lines", "polygon", "none"),
          col = NULL, alpha = 127, show.groups, label = FALSE,
          border = NULL, lty = NULL, lwd = NULL, ...)
ordiellipse(ord, groups, display="sites", kind = c("sd", "se", "ehull"),
             conf, draw = c("lines", "polygon", "none"),
             w = weights(ord, display), col = NULL, alpha = 127, show.groups,
             label = FALSE, border = NULL, lty = NULL, lwd = NULL, ...)
ordibar(ord, groups, display = "sites", kind = c("sd", "se"), conf,
        w = weights(ord, display), col = 1, show.groups, label = FALSE,
        lwd = NULL, length = 0, ...)
ordispider(ord, groups, display="sites", w = weights(ord, display),
           spiders = c("centroid", "median"), show.groups,
           label = FALSE, col = NULL, lty = NULL, lwd = NULL, ...)
ordicluster(ord, cluster, prune = 0, display = "sites",
             w = weights(ord, display), col = 1, draw = c("segments", "none"),
             ...)```

### S3 method for class 'ordihull'

```r
summary(object, ...)
```

### S3 method for class 'ordiellipse'

```r
summary(object, ...)
```

### ordiareatest(ord, groups, area = c("hull", "ellipse"), kind = "sd",
                 permutations = 999, parallel =getOption("mc.cores"), ...)

**Arguments**

- **ord**: An ordination object or an `ordiplot` object.
- **groups**: Factor giving the groups for which the graphical item is drawn.
- **display**: Item to be displayed.
- **draw**: character; how should objects be represented on the plot? For `ordihull` and `ordiellipse` use either `lines` or `polygon` to draw the lines. For `ordicluster`, line segments are drawn using `segments`. To suppress plotting, use "none". Graphical parameters are passed to both. The main difference is that polygons may be filled and non-transparent. With `none` nothing is drawn, but the function returns the `invisible` plotting.
col  Colour of hull or ellipse lines (if draw = "lines") or their fills (if draw = "polygon") in ordihull and ordiellipse. When draw = "polygon", the colour of bordering lines can be set with argument border of the polygon function. For other functions the effect depends on the underlining functions this argument is passed to. When multiple values of col are specified these are used for each element of names(table(groups)) (in that order), shorter vectors are recycled. Function ordicluster has no groups, and there the argument will be recycled for points, and the colour of connecting lines is a mixture of point s in the cluster.

alpha  Transparency of the fill colour with draw = "polygon" in ordihull and ordiellipse. The argument takes precedence over possible transparency definitions of the colour. The value must be in range 0...255, and low values are more transparent. Transparency is not available in all graphics devices or file formats.

show.groups  Show only given groups. This can be a vector, or TRUE if you want to show items for which condition is TRUE. This argument makes it possible to use different colours and line types for groups. The default is to show all groups.

label  Label the groups by their names in the centroid of the object. ordiellipse and ordihull use standard text, and others use ordilabel.

w  Weights used to find the average within group. Weights are used automatically for cca and decorana results, unless undone by the user. w=NULL sets equal weights to all points.

kind  Draw standard deviations of points (sd), standard errors (se) or ellipsoid hulls that enclose all points in the group (ehull).

conf  Confidence limit for ellipses, e.g. 0.95. If given, the corresponding sd or se is multiplied with the corresponding value found from the Chi-squared distribution with 2df.

spiders  Are centres or spider bodies calculated either as centroids (averages) or spatial medians.

cluster  Result of hierarchic cluster analysis, such as hclust or agnes.

prune  Number of upper level hierarchies removed from the dendrogram. If prune > 0, dendrogram will be disconnected.

object  A result object from ordihull or ordiellipse. The result is invisible, but it can be saved, and used for summaries (areas etc. of hulls and ellipses).

area  Evaluate the area of convex hulls of ordihull, or of ellipses of ordiellipse.

permutations  a list of control values for the permutations as returned by the function how, or the number of permutations required, or a permutation matrix where each row gives the permuted indices.

parallel  Number of parallel processes or a predefined socket cluster. With parallel = 1 uses ordinary, non-parallel processing. The parallel processing is done with parallel package.

lty, lwd, border

Vectors of these parameters can be supplied and will be applied (if appropriate) for each element of names(table(groups)) (in that order). Shorter vectors will be recycled.
ordihull

length

Width (in inches) of the small (“caps”) at the ends of the bar segment (passed to arrows).

... Parameters passed to graphical functions or to scores to select axes and scaling etc.

Details

Function ordihull draws lines or polygons for the convex hulls found by function chull encircling the items in the groups.

Function ordiellipse draws lines or polygons for ellipses by groups. The function can either draw standard deviation of points (kind="sd") or standard error of the (weighted) centroids (kind="se"), and the (weighted) correlation defines the direction of the principal axis of the ellipse. When kind = "se" is used together with argument conf, the ellipses will show the confidence regions for the locations of group centroids. With kind="ehull" the function draws an ellipse that encloses all points of a group using ellipsoidhull (cluster package).

Function ordibar draws crossed “error bars” using either either standard deviation of point scores or standard error of the (weighted) average of scores. These are the principal axes of the corresponding ordiellipse, and are found by principal component analysis of the (weighted) covariance matrix.

Functions ordihull and ordiellipse return invisibly an object that has a summary method that returns the coordinates of centroids and areas of the hulls or ellipses. Function ordiareatest studies the one-sided hypothesis that these areas are smaller than with randomized groups. Argument kind can be used to select the kind of ellipse, and has no effect with convex hulls.

Function ordispider draws a ‘spider’ diagram where each point is connected to the group centroid with segments. Weighted centroids are used in the correspondence analysis methods cca and decorana or if the user gives the weights in the call. If ordispider is called with cca or rda result without groups argument, the function connects each ‘WA’ scores to the corresponding ‘LC’ score. If the argument is a (invisible) ordihull object, the function will connect the points of the hull to their centroid.

Function ordicluster overlays a cluster dendrogram onto ordination. It needs the result from a hierarchic clustering such as hclust or agnes, or other with a similar structure. Function ordicluster connects cluster centroids to each other with line segments. Function uses centroids of all points in the clusters, and is therefore similar to average linkage methods.

Value

Functions ordihull, ordiellipse and ordispider return the invisible plotting structure.

Function ordispider return the coordinates to which each point is connected (centroids or ‘LC’ scores).

Function ordihull and ordiellipse return invisibly an object that has a summary method that returns the coordinates of centroids and areas of the hulls or ellipses. Function ordiareatest studies the one-sided hypothesis that these areas are smaller than with randomized groups.

Note

These functions add graphical items to ordination graph: You must draw a graph first. To draw line segments, grids or arrows, see ordisegments, ordigrid and ordiarrows.
Author(s)
Jari Oksanen

See Also
The functions pass parameters to basic graphical functions, and you may wish to change the default values in `lines`, `segments` and `polygon`. You can pass parameters to `scores` as well. Underlying functions for `ordihull` is `chull`. The underlying function for ellipsoid hulls in `ordiellipse` is `ellipsoidhull`.

Examples

```r
data(dune)
data(dune.env)
mod <- cca(dune ~ Management, dune.env)
plot(mod, type="n", scaling = "symmetric")
## Catch the invisible result of ordihull...
pl <- with(dune.env, ordihull(mod, Management,
   scaling = "symmetric", label = TRUE))
## ... and find centres and areas of the hulls
summary(pl)
## use more colours and add ellipsoid hulls
plot(mod, type = "n")
pl <- with(dune.env, ordihull(mod, Management,
   scaling = "symmetric", col = 1:4,
   draw="polygon", label =TRUE))
with(dune.env, ordiellipse(mod, Management, scaling = "symmetric",
   kind = "ehull", col = 1:4, lwd=3))
## ordispider to connect WA and LC scores
plot(mod, dis=c("wa","lc"), type="p")
ordispider(mod)
## Other types of plots
plot(mod, type = "p", display="sites")
c1 <- hclust(vegdist(dune))
ordicluster(mod, c1, prune=3, col = cutree(c1, 4))
## confidence ellipse: location of the class centroids
plot(mod, type="n", display = "sites")
with(dune.env, text(mod, display="sites", labels = as.character(Management),
   col=as.numeric(Management)))
pl <- with(dune.env, ordiellipse(mod, Management, kind="se", conf=0.95, lwd=2,
   draw = "polygon", col=1:4, border=1:4,
   alpha=63))
summary(pl)
## add confidence bars
with(dune.env, ordibar(mod, Management, kind="se", conf=0.95, lwd=2, col=1:4,
   label=TRUE))
```
ordilabel

Add Text on Non-transparent Label to an Ordination Plot.

Description

Function ordilabel is similar to text, but the text is on an opaque label. This can help in crowded ordination plots: you still cannot see all text labels, but at least the uppermost are readable. Argument priority helps to make the most important labels visible.

Usage

ordilabel(x, display, labels, choices = c(1, 2), priority, select,
        cex = 0.8, fill = "white", border = NULL, col = NULL, xpd = TRUE, ...)

Arguments

x An ordination object an any object known to scores.
display Kind of scores displayed (passed to scores).
labels Optional text used in plots. If this is not given, the text is found from the ordination object.
choices Axes shown (passed to scores).
priority Vector of the same length as the number of labels. The items with high priority will be plotted uppermost.
select Items to be displayed. This can either be a logical vector which is TRUE for displayed items or a vector of indices of displayed items.
cex Character expansion for the text (passed to text).
fill Background colour of the labels (the col argument of polygon).
border The colour and visibility of the border of the label as defined in polygon.
col Text colour. Default NULL will give the value of border or par("fg") if border is NULL.
xpd Draw labels also outside the plot region (see par).
... Other arguments (passed to text).

Details

The function may be useful with crowded ordination plots, in particular together with argument priority. You will not see all text labels, but at least some are readable. Other alternatives to crowded plots are identify.ordiplot, orditorp and orditkplot.

Author(s)

Jari Oksanen
See Also

scores, polygon, text. The function is modelled after s.label in ade4 package.

Examples

data(dune)
ord <- cca(dune)
plot(ord, type = "n")
ordilabel(ord, dis="sites", cex=1.2, font=3, fill="hotpink", col="blue")
## You may prefer separate plots, but here species as well
ordilabel(ord, dis="sp", font=2, priority=colSums(dune))

ordiplot

Alternative plot and identify Functions for Ordination

Description

Function ordiplot is an alternative plotting function which can be worked with any vegan ordination result and many non-vegan results. In addition, plot functions for vegan ordinations return invisibly an "ordiplot" result object, and this allows using ordiplot support functions with this result: identify can be used to add labels to selected site, species or constraint points, and points and text can add elements to the plot.

Usage

ordiplot(ord, choices = c(1, 2), type="points", display, xlim, ylim,
          cex = 0.7, ...)
## S3 method for class 'ordiplot'
identify(x, what, labels, ...)
## S3 method for class 'ordiplot'
points(x, what, select, arrows = FALSE, ...)
## S3 method for class 'ordiplot'
text(x, what, labels, select, arrows = FALSE, ...)

Arguments

ord         A result from an ordination.
choices     Axes shown.
type        The type of graph which may be "points", "text" or "none" for any ordination method.
display     Display only "sites" or "species". The default for most methods is to display both, but for cca, rda and capscale it is the same as in plot.cca.
xlim, ylim  the x and y limits (min,max) of the plot.
cex         Character expansion factor for points and text.
...         Other graphical parameters.
ordiplot

x         A result object from ordiplot.
what      Items identified in the ordination plot. The types depend on the kind of plot used. Most methods know sites and species, functions cca and rda know in addition constraints (for LC scores), centroids and biplot, and plot.procrustes ordination plot has heads and points.
labels    Optional text used for labels. Row names will be used if this is missing.
arrows    Draw arrows from the origin. This will always be TRUE for biplot scores and its value will be ignored. Setting this TRUE will draw arrows for any type of scores. This allows, e.g, using biplot arrows for species. The arrow head will be at the value of scores, and possible text is moved outwards.
select    Items to be displayed. This can either be a logical vector which is TRUE for displayed items or a vector of indices of displayed items.

Details

Function ordiplot draws an ordination diagram using black circles for sites and red crosses for species. It returns invisibly an object of class ordiplot which can be used by identify.ordiplot to label selected sites or species, or constraints in cca and rda.

The function can handle output from several alternative ordination methods. For cca, rda and decorana it uses their plot method with option type = "points". In addition, the plot functions of these methods return invisibly an ordiplot object which can be used by identify.ordiplot to label points. For other ordinations it relies on scores to extract the scores.

For full user control of plots, it is best to call ordiplot with type = "none" and save the result, and then add sites and species using points.ordiplot or text.ordiplot which both pass all their arguments to the corresponding default graphical functions.

Value

Function ordiplot returns invisibly an object of class ordiplot with used scores. In general, vegan plot functions for ordination results will also return an invisible ordiplot object. If the plot(..., type = "n") was used originally, the plot is empty, and items can be added with the invisible object. Functions points and text will return their input object without modification, which allows chaining these commands, for instance with magrittr pipes. Function identify.ordiplot uses this object to label the point.

Note

The purpose of these functions is to provide similar functionality as the plot, plotid and specid methods in library labdsv. The functions are somewhat limited in parametrization, but you can call directly the standard identify and plot functions for a better user control.

Author(s)

Jari Oksanen
See Also

identify for basic operations, plot.cca, plot.decorana, plot.procrustes which also produce objects for identify.ordiplot and scores for extracting scores from non-vegan ordinations.

Examples

# Draw a plot for a non-vegan ordination (cmdscale).
data(dune)
dune.dis <- vegdist(wisconsin(dune))
dune.mds <- cmdscale(dune.dis, eig = TRUE)
dune.mds$species <- wascores(dune.mds$points, dune, expand = TRUE)
pl <- ordiplot(dune.mds, type = "none")
points(pl, "sites", pch=21, col="red", bg="yellow")
text(pl, "species", col="blue", cex=0.9)
# Default plot of the previous using identify to label selected points
## Not run:
pl <- ordiplot(dune.mds)
identify(pl, "spec")
## End(Not run)

ordipointlabel

Ordination Plots with Points and Optimized Locations for Text

Description

The function ordipointlabel produces ordination plots with points and text label to the points. The points are in the exact location given by the ordination, but the function tries to optimize the location of the text labels to minimize overplotting text. The function may be useful with moderately crowded ordination plots.

Usage

ordipointlabel(x, display = c("sites", "species"), choices = c(1, 2),
col = c(1, 2), pch = c("o", "*"), font = c(1, 1),
cex = c(0.8, 0.8), add = FALSE, select, ...)

## S3 method for class 'ordipointlabel'
plot(x, ...)

Arguments

x For ordipointlabel() a result object from an ordination function. For plot.ordipointlabel an object resulting from a call to ordipointlabel().
display Scores displayed in the plot.
choices Axes shown.
col, pch, font, cex

   Colours, point types, font style and character expansion for each kind of scores
displayed in the plot. These should be vectors of the same length as the number
of items in display.

add

   Add to an existing plot.

select

   Items to be displayed. This can either be a logical vector which is TRUE for
displayed items or a vector of indices of displayed items. select is only used if
a single set of scores is being plotted (i.e. length(display) == 1), otherwise
it is ignored and a warning issued. If a logical vector is used, it must have the
same length as the scores plotted.

... Other arguments passed to points and text.

Details

The function uses simulated annealing (optim, method = "SANN") to optimize the location of the
text labels to the points. There are eight possible locations: up, down, sides and corners. There is a
weak preference to text right above the point, and a weak avoidance of corner positions. The exact
locations and the goodness of solution varies between runs, and there is no guarantee of finding the
global optimum. The optimization can take a long time in difficult cases with a high number of
potential overlaps. Several sets of scores can be displayed in one plot.

The function is modelled after pointLabel in maptools package (which has chained dependencies
of S4 packages).

Value

The function returns invisibly an object of class ordipointlabel with items xy for coordinates of
points, labels for coordinates of labels, items pch, cex and font for graphical parameters of each
point or label. In addition, it returns the result of optim as an attribute "optim". The unit of overlap
is the area of character "m", and with variable cex it is the smallest alternative.

There is a plot method based on orditkplot but which does not alter nor reset the graphical
parameters via par.

The result object from ordipointlabel inherits from orditkplot, and can also be replotted with
its plot method. It may be possible to further edit the result object with orditkplot, but for good
results it is necessary that the points span the whole horizontal axis without empty margins.

Note

The function is designed for ordination graphics, and the optimization works properly with plots of
isometric aspect ratio.

Author(s)

Jari Oksanen

References

See pointLabel for references.
See Also

pointLabel for the model implementation, and optim for the optimization.

Examples

data(dune)
ord <- cca(dune)
plt <- ordipointlabel(ord)

## set scaling - should be no warnings!
ordipointlabel(ord, scaling = "sites")

## plot then add
plot(ord, scaling = "symmetric", type = "n")
ordipointlabel(ord, display = "species", scaling = "symm", add = TRUE)
ordipointlabel(ord, display = "sites", scaling = "symm", add = TRUE)

## redraw plot without rerunning SANN optimisation
plot(plt)

ordiresids

Plots of Residuals and Fitted Values for Constrained Ordination

Description

The function provides plot.lm style diagnostic plots for the results of constrained ordination from cca, rda and capscale. Normally you do not need these plots, because ordination is descriptive and does not make assumptions on the distribution of the residuals. However, if you permute residuals in significance tests (anova.cca), you may be interested in inspecting that the residuals really are exchangeable and independent of fitted values.

Usage

ordiresids(x, kind = c("residuals", "scale", "qqmath"),
  residuals = "working", type = c("p", "smooth", "g"),
  formula, ...)

Arguments

x Ordination result from cca, rda or capscale.

kind The type of plot: "residuals" plot residuals against fitted values, "scale" the square root of absolute residuals against fitted values, and "qqmath" the residuals against expected distribution (defaults qnorm), unless defined differently in the formula argument.

residuals The kind of residuals and fitted values, with alternatives "working", "response", "standardized" and "studentized" (see Details).

type The type of plot. The argument is passed on to lattice functions.
ordistep

formula

Formula to override the default plot. The formula can contain items Fitted, Residuals, Species and Sites (provided that names of species and sites are available in the ordination result).

... Other arguments passed to lattice functions.

Details

The default plots are similar as in plot.lm, but they use Lattice functions xyplot and qqmath. The alternatives have default formulae but these can be replaced by the user. The elements available in formula or in the groups argument are Fitted, Residuals, Species and Sites.

With residuals = "response" and residuals = "working" the fitted values and residuals are found with functions fitted.cca and residuals.cca. With residuals = "standardized" the residuals are found with rstandard.cca, and with residuals = "studentized" they are found with rstudent.cca, and in both cases the fitted values are standardized with sigma.cca.

Value

The function returns a Lattice object that can displayed as plot.

Author(s)

Jari Oksanen

See Also

plot.lm, fitted.cca, residuals.cca, rstandard.cca, rstudent.cca, sigma.cca, Lattice, xyplot, qqmath.

Examples

data(varespec)
data(varechem)
mod <- cca(varespec ~ Al + P + K, varechem)
ordiresids(mod)
ordiresids(mod, formula = Residuals ~ Fitted | Species, residuals="standard", cex = 0.5)

ordistep

Choose a Model by Permutation Tests in Constrained Ordination

Description

Automatic stepwise model building for constrained ordination methods (cca, rda, capscale). The function ordistep is modelled after step and can do forward, backward and stepwise model selection using permutation tests. Function ordiR2step performs forward model choice solely on adjusted $R^2$ and P-value, for ordination objects created by rda or capscale.
Usage

ordistep(object, scope, direction = c("both", "backward", "forward"),
Pin = 0.05, Pout = 0.1, permutations = how(nperm = 199), steps = 50,
trace = TRUE, ...)
ordir2step(object, scope, Pin = 0.05, R2scope = TRUE,
permutations = how(nperm = 499), trace = TRUE, R2permutations = 1000, ...)

Arguments

object In ordistep, an ordination object inheriting from cca or rda. In ordir2step, the object must inherit from rda, that is, it must have been computed using rda or capscale.
scope Defines the range of models examined in the stepwise search. This can be a list containing components upper and lower, both formulae. If it is a single item, it is interpreted the target scope, depending on the direction. If direction is "forward", a single item is interpreted as the upper scope and the formula of the input object as the lower scope. See step for details. In ordir2step, this defines the upper scope; it can also be an ordination object from with the model is extracted.
direction The mode of stepwise search, can be one of "both", "backward", or "forward", with a default of "both". If the scope argument is missing, the default for direction is "backward" in ordistep (and ordir2step does not have this argument, but only works forward).
Pin, Pout Limits of permutation P-values for adding (Pin) a term to the model, or dropping (Pout) from the model. Term is added if $P \leq$ Pin, and removed if $P >$ Pout.
R2scope Use adjusted $R^2$ as the stopping criterion: only models with lower adjusted $R^2$ than scope are accepted.
permutations a list of control values for the permutations as returned by the function how, or the number of permutations required, or a permutation matrix where each row gives the permuted indices. This is passed to anova.cca: see there for details.
steps Maximum number of iteration steps of dropping and adding terms.
trace If positive, information is printed during the model building. Larger values may give more information.
R2permutations Number of permutations used in the estimation of adjusted $R^2$ for cca using RsquareAdj.
... Any additional arguments to add1.cca and drop1.cca.

Details

The basic functions for model choice in constrained ordination are add1.cca and drop1.cca. With these functions, ordination models can be chosen with standard R function step which bases the term choice on AIC. AIC-like statistics for ordination are provided by functions deviance.cca and extractAIC.cca (with similar functions for rda). Actually, constrained ordination methods do not have AIC, and therefore the step may not be trusted. This function provides an alternative using permutation P-values.
Function `ordistep` defines the model, scope of models considered, and direction of the procedure similarly as `step`. The function alternates with drop and add steps and stops when the model was not changed during one step. The − and + signs in the summary table indicate which stage is performed. It is often sensible to have \( \text{Pout} > \text{Pin} \) in stepwise models to avoid cyclic adds and drops of single terms.

Function `ordirRstep` builds model forward so that it maximizes adjusted \( R^2 \) (function `RsquareAdj`) at every step, and stopping when the adjusted \( R^2 \) starts to decrease, or the adjusted \( R^2 \) of the scope is exceeded, or the selected permutation \( P \)-value is exceeded (Blanchet et al. 2008). The second criterion is ignored with option `R2scope = FALSE`, and the third criterion can be ignored setting \( \text{Pin} = 1 \) (or higher). Adjusted \( R^2 \) cannot be calculated if the number of predictors is higher than the number of observations, but such models can be analysed with `R2scope = FALSE`. The \( R^2 \) of `cca` is based on simulations (see `RsquareAdj`) and different runs of `ordirRstep` can give different results.

Functions `ordistep` (based on \( P \) values) and `ordirRstep` (based on adjusted \( R^2 \) and hence on eigenvalues) can select variables in different order.

**Value**

Functions return the selected model with one additional component, `anova`, which contains brief information of steps taken. You can suppress voluminous output during model building by setting `trace = FALSE`, and find the summary of model history in the `anova` item.

**Author(s)**

Jari Oksanen

**References**


**See Also**

The function handles constrained ordination methods `cca`, `rda`, `dbrda` and `capscale`. The underlying functions are `add1.cca` and `drop1.cca`, and the function is modelled after standard `step` (which also can be used directly but uses AIC for model choice, see `extractAIC.cca`). Function `ordirRstep` builds upon `RsquareAdj`.

**Examples**

```r
## See add1.cca for another example

### Dune data
data(dune)
data(dune.env)
mod0 <- rda(dune - 1, dune.env) # Model with intercept only
mod1 <- rda(dune ~ ., dune.env) # Model with all explanatory variables

## With scope present, the default direction is "both"
ordistep(mod0, scope = formula(mod1))
```
ordisurf

## Description

Function `ordisurf` fits a smooth surface for given variable and plots the result on ordination diagram.

## Usage

```r
## Default S3 method:
ordisurf(x, y, choices = c(1, 2), knots = 10,
  family = "gaussian", col = "red", isotropic = TRUE,
  thinplate = TRUE, bs = "tp", fx = FALSE, add = FALSE,
  display = "sites", w = weights(x, display), main, nlevels = 10,
  levels, npoints = 31, labcex = 0.6, bubble = FALSE,
  cex = 1, select = TRUE, method = "REML", gamma = 1,
  plot = TRUE, lwd.cl = par("lwd"), ...)

## S3 method for class 'formula'
ordisurf(formula, data, ...)
```
## S3 method for class 'ordisurf'

`calibrate(object, newdata, ...)`

## S3 method for class 'ordisurf'

`plot(x, what = c("contour","persp","gam"),
    add = FALSE, bubble = FALSE, col = "red", cex = 1,
    nlevels = 10, levels, labcex = 0.6, lwd.cl = par("lwd"), ...)`

### Arguments

- **x**
  - For `ordisurf` an ordination configuration, either a matrix or a result known by `scores`. For `plot.ordisurf` an object of class "ordisurf" as returned by `ordisurf`.

- **y**
  - Variable to be plotted / modelled as a function of the ordination scores.

- **choices**
  - Ordination axes.

- **knots**
  - Number of initial knots in `gam` (one more than degrees of freedom). If knots = 0 or knots = 1 the function will fit a linear trend surface, and if knots = 2 the function will fit a quadratic trend surface instead of a smooth surface. A vector of length 2 is allowed when isotropic = FALSE, with the first and second elements of knots referring to the first and second of ordination dimensions (as indicated by choices) respectively.

- **family**
  - Error distribution in `gam`.

- **col**
  - Colour of contours.

- **isotropic, thinplate**
  - Fit an isotropic smooth surface (i.e. same smoothness in both ordination dimensions) via `gam`. Use of thinplate is deprecated and will be removed in a future version of the package.

- **bs**
  - a two letter character string indicating the smoothing basis to use. (e.g. "tp" for thin plate regression spline, "cr" for cubic regression spline). One of c("tp", "ts", "cr", "cs", ...). See `smooth.terms` for an over view of what these refer to. The default is to use thin plate splines: bs = "tp".

- **fx**
  - indicates whether the smoothers are fixed degree of freedom regression splines (fx = FALSE) or penalised regression splines (fx = TRUE). Can be a vector of length 2 for anisotropic surfaces (isotropic = FALSE). It doesn’t make sense to use fx = TRUE and select = TRUE and it is an error to do so. A warning is issued if you specify fx = TRUE and forget to use select = FALSE though fitting continues using select = FALSE.

- **add**
  - Add contours to an existing diagram or draw a new plot?

- **display**
  - Type of scores known by `scores`: typically "sites" for ordinary site scores or "lc" for linear combination scores.

- **w**
  - Prior weights on the data. Concerns mainly `cca` and `decorana` results which have nonconstant weights.

- **main**
  - The main title for the plot, or as default the name of plotted variable in a new plot.
nlevels, levels
Either a vector of levels for which contours are drawn, or suggested number
of contours in nlevels if levels are not supplied.

npoints numeric; the number of locations at which to evaluate the fitted surface. This
represents the number of locations in each dimension.

labcex Label size in contours. Setting this zero will suppress labels.
bubble Use a “bubble plot” for points, or vary the point diameter by the value of
the plotted variable. If bubble is numeric, its value is used for the maximum symbol
size (as in cex), or if bubble = TRUE, the value of cex gives the maximum. The
minimum size will always be cex = 0.4. The option only has an effect if
add = FALSE.
cex Character expansion of plotting symbols.
select Logical; specify gam argument "select". If this is TRUE then gam can add an
extra penalty to each term so that it can be penalized to zero. This means that
the smoothing parameter estimation that is part of fitting can completely remove
terms from the model. If the corresponding smoothing parameter is estimated
as zero then the extra penalty has no effect.
method character; the smoothing parameter estimation method. Options allowed are:
"GCV.Cp" uses GCV for models with unknown scale parameter and Mallows’
Cp/UBRE/AIC for models with known scale; "GACV.Cp" as for "GCV.Cp" but
uses GACV (Generalised Approximate CV) instead of GCV; "REML" and "ML"
use restricted maximum likelihood or maximum likelihood estimation for both
known and unknown scale; and "P-REML" and "P-ML" use REML or ML esti-
mation but use a Pearson estimate of the scale.
gamma Multiplier to inflate model degrees of freedom in GCV or UBRE/AIC score by.
This effectively places an extra penalty on complex models. An oft-used value
is gamma = 1.4.
plot logical; should any plotting be done by ordisurf? Useful if all you want is the
fitted response surface model.
lwdNcl numeric; the lwd (line width) parameter to use when drawing the contour lines.
formula, data Alternative definition of the fitted model as x ~ y, where left-hand side is the
ordination x and right-hand side the single fitted continuous variable y. The vari-
able y must be in the working environment or in the data frame or environment
given by data. All other arguments of are passed to the default method.
object An ordisurf result object.
newdata Coordinates in two-dimensional ordination for new points.
what character; what type of plot to produce. "contour" produces a contour plot of
the response surface, see contour for details. "persp" produces a perspective
plot of the same, see persp for details. "gam" plots the fitted GAM model, an
object that inherits from class "gam" returned by ordisurf, see plot.gam.
... Other parameters passed to scores, or to the graphical functions. See Note
below for exceptions.
**Details**

Function `ordisurf` fits a smooth surface using penalised splines (Wood 2003) in `gam`, and uses `predict.gam` to find fitted values in a regular grid. The smooth surface can be fitted with an extra penalty that allows the entire smoother to be penalized back to 0 degrees of freedom, effectively removing the term from the model (see Marra & Wood, 2011). The addition of this extra penalty is invoked by setting argument `select` to TRUE. An alternative is to use a spline basis that includes shrinkage (bs = "ts" or bs = "cs").

`ordisurf()` exposes a large number of options from `gam` for specifying the basis functions used for the surface. If you stray from the defaults, do read the Notes section below and relevant documentation in `s` and `smooth.terms`.

The function plots the fitted contours with convex hull of data points either over an existing ordination diagram or draws a new plot. If `select = TRUE` and the smooth is effectively penalised out of the model, no contours will be plotted.

`gam` determines the degree of smoothness for the fitted response surface during model fitting, unless `fx = TRUE`. Argument `method` controls how `gam` performs this smoothness selection. See `gam` for details of the available options. Using "REML" or "ML" yields p-values for smooths with the best coverage properties if such things matter to you.

The function uses `scores` to extract ordination scores, and `x` can be any result object known by that function.

The user can supply a vector of prior weights `w`. If the ordination object has weights, these will be used. In practise this means that the row totals are used as weights with `cca` or `decorana` results. If you do not like this, but want to give equal weights to all sites, you should set `w = NULL`. The behaviour is consistent with `envfit`. For complete accordance with constrained `cca`, you should set `display = "lc"`.

Function `calibrate` returns the fitted values of the response variable. The `newdata` must be coordinates of points for which the fitted values are desired. The function is based on `predict.gam` and will pass extra arguments to that function.

**Value**

`ordisurf` is usually called for its side effect of drawing the contour plot. The function returns a result object of class "ordisurf" that inherits from `gam` used internally to fit the surface, but adds an item `grid` that contains the data for the grid surface. The item `grid` has elements `x` and `y` which are vectors of axis coordinates, and element `z` that is a matrix of fitted values for `contour`. The values outside the convex hull of observed points are indicated as `NA` in `z`. The `gam` component of the result can be used for further analysis like predicting new values (see `predict.gam`).

**Warning**

The fitted GAM is a regression model and has the usual assumptions of such models. Of particular note is the assumption of independence of residuals. If the observations are not independent (e.g. they are repeat measures on a set of objects, or from an experimental design, *inter alia*) do not trust the `p`-values from the GAM output.

If you need further control (i.e. to add additional fixed effects to the model, or use more complex smoothers), extract the ordination scores using the `scores` function and then generate your own `gam` call.
Note

The default is to use an isotropic smoother via `s` employing thin plate regression splines (bs = "tp"). These make sense in ordination as they have equal smoothing in all directions and are rotation invariant. However, if different degrees of smoothness along dimensions are required, an anisotropic smooth surface may be more applicable. This can be achieved through the use of isotropic = FALSE, wherein the surface is fitted via a tensor product smoother via `te` (unless bs = "ad", in which case separate splines for each dimension are fitted using `s`).

Cubic regression splines and P splines can only be used with isotropic = FALSE.

Adaptive smooths (bs = "ad"), especially in two dimensions, require a large number of observations; without many hundreds of observations, the default complexities for the smoother will exceed the number of observations and fitting will fail.

To get the old behaviour of `ordisurf` use select = FALSE, method = "GCV.Cp", fx = FALSE, and bs = "tp". The latter two options are the current defaults.

Graphical arguments supplied to `plot.ordisurf` are passed on to the underlying plotting functions, `contour`, `persp`, and `plot.gam`. The exception to this is that arguments col and cex can not currently be passed to `plot.gam` because of a bug in the way that function evaluates arguments when arranging the plot.

A work-around is to call `plot.gam` directly on the result of a call to `ordisurf`. See the Examples for an illustration of this.

Author(s)

Dave Roberts, Jari Oksanen and Gavin L. Simpson

References


See Also

For basic routines `gam`, and `scores`. Function `envfit` provides a more traditional and compact alternative.

Examples

data(varespec)
data(varechem)
vare.dist <- vegdist(varespec)
vare.mds <- monoMDS(vare.dist)
ordisurf(vare.mds ~ Baresoil, varechem, bubble = 5)

## as above but without the extra penalties on smooth terms,
## and using GCV smoothness selection (old behaviour of `ordisurf()`):
ordisurf(vare.mds ~ Baresoil, varechem, col = "blue", add = TRUE,
          select = FALSE, method = "GCV.Cp")
orditkplot

## Cover of Cladina arbuscula
fit <- ordisurf(vare.mds ~ Cladaru, varespec, family=quasipoisson)
## Get fitted values
calibrate(fit)

## Variable selection via additional shrinkage penalties
## This allows non-significant smooths to be selected out
## of the model not just to a linear surface. There are 2
## options available:
## - option 1: `select = TRUE` --- the *default*
  ordisurf(vare.mds ~ Baesoil, varechem, method = "REML", select = TRUE)
## - option 2: use a basis with shrinkage
  ordisurf(vare.mds ~ Baesoil, varechem, method = "REML", bs = "ts")
## or bs = "cs" with `isotropic = FALSE`

## Plot method
plot(fit, what = "contour")

## Plotting the "gam" object
plot(fit, what = "gam") ## 'col' and 'cex' not passed on
## or via plot.gam directly
library(mgcv)
plot.gam(fit, cex = 2, pch = 1, col = "blue")
## 'col' effects all objects drawn...

### controlling the basis functions used
## Use Duchon splines
ordisurf(vare.mds ~ Baesoil, varechem, bs = "ds")

## A fixed degrees of freedom smooth, must use 'select = FALSE'
ordisurf(vare.mds ~ Baesoil, varechem, knots = 4,
  fx = TRUE, select = FALSE)

## An anisotropic smoother with cubic regression spline bases
ordisurf(vare.mds ~ Baesoil, varechem, isotropic = FALSE,
  bs = "cr", knots = 4)

## An anisotropic smoother with cubic regression spline with
## shrinkage bases & different degrees of freedom in each dimension
ordisurf(vare.mds ~ Baesoil, varechem, isotropic = FALSE,
  bs = "cs", knots = c(3,4), fx = TRUE,
  select = FALSE)

orditkplot

Ordination Plot with Movable Labels

**Description**

Function `orditkplot` produces an editable ordination plot with points and labels. The labels can be moved with mouse, and the edited plot can be saved as an encapsulated postscript file or exported via `R` `plot` function to other graphical formats, or saved in the `R` session for further processing.
Usage

orditkplot(x, display = "species", choices = 1:2, width, xlim, ylim,
tcem = 0.8, tcol, pch = 1, pcol, pbg, pce = 0.7, labels, ...)
## S3 method for class 'orditkplot'
plot(x, ...)
## S3 method for class 'orditkplot'
points(x, pch = x$arg$pch, cex = x$arg$pce,
       col = x$arg$pcol, bg = x$arg$pbg, ...)
## S3 method for class 'orditkplot'
text(x, cex = x$arg$tcem, col = x$arg$tcol,
      font = attr(x$labels, "font"), ...)  
## S3 method for class 'orditkplot'
scores(x, display, ...)

Arguments

x An ordination result or any other object that scores can handle, or for the plot
    function the object dumped from the interactive orditkplot session.
display Type of scores displayed. For ordination scores this typically is either "species"
or "sites", and for orditkplot result it is either "points" or "labels".
choices Axes displayed.
width Width of the plot in inches; defaults to the current width of the graphical device.
xlim, ylim x and y limits for plots: points outside these limits will be completely removed.
tcem Character expansion for text labels.
tcol Colour of text labels.
pch, pcol, pbg Point type and outline and fill colours. Defaults pcol="black" and pbg="transparent". 
    Argument pbg has an effect only in filled plotting characters pch = 21 to 25.
pce Expansion factor for point size.
labels Labels used instead of row names.
cex, col, bg, font graphical parameters used in the points and text methods. See par.
...
Other arguments passed to the function. These can be graphical parameters 
(see par) used in the plot, or extra arguments to scores. These arguments are 
ignored in plot, but honoured in text and points.

Details

Function orditkplot uses tcltk package to draw Tcl/Tk based ordination graphics with points and
labels. The function opens an editable canvas with fixed points, but the labels can be dragged with 
mouse to better positions or edited. In addition, it is possible to zoom to a part of the graph.

The function knows the following mouse operations:

- **Left mouse button** can be used to move labels to better positions. A line will connect a label 
to the corresponding point.
• **Double clicking left mouse button** opens a window where the label can be edited. After editing the label, hit the Return key.

• **Right mouse button** (or alternatively, Shift-Mouse button with one-button mouse) can be used for zooming to a part of the graph. Keeping the mouse button down and dragging will draw a box of the zoomed area, and after releasing the button, a new plot window will be created (this is still preliminary: all arguments are not passed to the new plot).

In addition there are buttons for the following tasks: **Copy to EPS** copies the current plot to an encapsulated postscript (eps) file using standard Tcl/Tk utilities. The faithfulness of this copy is system dependent. Button **Export plot** uses plot.orditkplot function to redraw the plot into graphical file formats. Depending on the system, the following graphical formats may be available: eps, pdf, svg, png, jpeg, tiff, bmp or xfig. Some of the output formats may be edited with external software: svg files with Illustrator or Inkscape, and xfig with the legacy program XFig. Button **Save to R** writes the edited coordinates of labels and points to the R session for further processing, and the plot.orditkplot function can be used to display the results. For faithful replication of the plot, the graph must have similar dimensions as the orditkplot canvas had originally. The plot function cannot be configured, but it uses the same settings as the original Tcl/Tk plot. However, points and text functions are fully configurable, but use the stored defaults for consistency with plot.orditkplot if none are supplied. Finally, button **Close** closes the window.

The produced plot will have equal aspect ratio. The width of the horizontal axis is fixed, but vertical axes will be scaled to needed height, and you can use scrollbar to move vertically if the whole canvas does not fit the window. If you use dumped labels in ordinary R plots, your plot must have the same dimensions as the orditkplot canvas to have identical location of the labels.

The function only displays one set of scores. However, you can use ordipointlabel to produce a result object that has different points and text types for several sets of scores and this can be further edited with orditkplot. For a good starting solution you need to scale the ordipointlabel result so that the points span over the whole horizontal axis. The function cannot show environmental variables or constraints, but it is limited to unconstrained ordination.

The plot is a Tcl/Tk canvas, but the function tries to replicate standard graphical device of the platform, and it honours several graphical parameters (see **par**). Many of the graphical parameters can be given on the command line, and they will be passed to the function without influencing other graphical devices in R. At the moment, the following graphical parameters are honoured: pch bg, cex, cex.axis, cex.lab, col (for labels), col.axis, col.lab, family (for font faces), fg, font, font.axis, font.lab, lheight, lwd (for the box), mar, mex, mgp, ps, tcl. These can be set with **par**, and they also will influence other plots similarly.

The **tkcanvas** text cannot be rotated, and therefore vertical axis is not labelled, and las parameter will not be honoured in the Tcl/Tk plot, but it will be honoured in the exported R plots and in plot.orditkplot.

**Value**

Function returns nothing useful directly, but you can save the edited graph to a file or save the edited positions to an R session for further processing and plotting.

**Note**

You need tcltk package and R must have been configured with capabilities for tcltk. Depending on your OS, you may need to start X11 and set the display before loading tcltk and starting the
function (for instance, withSys.setenv("DISPLAY":"\0")). See tcltk-package.

Author(s)
Jari Oksanen

See Also
Function ordipointlabel is an automatic procedure with similar goals of avoiding overplotting, and its output can be edited with orditkplot. See ordiplot, plot.cca, ordirgl and orditorp for alternative ordination plots, and scores for extracting ordination scores.

Examples
## The example needs user interaction and is not executed directly.
## It should work when pasted to the window.
## Not run:
data(varespec)
ord <- cca(varespec)
## Do something with the graph and end by clicking "Dismiss"
orditkplot(ord, mar = c(4,4,1,1)+.1, font=3)
## Use ordipointlabel to produce a plot that has both species and site
## scores in different colors and plotting symbols
pl <- ordipointlabel(ord)
orditkplot(pl)
## End(Not run)

---

orditorp

Add Text or Points to Ordination Plots

Description
The function adds text or points to ordination plots. Text will be used if this can be done without overwriting other text labels, and points will be used otherwise. The function can help in reducing clutter in ordination graphics, but manual editing may still be necessary.

Usage
orditorp(x, display, labels, choices = c(1, 2), priority, select, cex = 0.7, pcesx, col = par("col"), pcol, pch = par("pch"), air = 1, ...)

Arguments

- **x**: A result object from ordination or an ordiplot result.
- **display**: Items to be displayed in the plot. Only one alternative is allowed. Typically this is "sites" or "species".
**orditorp**

Optional text used for labels. Row names will be used if this is missing.

Axes shown.

Text will be used for items with higher priority if labels overlap. This should be vector of the same length as the number of items plotted.

Items to be displayed. This can either be a logical vector which is TRUE for displayed items or a vector of indices of displayed items. If a logical vector is used, it must have the same length as the scores plotted.

Text and point sizes, see `plot.default`.

Text and point colours, see `plot.default`.

Plotting character, see `points`.

Amount of empty space between text labels. Values <1 allow overlapping text.

Other arguments to `scores` (and its various methods), `text` and `points`.

**Details**

Function `orditorp` will add either text or points to an existing plot. The items with high priority will be added first and text will be used if this can be done without overwriting previous labels and points will be used otherwise. If priority is missing, labels will be added from the outskirts to the centre. Function `orditorp` can be used with most ordination results, or plotting results from `ordiplot` or ordination plot functions (`plot.cca`, `plot.decorana`, `plot.metaMDS`).

Arguments can be passed to the relevant `scores` method for the ordination object (x) being drawn. See the relevant `scores` help page for arguments that can be used.

**Value**

The function returns invisibly a logical vector where TRUE means that item was labelled with text and FALSE means that it was marked with a point. The returned vector can be used as the select argument in ordination text and points functions.

**Author(s)**

Jari Oksanen

**Examples**

```r
## A cluttered ordination plot:
data(BCI)
mod <- cca(BCI)
plot(mod, dis="sp", type="t")
# Now with orditorp and abbreviated species names
cnam <- make.cepnames(names(BCI))
plot(mod, dis="sp", type="n")
stems <- colSums(BCI)
orditorp(mod, "sp", label = cnam, priority = stems, pch="+", pcol="grey")

## show select in action
set.seed(1)
take <- sample(ncol(BCI), 50)
```
plot(mod, dis="sp", type="n")
stems <- colSums(BCI)
orditorp(mod, "sp", label = cnam, priority=stems, select = take,
        pch="+", pcol="grey")

---

**ordixyplot**  
*Trellis (Lattice) Plots for Ordination*

**Description**

Functions `ordicloud`, `ordisplom` and `ordixyplot` provide an interface to plot ordination results using Trellis functions `cloud`, `splom` and `xyplot` in package `lattice`.

**Usage**

```r
ordixyplot(x, data = NULL, formula, display = "sites", choices = 1:3,
            panel = "panel.ordi", aspect = "iso", envfit,
            type = c("p", "biplot"), ...)
ordisplom(x, data=NULL, formula = NULL, display = "sites", choices = 1:3,
          panel = "panel.ordi", type = "p", ...)
ordicloud(x, data = NULL, formula, display = "sites", choices = 1:3,
          panel = "panel.ordi3d", prepanel = "prepanel.ordi3d", ...)
```

**Arguments**

- `x`  
  An ordination result that `scores` knows: any ordination result in `vegan` and many others.

- `data`  
  Optional data to amend ordination results. The ordination results are found from `x`, but you may give here data for other variables needed in plots. Typically these are environmental data.

- `formula`  
  Formula to define the plots. A default formula will be used if this is omitted. The ordination axes must be called by the same names as in the ordination results (and these names vary among methods). In `ordisplom`, special character `.` refers to the ordination result.

- `display`  
  The kind of scores: an argument passed to `scores`.

- `choices`  
  The axes selected: an argument passed to `scores`.

- `panel, prepanel`  
  The names of the panel and prepanel functions.

- `aspect`  
  The aspect of the plot (passed to the `lattice` function).

- `envfit`  
  Result of `envfit` function displayed in `ordixyplot`. Please note that this needs same choices as `ordixyplot`.
type

The type of plot. This knows the same alternatives as panel.xyplot. In addition ordixyplot has alternatives "biplot", "arrows" and "polygon". The first displays fitted vectors and factor centroids of envfit, or in constrained ordination, the biplot arrows and factor centroids if envfit is not given. The second (type = "arrows") is a trellis variant of ordiarrows and draws arrows by groups. The line parameters are controlled by trellis.par.set for superpose.line, and the user can set length, angle and ends parameters of panel.arrows. The last one (type = "polygon") draws a polygon enclosing all points in a panel over a polygon enclosing all points in the data. The overall polygon is controlled by trellis.par.set for plot_polygon, and each panel polygon is controlled by superpose_polygon.

... Arguments passed to scores methods or lattice functions.

Details

The functions provide an interface to the corresponding lattice functions. All graphical parameters are passed to the lattice function so that these graphs are extremely configurable. See Lattice and xyplot, splom and cloud for details, usage and possibilities.

The argument x must always be an ordination result. The scores are extracted with vegan function scores so that these functions work with all vegan ordinations and many others.

The formula is used to define the models. All functions have simple default formulae which are used if formula is missing. If formula is omitted in ordisplom it produces a pairs plot of ordination axes and variables in data. If formula is given, ordination results must be referred to as . and other variables by their names. In other functions, the formula must use the names of ordination scores and names of data.

The ordination scores are found from x, and data is optional. The data should contain other variables than ordination scores to be used in plots. Typically, they are environmental variables (typically factors) to define panels or plot symbols.

The proper work is done by the panel function. The layout can be changed by defining own panel functions. See panel.xyplot, panel.splom and panel.cloud for details and survey of possibilities.

Ordination graphics should always be isometric: same scale should be used in all axes. This is controlled (and can be changed) with argument aspect in ordixyplot. In ordicloud the isometric scaling is defined in panel and prepanel functions. You must replace these functions if you want to have non-isometric scaling of graphs. You cannot select isometric scaling in ordisplom.

Value

The function return Lattice objects of class "trellis".

Author(s)

Jari Oksanen

See Also

Lattice, xyplot, splom, cloud, panel.splom, panel.cloud
Examples

data(dune, dune.env)
ord <- cca(dune)
## Pairs plots
ordisplom(ord)
ordisplom(ord, data=dune.env, choices=1:2)
ordisplom(ord, data=dune.env, form =~ . | Management, groups=Manure)
## Scatter plot with polygons
ordixyplot(ord, data=dune.env, form = CA1 ~ CA2 | Management,
            groups=Manure, type = c("p","polygon"))
## Choose a different scaling
ordixyplot(ord, scaling = "symmetric")
## ... Slices of third axis
ordixyplot(ord, form = CA1 ~ CA2 | equal.count(CA3, 4),
            type = c("g","p", "polygon"))
## Display environmental variables
ordixyplot(ord, envfit = envfit(ord ~ Management + A1, dune.env, choices=1:3))
## 3D Scatter plots
ordicloud(ord, form = CA2 ~ CA3*CA1, groups = Manure, data = dune.env)
ordicloud(ord, form = CA2 ~ CA3*CA1 | Management, groups = Manure,
          data = dune.env, auto.key = TRUE, type = c("p","h"))

---

pcnm

Principal Coordinates of Neighbourhood Matrix

Description

This function computed classical PCNM by the principal coordinate analysis of a truncated distance matrix. These are commonly used to transform (spatial) distances to rectangular data that suitable for constrained ordination or regression.

Usage

pcnm(dis, threshold, w, dist.ret = FALSE)

Arguments

dis A distance matrix.
threshold A threshold value or truncation distance. If missing, minimum distance giving connected network will be used. This is found as the longest distance in the minimum spanning tree of dis.
w Prior weights for rows.
dist.ret Return the distances used to calculate the PCNMs.
**Details**

Principal Coordinates of Neighbourhood Matrix (PCNM) map distances between rows onto rectangular matrix on rows using a truncation threshold for long distances (Borcard & Legendre 2002). If original distances were Euclidean distances in two dimensions (like normal spatial distances), they could be mapped onto two dimensions if there is no truncation of distances. Because of truncation, there will be a higher number of principal coordinates. The selection of truncation distance has a huge influence on the PCNM vectors. The default is to use the longest distance to keep data connected. The distances above truncation threshold are given an arbitrary value of 4 times threshold. For regular data, the first PCNM vectors show a wide scale variation and later PCNM vectors show smaller scale variation (Borcard & Legendre 2002), but for irregular data the interpretation is not as clear.

The PCNM functions are used to express distances in rectangular form that is similar to normal explanatory variables used in, e.g., constrained ordination (rda, cca and capscale) or univariate regression (lm) together with environmental variables (row weights should be supplied with cca; see Examples). This is regarded as a more powerful method than forcing rectangular environmental data into distances and using them in partial mantel analysis (mantel.partial) together with geographic distances (Legendre et al. 2008, but see Tuomisto & Ruokolainen 2008).

The function is based on pcnm function in Dray’s unreleased spacemakeR package. The differences are that the current function uses spantree as an internal support function. The current function also can use prior weights for rows by using weighted metric scaling of wcmdscale. The use of row weights allows finding orthonormal PCNMs also for correspondence analysis (e.g., cca).

**Value**

A list of the following elements:

- **values**: Eigenvalues obtained by the principal coordinates analysis.
- **vectors**: Eigenvectors obtained by the principal coordinates analysis. They are scaled to unit norm. The vectors can be extracted with scores function. The default is to return all PCNM vectors, but argument choices selects the given vectors.
- **threshold**: Truncation distance.
- **dist**: The distance matrix where values above threshold are replaced with arbitrary value of four times the threshold. String "pcnm" is added to the method attribute, and new attribute threshold is added to the distances. This is returned only when dist.ret = TRUE.

**Author(s)**

Jari Oksanen, based on the code of Stephane Dray.

**References**


See Also

spantree.

Examples

## Example from Borcard & Legendre (2002)
data(mite.xy)
model <- pcnm(dist(mite.xy))
par <- par(mfrow=c(1,3))
## Map of PCNM in the sample plot
ordisurf(mite.xy, scores(model, choi=1), bubble = 4, main = "PCNM 1")
ordisurf(mite.xy, scores(model, choi=2), bubble = 4, main = "PCNM 2")
ordisurf(mite.xy, scores(model, choi=3), bubble = 4, main = "PCNM 3")
par(op)
## Plot first PCNM against each other
ordisplom(model, choices=1:4)
## Weighted PCNM for CCA
data(mite)
model <- rowSums(mite)/sum(mite)
modelw <- pcnmw(dist(mite.xy), w = model)
ord <- cca(mite - scores(modelw))
## Multiscale ordination: residual variance should have no distance
## trend
msoplot(mso(ord, mite.xy))

permat

**Matrix Permutation Algorithms for Presence-Absence and Count Data**

Description

Individual (for count data) or incidence (for presence-absence data) based null models can be generated for community level simulations. Options for preserving characteristics of the original matrix (rows/columns sums, matrix fill) and restricted permutations (based on strata) are discussed in the Details section.

Usage

permatfull(m, fixedmar = "both", shuffle = "both", strata = NULL,
          mtype = "count", times = 99, ...)
permatswap(m, method = "quasiswap", fixedmar="both", shuffle = "both",
          strata = NULL, mtype = "count", times = 99,
          burnin = 0, thin = 1, ...)
## S3 method for class 'permat'
print(x, digits = 3, ...)
## S3 method for class 'permat'
summary(object, ...)
## S3 method for class 'summary.permat'
print(x, digits = 2, ...)
```
## S3 method for class 'permat'
plot(x, type = "bray", ylab, xlab, col, lty,
     lowess = TRUE, plot = TRUE, text = TRUE, ...)
## S3 method for class 'permat'
lines(x, type = "bray", ...)
## S3 method for class 'permat'
as.ts(x, type = "bray", ...)
## S3 method for class 'permat'
as.mcmc(x)
```

### Arguments

**m**
A community data matrix with plots (samples) as rows and species (taxa) as columns.

**fixedmar**
character, stating which of the row/column sums should be preserved ("none", "rows", "columns", "both").

**strata**
Numeric vector or factor with length same as nrow(m) for grouping rows within strata for restricted permutations. Unique values or levels are used.

**mtype**
Matrix data type, either "count" for count data, or "prab" for presence-absence type incidence data.

**times**
Number of permuted matrices.

**method**
Character for method used for the swap algorithm ("swap", "tswap", "quasiswap", "backtrack") as described for function `make.commsim`. If mtype="count" the "quasiswap", "swap", "swsh" and "abuswap" methods are available (see details).

**shuffle**
Character, indicating whether individuals ("ind"), samples ("samp") or both ("both") should be shuffled, see details.

**burnin**
Number of null communities discarded before proper analysis in sequential ("swap", "tswap") methods.

**thin**
Number of discarded permuted matrices between two evaluations in sequential ("swap", "tswap") methods.

**x**, **object**
Object of class "permat"

**digits**
Number of digits used for rounding.

**ylab, xlab, col, lty**
Graphical parameters for the plot method.

**type**
Character, type of plot to be displayed: "bray" for Bray-Curtis dissimilarities, "chisq" for Chi-squared values.

**lowess, plot, text**
Logical arguments for the plot method, whether a locally weighted regression curve should be drawn, the plot should be drawn, and statistic values should be printed on the plot.

**...**
Other arguments passed to `simulate.nullmodel` or methods.
Details

The function `permatfull` is useful when matrix fill is allowed to vary, and matrix type is count. The `fixedmar` argument is used to set constraints for permutation. If none of the margins are fixed, cells are randomised within the matrix. If rows or columns are fixed, cells within rows or columns are randomised, respectively. If both margins are fixed, the `r2dttable` function is used that is based on Patefield's (1981) algorithm. For presence absence data, matrix fill should be necessarily fixed, and `permatfull` is a wrapper for the function `make.commsim`. The `r00`, `r0`, `c0`, `quasiswap` algorithms of `make.commsim` are used for "none", "rows", "columns", "both" values of the `fixedmar` argument, respectively.

The shuffle argument only have effect if the `mtype` = "count" and `permatfull` function is used with "none", "rows", "columns" values of `fixedmar`. All other cases for count data are individual based randomisations. The "samp" and "both" options result fixed matrix fill. The "both" option means that individuals are shuffled among non zero cells ensuring that there are no cell with zeros as a result, then cell (zero and new valued cells) are shuffled.

The function `permatswap` is useful when with matrix fill (i.e. the proportion of empty cells) and row/columns sums should be kept constant. `permatswap` uses different kinds of swap algorithms, and row and columns sums are fixed in all cases. For presence-absence data, the swap and tswap methods of `make.commsim` can be used. For count data, a special swap algorithm ('swapcount') is implemented that results in permuted matrices with fixed marginals and matrix fill at the same time.

The 'quasiswapcount' algorithm (method="quasiswap" and `mtype="count"`) uses the same trick as Carsten Dormann's `swap.web` function in the package `bipartite`. First, a random matrix is generated by the `r2dttable` function retaining row and column sums. Then the original matrix fill is reconstructed by sequential steps to increase or decrease matrix fill in the random matrix. These steps are based on swapping 2x2 submatrices (see 'swapcount' algorithm for details) to maintain row and column totals. This algorithm generates independent matrices in each step, so burnin and thin arguments are not considered. This is the default method, because this is not sequential (as `swapcount` is) so independence of subsequent matrices does not have to be checked.

The `swapcount` algorithm (method="swap" and `mtype="count"`) tries to find 2x2 submatrices (identified by 2 random row and 2 random column indices), that can be swapped in order to leave column and row totals and fill unchanged. First, the algorithm finds the largest value in the submatrix that can be swapped (d) and whether in diagonal or antidiagonal way. Submatrices that contain values larger than zero in either diagonal or antidiagonal position can be swapped. Swap means that the values in diagonal or antidiagonal positions are decreased by d, while remaining cells are increased by d. A swap is made only if fill doesn't change. This algorithm is sequential, subsequent matrices are not independent, because swaps modify little if the matrix is large. In these cases many burnin steps and thinning is needed to get independent random matrices. Although this algorithm is implemented in C, large burnin and thin values can slow it down considerably. WARNING: according to simulations, this algorithm seems to be biased and non random, thus its use should be avoided!

The algorithm "swsh" in the function `permatswap` is a hybrid algorithm. First, it makes binary quasiswaps to keep row and column incidences constant, then non-zero values are modified according to the shuffle argument (only "samp" and "both" are available in this case, because it is applied only on non-zero values). It also recognizes the `fixedmar` argument which cannot be "both" ( vegam versions <= 2.0 had this algorithm with `fixedmar = "none"`).

The algorithm "abuswap" produces two kinds of null models (based on `fixedmar="columns"` or `fixedmar="rows"`) as described in Hardy (2008; randomization scheme 2x and 3x, respectively).
These preserve column and row occurrences, and column or row sums at the same time. (Note that similar constraints can be achieved by the non sequential "swsh" algorithm with fixedmar argument set to "columns" or "rows", respectively.)

Constraints on row/column sums, matrix fill, total sum and sums within strata can be checked by the summary method. plot method is for visually testing the randomness of the permuted matrices, especially for the sequential swap algorithms. If there are any tendency in the graph, higher burnin and thin values can help for sequential methods. New lines can be added to existing plot with the lines method.

Unrestricted and restricted permutations: if strata is NULL, functions perform unrestricted permutations. Otherwise, it is used for restricted permutations. Each strata should contain at least 2 rows in order to perform randomization (in case of low row numbers, swap algorithms can be rather slow). If the design is not well balanced (i.e. same number of observations within each stratum), permuted matrices may be biased because same constraints are forced on submatrices of different dimensions. This often means, that the number of potential permutations will decrease with their dimensions. So the more constraints we put, the less randomness can be expected.

The plot method is useful for graphically testing for trend and independence of permuted matrices. This is especially important when using sequential algorithms ("swap", "tswap", "abuswap"). The as . ts method can be used to extract Bray-Curtis dissimilarities or Chi-squared values as time series. This can further used in testing independence (see Examples). The method as . mcmc is useful for accessing diagnostic tools available in the coda package.

Value

Functions permatfull and permatswap return an object of class "permat" containing the function call (call), the original data matrix used for permutations (orig) and a list of permuted matrices with length times (perm).

The summary method returns various statistics as a list (including mean Bray-Curtis dissimilarities calculated pairwise among original and permuted matrices, Chi-square statistics, and check results of the constraints; see Examples). Note that when strata is used in the original call, summary calculation may take longer.

The plot creates a plot as a side effect.

The as . ts method returns an object of class "ts".

Author(s)

Péter Sólymos, <solymos@ualberta.ca> and Jari Oksanen

References

Original references for presence-absence algorithms are given on help page of make.commsim.


See Also

For other functions to permute matrices: make.commsim, r2dtable, sample, swap.web.

For the use of these permutation algorithms: oecosimu, adipart, hier.simu.

For time-series diagnostics: Box.test, lag.plot, tsdiag, ar, arima

For underlying low level implementation: commsim and nullmodel.

Examples

```r
## A simple artificial community data matrix.
m <- matrix(c(
    1,3,2,0,3,1,
    0,2,1,0,2,1,
    0,0,1,2,0,3,
    0,0,0,1,4,3
  ), 4, 6, byrow=TRUE)
## Using the quasiswap algorithm to create a
## list of permuted matrices, where
## row/columns sums and matrix fill are preserved:
x1 <- permatswap(m, "quasiswap")
summary(x1)
## Unrestricted permutation retaining
## row/columns sums but not matrix fill:
x2 <- permatfull(m)
summary(x2)
## Unrestricted permutation of presence-absence type
## not retaining row/columns sums:
x3$orig ## note: original matrix is binarized!
summary(x3)
## Restricted permutation,
## check sums within strata:
x4 <- permatfull(m, strata=c(1,1,2,2))
summary(x4)

## NOTE: 'times' argument usually needs to be >= 99
## here much lower value is used for demonstration

## Not sequential algorithm
data(BCI)
a <- permatswap(BCI, "quasiswap", times=19)
## Sequential algorithm
b <- permatswap(BCI, "abuswap", fixedmar="col",
    burnin=0, thin=100, times=19)
opar <- par(mfrow=c(2,2))
plot(a, main="Not sequential")
plot(b, main="Sequential")
plot(a, "chisq")
plot(b, "chisq")
par(opar)
## Extract Bray-Curtis dissimilarities
## as time series
```
permustats

Extract, Analyse and Display Permutation Results

Description

The permustats function extracts permutation results of `vegan` functions. Its support functions can find quantiles and standardized effect sizes, plot densities and Q-Q plots.

Usage

permustats(x, ...)  
# S3 method for class 'permustats'
summary(object, interval = 0.95, alternative, ...)
# S3 method for class 'permustats'
densityplot(x, data, xlab = "Permutations", ...)
# S3 method for class 'permustats'
density(x, observed = TRUE, ...)
# S3 method for class 'permustats'
qqnorm(y, observed = TRUE, ...)
# S3 method for class 'permustats'
qqmath(x, data, observed = TRUE, sd.scale = FALSE,
ylab = "Permutations", ...)
# S3 method for class 'permustats'
boxplot(x, scale = FALSE, names, ...)

Arguments

object, x, y The object to be handled.
interval numeric; the coverage interval reported.
alternative A character string specifying the limits used for the interval and the direction of the test when evaluating the p-values. Must be one of "two.sided" (both upper and lower limit), "greater" (upper limit), "less" (lower limit). Usually alternative is given in the result object, but it can be specified with this argument.
xlab, ylab Arguments of `densityplot` and `qqmath` functions.
observed Add observed statistic among permutations.

Example:

bc <- as.ts(b)
## Lag plot
lag.plot(bc)
## First order autoregressive model
mar <- arima(bc, c(1,0,0))
mar
## Ljung-Box test of residuals
Box.test(residuals(mar))
## Graphical diagnostics
tsdia(mar)
sd.scale Scale permutations to unit standard deviation and observed statistic to standardized effect size.
data Ignored.
scale Use standardized effect size (SES).
names Names of boxes (default: names of statistics).
... Other arguments passed to the function. In density these are passed to density.default, and in boxplot to boxplot.default.

Details

The permustats function extracts permutation results and observed statistics from several vegan functions that perform permutations or simulations.

The summary method of permustats estimates the standardized effect sizes (SES) as the difference of observed statistic and mean of permutations divided by the standard deviation of permutations (also known as z-values). It also prints the the mean, median, and limits which contain interval percent of permuted values. With the default (interval = 0.95), for two-sided test these are (2.5%, 97.5%) and for one-sided tests either 5% or 95% quantile and the p-value depending on the test direction. The mean, quantiles and z values are evaluated from permuted values without observed statistic, but the p-value is evaluated with the observed statistic. The intervals and the p-value are evaluated with the same test direction as in the original test, but this can be changed with argument alternative. Several permustats objects can be combined with c function. The c function checks that statistics are equal, but performs no other sanity tests.

The density and densityplot methods display the kernel density estimates of permuted values. When observed value of the statistic is included in the permuted values, the densityplot method marks the observed statistic as a vertical line. However the density method uses its standard plot method and cannot mark the observed value.

The qqnorm and qqmath display Q-Q plots of permutations, optionally together with the observed value (default) which is shown as horizontal line in plots. qqnorm plots permutation values against standard Normal variate. qqmath defaults to the standard Normal as well, but can accept other alternatives (see standard qqmath). The qqmath function can also plot observed statistic as standardized effect size (SES) with standandized permutations (argument sd.scale). The permutations are standardized without the observed statistic, similarly as in summary.

Functions density and qqnorm are based on standard R methods and accept their arguments. They only handle one statistic, and cannot be used when several test statistic were evaluated. The densityplot and qqmath are lattice graphics, and can be used either for one or for several statistics. All these functions pass arguments to their underlying functions; see their documentation.

Functions qqmath and densityplot default to use same axis scaling in all subplots of the lattice. You can use argument scales to set independent scaling for subplots when this is appropriate (see xyplot for an exhaustive list of arguments).

Function boxplot draws the box-and-whiskers plots of effect size, or the difference of permutations and observed statistic. If scale = TRUE, permutations are standardized to unit standard deviation, and the plot will show the standardized effect sizes.

The permustats can extract permutation statistics from the results of adonis2, adonis, anosim, anova.cca, mantel, mantel.partial, mrpp, oecosimu, ordiareatest, permutest.cca, protest, and permutest.betadisper.
Value

The permustats function returns an object of class "permustats". This is a list of items "statistic" for observed statistics, permutations which contains permuted values, and alternative which contains text defining the character of the test ("two.sided", "less" or "greater"). The qqnorm and density methods return their standard result objects.

Author(s)

Jari Oksanen with contributions from Gavin L. Simpson (permustats.permutest.betadisper method and related modifications to summary.permustats and the print method) and Eduard Szöcs (permustats.anova.cca).

See Also
density, densityplot, qqnorm, qqmath.

Examples

data(dune, dune.env)
mod <- adonis2(dune ~ Management + A1, data = dune.env)
  ## use permustats
perm <- permustats(mod)
summary(perm)
densityplot(perm)
qqmath(perm)
boxplot(perm, scale=TRUE, lty=1, pch=16, cex=0.6, col="hotpink", ylab="SES")
abline(h=0, col="skyblue")
  ## example of multiple types of statistic
mod <- with(dune.env, betadisper(vegdist(dune), Management))
pmod <- permutest(mod, nperm = 99, pairwise = TRUE)
perm <- permustats(pmod)
summary(perm, interval = 0.90)

permutations Permutation tests in Vegan

Description

From version 2.2-0, vegan has significantly improved access to restricted permutations which brings it into line with those offered by Canoco. The permutation designs are modelled after the permutation schemes of Canoco 3.1 (ter Braak, 1990).

vegan currently provides for the following features within permutation tests:

1. Free permutation of DATA, also known as randomisation,
2. Free permutation of DATA within the levels of a grouping variable,
3. Restricted permutations for line transects or time series,
4. Permutation of groups of samples whilst retaining the within-group ordering,
5. Restricted permutations for spatial grids,
6. Blocking, samples are never permuted between blocks, and
7. Split-plot designs, with permutation of whole plots, split plots, or both.

Above, we use \textit{DATA} to mean either the observed data themselves or some function of the data, for example the residuals of an ordination model in the presence of covariables.

These capabilities are provided by functions from the \textit{permute} package. The user can request a particular type of permutation by supplying the \textit{permutations} argument of a function with an object returned by \textit{how}, which defines how samples should be permuted. Alternatively, the user can simply specify the required number of permutations and a simple randomisation procedure will be performed. Finally, the user can supply a matrix of permutations (with number of rows equal to the number of permutations and number of columns equal to the number of observations in the data) and \textit{vegan} will use these permutations instead of generating new permutations.

The majority of functions in \textit{vegan} allow for the full range of possibilities outlined above. Exceptions include \textit{kendallNpost} and \textit{kendallNglobal}.

The Null hypothesis for the first two types of permutation test listed above assumes free exchangeability of \textit{DATA} (within the levels of the grouping variable, if specified). Dependence between observations, such as that which arises due to spatial or temporal autocorrelation, or more-complicated experimental designs, such as split-plot designs, violates this fundamental assumption of the test and requires more complex restricted permutation test designs. It is these designs that are available via the \textit{permute} package and to which \textit{vegan} provides access from version 2.2-0 onwards.

Unless otherwise stated in the help pages for specific functions, permutation tests in \textit{vegan} all follow the same format/structure:

1. An appropriate test statistic is chosen. Which statistic is chosen should be described on the help pages for individual functions.
2. The value of the test statistic is evaluate for the observed data and analysis/model and recorded. Denote this value \(x_0\).
3. The \textit{DATA} are randomly permuted according to one of the above schemes, and the value of the test statistic for this permutation is evaluated and recorded.
4. Step 3 is repeated a total of \(n\) times, where \(n\) is the number of permutations requested. Denote these values as \(x_i\), where \(i = 1, \ldots, n\)
5. Count the number of values of the test statistic, \(x_i\), in the Null distribution that are as extreme as test statistic for the observed data \(x_0\). Denote this count as \(N\).
   We use the phrase \textit{as extreme} to include cases where a two-sided test is performed and large negative values of the test statistic should be considered.
6. The permutation p-value is computed as
   \[
p = \frac{N + 1}{n + 1}
   \]

The above description illustrates why the default number of permutations specified in \textit{vegan} functions takes values of 199 or 999 for example. Pretty \(p\) values are achieved because the +1 in the denominator results in division by 200 or 1000, for the 199 or 999 random permutations used in the test.

The simple intuition behind the presence of +1 in the numerator and denominator is that these represent the inclusion of the observed value of the statistic in the Null distribution (e.g. Manly
Phipson & Smyth (2010) present a more compelling explanation for the inclusion of $+1$ in the numerator and denominator of the $p$ value calculation.

Fisher (1935) had in mind that a permutation test would involve enumeration of all possible permutations of the data yielding an exact test. However, doing this complete enumeration may not be feasible in practice owing to the potentially vast number of arrangements of the data, even in modestly-sized data sets with free permutation of samples. As a result we evaluate the $p$ value as the tail probability of the Null distribution of the test statistic directly from the random sample of possible permutations. Phipson & Smyth (2010) show that the naive calculation of the permutation $p$ value is

$$p = \frac{N}{n}$$

which leads to an invalid test with incorrect type I error rate. They go on to show that by replacing the unknown tail probability (the $p$ value) of the Null distribution with the biased estimator

$$p = \frac{N + 1}{n + 1}$$

that the positive bias induced is of just the right size to account for the uncertainty in the estimation of the tail probability from the set of randomly sampled permutations to yield a test with the correct type I error rate.

The estimator described above is correct for the situation where permutations of the data are samples randomly without replacement. This is not strictly what happens in vegan because permutations are drawn pseudo-randomly independent of one another. Note that the actual chance of this happening in practice is small but the functions in permute do not guarantee to generate a unique set of permutations unless complete enumeration of permutations is requested. This is not feasible for all but the smallest of data sets or restrictive of permutation designs, but in such cases the chance of drawing a set of permutations with repeats is lessened as the sample size, and thence the size of set of all possible permutations, increases.

Under the situation of sampling permutations with replacement then, the tail probability $p$ calculated from the biased estimator described above is somewhat conservative, being too large by an amount that depends on the number of possible values that the test statistic can take under permutation of the data (Phipson & Smyth, 2010). This represents a slight loss of statistical power for the conservative $p$ value calculation used here. However, unless sample sizes are small and the the permutation design such that the set of values that the test statistic can take is also small, this loss of power is unlikely to be critical.

The minimum achievable $p$-value is

$$p_{\text{min}} = \frac{1}{n + 1}$$

and hence depends on the number of permutations evaluated. However, one cannot simply increase the number of permutations ($n$) to achieve a potentially lower $p$-value unless the number of observations available permits such a number of permutations. This is unlikely to be a problem for all but the smallest data sets when free permutation (randomisation) is valid, but in restricted permutation designs with a low number of observations, there may not be as many unique permutations of the data as you might desire to reach the required level of significance.
It is currently the responsibility of the user to determine the total number of possible permutations for their *DATA*. The number of possible permutations allowed under the specified design can be calculated using `numPerms` from the *permute* package. Heuristics employed within the `shuffleSet` function used by *vegan* can be triggered to generate the entire set of permutations instead of a random set. The settings controlling the triggering of the complete enumeration step are contained within a permutation design created using `link[permute]{how}` and can be set by the user. See `how` for details.

Limits on the total number of permutations of *DATA* are more severe in temporally or spatially ordered data or experimental designs with low replication. For example, a time series of \( n = 100 \) observations has just 100 possible permutations including the observed ordering.

In situations where only a low number of permutations is possible due to the nature of *DATA* or the experimental design, enumeration of all permutations becomes important and achievable computationally.

Above, we have provided only a brief overview of the capabilities of *vegan* and *permute*. To get the best out of the new functionality and for details on how to set up permutation designs using `how`, consult the vignette *Restricted permutations; using the permute package* supplied with *permute* and accessible via vignette("permutations", package = "permute").

**Author(s)**

Gavin L. Simpson

**References**


See also:


**See Also**

`permutest` for the main interface in *vegan*. See also `how` for details on permutation design specification, `shuffleSet` for the code used to generate a set of permutations, `numPerms` for a function to return the size of the set of possible permutations under the current design.
Permutation test of multivariate homogeneity of groups dispersions (variances)

Description

Implements a permutation-based test of multivariate homogeneity of group dispersions (variances) for the results of a call to betadisper.

Usage

```r
## S3 method for class 'betadisper'
permutest(x, pairwise = FALSE,
          permutations = 999,
          parallel = getOption("mc.cores"),
          ...)```

Arguments

- `x`: an object of class "betadisper", the result of a call to betadisper.
- `pairwise`: logical; perform pairwise comparisons of group means?
- `permutations`: a list of control values for the permutations as returned by the function how, or the number of permutations required, or a permutation matrix where each row gives the permuted indices.
- `parallel`: Number of parallel processes or a predefined socket cluster. With parallel = 1 uses ordinary, non-parallel processing.
- `...`: Arguments passed to other methods.

Details

To test if one or more groups is more variable than the others, ANOVA of the distances to group centroids can be performed and parametric theory used to interpret the significance of F. An alternative is to use a permutation test. permutest.betadisper permutes model residuals to generate a permutation distribution of F under the Null hypothesis of no difference in dispersion between groups.

Pairwise comparisons of group mean dispersions can be performed by setting argument pairwise to TRUE. A classical t test is performed on the pairwise group dispersions. This is combined with a permutation test based on the t statistic calculated on pairwise group dispersions. An alternative to the classical comparison of group dispersions, is to calculate Tukey’s Honest Significant Differences between groups, via TukeyHSD.betadisper.

Value

permutest.betadisper returns a list of class "permutest.betadisper" with the following components:
tab

the ANOVA table which is an object inheriting from class "data.frame".

pairwise

a list with components observed and permuted containing the observed and
permuted p-values for pairwise comparisons of group mean distances (disper-
sions or variances).

groups

character; the levels of the grouping factor.

control

a list, the result of a call to how.

Author(s)

Gavin L. Simpson

References


See Also

For the main fitting function see *betadisper*. For an alternative approach to determining which
groups are more variable, see *TukeyHSD.betadisper*.

Examples

data(varespec)

```r
## Bray-Curtis distances between samples
dis <- vegdist(varespec)

## First 16 sites grazed, remaining 8 sites ungrazed
groups <- factor(c(rep(1,16), rep(2,8)), labels = c("grazed","ungrazed"))

## Calculate multivariate dispersions
mod <- betadisper(dis, groups)

## Perform test
anova(mod)

## Permutation test for F
pmod <- permutest(mod, permutations = 99, pairwise = TRUE)

## Tukey's Honest Significant Differences
(mod.HSD <- TukeyHSD(mod))
plot(mod.HSD)

## Has permustats() method
pstat <- permustats(pmod)
densityplot(pstat, scales = list(x = list(relation = "free")))
qqmath(pstat, scales = list(relation = "free"))
```
plot.cca

Plot or Extract Results of Constrained Correspondence Analysis or Redundancy Analysis

Description

Functions to plot or extract results of constrained correspondence analysis (cca), redundancy analysis (rda) or constrained analysis of principal coordinates (capscale).

Usage

```r
## S3 method for class 'cca'
plot(x, choices = c(1, 2), display = c("sp", "wa", "cn"),
     scaling = "species", type, xlim, ylim, const,
     correlation = FALSE, hill = FALSE, ...)
## S3 method for class 'cca'
text(x, display = "sites", labels, choices = c(1, 2),
     scaling = "species", arrow.mul, head.arrow = 0.05, select, const,
     axis.bp = FALSE, correlation = FALSE, hill = FALSE, ...)
## S3 method for class 'cca'
points(x, display = "sites", choices = c(1, 2),
     scaling = "species", arrow.mul, head.arrow = 0.05, select, const,
     axis.bp = FALSE, correlation = FALSE, hill = FALSE, ...)
## S3 method for class 'cca'
scores(x, choices = c(1,2), display = c("sp","wa","cn"),
     scaling = "species", hill = FALSE, ...)
## S3 method for class 'rda'
scores(x, choices = c(1,2), display = c("sp","wa","cn"),
     scaling = "species", const, correlation = FALSE, ...)
## S3 method for class 'cca'
summary(object, scaling = "species", axes = 6,
     display = c("sp", "wa", "1c", "bp", "cn"),
     digits = max(3,getOption("digits") - 3),
     correlation = FALSE, hill = FALSE, ...)
## S3 method for class 'summary.cca'
print(x, digits = x$digits, head = NA, tail = head, ...)
## S3 method for class 'summary.cca'
head(x, n = 6, tail = 0, ...)
## S3 method for class 'summary.cca'
tail(x, n = 6, head = 0, ...)
```

Arguments

- **x**, `object` A cca result object.
- **choices** Axes shown.
Scores shown. These must include some of the alternatives species or sp for species scores, sites or wa for site scores, 1c for linear constraints or LC scores, or bp for biplot arrows or cn for centroids of factor constraints instead of an arrow, and reg for regression coefficients (a.k.a. canonical coefficients).

Scaling for species and site scores. Either species (2) or site (1) scores are scaled by eigenvalues, and the other set of scores is left unscaled, or with 3 both are scaled symmetrically by square root of eigenvalues. Corresponding negative values can be used in cca to additionally multiply results with $\sqrt{1/(1 - \lambda)}$). This scaling is know as Hill scaling (although it has nothing to do with Hill's rescaling of decorana). With corresponding negative values in rda, species scores are divided by standard deviation of each species and multiplied with an equalizing constant. Unscaled raw scores stored in the result can be accessed with scaling = 0.

The type of scores can also be specified as one of "none", "sites", "species", or "symmetric", which correspond to the values 0, 1, 2, and 3 respectively. Arguments correlation and hill in scores.rda and scores.cca respectively can be used in combination with these character descriptions to get the corresponding negative value.

correlation, hill

logical; if scaling is a character description of the scaling type, correlation or hill are used to select the corresponding negative scaling type; either correlation-like scores or Hill’s scaling for PCA/RDA and CA/CCA respectively. See argument scaling for details.

type

Type of plot: partial match to text for text labels, points for points, and none for setting frames only. If omitted, text is selected for smaller data sets, and points for larger.

xlim, ylim

the x and y limits (min,max) of the plot.

labels

Optional text to be used instead of row names.

arrow.mul

Factor to expand arrows in the graph. Arrows will be scaled automatically to fit the graph if this is missing.

head.arrow

Default length of arrow heads.

select

Items to be displayed. This can either be a logical vector which is TRUE for displayed items or a vector of indices of displayed items.

const

General scaling constant to rda scores. The default is to use a constant that gives biplot scores, that is, scores that approximate original data (see vignette on ‘Design Decisions’ with browseVignettes("vegan") for details and discussion). If const is a vector of two items, the first is used for species, and the second item for site scores.

axis.bp

Draw axis for biplot arrows.

axes

Number of axes in summaries.

digits

Number of digits in output.

n, head, tail

Number of rows printed from the head and tail of species and site scores. Default NA prints all.

... Parameters passed to other functions.
Details

Same plot function will be used for `cca` and `rda`. This produces a quick, standard plot with current scaling.

The plot function sets colours (col), plotting characters (pch) and character sizes (cex) to certain standard values. For a fuller control of produced plot, it is best to call plot with type="none" first, and then add each plotting item separately using text.cca or points.cca functions. These use the default settings of standard text and points functions and accept all their parameters, allowing a full user control of produced plots.

Environmental variables receive a special treatment. With display="bp", arrows will be drawn. These are labelled with text and unlabelled with points. The arrows have basically unit scaling, but if sites were scaled (scaling "sites" or "symmetric"), the scores of requested axes are adjusted relative to the axis with highest eigenvalue. With scaling = "species" or scaling = "none", the arrows will be consistent with vectors fitted to linear combination scores (display = "lc" in function envfit), but with other scaling alternatives they will differ. The basic plot function uses a simple heuristics for adjusting the unit-length arrows to the current plot area, but the user can give the expansion factor in $mulNarrow$. With display="cn" the centroids of levels of factor variables are displayed (these are available only if there were factors and a formula interface was used in cca or rda). With this option continuous variables still are presented as arrows and ordered factors as arrows and centroids. With display = "reg" arrows will be drawn for regression coefficients (a.k.a. canonical coefficients) of constraints and conditions. Biplot arrows can be interpreted individually, but regression coefficients must be interpreted all together: the LC score for each site is the sum of regressions displayed by arrows. The partialled out conditions are zero and not shown in biplot arrows, but they are shown for regressions, and show the effect that must be partialled out to get the LC scores. The biplot arrows are more standard and more easily interpreted, and regression arrows should be used only if you know that you need them.

If you want to have a better control of plots, it is best to construct the plot text and points commands which accept graphical parameters. It is important to remember to use the same scaling, correlation and hill arguments in all calls. The plot.cca command returns invisibly an ordiplot result object, and this will have consistent scaling for all its elements. The easiest way for full control of graphics is to first set up the plot frame using plot with type = "n" and all needed scores in display and save this result. The points and text commands for ordiplot will allow full graphical control (see section Examples).

Function summary lists all scores and the output can be very long. You can suppress scores by setting axes = 0 or display = NA or display = NULL. You can display some first or last (or both) rows of scores by using head or tail or explicit print command for the summary.

Palmer (1993) suggested using linear constraints ("LC scores") in ordination diagrams, because these gave better results in simulations and site scores ("WA scores") are a step from constrained to unconstrained analysis. However, McCune (1997) showed that noisy environmental variables (and all environmental measurements are noisy) destroy "LC scores" whereas "WA scores" were little affected. Therefore the plot function uses site scores ("WA scores") as the default. This is consistent with the usage in statistics and other functions in R (lda, cancor).

Value

The plot function returns invisibly a plotting structure which can be used by function identify.ordiplot to identify the points or other functions in the ordiplot family.
Author(s)

Jari Oksanen

See Also

cca, rda and capscale for getting something to plot, ordiplot for an alternative plotting routine and more support functions, and text, points and arrows for the basic routines.

Examples

data(dune)
data(dune.env)
mod <- cca(dune ~ A1 + Moisture + Management, dune.env)
## better control -- remember to set scaling etc identically
plot(mod, type="n", scaling="sites")
text(mod, dis="cn", scaling="sites")
points(mod, pch=21, col="red", bg="yellow", cex=1.2, scaling="sites")
text(mod, "species", col="blue", cex=0.8, scaling="sites")
## catch the invisible result and use ordiplot support - the example
## will make a biplot with arrows for species and correlation scaling
pca <- rda(dune)
pl <- plot(pca, type="n", scaling="sites", correlation=TRUE)
with(dune.env, points(pl, "site", pch=21, col=1, bg=Management))
text(pl, "sp", arrow=TRUE, length=0.05, col=4, cex=0.6, xpd=TRUE)
with(dune.env, legend("bottomleft", levels(Management), pch=21, pt.bg=1:4, bty="n"))
## Limited output of 'summary'
head(summary(mod), tail=2)
## Scaling can be numeric or more user-friendly names
## e.g. Hill's scaling for (C)CA
scrs <- scores(mod, scaling = "sites", hill = TRUE)
## or correlation-based scores in PCA/RDA
scrs <- scores(rda(dune ~ A1 + Moisture + Management, dune.env),
               scaling = "sites", correlation = TRUE)

prc

Principal Response Curves for Treatments with Repeated Observations

Description

Principal Response Curves (PRC) are a special case of Redundancy Analysis (rda) for multivariate responses in repeated observation design. They were originally suggested for ecological communities. They should be easier to interpret than traditional constrained ordination. They can also be used to study how the effects of a factor A depend on the levels of a factor B, that is A + A:B, in a multivariate response experiment.
Usage

prc(response, treatment, time, ...)  
## S3 method for class 'prc'
summary(object, axis = 1, scaling = "symmetric", const, digits = 4, correlation = FALSE, ...)
## S3 method for class 'prc'
plot(x, species = TRUE, select, scaling = "symmetric",  
axis = 1, correlation = FALSE, const, type = "l", xlab, ylab, ylim,  
lty = 1:5, col = 1:6, pch, legpos, cex = 0.8, ...)

Arguments

response Multivariate response data. Typically these are community (species) data. If the data are counts, they probably should be log transformed prior to the analysis.
treatment A factor for treatments.
time An unordered factor defining the observations times in the repeated design.
object, x An prc result object.
axis Axis shown (only one axis can be selected).
scaling Scaling of species scores, identical to the scaling in scores.rda. The type of scores can also be specified as one of "none", "sites", "species", or "symmetric", which correspond to the values 0, 1, 2, and 3 respectively. Argument correlation can be used in combination with these character descriptions to get the corresponding negative value.
const General scaling constant for species scores (see scores.rda for details). Lower values will reduce the range of species scores, but will not influence the regression coefficients.
digits Number of significant digits displayed.
correlation logical; if scaling is a character description of the scaling type, correlation can be used to select correlation-like scores for PCA. See argument scaling for details.
species Display species scores.
select Vector to select displayed species. This can be a vector of indices or a logical vector which is TRUE for the selected species
type Type of plot: "l" for lines, "p" for points or "b" for both.
xlab, ylab Text to replace default axis labels.
ylim Limits for the vertical axis.
lty, col, pch Line type, colour and plotting characters (defaults supplied).
legpos The position of the legend. A guess is made if this is not supplied, and NA will suppress legend.
cex Character expansion for symbols and species labels.
... Other parameters passed to functions.
Details

PRC is a special case of rda with a single factor for treatment and a single factor for time points in repeated observations. In vegan, the corresponding rda model is defined as rda(response ~ treatment * time + Condition(time)). Since the time appears twice in the model formula, its main effects will be aliased, and only the main effect of treatment and interaction terms are available, and will be used in PRC. Instead of usual multivariate ordination diagrams, PRC uses canonical (regression) coefficients and species scores for a single axis. All that the current functions do is to provide a special summary and plot methods that display the rda results in the PRC fashion. The current version only works with default contrasts (contr.treatment) in which the coefficients are contrasts against the first level, and the levels must be arranged so that the first level is the control (or a baseline). If necessary, you must change the baseline level with function relevel.

Function summary prints the species scores and the coefficients. Function plot plots coefficients against time using matplot, and has similar defaults. The graph (and PRC) is meaningful only if the first treatment level is the control, as the results are contrasts to the first level when unordered factors are used. The plot also displays species scores on the right vertical axis using function linestack. Typically the number of species is so high that not all can be displayed with the default settings, but users can reduce character size or padding (air) in linestack, or select only a subset of the species. A legend will be displayed unless suppressed with legpos = NA, and the functions tries to guess where to put the legend if legpos is not supplied.

Value

The function is a special case of rda and returns its result object (see cca.object). However, a special summary and plot methods display returns differently than in rda.

Warning

The first level of treatment must be the control: use function relevel to guarantee the correct reference level. The current version will ignore user setting of contrasts and always use treatment contrasts (contr.treatment). The time must be an unordered factor.

Author(s)

Jari Oksanen and Cajo ter Braak

References


See Also

rda, anova.cca.
Examples

```r
# Chlorpyrifos experiment and experimental design: Pesticide
# treatment in ditches (replicated) and followed over from 4 weeks
# before to 24 weeks after exposure
data(pyrifos)
week <- gl(11, 12, labels=LETTERS[1:12], labels=rep(c(-4, -1, 0, 1, 2, 4, 8, 12, 15, 19, 24))

dose <- factor(rep(c(0.1, 0.5, 0.9, 0.44, 0.6, 0.1, 0.44, 0.9, 0.6, 1)), labels=LETTERS[1:12])
ditch <- gl(12, 1, length=132)

# PRC
mod <- prc(pyrifos, dose, week)

summary(mod)  # RDA
logabu <- colSums(pyrifos)
plot(mod, select = logabu > 100)

# Ditches are randomized, we have a time series, and are only
# interested in the first axis
ctrl <- how(plots = Plots(strata = ditch, type = "free"),
within = Within(type = "series"), nperm = 99)
anova(mod, permutations = ctrl, first=TRUE)
```

---

**predict.cca**

**Prediction Tools for [Constrained] Ordination (CCA, RDA, DCA, CA, PCA)**

**Description**

Function `predict` can be used to find site and species scores or estimates of the response data with new data sets. Function `calibrate` estimates values of constraints with new data set. Functions `fitted` and `residuals` return estimates of response data.

**Usage**

```r
# S3 method for class 'cca'
fitted(object, model = c("CCA", "CA", "pCCA"),
    type = c("response", "working"), ...)
# S3 method for class 'capscale'
fitted(object, model = c("CCA", "CA", "pCCA", "Imaginary"),
    type = c("response", "working"), ...)
# S3 method for class 'cca'
residuals(object, ...)
# S3 method for class 'cca'
predict(object, newdata, type = c("response", "wa", "sp", "lc", "working"),
    rank = "full", model = c("CCA", "CA"), scaling = "none",
    hill = FALSE, ...)
# S3 method for class 'rda'
predict(object, newdata, type = c("response", "wa", "sp", "lc", "working"),
    rank = "full", model = c("CCA", "CA"), scaling = "none",
    correlation = FALSE, const, ...)
```
## predict.cca

```r
## S3 method for class 'cca'
calibrate(object, newdata, rank = "full", ...)
## S3 method for class 'cca'
coef(object, norm = FALSE, ...)
## S3 method for class 'decorana'
predict(object, newdata, rank = 4, ...)
```

### Arguments

- **object**: A result object from `cca`, `rda`, `dbrda`, `capscale` or `decorana`.
- **model**: Show constrained ("CCA"), unconstrained ("CA") or conditioned "partial" ("pCCA") results. For fitted method of `capscale` this can also be "Imaginary" for imaginary components with negative eigenvalues.
- **newdata**: New data frame to be used in prediction or in calibration. Usually this a new community data frame, but with type = "lc" and for constrained component with type = "response" and type = "working" it must be a data frame of constraints. The newdata must have the same number of rows as the original community data for a `cca` result with type = "response" or type = "working". If the original model had row or column names, then new data must contain rows or columns with the same names (row names for species scores, column names for "wa" scores and constraint names of "lc" scores). In other cases the rows or columns must match directly.
- **type**: The type of prediction, fitted values or residuals: "response" scales results so that the same ordination gives the same results, and "working" gives the values used internally, that is after Chi-square standardization in `cca` and scaling and centring in `rda`. In `capscale` and `dbrda` the "response" gives the dissimilarities, and "working" the internal data structure analysed in the ordination. Alternative "wa" gives the site scores as weighted averages of the community data, "lc" the site scores as linear combinations of environmental data, and "sp" the species scores. In `predict.decorana` the alternatives are scores for "sites" or "species".
- **rank**: The rank or the number of axes used in the approximation. The default is to use all axes (full rank) of the "model" or all available four axes in `predict.decorana`.
- **scaling**: logical, character, or numeric; Scaling or predicted scores with the same meaning as in `cca`, `rda`, `dbrda`, and `capscale`. See `scores.cca` for further details on acceptable values.
- **correlation**: logical; correlation-like scores or Hill’s scaling as appropriate for RDA and CCA respectively. See `scores.cca` for additional details.
- **const**: Constant multiplier for RDA scores. This will be used only when scaling is not FALSE, and the default value will give similar scaling as in `scores.rda`.
- **norm**: Coefficients for variables that are centred and scaled to unit norm.
- **...**: Other parameters to the functions.
Details

Function `fitted` gives the approximation of the original data matrix or dissimilarities from the ordination result either in the scale of the response or as scaled internally by the function. Function `residuals` gives the approximation of the original data matrix, and fitted and residuals do not add up to the original data. Functions `fitted` and `residuals` for `dbrda` and `capscale` give the dissimilarities with type = "response", but these are not additive. However, the "working" scores are additive for `capscale` (but not for `dbrda`). The fitted and residuals for `capscale` and `dbrda` will include the additive constant if that was requested in the function call. All variants of `fitted` and `residuals` are defined so that for model `mod <- cca(y ~ x)`, `cca(fitted(mod))` is equal to constrained ordination, and `cca(residuals(mod))` is equal to unconstrained part of the ordination.

Function `predict` can find the estimate of the original data matrix or dissimilarities (type = "response") with any rank. With `rank = "full"` it is identical to `fitted`. In addition, the function can find the species scores or site scores from the community data matrix for `cca` or `rda`. The function can be used with new data, and it can be used to add new species or site scores to existing ordinations. The function returns (weighted) orthonormal scores by default, and you must specify explicit scaling to add those scores to ordination diagrams. With type = "wa" the function finds the site scores from species scores. In that case, the new data can contain new sites, but species must match in the original and new data. With type="sp" the function finds species scores from site constraints (linear combination scores). In that case the new data can contain new species, but sites must match in the original and new data. With type = "lc" the function finds the linear combination scores for sites from environmental data. In that case the new data frame must contain all constraining and conditioning environmental variables of the model formula. With type = "response" or type = "working" the new data must contain environmental variables if constrained component is desired, and community data matrix if residual or unconstrained component is desired. With these types, the function uses `newdata` to find new "lc" (constrained) or "wa" scores (unconstrained) and then finds the response or working data from these new row scores and species scores. The original site (row) and species (column) weights are used for type = "response" and type = "working" in correspondence analysis (`cca`) and therefore the number of rows must match in the original data and `newdata`.

If a completely new data frame is created, extreme care is needed defining variables similarly as in the original model, in particular with (ordered) factors. If ordination was performed with the formula interface, the `newdata` can be a data frame or matrix, but extreme care is needed that the columns match in the original and `newdata`.

Function `calibrate.cca` finds estimates of constraints from community ordination or "wa" scores from `cca`, `rda` and `capscale`. This is often known as calibration, bioindication or environmental reconstruction. Basically, the method is similar to projecting site scores onto biplot arrows, but it uses regression coefficients. The function can be called with `newdata` so that cross-validation is possible. The `newdata` may contain new sites, but species must match in the original and new data. The function does not work with 'partial' models with `Condition` term, and it cannot be used with `newdata` for `capscale` or `dbrda` results. The results may only be interpretable for continuous variables.

Function `coef` will give the regression coefficients from centred environmental variables (constraints and conditions) to linear combination scores. The coefficients are for unstandardized environmental variables. The coefficients will be NA for aliased effects.
Function `predict.decorana` is similar to `predict.cca`. However, `type = "species"` is not available in detrended correspondence analysis (DCA), because detrending destroys the mutual reciprocal averaging (except for the first axis when rescaling is not used). Detrended CA does not attempt to approximate the original data matrix, so `type = "response"` has no meaning in detrended analysis (except with `rank = 1`).

**Value**

The functions return matrices, vectors or dissimilarities as is appropriate.

**Author(s)**

Jari Oksanen.

**References**


**See Also**

`cca`, `rda`, `dbrda`, `capscale`, `decorana`, `vif`, `goodness.cca`.

**Examples**

data(dune)
data(dune.env)
mod <- cca(dune ~ A1 + Management + Condition(Moisture), data=dune.env)
# Definition of the concepts 'fitted' and 'residuals'
mod
cca(fitted(mod))
cca(residuals(mod))
# Remove rare species (freq=1) from 'cca' and find their scores
# 'passively'.
freq <- specnumber(dune, MARGIN=2)
freq
mod <- cca(dune[, freq>1] ~ A1 + Management + Condition(Moisture), dune.env)
predict(mod, type="sp", newdata=dune[, freq==1], scaling="species")
# New sites
predict(mod, type="lc", new=data.frame(A1 = 3, Management="NM", Moisture="2"), scal=2)
# Calibration and residual plot
mod <- cca(dune ~ A1 + Moisture, dune.env)
pred <- calibrate(mod)
pred
with(dune.env, plot(A1, pred[,"A1"] - A1, ylab="Prediction Error"))
abline(h=0)
Procrustes Rotation of Two Configurations and PROTEST

Description

Function procrustes rotates a configuration to maximum similarity with another configuration. Function protest tests the non-randomness (significance) between two configurations.

Usage

procrustes(X, Y, scale = TRUE, symmetric = FALSE, scores = "sites", ...)
## S3 method for class 'procrustes'
summary(object, digits = getOption("digits"), ...)
## S3 method for class 'procrustes'
plot(x, kind=1, choices=c(1,2), to.target = TRUE,
      type = "p", xlab, ylab, main, ar.col = "blue", len=0.05,
      cex = 0.7, ...)
## S3 method for class 'procrustes'
points(x, display = c("target", "rotated"),
       choices = c(1,2), truemean = FALSE, ...)
## S3 method for class 'procrustes'
text(x, display = c("target", "rotated"),
      choices = c(1,2), labels, truemean = FALSE, ...)
## S3 method for class 'procrustes'
lines(x, type = c("segments", "arrows"),
      choices = c(1, 2), truemean = FALSE, ...)
## S3 method for class 'procrustes'
residuals(object, ...)
## S3 method for class 'procrustes'
fitted(object, truemean = TRUE, ...)
## S3 method for class 'procrustes'
predict(object, newdata, truemean = TRUE, ...)
## S3 method for class 'procrustes'
protest(X, Y, scores = "sites", permutations = how(nperm = 999), ...)

Arguments

X Target matrix
Y Matrix to be rotated.
scale Allow scaling of axes of Y.
symmetric Use symmetric Procrustes statistic (the rotation will still be non-symmetric).
scores Kind of scores used. This is the display argument used with the corresponding scores function: see scores, scores.cca and scores.cca for alternatives.
x, object An object of class procrustes.
digits Number of digits in the output.
kind  For plot function, the kind of plot produced: kind = 1 plots shifts in two configurations, kind = 0 draws a corresponding empty plot, and kind = 2 plots an impulse diagram of residuals.

choices  Axes (dimensions) plotted.

display  Axes (dimensions) plotted.

xlab, ylab  Axis labels, if defaults unacceptable.

main  Plot title, if default unacceptable.

to.target  Show only the "target" or "rotated" matrix as points.

type  The type of plot drawn. In plot, the type can be "points" or "text" to select the marker for the tail of the arrow, or "none" for drawing an empty plot. In lines the type selects either arrows or line segments to connect target and rotated configuration.

truemean  Use the original range of target matrix instead of centring the fitted values. Function plot.procrustes needs truemean = FALSE, and adding graphical items to the plots from the original results may need truemean = TRUE.

newdata  Matrix of coordinates to be rotated and translated to the target.

permutations  a list of control values for the permutations as returned by the function how, or the number of permutations required, or a permutation matrix where each row gives the permuted indices.

ar.col  Arrow colour.

len  Width of the arrow head.

labels  Character vector of text labels. Rownames of the result object are used as default.

...  Other parameters passed to functions. In procrustes and protest parameters are passed to scores, in graphical functions to underlying graphical functions.

Details

Procrustes rotation rotates a matrix to maximum similarity with a target matrix minimizing sum of squared differences. Procrustes rotation is typically used in comparison of ordination results. It is particularly useful in comparing alternative solutions in multidimensional scaling. If scale=FALSE, the function only rotates matrix Y. If scale=TRUE, it scales linearly configuration Y for maximum similarity. Since Y is scaled to fit X, the scaling is non-symmetric. However, with symmetric=TRUE, the configurations are scaled to equal dispersions and a symmetric version of the Procrustes statistic is computed.

Instead of matrix, X and Y can be results from an ordination from which scores can extract results. Function procrustes passes extra arguments to scores, scores.cca etc. so that you can specify arguments such as scaling.

Function plot plots a procrustes object and returns invisibly an ordiplot object so that function identify.ordiplot can be used for identifying points. The items in the ordiplot object are called heads and points with kind=1 (ordination diagram) and sites with kind=2 (residuals). In ordination diagrams, the arrow heads point to the target configuration if to.target = TRUE, and to rotated configuration if to.target = FALSE. Target and original rotated axes are shown as cross.
hairs in two-dimensional Procrustes analysis, and with a higher number of dimensions, the rotated axes are projected onto plot with their scaled and centred range. Function plot passes parameters to underlying plotting functions. For full control of plots, you can draw the axes using plot with kind = 0, and then add items with points or lines. These functions pass all parameters to the underlying functions so that you can select the plotting characters, their size, colours etc., or you can select the width, colour and type of line segments or arrows, or you can select the orientation and head width of arrows.

Function residuals returns the pointwise residuals, and fitted the fitted values, either centred to zero mean (if truemean=FALSE) or with the original scale (these hardly make sense if symmetric = TRUE). In addition, there are summary and print methods.

If matrix X has a lower number of columns than matrix Y, then matrix X will be filled with zero columns to match dimensions. This means that the function can be used to rotate an ordination configuration to an environmental variable (most practically extracting the result with the fitted function). Function predict can be used to add new rotated coordinates to the target. The predict function will always translate coordinates to the original non-centred matrix. The function cannot be used with newdata for symmetric analysis.

Function protest performs symmetric Procrustes analysis repeatedly to estimate the significance of the Procrustes statistic. Function protest uses a correlation-like statistic derived from the symmetric Procrustes sum of squares ss as $r = \sqrt{1 - ss}$, and also prints the sum of squares of the symmetric analysis, sometimes called $m_{12}^2$. Function protest has own print method, but otherwise uses procrustes methods. Thus plot with a protest object yields a Procrustean superimposition plot.

**Value**

Function procrustes returns an object of class procrustes with items. Function protest inherits from procrustes, but amends that with some new items:

- **Yrot** Rotated matrix Y.
- **X** Target matrix.
- **ss** Sum of squared differences between X and Yrot.
- **rotation** Orthogonal rotation matrix.
- **translation** Translation of the origin.
- **scale** Scaling factor.
- **xmean** The centroid of the target.
- **symmetric** Type of ss statistic.
- **call** Function call.
- **t0** This and the following items are only in class protest: Procrustes correlation from non-permuted solution.
- **t** Procrustes correlations from permutations. The distribution of these correlations can be inspected with permustats function.
- **signif** Significance of t
- **permutations** Number of permutations.
- **control** A list of control values for the permutations as returned by the function how.
- **control** the list passed to argument control describing the permutation design.
Note

The function protest follows Peres-Neto & Jackson (2001), but the implementation is still after Mardia et al. (1979).

Author(s)

Jari Oksanen

References


See Also

 monoMDS, for obtaining objects for procrustes, and mantel for an alternative to protest without need of dimension reduction. See how for details on specifying the type of permutation required.

Examples

data(varespec)
vare.dist <- vegdist(wisconsin(varespec))
mds.null <- monoMDS(vare.dist, y = cmdscale(vare.dist))
mds.alt <- monoMDS(vare.dist)
vare.proc <- procrustes(mds.alt, mds.null)
vare.proc
summary(vare.proc)
plot(vare.proc)
plot(vare.proc, kind=2)
residuals(vare.proc)

---

**pyrifos**

*Response of Aquatic Invertebrates to Insecticide Treatment*

Description

The data are log transformed abundances of aquatic invertebrate in twelve ditches studied in eleven times before and after an insecticide treatment.

Usage

data(pyrifos)
Format

A data frame with 132 observations on the log-transformed \((\log(10 \times x + 1))\) abundances of 178 species. There are only twelve sites (ditches, mesocosms), but these were studied repeatedly in eleven occasions. The treatment levels, treatment times, or ditch ID’s are not in the data frame, but the data are very regular, and the example below shows how to obtain these external variables.

Details

This data set was obtained from an experiment in outdoor experimental ditches. Twelve mesocosms were allocated at random to treatments; four served as controls, and the remaining eight were treated once with the insecticide chlorpyrifos, with nominal dose levels of 0.1, 0.9, 6, and 44 \(\mu g/\text{L}\) in two mesocosms each. The example data set invertebrates. Sampling was done 11 times, from week -4 pre-treatment through week 24 post-treatment, giving a total of 132 samples (12 mesocosms times 11 sampling dates), see van den Brink & ter Braak (1999) for details. The data set contains only the species data, but the example below shows how to obtain the treatment, time and ditch ID variables.

Source

CANOCO 4 example data, with the permission of Cajo J. F. ter Braak.

References


Examples

data(pyrifos)
ditch <- gl(12, 1, length=132)
week <- gl(11, 12, labels=c(-4, -1, 0.1, 1, 2, 4, 8, 12, 15, 19, 24))
dose <- factor(rep(c(0.1, 0, 0, 0.9, 0, 44, 6, 0.1, 44, 0.9, 0, 6), 11))

radfit

Rank – Abundance or Dominance / Diversity Models

Description

Functions construct rank – abundance or dominance / diversity or Whittaker plots and fit broken-stick, preemption, log-Normal, Zipf and Zipf-Mandelbrot models of species abundance.

Usage

## Default S3 method:
radfit(x, ...)
rad.null(x, family=poisson, ...)
rad.preempt(x, family = poisson, ...)
rad.lognormal(x, family = poisson, ...)
Arguments

- **x**: Data frame, matrix or a vector giving species abundances, or an object to be plotted.
- **family**: Error distribution (passed to `glm`). All alternatives accepting `link = "log"` in `family` can be used, although not all make sense.
- **object**: A fitted result object.
- **newdata**: Ranks used for ordinations. All models can interpolate to non-integer “ranks” (although this may be approximate), but extrapolation may fail.
- **total**: The new total used for predicting abundance. Observed total count is used if this is omitted.
- **order.by**: A vector used for ordering sites in plots.
- **BIC**: Use Bayesian Information Criterion, BIC, instead of Akaike’s AIC. The penalty in BIC is \( k = \log(S) \) where \( S \) is the number of species, whereas AIC uses \( k = 2 \).
- **model**: Show only the specified model. If missing, AIC is used to select the model. The model names (which can be abbreviated) are Null, Preemption, Lognormal, Zipf, Mandelbrot.
- **legend**: Add legend of line colours.
- **as.table**: Arrange panels starting from upper left corner (passed to `xyplot`).
- **xlab, ylab**: Labels for x and y axes.
- **type**: Type of the plot, "b" for plotting both observed points and fitted lines, "p" for only points, "l" for only fitted lines, and "n" for only setting the frame.
- **log**: Use logarithmic scale for given axis. The default \( \log = "y" \) gives the traditional plot of community ecology where the preemption model is a straight line, and
with \( \log = "xy" \) Zipf model is a straight line. With \( \log = "" \) both axes are in the original arithmetic scale.

... Other parameters to functions.

Details

Rank–Abundance Dominance (RAD) or Dominance/Diversity plots (Whittaker 1965) display logarithmic species abundances against species rank order. These plots are supposed to be effective in analysing types of abundance distributions in communities. These functions fit some of the most popular models mainly following Wilson (1991).

Functions \texttt{rad.null}, \texttt{rad.preempt}, \texttt{rad.lognormal}, \texttt{rad.zipf} and \texttt{zipfbrot} fit the individual models (described below) for a single vector (row of data frame), and function \texttt{radfit} fits all models. The argument of the function \texttt{radfit} can be either a vector for a single community or a data frame where each row represents a distinct community.

Function \texttt{rad.null} fits a brokenstick model where the expected abundance of species at rank \( r \) is \( a_r = (J/S) \sum_{x=r}^{S} (1/x) \) (Pielou 1975), where \( J \) is the total number of individuals (site total) and \( S \) is the total number of species in the community. This gives a Null model where the individuals are randomly distributed among observed species, and there are no fitted parameters. Function \texttt{rad.preempt} fits the niche preemption model, a.k.a. geometric series or Motomura model, where the expected abundance \( a_r \) of species at rank \( r \) is \( a_r = J\alpha(1-\alpha)^{r-1} \). The only estimated parameter is the preemption coefficient \( \alpha \) which gives the decay rate of abundance per rank. The niche preemption model is a straight line in a RAD plot. Function \texttt{rad.lognormal} fits a log-Normal model which assumes that the logarithmic abundances are distributed Normally, or \( a_r = \exp(\log \mu + \log \sigma N) \), where \( N \) is a Normal deviate. Function \texttt{rad.zipf} fits the Zipf model \( a_r = Jp1r^\gamma \) where \( p1 \) is the fitted proportion of the most abundant species, and \( \gamma \) is a decay coefficient. The Zipf–Mandelbrot model (\texttt{rad.zipfbrot}) adds one parameter: \( a_r = Jc(r+\beta)^\gamma \) after which \( p1 \) of the Zipf model changes into a meaningless scaling constant \( c \).

Log-Normal and Zipf models are generalized linear models (\texttt{glm}) with logarithmic link function. Zipf–Mandelbrot adds one nonlinear parameter to the Zipf model, and is fitted using \texttt{nlm} for the nonlinear parameter and estimating other parameters and log-Likelihood with \texttt{glm}. Preemption model is fitted as a purely nonlinear model. There are no estimated parameters in the Null model.

The default \texttt{family} is \texttt{poisson} which is appropriate only for genuine counts (integers), but other families that accept link = "log" can be used. Families \texttt{Gamma} or \texttt{gaussian} may be appropriate for abundance data, such as cover. The best model is selected by \texttt{AIC}. Therefore ‘quasi’ families such as \texttt{quasipoisson} cannot be used: they do not have \texttt{AIC} nor log-Likelihood needed in non-linear models.

All these functions have their own \texttt{plot} functions. When \texttt{radfit} was applied for a data frame, \texttt{plot} uses \texttt{Lattice} graphics, and other \texttt{plot} functions use ordinary graphics. The ordinary graphics functions return invisibly an \texttt{ordiplot} object for observed points, and function \texttt{identify.ordiplot} can be used to label selected species. Alternatively, \texttt{radlattice} uses \texttt{Lattice} graphics to display each \texttt{radfit} model of a single site in a separate panel together with their AIC or BIC values.

Function \texttt{as.rad} is a base function to construct ordered RAD data. Its \texttt{plot} is used by other RAD \texttt{plot} functions which pass extra arguments (such as \texttt{xlab} and \texttt{log}) to this function. The function returns an ordered vector of taxa occurring in a site, and a corresponding attribute “index” of included taxa.
Value

Functions `rad.null`, `rad.preempt`, `rad.lognormal`, `zipf` and `zipfbrot` fit each a single RAD model to a single site. The result object has class "radline" and inherits from `glm`, and can be handled by some (but not all) `glm` methods.

Function `radfit` fits all models either to a single site or to all rows of a data frame or a matrix. When fitted to a single site, the function returns an object of class "radfit" with items `y` (observed values), `family`, and `models` which is a list of fitted "radline" models. When applied for a data frame or matrix, `radfit` function returns an object of class "radfit.frame" which is a list of "radfit" objects, each item names by the corresponding row name.

All result objects ("radline", "radfit", "radfit.frame") can be accessed with same method functions. The following methods are available: `AIC`, `coef`, `deviance`, `logLik`. In addition the fit results can be accessed with `fitted`, `predict` and `residuals` (inherting from `residuals.glm`). The graphical functions were discussed above in Details.

Note

The RAD models are usually fitted for proportions instead of original abundances. However, nothing in these models seems to require division of abundances by site totals, and original observations are used in these functions. If you wish to use proportions, you must standardize your data by site totals, e.g. with `decostand` and use appropriate `family` such as `Gamma`.

The lognormal model is fitted in a standard way, but I do think this is not quite correct – at least it is not equivalent to fitting Normal density to log abundances like originally suggested (Preston 1948).

Some models may fail. In particular, estimation of the Zipf-Mandelbrot model is difficult. If the fitting fails, `NA` is returned.

Wilson (1991) defined preemption model as \( r = J p_1 (1 - \alpha)^{r-1} \), where \( p_1 \) is the fitted proportion of the first species. However, parameter \( p_1 \) is completely defined by \( \alpha \) since the fitted proportions must add to one, and therefore I handle preemption as a one-parameter model.

Veiled log-Normal model was included in earlier releases of this function, but it was removed because it was flawed: an implicit veil line also appears in the ordinary log-Normal. The latest release version with `rad.veil` was 1.6-10.

Author(s)

Jari Oksanen

References


See Also

`fisherfit` and `prestonfit`. An alternative approach is to use `qqnorm` or `qqplot` with any distribution. For controlling graphics: `Lattice`, `xyplot`, `lset`. 
rankindex

**Examples**

```r
data(BCI)
mod <- rad.lognormal(BCI[5,])
mod.plot(mod)
mod <- radfit(BCI[1,])
## Standard plot overlaid for all models
## Preemption model is a line
plot(mod)
## log for both axes: Zipf model is a line
plot(mod, log = "xy")
## Lattice graphics separately for each model
radlattice(mod)
# Take a subset of BCI to save time and nerves
mod <- radfit(BCI[3:5,])
mod.plot(mod, pch="."
```

---

**rankindex**  
*Compares Dissimilarity Indices for Gradient Detection*

**Description**

Rank correlations between dissimilarity indices and gradient separation.

**Usage**

```r
rankindex(grad, veg, indices = c("euc", "man", "gow", "bra", "kul"),
    stepacross = FALSE, method = "spearman",
    metric = c("euclidean", "mahalanobis", "manhattan", "gower"),
    ...)
```

**Arguments**

- `grad` The gradient variable or matrix.
- `veg` The community data matrix.
- `indices` Dissimilarity indices compared, partial matches to alternatives in `vegdist`. Alternatively, it can be a (named) list of functions returning objects of class `dist`.
- `stepacross` Use `stepacross` to find a shorter path dissimilarity. The dissimilarities for site pairs with no shared species are set NA using `no.shared` so that indices with no fixed upper limit can also be analysed.
- `method` Correlation method used.
- `metric` Metric to evaluate the gradient separation. See Details.
- `...` Other parameters to `stepacross`.
Details

A good dissimilarity index for multidimensional scaling should have a high rank-order similarity with gradient separation. The function compares most indices in `vegdist` against gradient separation using rank correlation coefficients in `cor`. The gradient separation between each point is assessed using given metric. The default is to use Euclidean distance of continuous variables scaled to unit variance, or to use Gower metric for mixed data using function `daisy` when `grad` has factors. The other alternatives are Mahalanabis distances which are based on `grad` matrix scaled so that columns are orthogonal (uncorrelated) and have unit variance, or Manhattan distances of `grad` variables scaled to unit range.

The `indices` argument can accept any dissimilarity indices besides the ones calculated by the `vegdist` function. For this, the argument value should be a (possibly named) list of functions. Each function must return a valid ‘dist’ object with dissimilarities, similarities are not accepted and should be converted into dissimilarities beforehand.

Value

Returns a named vector of rank correlations.

Note

There are several problems in using rank correlation coefficients. Typically there are very many ties when \( n(n - 1)/2 \) gradient separation values are derived from just \( n \) observations. Due to floating point arithmetics, many tied values differ by machine epsilon and are arbitrarily ranked differently by `rank` used in `cor.test`. Two indices which are identical with certain transformation or standardization may differ slightly (magnitude \( 10^{-15} \)) and this may lead into third or fourth decimal instability in rank correlations. Small differences in rank correlations should not be taken too seriously. Probably this method should be replaced with a sounder method, but I do not yet know which... You may experiment with `mantel`, `anosim` or even `protest`.

Earlier version of this function used `method = "kendall"`, but that is far too slow in large data sets.

The functions returning dissimilarity objects should be self contained, because the ... argument passes additional parameters to `stepacross` and not to the functions supplied via the `indices` argument.

Author(s)

Jari Oksanen, with additions from Peter Solymos

References


See Also

`vegdist`, `stepacross`, `no.shared`, `monoMDS`, `cor`, `Machine`, and for alternatives `anosim`, `mantel` and `protest`. 
Examples

```r
data(varespec)
data(varechem)
## The variables are automatically scaled
rankindex(varechem, varespec)
rankindex(varechem, wisconsin(varespec))
## Using non vegdist indices as functions
funs <- list(Manhattan=function(x) dist(x, "manhattan"),
             Gower=function(x) cluster::daisy(x, "gower"),
             Ochiai=function(x) designdist(x, "1-1/sqrt(A*B)"))
rankindex(scale(varechem), varespec, funs)
```

---

**rarefy**

**Rarefaction Species Richness**

Description

Rarefied species richness for community ecologists.

Usage

```r
rarefy(x, sample, se = FALSE, MARGIN = 1)
rarefy(x, sample)
rrarefy(x, sample)
drarefy(x, sample)
rarecurve(x, step = 1, sample, xlab = "Sample Size", ylab = "Species",
          label = TRUE, col, lty, ...)
rareslope(x, sample)
```

Arguments

- **x**: Community data, a matrix-like object or a vector.
- **MARGIN**: Margin for which the index is computed.
- **sample**: Subsample size for rarefying community, either a single value or a vector.
- **se**: Estimate standard errors.
- **step**: Step size for sample sizes in rarefaction curves.
- **xlab, ylab**: Axis labels in plots of rarefaction curves.
- **label**: Label rarefaction curves by rownames of x (logical).
- **col, lty**: Plotting colour and line type, see `par`. Can be a vector of length `nrow(x)`, one per sample, and will be extended to such a length internally.
- **...**: Parameters passed to `nlm`, or to `plot, lines` and `ordilabel` in `rarecurve`.
Details

Function `rarefy` gives the expected species richness in random subsamples of size `sample` from the community. The size of `sample` should be smaller than total community size, but the function will work for larger `sample` as well (with a warning) and return non-rarefied species richness (and standard error = 0). If `sample` is a vector, rarefaction of all observations is performed for each `sample` size separately. Rarefaction can be performed only with genuine counts of individuals. The function `rarefy` is based on Hurlbert’s (1971) formulation, and the standard errors on Heck et al. (1975).

Function `rrarefy` generates one randomly rarefied community data frame or vector of given `sample` size. The `sample` can be a vector giving the sample sizes for each row. If the `sample` size is equal to or smaller than the observed number of individuals, the non-rarefied community will be returned. The random rarefaction is made without replacement so that the variance of rarefied communities is rather related to rarefaction proportion than to the size of the `sample`. Random rarefaction is sometimes used to remove the effects of different sample sizes. This is usually a bad idea: random rarefaction discards valid data, introduces random error and reduces the quality of the data (McMurdie & Holmes 2014). It is better to use normalizing transformations (`decostand` in `vegan`) possible with variance stabilization (`decostand` and `dispweight` in `vegan`) and methods that are not sensitive to sample sizes.

Function `drarefy` returns probabilities that species occur in a rarefied community of size `sample`. The `sample` can be a vector giving the sample sizes for each row. If the `sample` is equal to or smaller than the observed number of individuals, all observed species will have sampling probability 1.

Function `rarecurve` draws a rarefaction curve for each row of the input data. The rarefaction curves are evaluated using the interval of `step` sample sizes, always including 1 and total sample size. If `sample` is specified, a vertical line is drawn at `sample` with horizontal lines for the rarefied species richesses.

Function `rareslope` calculates the slope of `rarecurve` (derivative of `rarefy`) at given `sample` size; the `sample` need not be an integer.

Value

A vector of rarefied species richness values. With a single `sample` and `se = TRUE`, function `rarefy` returns a 2-row matrix with rarefied richness (`S`) and its standard error (`se`). If `sample` is a vector in `rarefy`, the function returns a matrix with a column for each `sample` size, and if `se = TRUE`, rarefied richness and its standard error are on consecutive lines.

Function `rarecurve` returns `invisible` list of `rarefy` results corresponding each drawn curve.

Author(s)

Jari Oksanen

References


**See Also**

Use `specaccum` for species accumulation curves where sites are sampled instead of individuals. `specpool` extrapolates richness to an unknown sample size.

**Examples**

```r
data(BCI)
S <- specnumber(BCI)  # observed number of species
(raremax <- min(rowSums(BCI)))
Srare <- rarefy(BCI, raremax)
plot(S, Srare, xlab = "Observed No. of Species", ylab = "Rarefied No. of Species")
abline(0, 1)
rarecurve(BCI, step = 20, sample = raremax, col = "blue", cex = 0.6)
```

---

**raupcrick**

*Raup-Crick Dissimilarity with Unequal Sampling Densities of Species*

**Description**

Function finds the Raup-Crick dissimilarity which is a probability of number of co-occurring species with species occurrence probabilities proportional to species frequencies.

**Usage**

```r
raupcrick(comm, null = "r1", nsimul = 999, chase = FALSE, ...)
```

**Arguments**

- `comm`: Community data which will be treated as presence/absence data.
- `null`: Null model used as the method in `oecosimu`.
- `nsimul`: Number of null communities for assessing the dissimilarity index.
- `chase`: Use the Chase et al. (2011) method of tie handling (not recommended except for comparing the results against the Chase script).
- `...`: Other parameters passed to `oecosimu`.

**Details**

Raup-Crick index is the probability that compared sampling units have non-identical species composition. This probability can be regarded as a dissimilarity, although it is not metric: identical sampling units can have dissimilarity slightly above 0, the dissimilarity can be nearly zero over a range of shared species, and sampling units with no shared species can have dissimilarity slightly below 1. Moreover, communities sharing rare species appear as more similar (lower probability of finding rare species together), than communities sharing the same number of common species.
The function will always treat the data as binary (presence/absence).

The probability is assessed using simulation with `oecosimu` where the test statistic is the observed number of shared species between sampling units evaluated against a community null model (see Examples). The default null model is "r1" where the probability of selecting species is proportional to the species frequencies.

The `vegdist` function implements a variant of the Raup-Crick index with equal sampling probabilities for species using exact analytic equations without simulation. This corresponds to null model "r0" which also can be used with the current function. All other null model methods of `oecosimu` can be used with the current function, but they are new unpublished methods.

Value

The function returns an object inheriting from `dist` which can be interpreted as a dissimilarity matrix.

Note

The test statistic is the number of shared species, and this is typically tied with a large number of simulation results. The tied values are handled differently in the current function and in the function published with Chase et al. (2011). In `vegan`, the index is the number of simulated values that are smaller or equal than the observed value, but smaller than observed value is used by Chase et al. (2011) with option split = FALSE in their script; this can be achieved with chase = TRUE in `vegan`. Chase et al. (2011) script with split = TRUE uses half of tied simulation values to calculate a distance measure, and that choice cannot be directly reproduced in vegan (it is the average of `vegan` raupcrick results with chase = TRUE and chase = FALSE).

Author(s)

The function was developed after Brian Inouye contacted us and informed us about the method in Chase et al. (2011), and the function takes its idea from the code that was published with their paper. The current function was written by Jari Oksanen.

References


See Also

The function is based on `oecosimu`. Function `vegdist` with method = "raup" implements a related index but with equal sampling densities of species, and `designdist` demonstrates its calculation.

Examples

```r
## data set with variable species richness
data(sipoo)
## default raupcrick
drl <- raupcrick(sipoo)
## use null model "r0" of oecosimu
```
dr0 <- raupcrick(sipoo, null = "r0")
## vegdist(..., method = "raup") corresponds to 'null = "r0"'
d <- vegdist(sipoo, "raup")
op <- par(mfrow=c(2,1), mar=c(4,4,1,1)+.1)
plot(dr1 - d, xlab = "Raup-Crick with Null R1", ylab="vegdist")
plot(dr0 - d, xlab = "Raup-Crick with Null R0", ylab="vegdist")
par(op)

## The calculation is essentially as in the following oecosimu() call,
## except that designdist() is replaced with faster code
## Not run:
oecosimu(sipoo, function(x) designdist(x, "J", "binary"), method = "r1")
## End(Not run)

---

**read.cerp**  
*Reads a CEP (Canoco) data file*

**Description**

read.cerp reads a file formatted with relaxed strict CEP format used in Canoco software, among others.

**Usage**

```r
read.cerp(file, positive=TRUE)
```

**Arguments**

- `file`  
  File name (character variable).
- `positive`  
  Only positive entries, like in community data.

**Details**

Cornell Ecology Programs (CEP) introduced several data formats designed for punched cards. One of these was the 'condensed strict' format which was adopted by popular software DECORANA and TWINSPAN. A relaxed variant of this format was later adopted in Canoco software (ter Braak 1984). Function read.cerp reads legacy files written in this format.

The condensed CEP and CANOCO formats have:

- Two or three title cards, most importantly specifying the format and the number of items per record.
- Data in condensed format: First number on the line is the site identifier (an integer), and it is followed by pairs ('couplets') of numbers identifying the species and its abundance (an integer and a floating point number).
- Species and site names, given in Fortran format (10A8): Ten names per line, eight columns for each.
With option positive = TRUE the function removes all rows and columns with zero or negative marginal sums. In community data with only positive entries, this removes empty sites and species. If data entries can be negative, this ruins data, and such data sets should be read in with option positive = FALSE.

Value

Returns a data frame, where columns are species and rows are sites. Column and row names are taken from the CEP file, and changed into unique R names by make.names after stripping the blanks.

Note

Function read.cep used Fortran to read data in vegan 2.4-5 and earlier, but Fortran I/O is no longer allowed in CRAN packages, and the function was re-written in R. The original Fortran code was more robust, and there are several legacy data sets that may fail with the current version, but could be read with the previous Fortran version. CRAN package cepreader makes available the original Fortran-based code run in a separate subprocess. The cepreader package can also read ‘free’ and ‘open’ Canoco formats that are not handled in this function.

The function is based on read.fortran. If the REAL format defines a decimal part for species abundances (such as F5.1), read.fortran divides the input with the corresponding power of 10 even when the input data had explicit decimal separator. With F5.1, 100 would become 10, and 0.1 become 0.01. Function read.cep tries to undo this division, but you should check the scaling of results after reading the data, and if necessary, multiply results to the original scale.

Author(s)

Jari Oksanen

References


Examples

```r
## Provided that you have the file "dune.spe"
## Not run:
theclassic <- read.cep("dune.spe")
## End(Not run)
```
**Description**

Function `renyi` finds Rényi diversities with any scale or the corresponding Hill number (Hill 1973). Function `renyiaccum` finds these statistics with accumulating sites.

**Usage**

```r
renyi(x, scales = c(0, 0.25, 0.5, 1, 2, 4, 8, 16, 32, 64, Inf),
    hill = FALSE)
```

```r
## S3 method for class 'renyi'
plot(x, ...)
```

```r
renyiaccum(x, scales = c(0, 0.5, 1, 2, 4, Inf), permutations = 100,
    raw = FALSE, collector = FALSE, subset, ...)
```

```r
## S3 method for class 'renyiaccum'
plot(x, what = c("Collector", "mean", "Qnt 0.025", "Qnt 0.975"),
    type = "1",
    ...)
```

```r
## S3 method for class 'renyiaccum'
persp(x, theta = 220, col = heat.colors(100), zlim, ...)
```

**Arguments**

- **x**: Community data matrix or plotting object.
- **scales**: Scales of Rényi diversity.
- **hill**: Calculate Hill numbers.
- **permutations**: Usually an integer giving the number permutations, but can also be a list of control values for the permutations as returned by the function `how`, or a permutation matrix where each row gives the permuted indices.
- **raw**: if FALSE then return summary statistics of permutations, and if TRUE then returns the individual permutations.
- **collector**: Accumulate the diversities in the order the sites are in the data set, and the collector curve can be plotted against summary of permutations. The argument is ignored if raw = TRUE.
- **subset**: logical expression indicating sites (rows) to keep: missing values are taken as FALSE.
- **what**: Items to be plotted.
- **type**: Type of plot, where type = "1" means lines.
- **theta**: Angle defining the viewing direction (azimuthal) in `persp`.
- **col**: Colours used for surface. Single colour will be passed on, and vector colours will be selected by the midpoint of a rectangle in `persp`.
- **zlim**: Limits of vertical axis.
- **...**: Other arguments which are passed to `renyi` and to graphical functions.
Details

Common diversity indices are special cases of Rényi diversity

\[ H_a = \frac{1}{1-a} \log \sum p_i^a \]

where \( a \) is a scale parameter, and Hill (1975) suggested to use so-called ‘Hill numbers’ defined as \( N_a = \exp(H_a) \). Some Hill numbers are the number of species with \( a = 0 \), \( \exp(H') \) or the exponent of Shannon diversity with \( a = 1 \), inverse Simpson with \( a = 2 \) and \( 1/\max(p_i) \) with \( a = \infty \). According to the theory of diversity ordering, one community can be regarded as more diverse than another only if its Rényi diversities are all higher (Tóthmérész 1995).

The plot method for renyi uses lattice graphics, and displays the diversity values against each scale in separate panel for each site together with minimum, maximum and median values in the complete data.

Function renyiaccum is similar to specaccum but finds Rényi or Hill diversities at given scales for random permutations of accumulated sites. Its plot function uses lattice function xyplot to display the accumulation curves for each value of scales in a separate panel. In addition, it has a persp method to plot the diversity surface against scale and number and sites. Similar dynamic graphics can be made with rgl.renyiaccum in vegan3d package.

Value

Function renyi returns a data frame of selected indices. Function renyiaccum with argument raw = FALSE returns a three-dimensional array, where the first dimension are the accumulated sites, second dimension are the diversity scales, and third dimension are the summary statistics mean, stdev, min, max, Qnt 0.025 and Qnt 0.975. With argument raw = TRUE the statistics on the third dimension are replaced with individual permutation results.

Author(s)

Roeland Kindt <r.kindt@cgiar.org> and Jari Oksanen

References

http://www.worldagroforestry.org/resources/databases/tree-diversity-analysis


See Also

diversity for diversity indices, and specaccum for ordinary species accumulation curves, and xyplot, persp and rgl.renyiaccum.
Examples

data(BCI)
i <- sample(nrow(BCI), 12)
mod <- renyi(BCI[i,])
plot(mod)
mod <- renyiaccum(BCI[i,])
plot(mod, as.table=TRUE, col = c(1, 2, 2))
persp(mod)

reorder.hclust  Reorder a Hierarchical Clustering Tree

Description

Function takes a hierarchical clustering tree from hclust and a vector of values and reorders the clustering tree in the order of the supplied vector, maintaining the constraints on the tree. This is a method of generic function reorder and an alternative to reordering a "dendrogram" object with reorder.dendrogram

Usage

## S3 method for class 'hclust'
reorder(x, wts,
        agglo.FUN = c("mean", "min", "max", "sum", "uwmean"), ...)

## S3 method for class 'hclust'
rev(x)

## S3 method for class 'hclust'
scores(x, display = "internal", ...)
cutreeord(tree, k = NULL, h = NULL)

Arguments

x, tree  hierarchical clustering from hclust.
wts  numeric vector for reordering.
agglo.FUN  a function for weights agglomeration, see below.
display  return "internal" nodes or "terminal" nodes (also called "leaves").
k, h  scalars or vectors giving the numbers of desired groups or the heights where the tree should be cut (passed to function cutree).
...  additional arguments (ignored).
Details

Dendrograms can be ordered in many ways. The reorder function reorders an \texttt{hclust} tree and provides an alternative to \texttt{reorder.dendrogram} which can reorder a \texttt{dendrogram}. The current function will also work differently when the \texttt{agglo.FUN} is "mean": the \texttt{reorder.dendrogram} will always take the direct mean of member groups ignoring their sizes, but this function will use \texttt{weighted.mean} weighted by group sizes, so that the group mean is always the mean of member leaves (terminal nodes). If you want to ignore group sizes, you can use unweighted mean with "\texttt{uwmean}"

The function accepts only a limited list of \texttt{agglo.FUN} functions for assessing the value of \texttt{wts} for groups. The ordering is always ascending, but the order of leaves can be reversed with \texttt{rev}.

Function \texttt{scores} finds the coordinates of nodes as a two-column matrix. For terminal nodes (leaves) this the value at which the item is merged to the tree, and the labels can still hang below this level (see \texttt{plot.hclust}).

Function \texttt{cutreeord} cuts a tree to groups numbered from left to right in the tree. It is based on the standard function \texttt{cutree} which numbers the groups in the order they appear in the input data instead of the order in the tree.

Value

Reordered \texttt{hclust} result object with added item \texttt{value} that gives the value of the statistic at each merge level.

Note

These functions should really be in base \texttt{R}.

Author(s)

Jari Oksanen

See Also

\texttt{hclust} for getting clustering trees, \texttt{as.hclust.spantree} to change a \texttt{vegan} minimum spanning tree to an \texttt{hclust} object, and \texttt{dendrogram} and \texttt{reorder.dendrogram} for an alternative implementation.

Examples

```r
## reorder by water content of soil
data(mite, mite.env)
hc <- hclust(vegdist(wisconsin(sqrt(mite))))
ohc <- with(mite.env, reorder(hc, WatrCont))
plot(hc)
plot(ohc)

## label leaves by the observed value, and each branching point
## (internal node) by the cluster mean
with(mite.env, plot(ohc, labels=round(WatrCont), cex=0.7))
ordilabel(scores(ohc), label=round(ohc$value), cex=0.7)
```
## Description

The functions finds the adjusted R-square.

### Usage

```r
## Default S3 method:
RsquareAdj(x, n, m, ...)  
## S3 method for class 'rda'
RsquareAdj(x, ...)  
## S3 method for class 'cca'
RsquareAdj(x, permutations = 1000, ...)
```

### Arguments

- `x`: Unadjusted R-squared or an object from which the terms for evaluation or adjusted R-squared can be found.
- `n, m`: Number of observations and number of degrees of freedom in the fitted model.
- `permutations`: Number of permutations to use when computing the adjusted R-squared for a cca. The permutations can be calculated in parallel by specifying the number of cores which is passed to `permutest`
- `...`: Other arguments (ignored) except in the cca method, where these are passed to `permutest.cca`.

### Details

The default method finds the adjusted $R^2$ from the unadjusted $R^2$, number of observations, and number of degrees of freedom in the fitted model. The specific methods find this information from the fitted result object. There are specific methods for `rda`, `cca`, `lm` and `glm`. Adjusted, or even unadjusted, $R^2$ may not be available in some cases, and then the functions will return NA. There is no adjusted $R^2$ in partial ordination, and $R^2$ values are available only for `gaussian` models in `glm`. The adjusted, $R^2$ of cca is computed using a permutation approach developed by Peres-Neto et al. (2006).

### Value

The functions return a list of items `r.squared` and `adj.r.squared`.

---

```r
## Slightly different from reordered 'dendrograms' which ignores group sizes in assessing means.
den <- as.dendrogram(hc)
den <- with(mite.env, reorder(den, WatrCont, agglo.FUN = mean))
plot(den)
```
References


See Also

`varpart` uses RsquareAdj.

Examples

data(mite)
data(mite.env)

```r
## rda
m <- rda(decostand(mite, "hell") ~ ., mite.env)
RsquareAdj(m)

## cca
m <- cca(decostand(mite, "hell") ~ ., mite.env)
RsquareAdj(m)

## default method
RsquareAdj(0.8, 20, 5)
```

---

**scores**

*Get Species or Site Scores from an Ordination*

Description

Function to access either species or site scores for specified axes in some ordination methods. The `scores` function is generic in `vegan`, and `vegan` ordination functions have their own `scores` functions that are documented separately with the method (see e.g. `scores.cca`, `scores.metaMDS`, `scores.decorana`). This help file documents the default `scores` method that is only used for non-`vegan` ordination objects.

Usage

```r
## Default S3 method:
scores(x, choices, display=c("sites", "species"), ...)
```

Arguments

- **x**: An ordination result.
- **choices**: Ordination axes. If missing, default method returns all axes.
- **display**: Partial match to access scores for sites or species.
- **...**: Other parameters (unused).
Details

Function `scores` is a generic method in `vegan`. Several `vegan` functions have their own `scores` methods with their own defaults and with some new arguments. This help page describes only the default method. For other methods, see, e.g., `scores.cca`, `scores.rda`, `scores.decorana`.

All `vegan` ordination functions should have a `scores` method which should be used to extract the scores instead of directly accessing them. Scaling and transformation of scores should also happen in the `scores` function. If the `scores` function is available, the results can be plotted using `ordiplot`, `ordixyplot` etc., and the ordination results can be compared in `procrustes` analysis.

The `scores.default` function is used to extract scores from non-`vegan` ordination results. Many standard ordination methods of libraries do not have a specific class, and no specific method can be written for them. However, `scores.default` guesses where some commonly used functions keep their site scores and possible species scores.

If `x` is a matrix, `scores.default` returns the chosen columns of that matrix, ignoring whether species or sites were requested (do not regard this as a bug but as a feature, please). Currently the function seems to work at least for `isoMDS`, `prcomp`, `princomp` and some `ade4` objects. It may work in other cases or fail mysteriously.

Value

The function returns a matrix of scores.

Author(s)

Jari Oksanen

See Also

Specific `scores` functions include (but are not limited to) `scores.cca`, `scores.rda`, `scores.decorana`, `scores.envfit`, `scores.metaMDS`, `scores.monoMDS` and `scores.pcmn`. These have somewhat different interface – `scores.cca` in particular – but all work with keywords `display="sites"` and return a matrix. However, they may also return a list of matrices, and some other `scores` methods will have quite different arguments.

Examples

```r
data(varespec)
vare.pca <- prcomp(varespec)
scores(vare.pca, choices=c(1,2))
```

Description

Screeplot methods for plotting variances of ordination axes/components and overlaying broken stick distributions. Also, provides alternative screeplot methods for `princomp` and `prcomp`. 
Usage

```r
## S3 method for class 'cca'
screeplot(x, bstick = FALSE, type = c("barplot", "lines"),
npcs = min(10, if (is.null(x$CCA) || x$CCA$rank == 0) x$CCA$rank else x$CCA$rank),
ptype = "o", bst.col = "red", bst.lty = "solid",
xlab = "Component", ylab = "Inertia",
main = deparse(substitute(x)), legend = bstick,
...)

## S3 method for class 'decorana'
screeplot(x, bstick = FALSE, type = c("barplot", "lines"),
npcs = 4,
ptype = "o", bst.col = "red", bst.lty = "solid",
xlab = "Component", ylab = "Inertia",
main = deparse(substitute(x)), legend = bstick,
...)

## S3 method for class 'prcomp'
screeplot(x, bstick = FALSE, type = c("barplot", "lines"),
npcs = min(10, length(x$sdev)),
ptype = "o", bst.col = "red", bst.lty = "solid",
xlab = "Component", ylab = "Inertia",
main = deparse(substitute(x)), legend = bstick,
...)

## S3 method for class 'princomp'
screeplot(x, bstick = FALSE, type = c("barplot", "lines"),
npcs = min(10, length(x$sdev)),
ptype = "o", bst.col = "red", bst.lty = "solid",
xlab = "Component", ylab = "Inertia",
main = deparse(substitute(x)), legend = bstick,
...)

bstick(n, ...)

## Default S3 method:
bstick(n, tot.var = 1, ...)

## S3 method for class 'cca'
bstick(n, ...)

## S3 method for class 'prcomp'
bstick(n, ...)

## S3 method for class 'princomp'
bstick(n, ...)

## S3 method for class 'decorana'
```
bstick(n, ...)

Arguments

\( x \)
an object from which the component variances can be determined.

\( \text{bstick} \)
logical; should the broken stick distribution be drawn?

\( \text{npcs} \)
the number of components to be plotted.

\( \text{type} \)
the type of plot.

\( \text{ptype} \)
if \( \text{type} == \text{"lines" or } \text{bstick} = \text{TRUE} \), a character indicating the type of plotting used for the lines; actually any of the types as in \text{plot.default}.

\( \text{bst.col, bst.lty} \)
the colour and line type used to draw the broken stick distribution.

\( \text{xlab, ylab, main} \)
graphics parameters.

\( \text{legend} \)
logical; draw a legend?

\( n \)
an object from which the variances can be extracted or the number of variances (components) in the case of \text{bstick.default}.

\( \text{tot.var} \)
the total variance to be split.

\( ... \)
arguments passed to other methods.

Details

The functions provide screeplots for most ordination methods in \text{vegan} and enhanced versions with broken stick for \text{prcomp} and \text{princomp}.

Function \text{bstick} gives the brokenstick values which are ordered random proportions, defined as
\[ p_i = \frac{\text{tot}}{n} \sum_{x=1}^{n} \frac{1}{x} \] (Legendre & Legendre 2012), where \text{tot} is the total and \( n \) is the number of brokenstick components (cf. \text{radfit}). Broken stick has been recommended as a stopping rule in principal component analysis (Jackson 1993): principal components should be retained as long as observed eigenvalues are higher than corresponding random broken stick components.

The \text{bstick} function is generic. The default needs the number of components and the total, and specific methods extract this information from ordination results. There also is a \text{bstick} method for \text{cca}. However, the broken stick model is not strictly valid for correspondence analysis (CA), because eigenvalues of CA are defined to be \( \leq 1 \), whereas brokenstick components have no such restrictions. The brokenstick components are not available for \text{decorana} where the sum of eigenvalues (total inertia) is unknown, and the eigenvalues of single axes are not additive in detrended analysis.

Value

Function \text{screeplot} draws a plot on the currently active device, and returns invisibly the \text{xy.coords} of the points or bars for the eigenvalues.

Function \text{bstick} returns a numeric vector of broken stick components.

Author(s)

Gavin L. Simpson
References


See Also

`cca`, `decorana`, `princomp` and `prcomp` for the ordination functions, and `screeplot` for the stock version.

Examples

data(varespec)
vare.pca <- rda(varespec, scale = TRUE)
bstick(vare.pca)
screepplot(vare.pca, bstick = TRUE, type = "lines")

---

**simper**

*Similarity Percentages*

**Description**

Discriminating species between two groups using Bray-Curtis dissimilarities

**Usage**

```r
simper(comm, group, permutations = 0, trace = FALSE,
parallel = getOption("mc.cores"), ...)
```

## S3 method for class 'simper'

```
summary(object, ordered = TRUE,
digits = max(3,getOption("digits") - 3), ...)
```

**Arguments**

- `comm` Community data matrix.
- `group` Factor describing the group structure. Must have at least 2 levels.
- `permutations` a list of control values for the permutations as returned by the function `how`, or the number of permutations required, or a permutation matrix where each row gives the permuted indices.
- `trace` Trace permutations.
- `object` an object returned by `simper`.
- `ordered` Logical; Should the species be ordered by their average contribution?
- `digits` Number of digits in output.
- `parallel` Number of parallel processes or a predefined socket cluster. With `parallel = 1` uses ordinary, non-parallel processing.
- `...` Parameters passed to other functions. In `simper` the extra parameters are passed to `shuffleSet` if permutations are used.
Details

Similarity percentage, simper (Clarke 1993) is based on the decomposition of Bray-Curtis dissimilarity index (see vegdist, designdist). The contribution of individual species $i$ to the overall Bray-Curtis dissimilarity $d_{jk}$ is given by

$$d_{ijk} = \frac{|x_{ij} - x_{ik}|}{\sum_{i=1}^{S}(x_{ij} + x_{ik})}$$

where $x$ is the abundance of species $i$ in sampling units $j$ and $k$. The overall index is the sum of the individual contributions over all $S$ species $d_{jk} = \sum_{i=1}^{S} d_{ijk}$.

The simper functions performs pairwise comparisons of groups of sampling units and finds the average contributions of each species to the average overall Bray-Curtis dissimilarity.

The function displays most important species for each pair of groups. These species contribute at least to 70 % of the differences between groups. The function returns much more extensive results which can be accessed directly from the result object (see section Value). Function summary transforms the result to a list of data frames. With argument ordered = TRUE the data frames also include the cumulative contributions and are ordered by species contribution.

The results of simper can be very difficult to interpret. The method very badly confounds the mean between group differences and within group variation, and seems to single out variable species instead of distinctive species (Warton et al. 2012). Even if you make groups that are copies of each other, the method will single out species with high contribution, but these are not contributions to non-existing between-group differences but to within-group variation in species abundance.

Value

A list of class "simper" with following items:

- species: The species names.
- average: Average contribution to overall dissimilarity.
- overall: The overall between-group dissimilarity.
- sd: Standard deviation of contribution.
- ratio: Average to sd ratio.
- ava, avb: Average abundances per group.
- ord: An index vector to order vectors by their contribution or order cusum back to the original data order.
- cusum: Ordered cumulative contribution.
- p: Permutation $p$-value. Probability of getting a larger or equal average contribution in random permutation of the group factor. These area only available if permutations were used (default: not calculated).

Author(s)

Eduard Szöcs <eduardszoecs@gmail.com>
References


Examples

```r
data(dune)
data(dune.env)
(sim <- with(dune.env, simper(dune, Management)))
summary(sim)
```

---

**simulate.rda**

*Simulate Responses with Gaussian Error or Permuted Residuals for Constrained Ordination*

**Description**

Function simulates a response data frame so that it adds Gaussian error to the fitted responses of Redundancy Analysis (*rda*), Constrained Correspondence Analysis (*cca*) or distance-based RDA (*capscale*). The function is a special case of generic *simulate*, and works similarly as *simulate.lm*.

**Usage**

```r
## S3 method for class 'rda'
simulate(object, nsim = 1, seed = NULL, indx = NULL,
    rank = "full", correlated = FALSE, ...)
```

**Arguments**

- **object**: An object representing a fitted *rda*, *cca* or *capscale* model.
- **nsim**: Number of response matrices to be simulated. Only one dissimilarity matrix is returned for *capscale*, and larger nsim is an error.
- **seed**: An object specifying if and how the random number generator should be initialized ("seeded"). See *simulate* for details.
- **indx**: Index of residuals added to the fitted values, such as produced by *shuffleSet* or *sample*. The index can have duplicate entries so that bootstrapping is allowed. If nsim > 1, the output should be compliant with *shuffleSet* with one line for each simulation. If nsim is missing, the number of rows of indx is used to define the number of simulations, but if nsim is given, it should match number of rows in indx. If null, parametric simulation is used and Gaussian error is added to the fitted values.
- **rank**: The rank of the constrained component: passed to *predict.rda* or *predict.cca*. 
correlated Are species regarded as correlated in parametric simulation or when \texttt{indx} is not given? If \texttt{correlated = TRUE}, multivariate Gaussian random error is generated, and if \texttt{FALSE}, Gaussian random error is generated separately for each species. The argument has no effect in \texttt{capscale} which has no information on species.

\[ \ldots \]

additional optional arguments (ignored).

\section*{Details}

The implementation follows "lm" method of \texttt{simulate}, and adds Gaussian (Normal) error to the fitted values (\texttt{fitted.rda}) using function \texttt{rnorm} if \texttt{correlated} = \texttt{FALSE} or \texttt{mvrnorm} if \texttt{correlated} = \texttt{TRUE}. The standard deviations (\texttt{rnorm}) or covariance matrices for species (\texttt{mvrnorm}) are estimated from the residuals after fitting the constraints. Alternatively, the function can take a permutation index that is used to add permuted residuals (unconstrained component) to the fitted values. Raw data are used in \texttt{rda}. Internal Chi-square transformed data are used in \texttt{cca} within the function, but the returned matrix is similar to the original input data. The simulation is performed on internal metric scaling data in \texttt{capscale}, but the function returns the Euclidean distances calculated from the simulated data. The simulation uses only the real components, and the imaginary dimensions are ignored.

\section*{Value}

If \texttt{nsim} = 1, returns a matrix or dissimilarities (in \texttt{capscale}) with similar additional arguments on random number seed as \texttt{simulate}. If \texttt{nsim} > 1, returns a similar array as returned by \texttt{simulate.nullmodel} with similar attributes.

\section*{Author(s)}

Jari Oksanen

\section*{See Also}

\texttt{simulate} for the generic case and for \texttt{lm} objects, and \texttt{simulate.nullmodel} for community null model simulation. Functions \texttt{fitted.rda} and \texttt{fitted.cca} return fitted values without the error component. See \texttt{rnorm} and \texttt{mvrnorm} (\texttt{MASS} package) for simulating Gaussian random error.

\section*{Examples}

data(dune)
data(dune.env)
mod <- rda(dune ~ Moisture + Management, dune.env)
## One simulation
update(mod, simulate(mod) ~ .)
## An impression of confidence regions of site scores
plot(mod, display="sites")
for (i in 1:5) lines(procrustes(mod, update(mod, simulate(mod) ~ .)), col="blue")
## Simulate a set of null communities with permutation of residuals
simulate(mod, indx = shuffleSet(nrow(dune), 99))
**Birds in the Archipelago of Sipoo (Sibbo and Borgå)**

**Description**

Land birds on islands covered by coniferous forest in the Sipoo Archipelago, southern Finland.

**Usage**

```r
data(sipoo)
data(sipoo.map)
```

**Format**

The `sipoo` data frame contains data of occurrences of 50 land bird species on 18 islands in the Sipoo Archipelago (Simberloff & Martin, 1991, Appendix 3). The species are referred by 4+4 letter abbreviation of their Latin names (but using five letters in two species names to make these unique).

The `sipoo.map` data contains the geographic coordinates of the islands in the ETRS89-TM35FIN coordinate system (EPSG:3067) and the areas of islands in hectares.

**Source**


**Examples**

```r
data(sipoo)
data(sipoo.map)
plot(N ~ E, data=sipoo.map, asp = 1)
```

---

**Minimum Spanning Tree**

**Description**

Function `spantree` finds a minimum spanning tree connecting all points, but disregarding dissimilarities that are at or above the threshold or NA.
Usage

```r
spantree(d, toolong = 0)
## S3 method for class 'spantree'
as.hclust(x, ...)
## S3 method for class 'spantree'
cophenetic(x)
spandepth(x)
## S3 method for class 'spantree'
plot(x, ord, cex = 0.7, type = "p", labels, dlim,
     FUN = sammon, ...)
## S3 method for class 'spantree'
lines(x, ord, display="sites", col = 1, ...)
```

Arguments

d  Dissimilarity data inheriting from class dist or an object, such as a matrix, that can be converted to a dissimilarity matrix. Functions `vegdist` and `dist` are some functions producing suitable dissimilarity data.

toolong  Shortest dissimilarity regarded as NA. The function uses a fuzz factor, so that dissimilarities close to the limit will be made NA, too. If toolong = 0 (or negative), no dissimilarity is regarded as too long.

x  A `spantree` result object.

ord  An ordination configuration, or an ordination result known by `scores`.

cex  Character expansion factor.

type  Observations are plotted as points with type="p" or type="b", or as text label with type="t". The tree (lines) will always be plotted.

labels  Text used with type="t" or node names if this is missing.

dlim  A ceiling value used to highest cophenetic dissimilarity.

FUN  Ordination function to find the configuration from cophenetic dissimilarities. If the supplied FUN does not work, supply ordination result as argument ord.

display  Type of `scores` used for ord.

col  Colour of line segments. This can be a vector which is recycled for points, and the line colour will be a mixture of two joined points.

...  Other parameters passed to functions.

Details

Function `spantree` finds a minimum spanning tree for dissimilarities (there may be several minimum spanning trees, but the function finds only one). Dissimilarities at or above the threshold toolong and NAs are disregarded, and the spanning tree is found through other dissimilarities. If the data are disconnected, the function will return a disconnected tree (or a forest), and the corresponding link is NA. Connected subtrees can be identified using `distconnected`.

Minimum spanning tree is closely related to single linkage clustering, a.k.a. nearest neighbour clustering, and in genetics as neighbour joining tree available in `hclust` and `agnes` functions. The most important practical difference is that minimum spanning tree has no concept of cluster membership,
but always joins individual points to each other. Function `as.hclust` can change the `spantree` result into a corresponding `hclust` object.

Function `cophenetic` finds distances between all points along the tree segments. Function `spandepth` returns the depth of each node. The nodes of a tree are either leaves (with one link) or internal nodes (more than one link). The leaves are recursively removed from the tree, and the depth is the layer at which the leaf was removed. In disconnected `spantree` object (in a forest) each tree is analysed separately and disconnected nodes not in any tree have depth zero.

Function `plot` displays the tree over a supplied ordination configuration, and `lines` adds a spanning tree to an ordination graph. If configuration is not supplied for `plot`, the function ordinates the cophenetic dissimilarities of the spanning tree and overlays the tree on this result. The default ordination function is `sammon` (package `MASS`), because Sammon scaling emphasizes structure in the neighbourhood of nodes and may be able to beautifully represent the tree (you may need to set `dlim`, and sometimes the results will remain twisted). These ordination methods do not work with disconnected trees, but you must supply the ordination configuration. Function `lines` will overlay the tree in an existing plot.

Function `spantree` uses Prim’s method implemented as priority-first search for dense graphs (Sedgewick 1990). Function `cophenetic` uses function `stepacross` with option `path = "extended"`. The `spantree` is very fast, but `cophenetic` is slow in very large data sets.

### Value

Function `spantree` returns an object of class `spantree` which is a list with two vectors, each of length \( n - 1 \). The number of links in a tree is one less the number of observations, and the first item is omitted. The items are

- `kid` The child node of the parent, starting from parent number two. If there is no link from the parent, value will be `NA` and tree is disconnected at the node.
- `dist` Corresponding distance. If `kid = NA`, then `dist = 0`.
- `labels` Names of nodes as found from the input dissimilarities.
- `call` The function call.

### Note

In principle, minimum spanning tree is equivalent to single linkage clustering that can be performed using `hclust` or `agnes`. However, these functions combine clusters to each other and the information of the actually connected points (the “single link”) cannot be recovered from the result. The graphical output of a single linkage clustering plotted with `ordicluster` will look very different from an equivalent spanning tree plotted with `lines.spantree`.

### Author(s)

Jari Oksanen

### References

specaccum

See Also

vecdist or dist for getting dissimilarities, and hclust or agnes for single linkage clustering.

Examples

data(dune)
dis <- vegdist(dune)
tr <- spantree(dis)
## Add tree to a metric scaling
plot(tr, cmdscale(dis), type = "t")
## Find a configuration to display the tree neatly
plot(tr, type = "t")
## Depths of nodes
depths <- spandepth(tr)
plot(tr, type = "t", label = depths)
## Plot as a dendogram
cl <- as.hclust(tr)
plot(cl)
## cut hclust tree to classes and show in colours in spantree
plot(tr, col = cutree(cl, 5), pch=16)

specaccum

Species Accumulation Curves

Description

Function specaccum finds species accumulation curves or the number of species for a certain number of sampled sites or individuals.

Usage

specaccum(comm, method = "exact", permutations = 100,
conditioned = TRUE, gamma = "jack1", w = NULL, subset, ...)
## S3 method for class 'specaccum'
plot(x, add = FALSE, random = FALSE, ci = 2,
   ci.type = c("bar", "line", "polygon"), col = par("fg"), lty = 1,
   ci.col = col, ci.lty = 1, ci.length = 0, xlab, ylab = x$method, ylim,
   xvar = c("sites", "individuals", "effort"), ...)
## S3 method for class 'specaccum'
boxplot(x, add = FALSE, ...)
fitspecaccum(object, model, method = "random", ...)
## S3 method for class 'fitspecaccum'
plot(x, col = par("fg"), lty = 1, xlab = "Sites",
   ylab = x$method, ...)
## S3 method for class 'specaccum'
predict(object, newdata, interpolation = c("linear", "spline"), ...)
## S3 method for class 'fitspecaccum'
predict(object, newdata, ...)
specslope(object, at)
Arguments

**comm**  Community data set.

**method**  Species accumulation method (partial match). Method "collector" adds sites in the order they happen to be in the data, "random" adds sites in random order, "exact" finds the expected (mean) species richness, "coleman" finds the expected richness following Coleman et al. 1982, and "rarefaction" finds the mean when accumulating individuals instead of sites.

**permutations**  Number of permutations with method = "random". Usually an integer giving the number permutations, but can also be a list of control values for the permutations as returned by the function *how*, or a permutation matrix where each row gives the permuted indices.

**conditioned**  Estimation of standard deviation is conditional on the empirical dataset for the exact SAC.

**gamma**  Method for estimating the total extrapolated number of species in the survey area by function *specpool*.

**w**  Weights giving the sampling effort.

**subset**  logical expression indicating sites (rows) to keep: missing values are taken as FALSE.

**x**  A specaccum result object

**add**  Add to an existing graph.

**random**  Draw each random simulation separately instead of drawing their average and confidence intervals.

**ci**  Multiplier used to get confidence intervals from standard deviation (standard error of the estimate). Value ci = 0 suppresses drawing confidence intervals.

**ci.type**  Type of confidence intervals in the graph: "bar" draws vertical bars, "line" draws lines, and "polygon" draws a shaded area.

**col**  Colour for drawing lines.

**lty**  line type (see *par*).

**ci.col**  Colour for drawing lines or filling the "polygon".

**ci.lty**  Line type for confidence intervals or border of the "polygon".

**ci.length**  Length of horizontal bars (in inches) at the end of vertical bars with ci.type = "bar".

**xlab,ylab**  Labels for x (defaults *xvar*) and y axis.

**ylim**  the y limits of the plot.

**xvar**  Variable used for the horizontal axis: "individuals" can be used only with method = "rarefaction".

**object**  Either a community data set or fitted specaccum model.

**model**  Nonlinear regression model (*nls*). See Details.

**newdata**  Optional data used in prediction interpreted as number of sampling units (sites). If missing, fitted values are returned.

**interpolation**  Interpolation method used with newdata.

**at**  Number of plots where the slope is evaluated. Can be a real number.

**...**  Other parameters to functions.
Details

Species accumulation curves (SAC) are used to compare diversity properties of community data sets using different accumulator functions. The classic method is "random" which finds the mean SAC and its standard deviation from random permutations of the data, or subsampling without replacement (Gotelli & Colwell 2001). The "exact" method finds the expected SAC using sample-based rarefaction method that has been independently developed numerous times (Chiarucci et al. 2008) and it is often known as Mao Tau estimate (Colwell et al. 2012). The unconditional standard deviation for the exact SAC represents a moment-based estimation that is not conditioned on the empirical data set (sd for all samples > 0). The unconditional standard deviation is based on an estimation of the extrapolated number of species in the survey area (a.k.a. gamma diversity), as estimated by function specpool. The conditional standard deviation that was developed by Jari Oksanen (not published, sd=0 for all samples). Method "coleman" finds the expected SAC and its standard deviation following Coleman et al. (1982). All these methods are based on sampling sites without replacement. In contrast, the "random" method the effort refers to the average effort per site, or sum of weights per number of sites. With weighted method = "random", the averaged species richness is found from linear interpolation of single random permutations. Therefore at least the first value (and often several first) have NA richness, because these values cannot be interpolated in all cases but should be extrapolated. The plot function defaults to display the results as scaled to sites, but this can be changed selecting xvar = "effort" (weighted methods) or xvar = "individuals" (with method = "rarefaction").

The summary and boxplot methods are available for method = "random".

Function predict for specaccum can return the values corresponding to newdata. With method "exact", "rarefaction" and "coleman" the function uses analytic equations for interpolated non-integer values, and for other methods linear (approx) or spline (spline) interpolation. If newdata is not given, the function returns the values corresponding to the data. NB., the fitted values with method="rarefaction" are based on rounded integer counts, but predict can use fractional non-integer counts with newdata and give slightly different results.

Function fitspecaccum fits a nonlinear (nls) self-starting species accumulation model. The input object can be a result of specaccum or a community in data frame. In the latter case the function first fits a specaccum model and then proceeds with fitting the nonlinear model. The function can apply a limited set of nonlinear regression models suggested for species-area relationship (Dengler 2009). All these are selfStart models. The permissible alternatives are "arrhenius" (SSarrhenius), "gleason" (SSgleason), "gitay" (SSgitay), "lomolino" (SSLomolino) of vegan package. In addition the following standard R models are available: "asymp" (SSasymp), "gompertz" (SSgompertz), "michaelis-menten" (SSmicmen), "logis" (SSlogis), "weibull" (SSweibull). See these functions for model specification and details.

When weights w were used the fit is based on accumulated effort and in model = "rarefaction"
on accumulated number of individuals. The plot is still based on sites, unless other alternative is selected with xvar.

Function predict for fitspecaccum uses predict.nls, and you can pass all arguments to that function. In addition, fitted, residuals, nobs, coef, AIC, loglik and deviance work on the result object.

Function specslope evaluates the derivative of the species accumulation curve at given number of sample plots, and gives the rate of increase in the number of species. The function works with specaccum result object when this is based on analytic models "exact", "rarefaction" or "coleman", and with non-linear regression results of fitspecaccum.

Nonlinear regression may fail for any reason, and some of the fitspecaccum models are fragile and may not succeed.

Value

Function specaccum returns an object of class "specaccum", and fitspecaccum a model of class "fitspecaccum" that adds a few items to the "specaccum" (see the end of the list below):

call Function call.
method Accumulator method.
sites Number of sites. For method = "rarefaction" this is the number of sites corresponding to a certain number of individuals and generally not an integer, and the average number of individuals is also returned in item individuals.
effort Average sum of weights corresponding to the number of sites when model was fitted with argument w
richness The number of species corresponding to number of sites. With method = "collector" this is the observed richness, for other methods the average or expected richness.
sd The standard deviation of SAC (or its standard error). This is NULL in method = "collector", and it is estimated from permutations in method = "random", and from analytic equations in other methods.
perm Permutation results with method = "random" and NULL in other cases. Each column in perm holds one permutation.
weights Matrix of accumulated weights corresponding to the columns of the perm matrix when model was fitted with argument w.
fitted, residuals, coefficients Only in fitspecaccum: fitted values, residuals and nonlinear model coefficients. For method = "random" these are matrices with a column for each random accumulation.
models Only in fitspecaccum: list of fitted nls models (see Examples on accessing these models).

Note

The SAC with method = "exact" was developed by Roeland Kindt, and its standard deviation by Jari Oksanen (both are unpublished). The method = "coleman" underestimates the SAC because it does not handle properly sampling without replacement. Further, its standard deviation does not take into account species correlations, and is generally too low.
Author(s)

Roeland Kindt <r.kindt@cgiar.org> and Jari Oksanen.

References


See Also

`rarefy` and `rrarefy` are related individual based models. Other accumulation models are `poolaccum` for extrapolated richness, and `renyiaccum` and `tsallisaccum` for diversity indices. Underlying graphical functions are `boxplot`, `matlines`, `segments` and `polygon`.

Examples

data(BCI)
s1 <- specaccum(BCI)
s2 <- specaccum(BCI, "random")
s2
summary(s2)
plot(s1, ci.type="poly", col="blue", lwd=2, ci.lty=0, ci.col="lightblue")
boxplot(s2, col="yellow", add=TRUE, pch="+")
## Fit Lomolino model to the exact accumulation
mod1 <- fitspecaccum(s1, "lomolino")
coef(mod1)
fitted(mod1)
plot(s1)

## Add Lomolino model using argument 'add'
plot(mod1, add = TRUE, col=2, lwd=2)
## Fit Arrhenius models to all random accumulations
mods <- fitspecaccum(s2, "arrrh")
plot(mods, col="hotpink")
boxplot(s2, col = "yellow", border = "blue", lty=1, cex=0.3, add= TRUE)
## Use nls() methods to the list of models
sapply(mods$models, AIC)
Extrapolated Species Richness in a Species Pool

Description

The functions estimate the extrapolated species richness in a species pool, or the number of unobserved species. Function `specpool` is based on incidences in sample sites, and gives a single estimate for a collection of sample sites (matrix). Function `estimater` is based on abundances (counts) on single sample site.

Usage

```r
specpool(x, pool, smallsample = TRUE)
estimater(x, ...)
specpool2vec(X, index = c("jack1", "jack2", "chao", "boot", "Species"))
poolaccum(x, permutations = 100, minsize = 3)
estaccumR(x, permutations = 100, parallel = getOption("mc.cores"))
## S3 method for class 'poolaccum'
summary(object, display, alpha = 0.05, ...)
## S3 method for class 'poolaccum'
plot(x, alpha = 0.05, type = c("l", "g"), ...)
```

Arguments

- `x`: Data frame or matrix with species data or the analysis result for `plot` function.
- `pool`: A vector giving a classification for pooling the sites in the species data. If missing, all sites are pooled together.
- `smallsample`: Use small sample correction \((N - 1)/N\), where \(N\) is the number of sites within the pool.
- `X`, `object`: A `specpool` result object.
- `index`: The selected index of extrapolated richness.
- `permutations`: Usually an integer giving the number permutations, but can also be a list of control values for the permutations as returned by the function `how`, or a permutation matrix where each row gives the permuted indices.
- `minsize`: Smallest number of sampling units reported.
- `parallel`: Number of parallel processes or a predefined socket cluster. With `parallel = 1` uses ordinary, non-parallel processing. The parallel processing is done with `parallel` package.
- `display`: Indices to be displayed.
- `alpha`: Level of quantiles shown. This proportion will be left outside symmetric limits.
- `type`: Type of graph produced in `xyplot`.
- `...`: Other parameters (not used).
Details

Many species will always remain unseen or undetected in a collection of sample plots. The function uses some popular ways of estimating the number of these unseen species and adding them to the observed species richness (Palmer 1990, Colwell & Coddington 1994).

The incidence-based estimates in specpool use the frequencies of species in a collection of sites. In the following, $S_P$ is the extrapolated richness in a pool, $S_0$ is the observed number of species in the collection, $a_1$ and $a_2$ are the number of species occurring only in one or only in two sites in the collection, $p_i$ is the frequency of species $i$, and $N$ is the number of sites in the collection. The variants of extrapolated richness in specpool are:

- **Chao**
  \[ S_P = S_0 + \frac{a_1^2}{2a_2} \frac{N-1}{N} \]

- **Chao bias-corrected**
  \[ S_P = S_0 + \frac{a_1(a_1-1)}{2(a_2+1)} \frac{N-1}{N} \]

- **First order jackknife**
  \[ S_P = S_0 + a_1 \frac{N-1}{N} \]

- **Second order jackknife**
  \[ S_P = S_0 + a_1^2 \frac{N-3}{N} - a_2 \frac{(N-2)^2}{N(N-1)} \]

- **Bootstrap**
  \[ S_P = S_0 + \sum_{i=1}^{S_0} (1-p_i) \frac{N}{N} \]

specpool normally uses basic Chao equation, but when there are no doubletons ($a_2 = 0$) it switches to bias-corrected version. In that case the Chao equation simplifies to $S_0 + \frac{1}{2} a_1 (a_1 - 1) \frac{N-1}{N}$.

The abundance-based estimates in estimater use counts (numbers of individuals) of species in a single site. If called for a matrix or data frame, the function will give separate estimates for each site. The two variants of extrapolated richness in estimater are bias-corrected Chao and ACE (O’Hara 2005, Chiu et al. 2014). The Chao estimate is similar as the bias corrected one above, but $a_i$ refers to the number of species with abundance $i$ instead of number of sites, and the small-sample correction is not used. The ACE estimate is defined as:

- **ACE**
  \[ S_P = S_{abund} + \frac{S_{rare}}{C_{ace}} + \frac{a_1}{C_{ace}} \gamma_{ace}^2 \]

  \[ C_{ace} = 1 - \frac{a_1}{S_{rare}} \frac{S_{rare}}{C_{ace} S_{rare} (N_{rare} - 1)} \]

  \[ \gamma_{ace}^2 = \max \left[ \frac{1}{C_{ace} N_{rare} (N_{rare} - 1)} - 1, 0 \right] \]

Here $a_i$ refers to number of species with abundance $i$ and $S_{rare}$ is the number of rare species, $S_{abund}$ is the number of abundant species, with an arbitrary threshold of abundance 10 for rare species, and $N_{rare}$ is the number of individuals in rare species.

Functions estimate the standard errors of the estimates. These only concern the number of added species, and assume that there is no variance in the observed richness. The equations of standard errors are too complicated to be reproduced in this help page, but they can be studied in the R source code of the function and are discussed in the vignette that can be read with the `browseVignettes("vegan")`. The standard error are based on the following sources: Chiu et al. (2014) for the Chao estimates and Smith and van Belle (1984) for the first-order Jackknife and the bootstrap (second-order jackknife is still missing). For the variance estimator of $S_{ace}$ see O’Hara (2005).

Functions poolaccum and estaccumR are similar to specaccum, but estimate extrapolated richness indices of specpool or estimater in addition to number of species for random ordering of sampling units. Function specpool uses presence data and estaccumR count data. The functions share summary and plot methods. The summary returns quantile envelopes of permutations correspond-
ing the given level of alpha and standard deviation of permutations for each sample size. NB., these are not based on standard deviations estimated within specpool or estimator, but they are based on permutations. The plot function shows the mean and envelope of permutations with given alpha for models. The selection of models can be restricted and order changes using the display argument in summary or plot. For configuration of plot command, see xyplot.

Value

Function specpool returns a data frame with entries for observed richness and each of the indices for each class in pool vector. The utility function specpool2vec maps the pooled values into a vector giving the value of selected index for each original site. Function estimater returns the estimates and their standard errors for each site. Functions poolaccum and estimater return matrices of permutation results for each richness estimator, the vector of sample sizes and a table of means of permutations for each estimator.

Note

The functions are based on assumption that there is a species pool: The community is closed so that there is a fixed pool size \( S_P \). In general, the functions give only the lower limit of species richness: the real richness is \( S > S_P \), and there is a consistent bias in the estimates. Even the bias-correction in Chao only reduces the bias, but does not remove it completely (Chiu et al. 2014).

Optional small sample correction was added to specpool in vegan 2.2-0. It was not used in the older literature (Chao 1987), but it is recommended recently (Chiu et al. 2014).

See [http://viceroy.eeb.uconn.edu/Estimates](http://viceroy.eeb.uconn.edu/Estimates) for a more complete (and positive) discussion and alternative software for some platforms.

Author(s)

Bob O’Hara (estimator) and Jari Oksanen.

References


See Also

veiledspec, diversity, beals, specaccum.
sppscores

Examples

```R
data(dune)
data(dune.env)
pool <- with(dune.env, specpool(dune, Management))
pool
op <- par(mfrow=c(1,2))
boxplot(specnumber(dune) ~ Management, data = dune.env,
       col = "hotpink", border = "cyan3")
boxplot(specnumber(dune)/specpool2vect(pool) ~ Management,
       data = dune.env, col = "hotpink", border = "cyan3")
par(op)
data(BCI)
## Accumulation model
pool <- poolaccum(BCI)
summary(pool, display = "chao")
plot(pool)
## Quantitative model
estimateR(BCI[1:5,])
```

Description

Distance-based ordination (`dbrda`, `capscale`, `metaMDS`) have no information on species, but some methods may add species scores if community data were available. However, the species scores may be missing (and they always are in `dbrda`), or they may not have a close relation to used dissimilarity index. This function will add the species scores or replace the existing species scores in distance-based methods.

Usage

```R
sppscores(object) <- value
```

Arguments

- `object`: Ordination result.
- `value`: Community data to find the species scores.

Details

Distances have no information on species (columns, variables), and hence distance-based ordination has no information on species scores. However, the species scores can be added as supplementary information after the analysis to help the interpretation of results. Some ordination methods (`capscale`, `metaMDS`) can supplement the species scores during the analysis if community data was available in the analysis.

In `capscale` the species scores are found by projecting the community data to site ordination (linear combination scores), and the scores are accurate if the analysis used Euclidean distances. If the
dissimilarity index can be expressed as Euclidean distances of transformed data (for instance, Chord and Hellinger Distances), the species scores based on transformed data will be accurate, but the function still finds the dissimilarities with untransformed data. Usually community dissimilarities differ in two significant ways from Euclidean distances: They are bound to maximum 1, and they use absolute differences instead of squared differences. In such cases, it may be better to use species scores that are transformed so that their Euclidean distances have a good linear relation to used dissimilarities. It is often useful to standardize data so that each row has unit total, and perform squareroot transformation to damp down the effect of squared differences (see Examples).

Function dbrda never finds the species scores, but it is mathematically similar to capscale, and similar rules should be followed when supplementing the species scores.

Function metaMDS uses weighted averages (wascores) to find the species scores. These have a better relationship with most dissimilarities than the projection scores used in metric ordination, but similar transformation of the community data should be used both in dissimilarities and in species scores.

Value

Replacement function adds the species scores or replaces the old scores in the ordination object.

Author(s)

Jari Oksanen

See Also

Function envfit finds similar scores, but based on correlations. The species scores for non-metric ordination use wascores which can also used directly on any ordination result.

Examples

data(BCI, BCI.env)
mod <- dbrda(vegdist(BCI) ~ Habitat, BCI.env)
## add species scores
sppscores(mod) <- BCI
## Euclidean distances of BCI differ from used dissimilarity
plot(vegdist(BCI), dist(BCI))
## more linear relationship
plot(vegdist(BCI), dist(sqrt(decostand(BCI, "total"))))
## better species scores
sppscores(mod) <- sqrt(decostand(BCI, "total"))
SSarrhenius

Usage

SSarrhenius(area, k, z)
SSgleason(area, k, slope)
SSgitay(area, k, slope)
SSLomolino(area, Asym, xmid, slope)

Arguments

area
Area or size of the sample: the independent variable.
k, z, slope, Asym, xmid
Estimated model parameters: see Details.

Details

All these functions are assumed to be used for species richness (number of species) as the independent variable, and area or sample size as the independent variable. Basically, these define least squares models of untransformed data, and will differ from models for transformed species richness or models with non-Gaussian error.

The Arrhenius model (SSarrhenius) is the expression $k \cdot \text{area}^z$. This is the most classical model that can be found in any textbook of ecology (and also in Dengler 2009). Parameter $z$ is the steepness of the species-area curve, and $k$ is the expected number of species in a unit area.

The Gleason model (SSgleason) is a linear expression $k + \text{slope} \cdot \log(\text{area})$ (Dengler 200). This is a linear model, and starting values give the final estimates; it is provided to ease comparison with other models.

The Gitay model (SSgitay) is a quadratic logarithmic expression $(k + \text{slope} \cdot \log(\text{area}))^2$ (Gitay et al. 1991, Dengler 2009). Parameter slope is the steepness of the species-area curve, and $k$ is the square root of expected richness in a unit area.

The Lomolino model (SSLomolino) is $\text{Asym}/(1 + \text{slope} \cdot \log(\text{xmid}/\text{area}))$ (Lomolino 2000, Dengler 2009). Parameter Asym is the asymptotic maximum number of species, slope is the maximum slope of increase of richness, and xmid is the area where half of the maximum richness is achieved.

In addition to these models, several other models studied by Dengler (2009) are available in standard R self-starting models: Michaelis-Menten (SSmicmen), Gompertz (SSgompertz), logistic (SSlogis), Weibull (SSweibull), and some others that may be useful.

Value

Numeric vector of the same length as area. It is the value of the expression of each model. If all arguments are names of objects the gradient matrix with respect to these names is attached as an attribute named gradient.

Author(s)

Jari Oksanen.
References


See Also

nls, fitspecaccum.

Examples

```r
## Get species area data: sipoo.map gives the areas of islands
data(sipoo, sipoo.map)
S <- specnumber(sipoo)
plot(S ~ area, sipoo.map, xlab = "Island Area (ha)",
ylab = "Number of Species", ylim = c(1, max(S))
## The Arrhenius model
marr <- nls(S ~ SSarrhenius(area, k, z), data=sipoo.map)
marr
## confidence limits from profile likelihood
confint(marr)
## draw a line
xtmp <- with(sipoo.map, seq(min(area), max(area), len=51))
lines(xtmp, predict(marr, newdata=data.frame(area = xtmp)), lwd=2)
## The normal way is to use linear regression on log-log data,
## but this will be different from the previous:
mloglog <- lm(log(S) ~ log(area), data=sipoo.map)
mloglog
lines(xtmp, exp(predict(mloglog, newdata=data.frame(area=xtmp))),
      lty=2)
## Gleason: log-linear
mgle <- nls(S ~ SSgleason(area, k, slope), sipoo.map)
lines(xtmp, predict(mgle, newdata=data.frame(area=xtmp)),
      lwd=2, col=2)
## Gitay: quadratic of log-linear
mgit <- nls(S ~ SSgitay(area, k, slope), sipoo.map)
lines(xtmp, predict(mgit, newdata=data.frame(area=xtmp)),
      lwd=2, col = 3)
## Lomolino: using original names of the parameters (Lomolino 2000):
mlom <- nls(S ~ SSLomolino(area, Smax, A50, Hill), sipoo.map)
mlom
lines(xtmp, predict(mlom, newdata=data.frame(area=xtmp)),
      lwd=2, col = 4)
## One canned model of standard R:
mmic <- nls(S ~ SSmicmen(area, slope, Asym), sipoo.map)
lines(xtmp, predict(mmic, newdata = data.frame(area=xtmp)),
      lwd =2, col = 5)
```
**Stepacross as Flexible Shortest Paths or Extended Dissimilarities**

**Description**

Function `stepacross` tries to replace dissimilarities with shortest paths stepping across intermediate sites while regarding dissimilarities above a threshold as missing data (NA). With `path = "shortest"` this is the flexible shortest path (Williamson 1978, Bradfield & Kenkel 1987), and with `path = "extended"` an approximation known as extended dissimilarities (De’ath 1999). The use of `stepacross` should improve the ordination with high beta diversity, when there are many sites with no species in common.

**Usage**

```r
stepacross(dis, path = "shortest", toolong = 1, trace = TRUE, ...)
```

**Arguments**

- **dis**: Dissimilarity data inheriting from class `dist` or a an object, such as a matrix, that can be converted to a dissimilarity matrix. Functions `vegdist` and `dist` are some functions producing suitable dissimilarity data.

- **path**: The method of stepping across (partial match) Alternative "shortest" finds the shortest paths, and "extended" their approximation known as extended dissimilarities.

- **toolong**: Shortest dissimilarity regarded as NA. The function uses a fuzz factor, so that dissimilarities close to the limit will be made NA, too.

- **trace**: Trace the calculations.

- **...**: Other parameters (ignored).

**Details**

Williamson (1978) suggested using flexible shortest paths to estimate dissimilarities between sites which have nothing in common, or no shared species. With `path = "shortest"` function `stepacross` replaces dissimilarities that are `toolong` or longer with NA, and tries to find shortest paths between all sites using remaining dissimilarities. Several dissimilarity indices are semi-metric which means that they do not obey the triangle inequality $d_{ij} \leq d_{ik} + d_{kj}$, and shortest path algorithm can replace these dissimilarities as well, even when they are shorter than `toolong`.

De’ath (1999) suggested a simplified method known as extended dissimilarities, which are calculated with `path = "extended"`. In this method, dissimilarities that are `toolong` or longer are first
made NA, and then the function tries to replace these NA dissimilarities with a path through single stepping stone points. If not all NA could be replaced with one pass, the function will make new passes with updated dissimilarities as long as all NA are replaced with extended dissimilarities. This means that in the second and further passes, the remaining NA dissimilarities are allowed to have more than one stepping stone site, but previously replaced dissimilarities are not updated. Further, the function does not consider dissimilarities shorter than tooLong, although some of these could be replaced with a shorter path in semi-metric indices, and used as a part of other paths. In optimal cases, the extended dissimilarities are equal to shortest paths, but they may be longer.

As an alternative to defining too long dissimilarities with parameter tooLong, the input dissimilarities can contain NAs. If tooLong is zero or negative, the function does not make any dissimilarities into NA. If there are no NAs in the input and tooLong = 0, path = "shortest" will find shorter paths for semi-metric indices, and path = "extended" will do nothing. Function no.shared can be used to set dissimilarities to NA.

If the data are disconnected or there is no path between all points, the result will contain NAs and a warning is issued. Several methods cannot handle NA dissimilarities, and this warning should be taken seriously. Function distconnected can be used to find connected groups and remove rare outlier observations or groups of observations.

Alternative path = "shortest" uses Dijkstra’s method for finding flexible shortest paths, implemented as priority-first search for dense graphs (Sedgewick 1990). Alternative path = "extended" follows De’ath (1999), but implementation is simpler than in his code.

Value

Function returns an object of class dist with extended dissimilarities (see functions vegdist and dist). The value of path is appended to the method attribute.

Note

The function changes the original dissimilarities, and not all like this. It may be best to use the function only when you really must: extremely high beta diversity where a large proportion of dissimilarities are at their upper limit (no species in common).

Semi-metric indices vary in their degree of violating the triangle inequality. Morisita and Horn–Morisita indices of vegdist may be very strongly semi-metric, and shortest paths can change these indices very much. Mountford index violates basic rules of dissimilarities: non-identical sites have zero dissimilarity if species composition of the poorer site is a subset of the richer. With Mountford index, you can find three sites $i, j, k$ so that $d_{ik} = 0$ and $d_{jk} = 0$, but $d_{ij} > 0$. The results of stepacross on Mountford index can be very weird. If stepacross is needed, it is best to try to use it with more metric indices only.

Author(s)

Jari Oksanen

References


See Also

Function `distconnected` can find connected groups in disconnected data, and function `no.shared` can be used to set dissimilarities as NA. See `swan` for an alternative approach. Function `stepacross` is an essential component in `isomap` and `cophenetic.spantree`.

Examples

```r
# There are no data sets with high beta diversity in vegan, but this
# should give an idea.

data(dune)
dis <- vegdist(dune)
edis <- stepacross(dis)
plot(edis, dis, xlab = "Shortest path", ylab = "Original")
## Manhatten distance have no fixed upper limit.
dis <- vegdist(dune, "manhattan")
is.na(dis) <- no.shared(dune)
dis <- stepacross(dis, tooLong=0)
```

---

### stressplot.wcmdscale

*Display Ordination Distances Against Observed Distances in Eigenvector Ordinations*

**Description**

Functions plot ordination distances in given number of dimensions against observed distances or distances in full space in eigenvector methods. The display is similar as the Shepard diagram (`stressplot` for non-metric multidimensional scaling with `metadms` or `monodms`), but shows the linear relationship of the eigenvector ordinations. The stressplot methods are available for `wcmdscale`, `rda`, `cca`, `capscale`, `dbrda`, `prcomp` and `princomp`.

**Usage**

```r
# S3 method for class 'wcmdscale'
stressplot(object, k = 2, pch, p.col = "blue", l.col = "red",
           lwd = 2, ...)
```

**Arguments**

- `object`: Result object from eigenvector ordination (`wcmdscale`, `rda`, `cca`, `capscale`)
- `k`: Number of dimensions for which the ordination distances are displayed.
- `pch, p.col, l.col, lwd`: Plotting character, point colour and line colour like in default `stressplot`
- `...`: Other parameters to functions, e.g. graphical parameters.
Details

The functions offer a similar display for eigenvector ordinations as the standard Shepard diagram (stressplot) in non-metric multidimensional scaling. The ordination distances in given number of dimensions are plotted against observed distances. With metric distances, the ordination distances in full space (with all ordination axes) are equal to observed distances, and the fit line shows this equality. In general, the fit line does not go through the points, but the points for observed distances approach the fit line from below. However, with non-Euclidean distances (in wcmdscale or capscale) with negative eigenvalues the ordination distances can exceed the observed distances in real dimensions; the imaginary dimensions with negative eigenvalues will correct these excess distances. If you have used capscale or wcmdscale with argument add to avoid negative eigenvalues, the ordination distances will exceed the observed dissimilarities.

In partial ordination (cca, rda and capscale with Condition in the formula), the distances in the partial component are included both in the observed distances and in ordination distances. With k=0, the ordination distances refer to the partial ordination.

Value

Functions draw a graph and return invisibly the ordination distances or the ordination distances.

Author(s)

Jari Oksanen.

See Also

stressplot and stressplot.monoMDS for standard Shepard diagrams.

Examples

data(dune, dune.env)
mod <- rda(dune)
stressplot(mod)
mod <- rda(dune ~ Management, dune.env)
stressplot(mod, k=3)

indices of taxonomic diversity and distinctness

Indices of Taxonomic Diversity and Distinctness

Description

Function finds indices of taxonomic diversity and distinctness, which are averaged taxonomic distances among species or individuals in the community (Clarke & Warwick 1998, 2001)

Usage

taxondive(comm, dis, match.force = FALSE)
taxa2dist(x, varstep = FALSE, check = TRUE, labels)
Arguments

- **comm**: Community data.
- **dis**: Taxonomic distances among taxa in `comm`. This should be a `dist` object or a symmetric square matrix.
- **match.force**: Force matching of column names in `comm` and labels in `dis`. If `FALSE`, matching only happens when dimensions differ, and in that case the species must be in identical order in both.
- **x**: Classification table with a row for each species or other basic taxon, and columns for identifiers of its classification at higher levels.
- **varstep**: Vary step lengths between successive levels relative to proportional loss of the number of distinct classes.
- **check**: If `TRUE`, remove all redundant levels which are different for all rows or constant for all rows and regard each row as a different basal taxon (species). If `FALSE` all levels are retained and basal taxa (species) also must be coded as variables (columns). You will get a warning if species are not coded, but you can ignore this if that was your intention.
- **labels**: The `labels` attribute of taxonomic distances. Row names will be used if this is not given. Species will be matched by these labels in `comm` and `dis` in `taxondive` if these have different dimensions.

Details

Clarke & Warwick (1998, 2001) suggested several alternative indices of taxonomic diversity or distinctness. Two basic indices are called taxonomic diversity ($\Delta$) and distinctness ($\Delta^*$):

$$
\Delta = \frac{\left(\sum_{i<j} \omega_{ij} x_i x_j\right)}{\left(n(n-1)/2\right)}
\Delta^* = \frac{\left(\sum_{i<j} \omega_{ij} x_i x_j\right)}{\left(\sum_{i<j} x_i x_j\right)}
$$

The equations give the index value for a single site, and summation goes over species $i$ and $j$. Here $\omega$ are taxonomic distances among taxa, and $x$ are species abundances, and $n$ is the total abundance for a site. With presence/absence data both indices reduce to the same index $\Delta^+$, and for this index Clarke & Warwick (1998) also have an estimate of its standard deviation. Clarke & Warwick (2001) presented two new indices: $s\Delta^+$ is the product of species richness and $\Delta^+$, and index of variation in taxonomic distinctness ($\Lambda^+$) defined as

$$
\Lambda^+ = \frac{\left(\sum_{i<j} \omega_{ij}^2\right)}{\left(n(n-1)/2\right)} - (\Delta^+)^2
$$

The `dis` argument must be species dissimilarities. These must be similar to dissimilarities produced by `dist`. It is customary to have integer steps of taxonomic hierarchies, but other kind of dissimilarities can be used, such as those from phylogenetic trees or genetic differences. Further, the `dis` need not be taxonomic, but other species classifications can be used.

Function `taxa2dist` can produce a suitable `dist` object from a classification table. Each species (or basic taxon) corresponds to a row of the classification table, and columns give the classification at different levels. With `varstep = FALSE` the successive levels will be separated by equal steps, and with `varstep = TRUE` the step length is relative to the proportional decrease in the number
of classes (Clarke & Warwick 1999). With check = TRUE, the function removes classes which are distinct for all species or which combine all species into one class, and assumes that each row presents a distinct basic taxon. The function scales the distances so that longest path length between taxa is 100 (not necessarily when check = FALSE).

Function plot.taxondive plots $\Delta^+$ against Number of species, together with expectation and its approximate $2*sd$ limits. Function summary.taxondive finds the $z$ values and their significances from Normal distribution for $\Delta^+$.

Value

Function returns an object of class taxondive with following items:

- **Species**: Number of species for each site.
- **D, Dstar, Dplus, SDplus, Lambda**: $\Delta$, $\Delta^*$, $\Delta^+$, $s\Delta^+$ and $\Lambda^+$ for each site.
- **sd.Dplus**: Standard deviation of $\Delta^+$.
- **ED, EDstar, EDplus**: Expected values of corresponding statistics.

Function taxa2dist returns an object of class "dist", with an attribute "steps" for the step lengths between successive levels.

Note

The function is still preliminary and may change. The scaling of taxonomic dissimilarities influences the results. If you multiply taxonomic distances (or step lengths) by a constant, the values of all Deltas will be multiplied with the same constant, and the value of $\Lambda^+$ by the square of the constant.

Author(s)

Jari Oksanen

References


See Also

diversity.
tolerance

Examples

## Preliminary: needs better data and some support functions
data(dune)
data(dune.taxon)
# Taxonomic distances from a classification table with variable step lengths.
taxdis <- taxa2dist(dune.taxon, varstep=TRUE)
plot(hclust(taxdis), hang = -1)
# Indices
mod <- taxondive(dune, taxdis)
mod
summary(mod)
plot(mod)

tolerance Species tolerances and sample heterogeneities

Description

Species tolerances and sample heterogeneities.

Usage

tolerance(x, ...)

## S3 method for class 'cca'
tolerance(x, choices = 1:2, which = c("species","sites"),
           scaling = "species", useN2 = TRUE, hill = FALSE, ...)

Arguments

x object of class "cca".
choices numeric; which ordination axes to compute tolerances and heterogeneities for. Defaults to axes 1 and 2.
which character; one of "species" or "sites", indicating whether species tolerances or sample heterogeneities respectively are computed.
scaling character or numeric; the ordination scaling to use. See scores.cca for details.
hill logical; if scaling is a character, these control whether Hill's scaling is used for (C)CA respectively. See scores.cca for details.
useN2 logical; should the bias in the tolerances / heterogeneities be reduced via scaling by Hill's N2?
... arguments passed to other methods.

Details

Value

Matrix of tolerances/heterogeneities with some additional attributes: which, scaling, and N2, the latter of which will be NA if useN2 = FALSE.

Author(s)

Gavin L. Simpson

Examples

data(dune)
data(dune.env)
mod <- cca(dune ~ ., data = dune.env)

## defaults to species tolerances
tolerance(mod)

## sample heterogeneities for CCA axes 1:6
tolerance(mod, which = "sites", choices = 1:6)

---

**treedive**  
*Functional Diversity and Community Distances from Species Trees*

Description

Functional diversity is defined as the total branch length in a trait dendrogram connecting all species, but excluding the unnecessary root segments of the tree (Petchey and Gaston 2006). Tree distance is the increase in total branch length when combining two sites.

Usage

treedive(comm, tree, match.force = TRUE, verbose = TRUE)
treedist(tree)
treedist(x, tree, relative = TRUE, match.force = TRUE, ...)

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>comm, x</td>
<td>Community data frame or matrix.</td>
</tr>
<tr>
<td>tree</td>
<td>A dendrogram which for treedive must be for species (columns).</td>
</tr>
<tr>
<td>match.force</td>
<td>Force matching of column names in data (comm, x) and labels in tree. If FALSE, matching only happens when dimensions differ (with a warning or message). The order of data must match to the order in tree if matching by names is not done.</td>
</tr>
<tr>
<td>verbose</td>
<td>Print diagnostic messages and warnings.</td>
</tr>
<tr>
<td>relative</td>
<td>Use distances relative to the height of combined tree.</td>
</tr>
<tr>
<td>...</td>
<td>Other arguments passed to functions (ignored).</td>
</tr>
</tbody>
</table>
**Details**

Function `treeheight` finds the sum of lengths of connecting segments in a dendrogram produced by `hclust`, or other dendrogram that can be coerced to a correct type using `as.hclust`. When applied to a clustering of species traits, this is a measure of functional diversity (Petchey and Gaston 2002, 2006), and when applied to phylogenetic trees this is phylogenetic diversity.

Function `treedive` finds the tree height for each site (row) of a community matrix. The function uses a subset of dendrogram for those species that occur in each site, and excludes the tree root if that is not needed to connect the species (Petchey and Gaston 2006). The subset of the dendrogram is found by first calculating cophenetic distances from the input dendrogram, then reconstructing the dendrogram for the subset of the cophenetic distance matrix for species occurring in each site. Diversity is 0 for one species, and `NA` for empty communities.

Function `treedist` finds the dissimilarities among trees. Pairwise dissimilarity of two trees is found by combining species in a common tree and seeing how much of the tree height is shared and how much is unique. With `relative = FALSE` the dissimilarity is defined as $2(A \cup B) - A - B$, where $A$ and $B$ are heights of component trees and $A \cup B$ is the height of the combined tree. With `relative = TRUE` the dissimilarity is $(2(A \cup B) - A - B)/(A \cup B)$. Although the latter formula is similar to Jaccard dissimilarity (see `vegdist`, `designdist`), it is not in the range $0 \ldots 1$, since combined tree can add a new root. When two zero-height trees are combined into a tree of above zero height, the relative index attains its maximum value 2. The dissimilarity is zero from a combined zero-height tree.

The functions need a dendrogram of species traits or phylogenies as an input. If species traits contain `factor` or `ordered` factor variables, it is recommended to use Gower distances for mixed data (function `daisy` in package `cluster`), and usually the recommended clustering method is UPGMA (method = "average" in function `hclust`) (Podani and Schmera 2006). Phylogenetic trees can be changed into dendrograms using `as.hclust.phylo` (package `ape`).

It is possible to analyse the non-randomness of tree diversity using `oecosimu`. This needs specifying an adequate Null model, and the results will change with this choice.

**Value**

A vector of diversity values or a single tree height, or a dissimilarity structure that inherits from `dist` and can be used similarly.

**Author(s)**

Jari Oksanen

**References**


See Also

Function treedive is similar to the phylogenetic diversity function \texttt{pd} in \texttt{picante}, but excludes tree root if that is not needed to connect species. Function treedist is similar to the phylogenetic similarity \texttt{phylosor} in \texttt{picante}, but excludes unneeded tree root and returns distances instead of similarities.

\texttt{taxondive} is something very similar from another world.

Examples

```r
## There is no data set on species properties yet, and we demonstrate
## the methods using phylogenetic trees
data(dune)
data(dune.phylodis)
cl <- hclust(dune.phylodis)
treedive(dune, cl)
## Significance test using Null model communities.
## The current choice fixes numbers of species and picks species
## proportionally to their overall frequency
oecosimu(dune, treedive, "r", tree = cl, verbose = FALSE)
## Phylogenetically ordered community table
dtree <- treedist(dune, cl)
tabasco(dune, hclust(dtree), cl)
## Use tree distances in capscale
capscale(dtree ~ 1, comm = dune)
```

---

**tsallis**

\textit{Tsallis Diversity and Corresponding Accumulation Curves}

Description

Function \texttt{tsallis} find Tsallis diversities with any scale or the corresponding evenness measures. Function \texttt{tsallisaccum} finds these statistics with accumulating sites.

Usage

```
tsallis(x, scales = seq(0, 2, 0.2), norm = FALSE, hill = FALSE)
tsallisaccum(x, scales = seq(0, 2, 0.2), permutations = 100, 
            raw = FALSE, subset, ...)
```

## S3 method for class 'tsallisaccum'
persp(x, theta = 220, phi = 15, col = heat.colors(100), zlim, ...)

Arguments

- \texttt{x} Community data matrix or plotting object.
- \texttt{scales} Scales of Tsallis diversity.
- \texttt{norm} Logical, if \texttt{TRUE} diversity values are normalized by their maximum (diversity value at equiprobability conditions).
Calculate Hill numbers.

Usually an integer giving the number permutations, but can also be a list of control values for the permutations as returned by the function `how`, or a permutation matrix where each row gives the permuted indices.

If FALSE then return summary statistics of permutations, and if TRUE then returns the individual permutations.

logical expression indicating sites (rows) to keep: missing values are taken as FALSE.

angles defining the viewing direction. theta gives the azimuthal direction and phi the colatitude.

Colours used for surface.

Limits of vertical axis.

Other arguments which are passed to `tsallis` and to graphical functions.

The Tsallis diversity (also equivalent to Patil and Taillie diversity) is a one-parametric generalised entropy function, defined as:

\[ H_q = \frac{1}{q-1} (1 - \sum_{i=1}^{S} p_i^q) \]

where \( q \) is a scale parameter, \( S \) the number of species in the sample (Tsallis 1988, Tothmeresz 1995). This diversity is concave for all \( q > 0 \), but non-additive (Keylock 2005). For \( q = 0 \) it gives the number of species minus one, as \( q \) tends to 1 this gives Shannon diversity, for \( q = 2 \) this gives the Simpson index (see function `diversity`).

If \( \text{norm} = \text{TRUE} \), `tsallis` gives values normalized by the maximum:

\[ H_q(\text{max}) = \frac{S^{1-q} - 1}{1 - q} \]

where \( S \) is the number of species. As \( q \) tends to 1, maximum is defined as \( \ln(S) \).

If \( \text{hill} = \text{TRUE} \), `tsallis` gives Hill numbers (numbers equivalents, see Jost 2007):

\[ D_q = (1 - (q - 1)H)^{1/(1-q)} \]

Details on plotting methods and accumulating values can be found on the help pages of the functions `renyi` and `renyiaccum`.

Function `tsallis` returns a data frame of selected indices. Function `tsallisaccum` with argument \( \text{raw} = \text{FALSE} \) returns a three-dimensional array, where the first dimension are the accumulated sites, second dimension are the diversity scales, and third dimension are the summary statistics mean, stdev, min, max, Qnt .025 and Qnt .975. With argument \( \text{raw} = \text{TRUE} \) the statistics on the third dimension are replaced with individual permutation results.
Author(s)

Péter Sólymos, <solymos@ualberta.ca>, based on the code of Roeland Kindt and Jari Oksanen written for renyi

References


See Also

Plotting methods and accumulation routines are based on functions renyi and renyiaccum. An object of class ‘tsallisaccum’ can be used with function rgl.renyiaccum as well. See also settings for persp.

Examples

data(BCI)
i <- sample(nrow(BCI), 12)
x1 <- tsallis(BCI[i,])
x1
diversity(BCI[i,],"simpson") == x1["2"]
plot(x1)
x2 <- tsallis(BCI[i,],norm=TRUE)
x2
plot(x2)
mod1 <- tsallisaccum(BCI[i,])
plot(mod1, as.table=TRUE, col = c(1, 2, 2))
persp(mod1)
mod2 <- tsallisaccum(BCI[i,], norm=TRUE)
persp(mod2,theta=100,phi=30)
varpart

Description

The varespec data frame has 24 rows and 44 columns. Columns are estimated cover values of 44 species. The variable names are formed from the scientific names, and are self explanatory for anybody familiar with the vegetation type. The varechem data frame has 24 rows and 14 columns, giving the soil characteristics of the very same sites as in the varespec data frame. The chemical measurements have obvious names. Baresoil gives the estimated cover of bare soil, humdepth the thickness of the humus layer.

Usage

```
data(varechem)
data(varespec)
```

References


Examples

```
data(varespec)
data(varechem)
```

---

**varpart**

*Partition the Variation of Community Matrix by 2, 3, or 4 Explanatory Matrices*

Description

The function partitions the variation in community data or community dissimilarities with respect to two, three, or four explanatory tables, using adjusted $R^2$ in redundancy analysis ordination (RDA) or distance-based redundancy analysis. If response is a single vector, partitioning is by partial regression. Collinear variables in the explanatory tables do NOT have to be removed prior to partitioning.

Usage

```
varpart(Y, X, ..., data, chisquare = FALSE, transfo, scale = FALSE,
          add = FALSE, sqrt.dist = FALSE, permutations)
showvarparts(parts, labels, bg = NULL, alpha = 63, Xnames,
             id.size = 1.2, ...)  
## S3 method for class 'varpart234'
plot(x, cutoff = 0, digits = 1, ...)
```
Arguments

Y  Data frame or matrix containing the response data table or dissimilarity structure inheriting from `dist`. In community ecology, that table is often a site-by-species table or a dissimilarity object.

X  Two to four explanatory models, variables or tables. These can be defined in three alternative ways: (1) one-sided model formulae beginning with `~` and then defining the model, (2) name of a single numeric or factor variable, or (3) name of matrix with numeric or data frame with numeric and factor variables. The model formulae can have factors, interaction terms and transformations of variables. The names of the variables in the model formula are found in data frame given in data argument, and if not found there, in the user environment. Single variables, data frames or matrices are found in the user environment. All entries till the next argument (data or transfo) are interpreted as explanatory models, and the names of these extra arguments cannot be abbreviated nor omitted.

... Other parameters passed to functions. NB, arguments after dots cannot be abbreviated but they must be spelt out completely.

data  The data frame with the variables used in the formulae in X.

chisquare  Partition Chi-square or the inertia of Correspondence Analysis (`cca`).

transfo  Transformation for Y (community data) using `decostand`. All alternatives in `decostand` can be used, and those preserving Euclidean metric include "hellinger", "chi.square", "total", "norm". Ignored if Y are dissimilarities.

scale  Should the columns of Y be standardized to unit variance. Ignored if Y are dissimilarities.

add  Add a constant to the non-diagonal values to euclidify dissimilarities (see `wcmdscale` for details). Choice "lingoes" (or TRUE) use the recommended method of Legendre & Anderson (1999: “method 1”) and "cailliez" uses their “method 2”. The argument has an effect only when Y are dissimilarities.

sqrt.dist  Take square root of dissimilarities. This often euclidifies dissimilarities. NB., the argument name cannot be abbreviated. The argument has an effect only when Y are dissimilarities.

permutations  If chisquare = TRUE, the adjusted $R^2$ is estimated by permutations, and this parameter can be a list of control values for the permutations as returned by the function `how`, or the number of permutations required, or a permutation matrix where each row gives the permuted indices.

parts  Number of explanatory tables (circles) displayed.

labels  Labels used for displayed fractions. Default is to use the same letters as in the printed output.

bg  Fill colours of circles or ellipses.

alpha  Transparency of the fill colour. The argument takes precedence over possible transparency definitions of the colour. The value must be in range 0...255, and low values are more transparent. Transparency is not available in all graphics devices or file formats.

Xnames  Names for sources of variation. Default names are X1, X2, X3 and X4. Xnames=NA, Xnames=NULL and Xnames="" produce no names. The names can be changed to other names. It is often best to use short names.
id.size  A numerical value giving the character expansion factor for the names of circles or ellipses.

x  The varpart result.

cutoff  The values below cutoff will not be displayed.

digits  The number of significant digits; the number of decimal places is at least one higher.

Details

The functions partition the variation in \( Y \) into components accounted for by two to four explanatory tables and their combined effects. If \( Y \) is a multicolumn data frame or matrix, the partitioning is based on redundancy analysis (RDA, see \texttt{rda}) or on constrained correspondence analysis if \texttt{chisquare = TRUE} (CCA, see \texttt{cca}). If \( Y \) is a single variable, the partitioning is based on linear regression. If \( Y \) are dissimilarities, the decomposition is based on distance-based redundancy analysis (db-RDA, see \texttt{capscale}) following McArdle & Anderson (2001). The input dissimilarities must be compatible to the results of \texttt{dist}. \texttt{Vegan} functions \texttt{vegdist, designdist, raupcrick} and \texttt{betadiver} produce such objects, as do many other dissimilarity functions in \texttt{R} packages. However, symmetric square matrices are not recognized as dissimilarities but must be transformed with \texttt{as.dist}. Partitioning will be made to squared dissimilarities analogously to using variance with rectangular data – unless \texttt{sqrt.dist = TRUE} was specified.

The function primarily uses adjusted \( R^2 \) to assess the partitions explained by the explanatory tables and their combinations (see \texttt{RsquareAdj}), because this is the only unbiased method (Peres-Neto et al., 2006). The raw \( R^2 \) for basic fractions are also displayed, but these are biased estimates of variation explained by the explanatory table. In correspondence analysis (\texttt{chisquare = TRUE}), the adjusted \( R^2 \) are found by permutation and they vary in repeated analyses.

The identifiable fractions are designated by lower case alphabets. The meaning of the symbols can be found in the separate document (use \texttt{browseVignettes(“vegan”)}), or can be displayed graphically using function \texttt{showvarparts}.

A fraction is testable if it can be directly expressed as an RDA or db-RDA model. In these cases the printed output also displays the corresponding RDA model using notation where explanatory tables after \( | \) are conditions (partialled out; see \texttt{rda} for details). Although single fractions can be testable, this does not mean that all fractions simultaneously can be tested, since the number of testable fractions is higher than the number of estimated models. The non-testable components are found as differences of testable components. The testable components have permutation variance in correspondence analysis (\texttt{chisquare = TRUE}), and the non-testable components have even higher variance.

An abridged explanation of the alphabetic symbols for the individual fractions follows, but computational details should be checked in the vignette (readable with \texttt{browseVignettes(“vegan”)}) or in the source code.

With two explanatory tables, the fractions explained uniquely by each of the two tables are \([a]\) and \([c]\), and their joint effect is \([b]\) following Borcard et al. (1992).

With three explanatory tables, the fractions explained uniquely by each of the three tables are \([a]\) to \([c]\), joint fractions between two tables are \([d]\) to \([f]\), and the joint fraction between all three tables is \([g]\).
With four explanatory tables, the fractions explained uniquely by each of the four tables are \([a]\) to \([d]\), joint fractions between two tables are \([e]\) to \([j]\), joint fractions between three variables are \([k]\) to \([n]\), and the joint fraction between all four tables is \([o]\).

There is a \texttt{plot} function that displays the Venn diagram and labels each intersection (individual fraction) with the adjusted R squared if this is higher than \texttt{cutoff}. A helper function \texttt{showvarpart} displays the fraction labels. The circles and ellipses are labelled by short default names or by names defined by the user in argument \texttt{Xnames}. Longer explanatory file names can be written on the \texttt{varpart} output plot as follows: use option \texttt{Xnames=NA}, then add new names using the \texttt{text} function. A bit of fiddling with coordinates (see \texttt{locator}) and character size should allow users to place names of reasonably short lengths on the \texttt{varpart} plot.

**Value**

Function \texttt{varpart} returns an object of class "\texttt{varpart}" with items \texttt{scale} and \texttt{transfo} (can be missing) which hold information on standardizations, \texttt{tables} which contains names of explanatory tables, and \texttt{call} with the function \texttt{call}. The function \texttt{varpart} calls function \texttt{varpart2}, \texttt{varpart3} or \texttt{varpart4} which return an object of class "\texttt{varpart234}" and saves its result in the item \texttt{part}. The items in this object are:

- \texttt{SS.Y}: Sum of squares of matrix Y.
- \texttt{n}: Number of observations (rows).
- \texttt{nsets}: Number of explanatory tables.
- \texttt{bigwarning}: Warnings on collinearity.
- \texttt{fract}: Basic fractions from all estimated constrained models.
- \texttt{indfract}: Individual fractions or all possible subsections in the Venn diagram (see \texttt{showvarparts}).
- \texttt{contr1}: Fractions that can be found after conditioning on single explanatory table in models with three or four explanatory tables.
- \texttt{contr2}: Fractions that can be found after conditioning on two explanatory tables in models with four explanatory tables.

**Fraction Data Frames**

Items \texttt{fract}, \texttt{indfract}, \texttt{contr1} and \texttt{contr2} are all data frames with items:

- \texttt{Df}: Degrees of freedom of numerator of the \(F\)-statistic for the fraction.
- \texttt{R.square}: Raw \(R^2\). This is calculated only for \texttt{fract} and this is \texttt{NA} in other items.
- \texttt{Adj.R.square}: Adjusted \(R^2\).
- \texttt{Testable}: If the fraction can be expressed as a (partial) RDA model, it is directly \texttt{Testable}, and this field is \texttt{TRUE}. In that case the fraction label also gives the specification of the testable RDA model.

**Note**

You can use command \texttt{browseVignettes("vegan"}) to display document which presents Venn diagrams showing the fraction names in partitioning the variation of Y with respect to 2, 3, and 4 tables of explanatory variables, as well as the equations used in variation partitioning.
The functions frequently give negative estimates of variation. Adjusted $R^2$ can be negative for any fraction; unadjusted $R^2$ of testable fractions of variances will be non-negative. Non-testable fractions cannot be found directly, but by subtracting different models, and these subtraction results can be negative. The fractions are orthogonal, or linearly independent, but more complicated or nonlinear dependencies can cause negative non-testable fractions. Any fraction can be negative for non-Euclidean dissimilarities because the underlying db-RDA model can yield negative eigenvalues (see `capscale`, `dbrda`). These negative eigenvalues in the underlying analysis can be avoided with arguments `sqrt.dist` and `add` which have a similar effect as in `capscale`: the square roots of several dissimilarities do not have negative eigenvalues, and no negative eigenvalues are produced after Lingoes or Cailliez adjustment, which in effect add random variation to the dissimilarities.

A simplified, fast version of RDA, CCA and dbRDA are used (functions `simplerda`, `simplecca` and `simpledbRDA`). The actual calculations are done in functions `varpartR` to `varpartT`, but these are not intended to be called directly by the user.

**Author(s)**

Pierre Legendre, Departement de Sciences Biologiques, Universite de Montreal, Canada. Further developed by Jari Oksanen.

**References**

(a) References on variation partitioning


(b) Reference on transformations for species data


(c) Reference on adjustment of the bimultivariate redundancy statistic


(d) References on partitioning of dissimilarities


**See Also**

For analysing testable fractions, see `rda` and `anova.cca`. For data transformation, see `decostand`. Function `inertcomp` gives (unadjusted) components of variation for each species or site separately. Function `rda` displays unadjusted components in its output, but `RsquareAdj` will give adjusted $R^2$ that are similar to the current function also for partial models.
Examples

data(mite)
data(mite.env)
data(mite.pcmn)

# Two explanatory data frames -- Hellinger-transform Y
mod <- varpart(mite, mite.env, mite.pcmn, transfo="hel")
mod

## Use fill colours
showvarparts(2, bg = c("hotpink","skyblue"))
plot(mod, bg = c("hotpink","skyblue"))
## Test fraction [a] using partial RDA, '-' in formula tells to use
## all variables of data mite.env.
aFrac <- rda(decostand(mite, "hel"), mite.env, mite.pcmn)
anova(aFrac)
## RsquareAdj gives the same result as component [a] of varpart
RsquareAdj(aFrac)

## Partition Bray-Curtis dissimilarities
varpart(vegdist(mite), mite.env, mite.pcmn)
## Three explanatory tables with formula interface
mod <- varpart(mite, ~ SubsDens + WatrCont, ~ Substrate + Shrub + Topo,
mite.pcmn, data=mite.env, transfo="hel")
mod
showvarparts(3, bg=2:4)
plot(mod, bg=2:4)

## Use RDA to test fraction [a]
## Matrix can be an argument in formula
rda.result <- rda(decostand(mite, "hell") ~ SubsDens + WatrCont +
Condition(Substrate + Shrub + Topo) +
Condition(as.matrix(mite.pcmn)), data = mite.env)
anova(rda.result)

## Four explanatory tables
mod <- varpart(mite, ~ SubsDens + WatrCont, ~Substrate + Shrub + Topo,
mite.pcmn[,1:11], mite.pcmn[,12:22], data=mite.env, transfo="hel")
mod
plot(mod, bg=2:5)
## Show values for all partitions by putting 'cutoff' low enough:
plot(mod, cutoff = -Inf, cex = 0.7, bg=2:5)

vegan-deprecated

Deprecation Functions in vegan package

Description

These functions are provided for compatibility with older versions of vegan only, and may be defunct as soon as the next release.
**Usage**

```r
as.mlm(x)
```

**Arguments**

- `x` Constrained ordination result.

**Details**

Function `as.mlm` refits results of constrained ordination (`cca`, `rda`, `capscale`) as a multiple response linear model (`lm`). This allows finding influence statistics (`influence.measures`). It is also possible to derive several other statistics, but most of these are biased and misleading, since refitting ignores a major component of variation in constrained ordination. This function is replaced with a set functions that can find the same statistics directly from the ordination result object: see `hatvalues.cca`, `rstandard.cca`, `rstudent.cca`, `cooks.distance.cca`, `vcov.cca`.

**See Also**

- `influence.measures`
- `hatvalues.cca`
- `rstandard.cca`
- `rstudent.cca`
- `cooks.distance.cca`
- `vcov.cca`
Dissimilarity Indices for Community Ecologists

Description

The function computes dissimilarity indices that are useful for or popular with community ecologists. All indices use quantitative data, although they would be named by the corresponding binary index, but you can calculate the binary index using an appropriate argument. If you do not find your favourite index here, you can see if it can be implemented using designdist. Gower, Bray–Curtis, Jaccard and Kulczynski indices are good in detecting underlying ecological gradients (Faith et al. 1987). Morisita, Horn–Morisita, Binomial, Cao and Chao indices should be able to handle different sample sizes (Wolda 1981, Krebs 1999, Anderson & Millar 2004), and Mountford (1962) and Raup-Crick indices for presence–absence data should be able to handle unknown (and variable) sample sizes. Most of these indices are discussed by Krebs (1999) and Legendre & Legendre (2012), and their properties further compared by Wolda (1981) and Legendre & De Cáceres (2012).

Usage

vegdist(x, method="bray", binary=FALSE, diag=FALSE, upper=FALSE, na.rm = FALSE, ...)

Arguments

x
Community data matrix.

method
Dissimilarity index, partial match to "manhattan", "euclidean", "canberra", "clark", "bray", "kulczynski", "jaccard", "gower", "altGower", "morisita", "horn", "mountford", "raup", "binomial", "chao", "cao" or "mahalanobis".

binary
Perform presence/absence standardization before analysis using decostand.

diag
Compute diagonals.

upper
Return only the upper diagonal.

na.rm
Pairwise deletion of missing observations when computing dissimilarities.

...
Other parameters. These are ignored, except in method ="gower" which accepts rangeNglobal parameter of decostand.

Details

Jaccard ("jaccard"), Mountford ("mountford"), Raup–Crick ("raup"), Binomial and Chao indices are discussed later in this section. The function also finds indices for presence/absence data by setting binary = TRUE. The following overview gives first the quantitative version, where \( x_{ij} \) and \( x_{ik} \) refer to the quantity on species (column) \( i \) and sites (rows) \( j \) and \( k \). In binary versions \( A \) and \( B \) are the numbers of species on compared sites, and \( J \) is the number of species that occur on both compared sites similarly as in designdist (many indices produce identical binary versions):

euclidean
\[
d_{jk} = \sqrt{\sum (x_{ij} - x_{ik})^2}
\]
binary:
\[
\sqrt{A + B - 2J}
\]
manhattan  \( d_{jk} = \sum_i |x_{ij} - x_{ik}| \)  
binary: \( A + B - 2J \)
gower  \( d_{jk} = (1/M) \sum_i \frac{|x_{ij} - x_{ik}|}{\max x_{ij} - \min x_{ij}} \)  
binary: \((A + B - 2J)/M\)
where \( M \) is the number of columns (excluding missing values)
altGower  \( d_{jk} = (1/NZ) \sum_i |x_{ij} - x_{ik}| \)  
where \( NZ \) is the number of non-zero columns excluding double-zeros (Anderson et al. 2006).
binary: \( A + B - 2J \)
canberra  \( d_{jk} = \frac{1}{NZ} \sum_i \frac{|x_{ij} - x_{ik}|}{|x_{ij}| + |x_{ik}|} \)  
where \( NZ \) is the number of non-zero entries.
binary: \( A + B - 2J \)
clark  \( d_{jk} = \sqrt{\frac{1}{NZ} \sum_i (\frac{x_{ij} - x_{ik}}{x_{ij} + x_{ik}})^2} \)  
where \( NZ \) is the number of non-zero entries.
binary: \( A + B - 2J \)
bray  \( d_{jk} = \frac{\sum_i |x_{ij} - x_{ik}|}{\sum (x_{ij} + x_{ik})} \)  
binary: \( A + B - 2J \)
kulczynski  \( d_{jk} = 1 - 0.5(\sum_i \min(x_{ij},x_{ik})/\sum_i x_{ij} + \sum_i \min(x_{ij},x_{ik})/\sum_i x_{ik}) \)  
binary: \( 1 - (J/A + J/B)/2 \)
Morisita  \( d_{jk} = 1 - \frac{2\sum_i x_{ij}x_{ik}}{(\lambda_j + \lambda_k)\sum_i x_{ij}\sum_i x_{ik}} \), where \( \lambda_j = \sum_i x_{ij}/(\sum_i x_{ij}) \)  
binary: cannot be calculated
horn  Like morisita, but \( \lambda_j = \sum_i x_{ij}^2/((\sum_i x_{ij})^2) \)  
binary: \( A + B - 2J \)
binomial  \( d_{jk} = \sum_i [x_{ij} \log(\frac{x_{ij}}{n_i}) + x_{ik} \log(\frac{x_{ik}}{n_i}) - n_i \log(\frac{1}{2})]/n_i \), where \( n_i = x_{ij} + x_{ik} \)  
binary: log \((2) \times (A + B - 2J) \)
Cao  \( d_{jk} = \frac{1}{S} \sum_i \log(\frac{n_i}{2}) - (x_{ij} \log(x_{ij}) + x_{ik} \log(x_{ik}))/n_i \), where \( S \) is the number of species in compared sites and \( n_i = x_{ij} + x_{ik} \)

Jaccard index is computed as \( 2B/(1 + B) \), where \( B \) is Bray–Curtis dissimilarity.

Binomial index is derived from Binomial deviance under null hypothesis that the two compared communities are equal. It should be able to handle variable sample sizes. The index does not have a fixed upper limit, but can vary among sites with no shared species. For further discussion, see Anderson & Millar (2004).

Cao index or CYd index (Cao et al. 1997) was suggested as a minimally biased index for high beta diversity and variable sampling intensity. Cao index does not have a fixed upper limit, but can vary among sites with no shared species. The index is intended for count (integer) data, and it is undefined for zero abundances; these are replaced with arbitrary value 0.1 following Cao et al. (1997). Cao et al. (1997) used \( \log_{10} \), but the current function uses natural logarithms so that the values are approximately 2.30 times higher than with 10-based logarithms. Anderson & Thompson (2004) give an alternative formulation of Cao index to highlight its relationship with Binomial index (above).
Mountford index is defined as $M = 1/\alpha$ where $\alpha$ is the parameter of Fisher’s logseries assuming that the compared communities are samples from the same community (cf. fisherfit, fisher.alpha). The index $M$ is found as the positive root of equation $\exp(aM) + \exp(bM) = 1 + \exp[(a + b - j)M]$, where $j$ is the number of species occurring in both communities, and $a$ and $b$ are the number of species in each separate community (so the index uses presence–absence information). Mountford index is usually misrepresented in the literature: indeed Mountford (1962) suggested an approximation to be used as starting value in iterations, but the proper index is defined as the root of the equation above. The function vegdist solves $M$ with the Newton method. Please note that if either $a$ or $b$ are equal to $j$, one of the communities could be a subset of other, and the dissimilarity is 0 meaning that non-identical objects may be regarded as similar and the index is non-metric. The Mountford index is in the range $0 \ldots \log(2)$, but the dissimilarities are divided by $\log(2)$ so that the results will be in the conventional range $0 \ldots 1$.

Raup–Crick dissimilarity (method = “raup”) is a probabilistic index based on presence/absence data. It is defined as $1 - \text{prob}(j)$, or based on the probability of observing at least $j$ species in shared in compared communities. The current function uses analytic result from hypergeometric distribution (phyper) to find the probabilities. This probability (and the index) is dependent on the number of species missing in both sites, and adding all-zero species to the data or removing missing species from the data will influence the index. The probability (and the index) may be almost zero or almost one for a wide range of parameter values. The index is nonmetric: two communities with no shared species may have a dissimilarity slightly below one, and two identical communities may have dissimilarity slightly above zero. The index uses equal occurrence probabilities for all species, but Raup and Crick originally suggested that sampling probabilities should be proportional to species frequencies (Chase et al. 2011). A simulation approach with unequal species sampling probabilities is implemented in raupcrick function following Chase et al. (2011). The index can be also used for transposed data to give a probabilistic dissimilarity index of species co-occurrence (identical to Veech 2013).

Chao index tries to take into account the number of unseen species pairs, similarly as in method = “chao” in specpool. Function vegdist implements a Jaccard, index defined as $1 - \frac{C_j}{N_j} + \frac{(N_k - 1)/N_k \times a_1/(2a_2) \times S_j/N_j}$, and $V$ is similar except for site index $k$. $C_j$ is the total number of individuals in the species of site $j$ that are shared with site $k$, $N_j$ is the total number of individuals at site $j$, $a_1$ (and $a_2$) are the number of species occurring in site $j$ that have only one (or two) individuals in site $k$, and $S_j$ is the total number of individuals in the species present at site $j$ that occur with only one individual in site $k$ (Chao et al. 2005).

Morisita index can be used with genuine count data (integers) only. Its Horn–Morisita variant is able to handle any abundance data.

Mahalanobis distances are Euclidean distances of a matrix where columns are centred, have unit variance, and are uncorrelated. The index is not commonly used for community data, but it is sometimes used for environmental variables. The calculation is based on transforming data matrix and then using Euclidean distances following Mardia et al. (1979). The Mahalanobis transformation usually fails when the number of columns is larger than the number of rows (sampling units). When the transformation fails, the distances are nearly constant except for small numeric noise. Users must check that the returned Mahalanobis distances are meaningful.

Euclidean and Manhattan dissimilarities are not good in gradient separation without proper standardization but are still included for comparison and special needs.

Bray–Curtis and Jaccard indices are rank-order similar, and some other indices become identical or rank-order similar after some standardizations, especially with presence/absence transformation of
equalizing site totals with \texttt{decostand}. Jaccard index is metric, and probably should be preferred instead of the default Bray-Curtis which is semimetric.

The naming conventions vary. The one adopted here is traditional rather than truthful to priority. The function finds either quantitative or binary variants of the indices under the same name, which correctly may refer only to one of these alternatives. For instance, the Bray index is known also as Steinhaus, Czekanowski and Sørensen index. The quantitative version of Jaccard should probably called Ružička index. The abbreviation "horn" for the Horn–Morisita index is misleading, since there is a separate Horn index. The abbreviation will be changed if that index is implemented in \texttt{vegan}.

**Value**

Should provide a drop-in replacement for \texttt{dist} and return a distance object of the same type.

**Note**

The function is an alternative to \texttt{dist} adding some ecologically meaningful indices. Both methods should produce similar types of objects which can be interchanged in any method accepting either. Manhattan and Euclidean dissimilarities should be identical in both methods. Canberra index is divided by the number of variables in \texttt{vegdist}, but not in \texttt{dist}. So these differ by a constant multiplier, and the alternative in \texttt{vegdist} is in range \((0,1)\). Function \texttt{daisy} (package \texttt{cluster}) provides alternative implementation of Gower index that also can handle mixed data of numeric and class variables. There are two versions of Gower distance ("\texttt{gower}", "\texttt{altGower}") which differ in scaling: "\texttt{gower}" divides all distances by the number of observations (rows) and scales each column to unit range, but "\texttt{altGower}" omits double-zeros and divides by the number of pairs with at least one above-zero value, and does not scale columns (Anderson et al. 2006). You can use \texttt{decostand} to add range standardization to "\texttt{altGower}" (see Examples). Gower (1971) suggested omitting double zeros for presences, but it is often taken as the general feature of the Gower distances. See Examples for implementing the Anderson et al. (2006) variant of the Gower index.

Most dissimilarity indices in \texttt{vegdist} are designed for community data, and they will give misleading values if there are negative data entries. The results may also be misleading or \texttt{NA} or \texttt{NaN} if there are empty sites. In principle, you cannot study species composition without species and you should remove empty sites from community data.

**Author(s)**

Jari Oksanen, with contributions from Tyler Smith (Gower index) and Michael Bedward (Raup–Crick index).

**References**


See Also

Function `designdist` can be used for defining your own dissimilarity index. Alternative dissimilarity functions include `dist` in base R, `daisy` (package `cluster`), and `dsvdis` (package `labdsv`). Function `betadiver` provides indices intended for the analysis of beta diversity.

Examples

data(varespec)
vare.dist <- vegdist(varespec)
# Orlóci's Chord distance: range 0 .. sqrt(2)
vare.dist <- vegdist(decostand(varespec, "norm"), "euclidean")
# Anderson et al. (2006) version of Gower
vare.dist <- vegdist(decostand(varespec, "log"), "altGower")
# Range standardization with "altGower" (that excludes double-zeros)
vare.dist <- vegdist(decostand(varespec, "range"), "altGower")
**Description**

Functions *vegemite* and *tabasco* display compact community tables. Function *vegemite* prints text tables where species are rows, and each site takes only one column without spaces. Function *tabasco* provides interface for *heatmap* for a colour image of the data. The community table can be ordered by explicit indexing, by environmental variables or results from an ordination or cluster analysis.

**Usage**

```r
vegemite(x, use, scale, sp.ind, site.ind, zero=".", select,...)
tabasco(x, use, sp.ind = NULL, site.ind = NULL, select,
  Rowv = TRUE, Colv = TRUE, labRow = NULL, labCol = NULL,
  scale, col = heat.colors(12), ...)
coverscale(x, scale=c("Braun.Blanquet", "Domin", "Hult", "Hill", "fix","log"),
  maxabund, character = TRUE)
```

**Arguments**

- **x**  
  Community data.

- **use**  
  Either a vector, or an object from *cca*, *decorana* etc. or *hclust* or a *dendrogram* for ordering sites and species.

- **sp.ind, site.ind**  
  Species and site indices. In *tabasco*, these can also be *hclust* tree, *agnes* clusterings or *dendrograms*.

- **zero**  
  Character used for zeros.

- **select**  
  Select a subset of sites. This can be a logical vector (TRUE for selected sites), or a vector of indices of selected sites. The order of indices does not influence results, but you must specify use or site.ind to reorder sites.

- **Rowv, Colv**  
  Re-order dendrograms for the rows (sites) or columns (species) of *x*. If the Rowv = TRUE, row dendrograms are ordered by the first axis of correspondence analysis, and when Colv = TRUE column dendrograms by the weighted average (*wascores*) of the row order. Alternatively, the arguments can be vectors that are used to reorder the dendrogram.

- **labRow, labCol**  
  Character vectors with row and column labels used in the *heatmap* instead of the default. NB., the input matrix is transposed so that row labels will be used for data columns.

- **scale**  
  In *vegemite* and *coverscale*: cover scale used (can be abbreviated). In *tabasco*: scaling of colours in *heatmap*. The alternatives of *coverscale* can be used in *tabasco*, and in addition "column" or "row" scale columns or rows to equal maxima (NB., these refer to the transposed data of the *heatmap*), while "none" uses original values.
col  A vector of colours used for above-zero abundance values.
maxabund Maximum abundance used with scale = "log". Data maximum in the selected subset will be used if this is missing.
character Return character codes suitable for vegemite. If FALSE, returns corresponding integers.
... Arguments passed to coverscale (i.e., maxabund) in vegemite and to heatmap in tabasco.

Details

The function vegemite prints a traditional community table. The display is transposed, so that species are in rows and sites in columns. The table is printed in compact form: only one character can be used for abundance, and there are no spaces between columns. Species with no occurrences are dropped from the table.

Function tabasco produces a similar table as vegemite using heatmap, where abundances are coded by colours. The function scales the abundances to equal intervals for colour palette, but either rows or columns can be scaled to equal maxima, or the coverscale class systems can be used. The function can also display dendrograms for sites (columns) or species if these are given as an argument (use for sites, sp.ind for species).

The parameter use will be used to re-order output. The use can be a vector or an object from hclust or agnes, a dendrogram or any ordination result recognized by scores (all ordination methods in vegan and some of those not in vegan). The hclust, agnes and dendrogram must be for sites. The dendrogram is displayed above the sites in tabasco, but is not shown in vegemite. No dendrogram for species is displayed, except when given in sp.ind.

If use is a vector, it is used for ordering sites. If use is an object from ordination, both sites and species are arranged by the first axis (provided that results are available both also for species). When use is an object from hclust, agnes or a dendrogram, the sites are ordered similarly as in the cluster dendrogram. Function tabasco re-orders the dendrogram if Rowv = TRUE or Rowv is a vector. Such re-ordering is not available for vegemite, but it can be done by hand using reorder.dendrogram or reorder.hclust. Please note that dendrogram and hclust reordering can differ: unweighted means of merged branches are used in dendrogram, but weighted means (= means of leaves of the cluster) are used in reorder.hclust. In all cases where species scores are missing, species are ordered by their weighted averages (wascores) on site order.

Species and sites can be ordered explicitly giving their indices or names in parameters sp.ind and site.ind. If these are given, they take precedence over use. A subset of sites can be displayed using argument select, but this cannot be used to order sites, but you still must give use or site.ind. However, tabasco makes two exceptions: site.ind and select cannot be used when use is a dendrogram (clustering result). In addition, the sp.ind can be an hclust tree, agnes clustering or a dendrogram, and in that case the dendrogram is plotted on the left side of the heatmap. Phylogenetic trees cannot be directly used, as hclust.phylo (package ape) can transform these to hclust trees.

If scale is given, vegemite calls coverscale to transform percent cover scale or some other scales into traditional class scales used in vegetation science (coverscale can be called directly, too). Function tabasco can also use these traditional class scales, but it treats the transformed values as corresponding integers. Braun-Blanquet and Domin scales are actually not strict cover scales, and the limits used for codes r and + are arbitrary. Scale hill may be inappropriately named,
since Mark O. Hill probably never intended this as a cover scale. However, it is used as default “cut levels” in his TWINSPLAN, and surprisingly many users stick to this default, and this is a de facto standard in publications. All traditional scales assume that values are cover percentages with maximum 100. However, non-traditional alternative log can be used with any scale range. Its class limits are integer powers of 1/2 of the maximum (argument maxabund), with + used for non-zero entries less than 1/512 of the maximum (log stands alternatively for logarithmic or logical). Scale fix is intended for “fixing” 10-point scales: it truncates scale values to integers, and replaces 10 with X and positive values below 1 with +.

Value

The functions are used mainly to display a table, but they return (invisibly) a list with items:

species  Ordered species indices
sites    Ordered site indices

These items can be used as arguments sp.ind and site.ind to reproduce the table. In addition to the proper table, vegemite prints the numbers of species and sites and the name of the used cover scale at the end.

Note

The name vegemite was chosen because the output is so compact, and the tabasco because it is just as compact, but uses heat colours.

Author(s)

Jari Oksanen

References

The cover scales are presented in many textbooks of vegetation science; I used:

See Also

cut and approx for making your own ‘cover scales’ for vegemite. Function tabasco is based on heatmap which in turn is based on image. Both functions order species with weighted averages using wascores.

Examples

data(varespec)
## Print only more common species
freq <- apply(varespec > 0, 2, sum)
vegemite(varespec, scale="Hult", sp.ind = freq > 10)
## Order by correspondence analysis, use Hill scaling and layout:
dca <- decorana(varespec)
vegemite(varespec, dca, "Hill", zero="-")
## Show one class from cluster analysis, but retain the ordering above
clus <- hclust(vegdist(varespec))
wascores

Weighted Averages Scores for Species

Description

Computes Weighted Averages scores of species for ordination configuration or for environmental variables.

Usage

```r
wascores(x, w, expand=FALSE)
eigengrad(x, w)
```

Arguments

- `x`: Environmental variables or ordination scores.
- `w`: Weights: species abundances.
- `expand`: Expand weighted averages so that they have the same weighted variance as the corresponding environmental variables.
Details

Function `wascores` computes weighted averages. Weighted averages “shrink”: they cannot be more extreme than values used for calculating the averages. With `expand = TRUE`, the function “deshrink”s the weighted averages by making their biased weighted variance equal to the biased weighted variance of the corresponding environmental variable. Function `eigengrad` returns the inverses of squared expansion factors or the attribute shrinkage of the `wascores` result for each environmental gradient. This is equal to the constrained eigenvalue of `cca` when only this one gradient was used as a constraint, and describes the strength of the gradient.

Value

Function `wascores` returns a matrix where species define rows and ordination axes or environmental variables define columns. If `expand = TRUE`, attribute shrinkage has the inverses of squared expansion factors or `cca` eigenvalues for the variable. Function `eigengrad` returns only the shrinkage attribute.

Author(s)

Jari Oksanen

See Also

`monomDS`, `cca`.

Examples

data(varespec)
data(varechem)
vare.dist <- vegdist(wisconsin(varespec))
vare.mds <- monomDS(vare.dist)
vare.points <- postMDS(vare.mds$points, vare.dist)
vare.wa <- wascores(vare.points, varespec)
plot(scores(vare.points), pch="++", asp=1)
text(vare.wa, rownames(vare.wa), cex=0.8, col="blue")
## Omit rare species (frequency <= 4)
freq <- apply(varespec[, 2, sum)
plot(scores(vare.points), pch="++", asp=1)
text(vare.wa[freq > 4, ], rownames(vare.wa)[freq > 4], cex=0.8, col="blue")
## Works for environmental variables, too.
wascores(varechem, varespec)
## And the strengths of these variables are:
eigengrad(varechem, varespec)
Description

Weighted classical multidimensional scaling, also known as weighted principal coordinates analysis.

Usage

```
wcmdscale(d, k, eig = FALSE, add = FALSE, x.ret = FALSE, w)
```

## S3 method for class 'wcmdscale'

```
plot(x, choices = c(1, 2), type = "t", ...)
```

## S3 method for class 'wcmdscale'

```
scores(x, choices = NA, ...)
```

Arguments

- **d**: a distance structure such as that returned by `dist` or a full symmetric matrix containing the dissimilarities.
- **k**: the dimension of the space which the data are to be represented in; must be in \{1, 2, ..., n - 1\}. If missing, all dimensions with above zero eigenvalue.
- **eig**: indicates whether eigenvalues should be returned.
- **add**: an additive constant \(c\) is added to the non-diagonal dissimilarities such that all \(n - 1\) eigenvalues are non-negative. Alternatives are "lingoes" (default, also used with `TRUE`) and "cailliez" (which is the only alternative in `cmdscale`). See Legendre & Anderson (1999).
- **x.ret**: indicates whether the doubly centred symmetric distance matrix should be returned.
- **w**: Weights of points.
- **x**: The `wcmdscale` result object when the function was called with options `eig = TRUE` or `x.ret = TRUE` (See Details).
- **choices**: Axes to be returned; `NA` returns all real axes.
- **type**: Type of graph which may be "t"ext, "p"oints or "n"one.
- **...**: Other arguments passed to graphical functions.

Details

Function `wcmdscale` is based on function `cmdscale` (package `stats` of base \(R\)), but it uses point weights. Points with high weights will have a stronger influence on the result than those with low weights. Setting equal weights \(w = 1\) will give ordinary multidimensional scaling.

With default options, the function returns only a matrix of scores scaled by eigenvalues for all real axes. If the function is called with `eig = TRUE` or `x.ret = TRUE`, the function returns an
object of class "wcmdscale" with print, plot, scores, eigenvals and stressplot methods, and described in section Value.

The method is Euclidean, and with non-Euclidean dissimilarities some eigenvalues can be negative. If this disturbs you, this can be avoided by adding a constant to non-diagonal dissimilarities making all eigenvalues non-negative. The function implements methods discussed by Legendre & Anderson (1999): The method of Lingoes (add="lingoes") adds the constant $c$ to squared dissimilarities $d$ using $\sqrt{d^2 + 2c}$ and the method of Cailliez (add="cailliez") to dissimilarities using $d + c$. Legendre & Anderson (1999) recommend the method of Lingoes, and base R function cmdscale implements the method of Cailliez.

**Value**

If `eig = FALSE` and `x.ret = FALSE` (default), a matrix with k columns whose rows give the coordinates of points corresponding to positive eigenvalues. Otherwise, an object of class wcmdscale containing the components that are mostly similar as in cmdscale:

- **points** a matrix with k columns whose rows give the coordinates of the points chosen to represent the dissimilarities.
- **eig** the $n - 1$ eigenvalues computed during the scaling process if `eig` is true.
- **x** the doubly centred and weighted distance matrix if `x.ret` is true.
- **ac, add** additive constant and adjustment method used to avoid negative eigenvalues. These are NA and FALSE if no adjustment was done.
- **GOF** Goodness of fit statistics for k axes. The first value is based on the sum of absolute values of all eigenvalues, and the second value is based on the sum of positive eigenvalues.
- **weights** Weights.
- **negaxes** A matrix of scores for axes with negative eigenvalues scaled by the absolute eigenvalues similarly as **points**. This is NULL if there are no negative eigenvalues or k was specified, and would not include negative eigenvalues.
- **call** Function call.

**References**


**See Also**

The function is modelled after cmdscale, but adds weights (hence name) and handles negative eigenvalues differently. eigenvals.wcmdscale and stressplot.wcmdscale are some specific methods. Other multidimensional scaling methods are monoMDS, and isoMDS and sammon in package MASS.
Examples

```r
## Correspondence analysis as a weighted principal coordinates
## analysis of Euclidean distances of Chi-square transformed data
data(dune)
rs <- rowSums(dune)/sum(dune)
d <- dist(decosstand(dune, "chi"))
ord <- wcmdscale(d, w = rs, eig = TRUE)
## Ordinary CA
ca <- cca(dune)
## Eigenvalues are numerically similar
cia$CA$eig - ord$eig
## Configurations are similar when site scores are scaled by
## eigenvalues in CA
procrustes(ord, ca, choices=1:19, scaling = "sites")
plot(procrustes(ord, ca, choices=1:2, scaling="sites"))
## Reconstruction of non-Euclidean distances with negative eigenvalues
d <- vegdist(dune)
ord <- wcmdscale(d, eig = TRUE)
## Only positive eigenvalues:
cor(d, dist(ord$points))
## Correction with negative eigenvalues:
cor(d, sqrt(dist(ord$points)^2 - dist(ord$negaxes)^2))
```
Index

*Topic **IO**
  read.cep, 223
*Topic **aplot**
  envfit, 87
  linestack, 112
  ordiarrows, 155
  ordihull, 159
  ordilabel, 163
  ordiplot, 164
  ordipointlabel, 166
  ordisurf, 172
  orditorp, 180
  plot. cca, 199
*Topic **character**
  make. cepnames, 114
*Topic **cluster**
  cascadeKM, 42
*Topic **datagen**
  commsim, 57
  nullmodel, 147
  oecosimu, 151
  permat, 186
  simulate. rda, 236
*Topic **datasets**
  BCI, 21
  dune, 84
  dune. taxon, 85
  mite, 127
  pyrifos, 212
  sipoo, 238
  varespec, 264
*Topic **distribution**
  fisherfit, 93
  permutats, 191
  radfit, 213
*Topic **documentation**
  vegandocs, 271
*Topic **dynamic**
  orditkplot, 177

*Topic **file**
  read.cep, 223
*Topic **hplot**
  betadisper, 25
  biplot.rda, 37
  linestack, 112
  ordiplot, 164
  ordipointlabel, 166
  ordiresids, 168
  orditorp, 180
  ordixyplot, 182
  plot. cca, 199
  vegemite, 277
*Topic **htest**
  anosim, 14
  anova. cca, 16
  clamtest, 55
  envfit, 87
  mantel, 115
  mrpp, 134
  procrustes, 209
  vegan-package, 4
*Topic **iplot**
  ordiplot, 164
  orditkplot, 177
*Topic **manip**
  beals, 22
  decostand, 69
  dispweight, 78
  vegemite, 277
*Topic **methods**
  betadisper, 25
  permutest. betadisper, 197
*Topic **misc**
  vegan-deprecated, 270
*Topic **models**
  add1. cca, 5
  cca.object, 49
  deviance. cca, 75
humpfit, 100
influence.cca, 105
MOStest, 131
nobs.adonis, 146
ordistep, 169
simulate.rda, 236
specaccum, 241
SSarrhenius, 250
vegan-package, 4

+Topic multivariate
  add1.cca, 5
  adipart, 7
  adonis, 11
  anosim, 14
 anova.cca, 16
  avgdist, 19
  betadisper, 25
  betadiver, 30
  bgdispersal, 33
  bioenv, 34
  capscale, 39
  cca, 45
  cca.object, 49
  CCorA, 52
  commsim, 57
  contribdiv, 64
  decorana, 66
  decostand, 69
  designdist, 72
  deviance.cca, 75
  dispindmorisita, 76
  dispweight, 78
  distconnected, 80
  eigenvals, 86
  envfit, 87
  eventstar, 91
  goodness.cca, 96
  goodness.metaMDS, 99
  indpower, 103
  influence.cca, 105
  isomap, 107
  kendall.global, 110
  mantel, 115
  mantel.correlog, 117
  MDSrotate, 120
  metaMDS, 121
  monoMDS, 128
  mrpp, 134
  mso, 138
  multipart, 141
  nullmodel, 147
  oecosimu, 151
  ordcosimu, 169
  ordisurf, 172
  pcnm, 184
  permat, 186
  permutations, 193
  permute.test.betadisper, 197
  pcr, 202
  predict.cca, 205
  procrustes, 209
  rankindex, 217
  raupcrick, 221
  reorder.hclust, 227
  RsquareAdj, 229
  scores, 230
  screeplot.cca, 231
  simper, 234
  simulate.rda, 236
  spantree, 238
  sppscores, 249
  stepacross, 253
  stressplot.wcmdscale, 255
  tsallis, 262
  varpart, 265
  vegan-package, 4
  vegdist, 272
  wascores, 280
  wcmdscale, 282

+Topic nonlinear
  humpfit, 100

+Topic nonparametric
  adonis, 11
  anosim, 14
  bgdispersal, 33
  kendall.global, 110
  mrpp, 134
  oecosimu, 151
  vegan-package, 4

+Topic optimize
  eventstar, 91

+Topic package
  vegan-package, 4

+Topic print
  vegemite, 277

+Topic regression
humpfit, 100
MOSTest, 131
vegan-package, 4

*Topic smooth
  beals, 22
  permustats, 191

*Topic spatial
  dispindmorisita, 76
  mso, 138
  pcnm, 184
  vegan-package, 4

*Topic univar
  diversity, 82
  fisherfit, 93
  nestedtemp, 143
  radfit, 213
  rarefy, 219
  renyi, 225
  RsquareAdj, 229
  specaccum, 241
  specpool, 246
  taxondiv, 256
  treedive, 260
  vegan-package, 4
  wascores, 280

*Topic utilities
  eventstar, 91
  ordiArrowTextXY, 157
  vegandocs, 271
  Random.seed, 148

abbreviate, 114, 115
add1, 6, 7
add1.cca, 5, 19, 49, 52, 76, 170, 171
add1.default, 6
adipart, 7, 64, 65, 83, 142, 190
adonis, 11, 147, 192
adonis2 (adonis), 11
agnes, 160, 161, 239–241, 277, 278
AIC, 75, 102, 215, 216
AIC.fitspecaccum (specaccum), 241
AIC.radfit (radfit), 213
alias.cca, 50, 52
alias.cca (goodness.cca), 96
alias.lm, 98
amova, 12
anosim, 13, 14, 117, 137, 192, 218
anova, 18
anova.betadisper (betadisper), 25
anova.cca, 6, 7, 11, 12, 16, 19, 42, 47, 49, 52,
  76, 168, 170, 192, 204, 269
anova.glm, 132
anova.lm, 29
approx, 243, 279
ar, 190
arima, 190
arrows, 38, 156, 161, 202, 210, 211
as.dist, 267
as.factor, 26
as.fisher (fisherfit), 93
as.hclust, 261
as.hclust.phylo, 261, 278
as.hclust.spantree, 228
as.hclust.spantree (spantree), 238
as.mcmc.oecosimu (oecosimu), 151
as.mcmc.permat (permat), 186
as.mlm, 106
as.mlm (vegan-deprecated), 270
as.mlm.cca, 49, 52, 107
as.preston (fisherfit), 93
as.rad (radfit), 213
as.ts.oecosimu (oecosimu), 151
as.ts.permat (permat), 186
avgdist, 19
axis, 200

BCI, 21
beals, 22, 104, 248
betadisper, 13, 25, 31, 32, 147, 197, 198
betadiver, 26, 27, 29, 30, 73, 74, 83, 145,
  267, 276
bgdispersal, 33
bioenv, 34, 117
bioenvdist (bioenv), 34
biplot, 53
biplot.cca, 52
biplot.cca (biplot.rda), 37
biplot.CCorA (CCorA), 52
biplot.default, 53
biplot.rda, 37, 38, 48
Box.test, 190
boxplot, 15, 28, 29, 245
boxplot.betadisper (betadisper), 25
boxplot.default, 192
boxplot.permutest (permustats), 191
boxplot.specaccum (specaccum), 241
browseVignettes, 144, 271
kendall.post, 194
kendall.post (kendall.global), 110
kmeans, 43, 45, 112
labels.envfit (envfit), 87
lag.plot, 190
Lattice, 169, 183, 215, 216
lda, 120, 201
legend, 56, 139, 203
lines, 156, 159, 161, 162, 219
lines.specaccum (specaccum), 241
lines.humpfit (humpfit), 100
lines.permat (permat), 186
lines.preston (fisherfit), 93
lines.prestonfit (fisherfit), 93
lines.procrustes (procrustes), 209
lines.radfit (radfit), 213
lines.radline (radfit), 213
lines.spantree (spantree), 238
lines.specaccum (specaccum), 241
lines.stack, 112, 204
lm, 49, 105–107, 185, 229, 237, 271
lm.influence, 49
locator, 268
logLik, 216
logLik, radfit (radfit), 213
logLik.specaccum (specaccum), 241
lset, 216

Machine, 218
make.cepnames, 84, 114
make.commsim, 8, 141, 149, 151, 152, 187–190
make.commsim (commsim), 57
make.names, 114, 115, 224
make.unique, 114
mantel, 13, 16, 31, 32, 36, 115, 118, 137, 139, 192, 212, 218
mantel.correlog, 117
mantel.partial, 36, 185, 192
matlines, 245
matplot, 204
mcmc, 153
mcnenmar.test, 34
MDSrotate, 120, 124, 126, 130
meandist (mrpp), 134
metaMDS, 32, 37, 40, 41, 47, 69, 99, 100, 109, 120, 121, 130, 131, 147, 249, 250, 255
metaMDSdist, 40
metaMDSdist (metaMDS), 121
metaMDSiter (metaMDS), 121
metaMDSredist, 100
metaMDSredist (metaMDS), 121
mite, 127
model.frame.cca, 52
model.frame.cca (cca.object), 49
model.matrix, 13, 46
model.matrix.cca, 52
model.matrix.cca (cca.object), 49
monoMDS, 15, 37, 69, 99, 100, 120–126, 128, 212, 218, 255, 281, 283
MOStest, 131
mrpp, 13, 16, 117, 134, 192
mso, 52, 138
msoplot (mso), 138
multipart, 10, 83, 141
mvnorm, 237
na.action, 51
na.exclude, 40, 46
na.fail, 40, 46
na.omit, 40, 46
nestedbetajac, 32
nestedbetajac (nestedtemp), 143
nestedbetasor, 32
nestedbetasor (nestedtemp), 143
nestedchecker, 152
nestedchecker (nestedtemp), 143
nesteddisc, 152
nesteddisc (nestedtemp), 143
nesteddisc (nestedtemp), 143
nesteddisc (nestedtemp), 143
nesteddisc, 152
nesteddisc (nestedtemp), 143
nesteddisc (nestedtemp), 143
nesteddisc, 152
nesteddisc (nestedtemp), 143
nesteddisc (nestedtemp), 143
news, 271
nlm, 95, 102, 215, 219
nls, 242–244, 250, 252
no.shared, 122, 217, 218, 254, 255
no.shared (distconnected), 80
nobs.adonis, 146
nobs.betadisper (nobs.adonis), 146
nobs.cca, 51, 52
nobs.cca (nobs.adonis), 146
nobs.CCorA (nobs.adonis), 146
nobs.decorana (nobs.adonis), 146
INDEX

nobs.fitspecaccum (specaccum), 241
nobs.isomap (nobs.adonis), 146
nobs.metaMDS (nobs.adonis), 146
nobs.pcmn (nobs.adonis), 146
nobs.procrustes (nobs.adonis), 146
nobs.rad (nobs.adonis), 146
nobs.varpart (nobs.adonis), 146
nobs.wcmdscale (nobs.adonis), 146
nullmodel, 57, 62, 79, 145, 147, 151–154, 190
numPerms, 196

object.size, 153
ocecosimu, 8–10, 62, 141, 142, 144–146, 151, 190, 192, 221, 222, 261
optim, 167, 168
optimize, 92
ordConstrained, 52
ordConstrained (cca.object), 49
ordered, 261
orderingKM (cascadeKM), 42
ordiareatest, 192
ordiareatest (ordihull), 159
ordiArrowMul, 90
ordiArrowMul (ordiArrowTextXY), 157
ordiarrows, 155, 161, 183
ordiArrowTextXY, 90, 157
ordiBar (ordihull), 159
ordicloud (ordixyplot), 182
ordicluster, 240
ordicluster (ordihull), 159
ordiellipse (ordihull), 159
ordigrid, 161
ordigrid (ordiarrows), 155
ordihull, 159
ordilabel, 88, 108, 156, 160, 163, 219
ordilattice.getEnvfit (ordixyplot), 182
ordimedian (betadisper), 25
ordiplot, 28, 31, 38, 108, 109, 126, 155, 159,
164, 180, 181, 201, 202, 215, 231
ordipointlabel, 166, 179, 180
ordiR2step, 6, 49, 52
ordiR2step (ordistep), 169
ordiresids, 52, 107, 168
ordirgl, 180
ordisegments, 161
ordisegments (ordiarrows), 155
ordispider, 97
ordispider (ordihull), 159
ordisplom (ordixyplot), 182
ordistep, 6, 7, 49, 52, 169
 ordisurf, 47, 90, 172
 orditkplot, 163, 167, 177
 orditorp, 163, 180, 180
 ordixyplot, 182, 231
 ordiYbar (cca.object), 49
p.adjust, 110, 111, 118
pairs.profile, 102
panel.arrows, 183
panel.cloud, 183
panel.ordi (ordixyplot), 182
panel.ordi3d (ordixyplot), 182
panel.ordiarrows (ordixyplot), 182
panel.splom, 183
panel.xyplot, 183
par, 163, 178, 179, 219, 242
paste, 115
pca, 86
pcaiv, 49
pchisq, 79
pcnm, 86, 147, 184
pco, 86
pd, 262
permat, 186
permatfull, 62, 149, 151, 152
permatfull (permat), 186
permatswap, 60–62, 149, 151, 152
permatswap (permat), 186
permutest, 13, 15, 116, 136, 153, 191, 211
permutations, 15, 89, 116, 193
permutest, 196, 229
permutest (anova.cca), 16
permutest.betadisper, 27, 29, 192, 197
permutest.cca, 51, 52, 192, 229
persp, 174, 225, 226, 264
persp.renyiaccum (renyi), 225
persp.tsallisaccum (tsallis), 262
phylosor, 262
phyper, 274
plot, 113, 165, 219
plot.anosim (anosim), 14
plot.betadisper (betadisper), 25
plot.betadiver (betadiver), 30
plot.cascadeKM (cascadeKM), 42
plot.cca, 38, 42, 48, 52, 88, 155, 159, 164,
166, 180, 181, 199
plot.clamtest (clamtest), 55
plot.contribdiv (contribdiv), 64
plot.decorana, 155, 159, 166, 181
plot.decorana(decorana), 66
plot.default, 26, 181, 233
plot.envfit(envfit), 87
plot.fisher(fisherfit), 93
plot.fisherfit(fisherfit), 93
plot.fitspecaccum(specaccum), 241
plot.gam, 174, 176
plot.hclust, 228
plot.humpfit(humpfit), 100
plot.isomap(isomap), 107
plot.lm, 132, 168, 169
plot.mantel.correlog(mantel.correlog), 117
plot.meandist(mrpp), 134
plot.metaMDS, 181
plot.metaMDS(metaMDS), 121
plot.monomMDS(monomMDS), 128
plot.MOSTest(MOSTest), 131
plot.nestedndof(nestedndof), 143
plot.nestedtemp(nestedtemp), 143
plot.ordiplot(ordiplotlabel), 166
plot.ordisurf(ordisurf), 172
plot.orditkplot(orditkplot), 177
plot.permat(permat), 186
plot.poolaccum(specpool), 246
plot.prc(prc), 202
plot.preston(fisherfit), 93
plot.prestonfit(fisherfit), 93
plot.procrustes, 165, 166
plot.procrustes(procrustes), 209
plot.profile, 102
plot.rad(radfit), 213
plot.radfit(radfit), 213
plot.radline(radfit), 213
plot.renyi(renyi), 225
plot.renyiaccum(renyi), 225
plot.spantree(spantree), 238
plot.specaccum(specaccum), 241
plot.taxondive(taxondive), 256
plot.varpart(varpart), 265
plot.varpart234(varpart), 265
plot.wcmdscale(wcmdscale), 282
plotmath, 113
pointLabel, 167, 168
points, 38, 156, 167, 180, 181, 201, 202
points.cca, 52
points.cca(plot.cca), 199
points.decorana(decorana), 66
points.humpfit(humpfit), 100
points.metaMDS(metaMDS), 121
points.monomMDS(monomMDS), 128
points.ordiplot(ordiplot), 164
points.orditkplot(orditkplot), 177
points.procrustes(procrustes), 209
points.radfit(radfit), 213
points.radline(radfit), 213
polygon, 159–164, 245
poolaccum, 245
poolaccum(specpool), 246
postMDS(metaMDS), 121
prc, 49, 202
prcomp, 86, 87, 231, 233, 234, 255
predict, 216
predict.cca, 49–52, 205, 236
predict.decorana, 69
predict.decorana(predict.cca), 205
predict.fitspecaccum(specaccum), 241
predict.gam, 175
predict.humpfit(humpfit), 100
predict.nls, 244
predict.procrustes(procrustes), 209
predict.radfit(radfit), 213
predict.radline(radfit), 213
predict.rda, 51, 236
predict.rda(predict.cca), 205
predict.specaccum(specaccum), 241
pregraphKM(cascadeKM), 42
prepanel.ordi3d(ordixyplot), 182
prestondist(fisherfit), 93
prestonfit, 216
prestonfit(fisherfit), 93
princomp, 86, 87, 231, 233, 234, 255
print.cca, 52
print.cca(cca.object), 49
print.commsim(commsim), 57
print.nullmodel(nullmodel), 147
print.permat(permat), 186
print.simmat(nullmodel), 147
print.specaccum(specaccum), 241
print.summary.cca(plot.cca), 199
print.summary.decorana(decorana), 66
print.summary.permat(permat), 186
procrustes, 37, 124, 126, 147, 209, 231
profile.glm, 103, 133
<table>
<thead>
<tr>
<th>Function</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>profile.humpfit (humpfit), qnorm, qr, r2dtable, rda, read.cex</td>
<td>293</td>
</tr>
<tr>
<td>profile.MOSTest (MOSTest)</td>
<td>131</td>
</tr>
<tr>
<td>protest, pyrifos, qnorm, qr, qqmath, rankindex, rank, rarefy, rarecurve, rda, read.fortran, read.mat, relevel, renyi, renyiaccum, renyiaccum (renyi), reorder, reorder.dendrogram, reorder.hclust, residuals, residuals.cca (predict.cca), residuals.glm, residuals.procrustes (procrustes), rev.hclust (reorder.hclust), rgl.isomap, rgl.renyiaccum, rrarefy, rstudent, scores, scale, scores.cca (plot.cca), scores.decorana, scores.decorana (decorana), scores.envfit, scores.envfit (envfit), scores.hclust (reorder.hclust), scores.metaMDS, scores.metaMDS (metaMDS), scores.monosY, scores.monosY (monosY), scores.ordihull (ordihull), scores.ordiplot (ordiplot), scores.orditkplot (orditkplot), scores.pcmn, scores.pcxn (pcmn), scores.rda, rug, s</td>
<td>205</td>
</tr>
</tbody>
</table>

INDEX
summary.glm, 132
summary.humpfit (humpfit), 100
summary.isomap (isomap), 107
summary.meanard (mrpp), 134
summary.oridiellipse (oridiellipse), 159
summary.oridiellipse (oridiellipse), 159
summary.permat (permats), 186
summary.permustats (permustats), 191
summary.poolaccum (specpool), 246
summary.prc (prc), 202
summary.procrustes (procrustes), 209
summary.radfit.frame (radfit), 213
summary.simper (simper), 234
summary.specaccum (specaccum), 241
summary.taxondive (taxondive), 256
svd, 47, 86, 87
swan, 255
swan (beals), 22
swap.web, 60, 188, 190
symbols, 90

tabasco (vegemite), 277
tail.summary.cca (plot.cca), 199
taxa2dist (taxondive), 256
taxondive, 85, 256, 262
tc, 176
terms, 13, 51
text, 38, 156, 158, 160, 163, 164, 167, 180, 181, 201, 202
text.cca, 52
text.cca (plot.cca), 199
text.decorana (decorana), 66
text.metaMDS (metaMDS), 121
text.monodMDS (monodMDS), 128
text.ordiplot (ordiplot), 164
text.orditkplot (orditkplot), 177
text.procrustes (procrustes), 209
tkcanvas, 179
tolerance, 259
tolerance.cca, 49, 52
treedist, 85
treedist (treedive), 260
treedive, 85, 152, 154, 260
treeheight (treedive), 260
trellis.par.set, 183
ts, 153
tsalis, 83, 92, 93, 142, 262
tsalisaccum, 245
tsalisaccum (tsallis), 262
tsdiaq, 190
TukeyHSD, 26–29
TukeyHSD.betadisper, 197, 198
TukeyHSD.betadisper (betadisper), 25
update, 147
update.nullmodel (nullmodel), 147
varechem (varespec), 264
varespec, 264
varpart, 13, 49, 147, 230, 265
varpart2 (varpart), 265
varpart3 (varpart), 265
varpart4 (varpart), 265
vcov, 105, 106
vcov.cca, 52, 271
vcov.cca (influence.cca), 105
vectorfit (envfit), 87
vegan (vegan-package), 4
vegan-deprecated, 270
vegan-package, 4
vegandocs, 271
vegdist, 11, 15, 16, 19, 20, 26, 31, 32, 35, 37,
39, 40, 42, 70, 72–74, 81, 117, 122,
124–126, 135–137, 217, 218, 222,
235, 239, 241, 253, 254, 261, 267, 272
vegemite, 277
veiledspec, 248
veiledspec (fisherfit), 93
vif, 98, 208
vif.cca, 49, 51
vif.cca (goodness.cca), 96
vignette, 49, 144, 200, 247, 271
wascores, 122, 125, 126, 250, 277–279, 280
wcmdscale, 11, 26, 40, 42, 86, 87, 130, 147,
185, 255, 266, 282
weighted.mean, 228
weights.cca, 52
weights.cca (cca.object), 49
weights.decorana (cca.object), 49
weights.rda (cca.object), 49
wisconsin, 124, 126
wisconsin (decostand), 69
xy.coords, 233
xyplot, 169, 182, 183, 192, 214, 216, 226,
246, 248