Package ‘betaper’

June 18, 2025

Type Package

Title Taxonomic Uncertainty on Multivariate Analyses of Ecological Data

Version 1.1-3

Date 2025-06-17

Author Luis Cayuela [aut, cre],
Marcelino de la Cruz [aut]

Maintainer Luis Cayuela <`luis.cayuela@urjc.es`>

Depends vegan, parallel

Description Permutational method to incorporate taxonomic uncertainty and some functions to assess its effects on parameters of some widely used multivariate methods in ecology, as explained in Cayuela et al. (2011) <doi:10.1111/j.1600-0587.2009.05899.x>.

License GPL (>= 2)

LazyLoad yes

NeedsCompilation no

Repository CRAN

Date/Publication 2025-06-17 23:00:06 UTC

Encoding UTF-8

Contents

`adonis_pertables` *Function to assess the efects of taxonomic uncertainty on permutational multivariate analysis of variance using distance matrices*

Description

This function asses the effects of taxonomic uncertainty on the R2 coefficients and the p-values of a permutational multivariate analysis of variance using distance matrices.

Usage

```
adonis_pertables(formula = X ~ ., data, permutations = 5, method = "bray", by=NULL)
## S3 method for class 'adonis_pertables'
plot(x, ...)
```

Arguments

`formula`	A typical model formula such as 'Y ~ A + B*C', but where 'Y' is a pertables object (i.e. a list of simulated community data matrices obtained with `pertables`; 'A', 'B', and 'C' may be factors or continuous variables.
`data`	The data frame from which 'A', 'B', and 'C' would be drawn.
`permutations`	Number of replicate permutations used for the hypothesis tests (F tests) for each simulated community data matrices obtained with `pertables`.
`method`	The name of any method used in 'vegdist' to calculate pairwise distances.
`by`	by = NULL will assess the overall significance of all terms together, 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), by = "onedf" will analyse one-degree-of-freedom contrasts sequentially.
`x`	`adonis_pertables` object to plot.
`...`	Additional graphical parameters passed to plot.

Value

`adonis_pertables` returns an object of class`adonis_pertables`, basically a list with the following components:

`raw`	An object of class `adonis`, i.e. the results of applying `mantel` to the original biological data table without the unidentified species. This includes p-values for each explanatory variable showing the probability of obtaining the same F statistic under different scenarios of taxonomic uncertainty.
`simulation`	A list with the results of the simulation: `F`, i.e. a data.frame with all the simulated pseudo-F (columns) for each explanatory variable (rows); `R2`, i.e. a data.frame with all the simulated R2 coefficients (columns) for each explanatory variable (rows); `pvalue`, i.e. a data.frame with all the simulated p-values (columns) for

each explanatory variable (rows); `R2.quant`, i.e. a data.frame with the summary of `R2` by quantiles; `p.quant`,, i.e. a data.frame with the summary of `pvalue` by quantiles.

The objects of class `adonis_pertables` have `print` and `plot` S3 methods for a simple access to results. See the examples.

Author(s)

Luis Cayuela and Marcelino de la Cruz

References

Cayuela, L., De la Cruz, M. and Ruokolainen, K. (2011). A method to incorporate the effect of taxonomic uncertainty on multivariate analyses of ecological data. *Ecography*, 34: 94-102. doi:10.1111/j.16000587.2009.05899.x.

See Also

`pertables`, `adonis2`

Examples

```
data(Amazonia)
data(soils)

# Define a new index that includes the terms used in the \code{Amazonia} dataset to define
# undetermined taxa at different taxonomic levels

index.Amazon <- c(paste("sp.", rep(1:20), sep=""), "Indet.", "indet.")

# Generate a pertables object (i.e. a list of biological data tables simulated from taxonomic
# uncertainty)
 ## Not run:
Amazonia100 <- pertables(Amazonia, index=index.Amazon, nsim=100)

# Assess the effects of taxonomic uncertainty on a PERMANOVA (i.e., adonis) test:

Amazonia.adonis <- adonis_pertables(Amazonia100 ~ Ca + K + Mg + Na, data=soils, by="terms")

Amazonia.adonis

plot(Amazonia.adonis)

## End(Not run)
# Fast example for Rcheck

Amazonia4.p2 <- pertables.p2(Amazonia[1:50,], index=index.Amazon, nsim=4, ncl=2, iseed=4)
set.seed(2)
Amazonia.adonis <- adonis_pertables(Amazonia4.p2  ~ Ca + K + Mg + Na, data=soils, by="terms")
```

```
Amazonia.adonis

plot(Amazonia.adonis)
```

`Amazonia` *Tree abundance and soil data in Western Amazonia*

Description

The `Amazonia` data frame has tree counts in nine 0.16-hectare inventory plots in Western Amazonia. `soils` contains data on soil cations at each location.

Usage

```
data(Amazonia)
data(soils)
```

Format

`Amazonia` is a data frame with 1188 observations (species) and 12 columns (taxonomic descripcion and sites). The three first columns refer to family, genus and specific species Latin names. Columns 4 to 12 have tree abundance data for nine inventory plots.

`soils` is a data frame with 9 observations (inventory plots) and 4 columns (variables). Soil variables (Ca, K, Mg, Na) are given in cmol/kg.

Details

Data from Western Amazonia includes tree inventories at nine lowland sites (approximately 100-150 m above sea level) near Iquitos, Peru. The sites were selected to represent regional variations in geology and were distributed along a soil nutrient gradient ranging from poor loamy soils to richer clayey soils. Each inventory consisted of 20 x 20 m plots (0.16 ha total area) distributed along 1.3-km transects. At each site, K. Ruokolainen and colleagues identified to species or morphospecies all woody, free-standing stems of > 2.5 cm dbh. The full inventories sampled 3980 individuals from 1188 species or morphospecies.

References

Higgins, M.A. & Ruokolainen, K. 2004. Rapid tropical forest inventory: a comparison of techniques based on inventory data from western Amazonia. *Conservation Biology* 18(3): 799-811.

Ruokolainen, K., Tuomisto, H., Macia, M.J., Higgins, M.A. & Yli-Halla, M. 2007. Are floristic and edaphic patterns in Amazonian rain forests congruent for trees, pteridophytes and Melastomataceae? *Journal of Tropical Ecology* 23: 13-25.

Examples

```
data(Amazonia)
data(soils)
```

`cca_pertables` *Function to assess the efects of taxonomic uncertainty on [Partial] Constrained Correspondence Analysis*

Description

This function asses the effects of taxonomic uncertainty on two widely used parameters of a [Partial] Constrained Correspondence Analysis, i.e. the 'percentage explained variance' (sometimes referred to as *R-squared*) and the *pseudo-F*.

Usage

```
cca_pertables(fml, data,  scale = FALSE,...)
## S3 method for class 'cca_pertables'
 plot(x, pch = 18, ...)
```

Arguments

`fml`	Model formula, where the left hand side gives a pertables object (i.e. a list of simulated community data matrices obtained with `pertables`, 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.
`scale`	Scale species to unit variance (like correlations).
`x`	`cca_pertables` object to plot.
`pch`	Plotting 'character', i.e., symbol to use in the CCA plot. See `points` for examples of use of this graphical argument.
`...`	Additional graphical parameters passed to plot.

Details

This function is a wrapper to submit a `pertables` object to `cca` function of the **vegan** package. See the documentation of `cca` for details about `formula` and `Condition` use.

Value

`cca_pertables` returns an object of class `cca_pertables`, basically a list with the following components:

`raw`	An object of class classcca. The results of applying `cca` to the original biological data table without the unidentified species.
`simulation`	A list with the results of the simulation: `results`, i.e. a data.frame with all the simulated R-squared and pseudo-F values; `cca.quant`, i.e. a data.frame with the summary of `results` by quantiles; `sites` i.e. a list with the scores of the sites of all the simulated data tables and `biplot`, i.e. a list with the scores of the environmental data in all the analyses

The objects of class `cca_pertables` have `print` and `plot` S3 methods for a simple access to results. See the examples.

Author(s)

Luis Cayuela and Marcelino de la Cruz

References

Cayuela, L., De la Cruz, M. and Ruokolainen, K. (2011). A method to incorporate the effect of taxonomic uncertainty on multivariate analyses of ecological data. *Ecography*, 34: 94-102. doi:10.1111/j.16000587.2009.05899.x.

See Also

`pertables`, `cca`

Examples

```
data(Amazonia)
data(soils)

# Define a new index that includes the terms used in the \code{Amazonia} dataset to define
# undetermined taxa at different taxonomic levels

index.Amazon <- c(paste("sp.", rep(1:20), sep=""), "Indet.", "indet.")

# Generate a pertables object (i.e. a list of biological data tables simulated from taxonomic
# uncertainty)
  ## Not run:

Amazonia100 <- pertables(Amazonia, index=index.Amazon, nsim=100)

# Assess the effects of taxonomic uncertainty on a CCA analysis of biological data explained
# by all the environmental variables of the soil data:

Amazonia.cca <- cca_pertables(Amazonia100 ~., data=soils)

Amazonia.cca

plot(Amazonia.cca)

## End(Not run)
# Fast example for Rcheck

Amazonia4.p2 <- pertables.p2(Amazonia[1:50,], index=index.Amazon, nsim=4, ncl=2, iseed=4)
set.seed(2)
Amazonia.cca <- cca_pertables(Amazonia4.p2 ~., data=soils)
Amazonia.cca

plot(Amazonia.cca)
```

`HCP` *Tree counts in tropical montane forest fragments*

Description

`HCP` has tree abundance data from 16 forest fragments located in the Highlands of Chiapas, southern Mexico. `HCP.coords` contains the geographical UTM coordinates for the 16 forest fragments' centroids.

Usage

```
data(HCP)
data(HCP.coords)
```

Format

`HCP` is a data frame with 231 observations and 19 variables. The three first columns contain family, genus and specific species Latin names. Columns 4 to 19 have tree abundance data for the 16 forest fragments. `HCP.coords` is a data frame with two columns and 16 rows.

References

Cayuela, L., Golicher, D.J., Rey Benayas, J.M., Gonzalez-Espinosa, M. & Ramirez-Marcial, N. 2006. Fragmentation, disturbance and tree diversity conservation in tropical montane forests. *Journal of Applied Ecology* 43: 1172-1181.

Examples

```
data(HCP)
data(HCP.coords)
```

`mantel_pertables` *Function to assess the efects of taxonomic uncertainty on Mantel tests*

Description

This function asses the effects of taxonomic uncertainty on the coefficient of correlation and the p-values of a Mantel test.

Usage

```
mantel_pertables(pertab, env, dist.method = "bray", binary = FALSE,
cor.method = "pearson", permutations = 100)
## S3 method for class 'mantel_pertables'
 plot(x, xlab = "Environmental distance",
ylab = "Sorensen's similarity index", pch = 19, ...)
```

Arguments

`pertab`	A pertables object (i.e. a list of simulated community data matrices obtained with `pertables`.
`env`	Data frame with the environmental variables.
`dist.method`	Method to compute the dissimilarity matrices from the biological and environmental data tables. One of the methods described in function `vegdist` of the package **vegan**.
`binary`	Value for the argument `binary` in the function `vegdist` of the package **vegan**.
`cor.method`	Correlation method, as accepted by `cor`: "pearson", "spearman" or "kendall".
`permutations`	Number of permutations in assessing significance.
`x`	`mantel_pertables` object to plot.
`xlab`	Label to name x-axis
`ylab`	Label to name y-axis
`pch`	Plotting 'character', i.e., symbol to use in the distance decay plot. See `points` for examples of use of this graphical argument.
`...`	Additional graphical parameters passed to plot.

Value

`mantel_pertables` returns an object of class`mantel_pertables`, basically a list with the following components:

`mantel`	A list with two components: `mantel.raw`, an object of class `'mantel'`, i.e. the results of applying `mantel` to the original biological data table without the unidentified species, and `ptax`, a p-value showing the probability of obtaining the same mantel statistic under different scenarios of taxonomic uncertainty.
`simulation`	A list with the results of the simulation: `results`, i.e. a data.frame with all the simulated mantel statistics and p-values; `mantel.quant`, i.e. a data.frame with the summary of `results` by quantiles; `vegdist`, i.e. a list with all the dissimilarity matrices employed.

The objects of class `mantel_pertables` have `print` and `plot` S3 methods for a simple access to results. See the examples.

Author(s)

Luis Cayuela and Marcelino de la Cruz

References

Cayuela, L., De la Cruz, M. and Ruokolainen, K. (2011). A method to incorporate the effect of taxonomic uncertainty on multivariate analyses of ecological data. *Ecography*, 34: 94-102. doi:10.1111/j.16000587.2009.05899.x.

See Also

`pertables`, `mantel`

Examples

```
data(Amazonia)
data(soils)

# Define a new index that includes the terms used in the \code{Amazonia} dataset to define
# undetermined taxa at different taxonomic levels

index.Amazon <- c(paste("sp.", rep(1:20), sep=""), "Indet.", "indet.")

## Not run:
# Generate a pertables object (i.e. a list of biological data tables simulated from taxonomic
# uncertainty)

Amazonia100 <- pertables(Amazonia, index=index.Amazon, nsim=100)

# Assess the effects of taxonomic uncertainty on a Mantel test of biological dissimilarity
# correlated to soil dissimilarity among sites:

Amazonia.mantel <- mantel_pertables(pertab=Amazonia100, env=soils, dist.method = "bray")

Amazonia.mantel

plot(Amazonia.mantel)

## End(Not run)
# Fast example for Rcheck

Amazonia4.p2 <- pertables.p2(Amazonia[1:50,], index=index.Amazon, nsim=4, ncl=2, iseed=4)
set.seed(2)
Amazonia.mantel <- mantel_pertables(pertab=Amazonia4.p2, env=soils, dist.method = "bray")

Amazonia.mantel

plot(Amazonia.mantel)
```

`pertables` *Function to incorporate the effect of taxonomic uncertainty on multivariate analyses of ecological data.*

Description

This function implements a permutational method to incorporate taxonomic uncertainty on multivariate analyses typically used in the analysis of ecological data. The procedure is based on iterative randomizations that randomly re-assign non identified species in each site to any of the other species found in the remaining sites.

Usage

```
pertables(data, index = NULL, nsim = 100)
pertables.p2(data, index = NULL, nsim = 100, ncl=2, iseed = NULL)
```

Arguments

`data`	Community data matrix. The three first columns are factors referring to the family, genus and species specific names. The remaining columns are numeric vectors indicating species abundances at each site.
`index`	List of additional parameters to determine the level at which species have been identified. Default values include ’Indet’, ’indet’, ’sp’, ’sp1’ to ’sp100’, ’sp 1’ to ’sp 100’, ”, and ’ ’.
`nsim`	Number of simulations of species’ identities, i.e., number of data tables to simulate.
`ncl`	Number of clusters for parallel simulation.
`iseed`	An integer to be supplied to clusterSetRNGStream, or NULL not to set reproducible seeds.

Details

The procedure is implemented in two sequential steps:

Step 1. Morphospecies identified only to genus are randomly re-assigned with the same probability within the group of species and morphospecies that share the same genus, provided they are not found in the same sites. In the re-assignment of the species identity, the species considered can also receive its own identity. For instance, let’s assume we have three floristic inventories. In site A we have *Eugenia sp1* and *E. nesiotica*. In site B we have *Eugenia nesiotica*, *E. principium* and *E. salamensis*. In site C we have *Eugenia sp2* and *E. salamensis*. *Eugenia sp1* can be thus re-identified with equal probability as *Eugenia sp2, E. principium, E. salamensis* or just maintain its own identity (*Eugenia sp1*). In the latter case, this means that we assume that *E. sp1* is a completely different species, although we do not know its true identity. On the contrary, we cannot re-identify *E. sp1* as *E. nesiotica* because they were found in the same site, so we are quite certain that *E. sp1* is different from *E. nesiotica*. The same is applied to species identified only to family and fully unidentified species. Note that when collating inventories from different researchers, we must rename all unidentified species. This is because two researchers can use the same label, e.g. *Eugenia sp1*, even though this name does not necessarily refer to the same species. For a verification of the biological identity of *Eugenia sp1* one would need to cross-check the vouchers bearing the same name.

Step 2. Step 1 is iterated *nsim* times. As a result, *nsim* matrices are obtained, all of which contain the same number of sites but variable number of species depending on the resulting re-assignment of morphospecies, The process can be time-consuming if community data matrices are large.

Function `pertables.p2` implements a parallelized version which considerably reduces computation time.

Value

The function return a list of class `pertables` with the following components

`taxunc`	Summary of the number of species fully identified (0), identified to genus (1), identified to family (2), or fully undetermined (3).
`pertables`	A list with all the simulated data matrices.
`raw`	The raw data matrix, without the unidentified especies.

Author(s)

Luis Cayuela and Marcelino de la Cruz

References

Cayuela, L., De la Cruz, M. and Ruokolainen, K. (2011). A method to incorporate the effect of taxonomic uncertainty on multivariate analyses of ecological data. *Ecography*, 34: 94-102. doi:10.1111/j.16000587.2009.05899.x.

Examples

```
data(Amazonia)
data(soils)

# Define a new index that includes the terms used in the \code{Amazonia} dataset to define
# undetermined taxa at different taxonomic levels

index.Amazon <- c(paste("sp.", rep(1:20), sep=""), "Indet.", "indet.")

#Generate a pertables object (i.e. a list of biological data tables simulated from taxonomic
# uncertainty)

 ## Not run:
# compare prformance of pertables and pertables.p2
nsim <-100
ncl <-2
gc()
t0<- Sys.time()
 Amazonia100<- pertables(Amazonia, index=index.Amazon, nsim=nsim)
 Sys.time()-t0
gc()
t0<- Sys.time()
 Amazonia100.p2<- pertables.p2(Amazonia, index=index.Amazon, nsim=nsim, ncl=ncl)
 Sys.time()-t0

## End(Not run)
# Example for Rcheck

Amazonia4.p2<- pertables.p2(Amazonia, index=index.Amazon, nsim=4, ncl=2)
```

`rda_pertables` *Function to assess the efects of taxonomic uncertainty on [Partial] Redundance Analysis*

Description

This function asses the effects of taxonomic uncertainty on two widely used parameters of a [Partial] Redundance Analysis, i.e. the 'percentage explained variance' (sometimes referred to as *R-squared*) and the *'pseudo-F'* .

Usage

```
rda_pertables(fml, data, scale=FALSE,...)
## S3 method for class 'rda_pertables'
 plot(x, pch = 18, ...)
```

Arguments

`fml`	Model formula, where the left hand side gives a pertables object (i.e. a list of simulated community data matrices obtained with `pertables`, 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.
`scale`	Scale species to unit variance (like correlations).
`x`	rda_pertables object to plot.
`pch`	Plotting 'character', i.e., symbol to use in the RDA plot. See `points` for examples of use of this graphical argument.
`...`	Additional graphical parameters passed to plot.

Details

This function is a wrapper to submit a `pertables` object to `rda` function of the **vegan** package. See the documentation of `cca` for details about `formula` and `Condition` use.

Value

`rda_pertables` returns an object of class`'rda_pertables'`, basically a list with the following components:

`raw`	An object of class class`'rda'`. The results of applying `rda` to the original biological data table without the unidentified species.
`simulation`	A list with the results of the simulation: `'results'`, i.e. a data.frame with all the simulated R-squared and pseudo-F values; `'rda.quant'`, i.e. a data.frame with the summary of `'results'` by quantiles; `'sites'` i.e. a list with the scores of the sites of all the simulated data tables and `'biplot'`, i.e. a list with the scores of the environmental data in all the analyses

The objects of class `'rda_pertables'` have `print` and `plot` S3 methods for a simple access to results. See the examples.

Author(s)

Luis Cayuela and Marcelino de la Cruz

References

Cayuela, L., De la Cruz, M. and Ruokolainen, K. (2011). A method to incorporate the effect of taxonomic uncertainty on multivariate analyses of ecological data. *Ecography*, 34: 94-102. doi:10.1111/j.16000587.2009.05899.x.

See Also

`pertables`, `cca`

Examples

```
data(Amazonia)
data(soils)

# Define a new index that includes the terms used in the \code{Amazonia} dataset to define
# undetermined taxa at different taxonomic levels

index.Amazon <- c(paste("sp.", rep(1:20), sep=""), "Indet.", "indet.")


#Generate a pertables object (i.e. a list of biological data tables simulated from taxonomic
# uncertainty)
 ## Not run:
Amazonia100 <- pertables(Amazonia, index=index.Amazon, nsim=100)

# Assess the effects of taxonomic uncertainty on a RDA analysis of biological data explained
# by all the environmental variables of the soil data:

Amazonia.rda <- rda_pertables(Amazonia100 ~., data=soils)

Amazonia.rda

plot(Amazonia.rda)


## End(Not run)

# Fast example for Rcheck

Amazonia4.p2 <- pertables.p2(Amazonia[1:50,], index=index.Amazon, nsim=4, ncl=2, iseed=4)
set.seed(2)
Amazonia.rda <- rda_pertables(Amazonia4.p2 ~., data=soils)
```

```
Amazonia.rda

plot(Amazonia.rda)
```

Index