

Package ‘ecocbo’

August 21, 2024

Title Calculating Optimum Sampling Effort in Community Ecology

Version 0.12.0

Description A system for calculating the optimal sampling effort, based on the ideas of “Ecological cost-benefit optimization” as developed by A. Underwood (1997, ISBN 0 521 55696 1). Data is obtained from simulated ecological communities with `prep_data()` which formats and arranges the initial data, and then the optimization follows the following procedure of four functions: (1) `scompvar()` calculates the variation components necessary for (2) `sim_cbo()` to calculate the optimal combination of number of sites and samples depending on either an economic budget or on a desired statistical accuracy. Additionally, (3) `sim_beta()` estimates statistical power and type 2 error by using Permutational Multivariate Analysis of Variance, and (6) `plot_power()` represents the results of the previous function.

License GPL (>= 3)

Encoding UTF-8

RoxygenNote 7.3.1

Imports ggplot2, ggpubr, sampling, stats, rlang, foreach, parallel, doParallel, doSNOW, vegan, SSP

Depends R (>= 2.10)

LazyData true

Suggests knitr, rmarkdown, testthat (>= 3.0.0)

Config/testthat/edition 3

VignetteBuilder knitr

NeedsCompilation no

Author Edlin Guerra-Castro [aut, cph]
(<https://orcid.org/0000-0003-3529-4507>),
Arturo Sanchez-Porras [aut, cre]
(<https://orcid.org/0000-0002-1691-286X>)

Maintainer Arturo Sanchez-Porras <sp.arturo@gmail.com>

Repository CRAN

Date/Publication 2024-08-21 08:00:02 UTC

Contents

ecocbo-package	2
epiDat	3
plot_power	4
prep_data	5
prints	7
scompvar	8
simResults	9
sim_beta	10
sim_cbo	11

Index	13
--------------	-----------

ecocbo-package	<i>ecocbo: Calculating Optimum Sampling Effort in Community Ecology</i>
----------------	---

Description

A system for calculating the optimal sampling effort, based on the ideas of "Ecological cost-benefit optimization" as developed by A. Underwood (1997, ISBN 0 521 55696 1). Data is obtained from simulated ecological communities, and the optimization follows the following procedure of two functions (1) `scompvar()` calculates the variation components necessary for (2) `sim_cbo()` to calculate the optimal combination of number of sites and samples depending on either an economical budget or on a desired statistical accuracy. Additionally, (3) `sim_beta()` estimates statistical power and type 2 error by using Permutational Multivariate Analysis of Variance, and (4) `plot_power()` represents the results of the previous function.

Details

The functions in **ecocbo** package can be used to identify the optimal number of sites and samples that must be considered in a community ecology study by using simulated data. Together with **SSP** package, **ecocbo** proposes a novel approach to the determination of the appropriate sampling effort in community ecology studies.

ecocbo is composed by five functions: `prep_data` gives the appropriate format to the data so that it can be used by the other functions in the package. `scompvar` calculates the components of variation for the analyzed dataset, and finally, `sim_cbo` determines an estimate of the number of sites and samples to consider to optimize the cost-benefit for an ecological sampling study. For getting more information on the data, `sim_beta` calculates statistical power for different sampling efforts and `plot_power` plots those results to help the user define the a combination of sampling effort and power to move on.

ecocbo is being developed at Github(<https://github.com/arturoSP/ecocbo>), where up-to-date versions can be found.

Author(s)

The **ecocbo** development team is Edlin Guerra-Castro and Arturo Sanchez-Porras.

References

- Underwood, A. J. (1997). Experiments in ecology: their logical design and interpretation using analysis of variance. Cambridge university press.
- Underwood, A. J., & Chapman, M. G. (2003). Power, precaution, Type II error and sampling design in assessment of environmental impacts. *Journal of Experimental Marine Biology and Ecology*, 296(1), 49-70.
- Anderson, M. J. (2014). Permutational multivariate analysis of variance (PERMANOVA). Wiley statsref: statistics reference online, 1-15.
- Guerra-Castro, E. J., Cajas, J. C., Simões, N., Cruz-Motta, J.J., & Mascaró, M. (2021). SSP: an R package to estimate sampling effort in studies of ecological communities. *Ecography*, 44(4), 561-573.

Examples

```
# Load and adjust data.
data(epiDat)

simResults <- prep_data(data = epiDat, type = "counts", Sest.method = "average",
                        cases = 5, N = 100, sites = 10,
                        n = 5, m = 5, k = 30,
                        transformation = "none", method = "bray",
                        dummy = FALSE, useParallel = FALSE,
                        model = "single.factor")

simResults

# Computing components of variation
compVar <- scompvar(data = simResults)
compVar

# Cost-benefit optimization
cboResult <- sim_cbo(comp.var = compVar, ct = 20000, ck = 100, cj = 2500)
cboResult

# Determination of statistical power
epiBetaR <- sim_beta(simResults, alpha = 0.05)
epiBetaR

# Visualization of statistical power
plot_power(data = epiBetaR, n = NULL, m = 3, method = "both")
```

epiDat

Dataset on species count of marine communities

Description

This is a dataset containing a subset from the epibionts dataset from ‘SSP’ which was made by using the three local communities that differ the most.

Usage

```
data("epiDat")
```

Format

A data frame with count of individuals for 24 observations on 151 species.

Source

Data available from the Dryad Digital Repository: <<http://dx.doi.org/10.5061/dryad.3bk3j9kj5>> (Guerra-Castro et al. 2020).

References

Guerra-Castro, E. J. et al. 2016. Scales of spatial variation in tropical benthic assemblages and their ecological relevance: epibionts on Caribbean mangrove roots as a model system. – Mar. Ecol. Prog. Ser. 548: 97–110.

Examples

```
data("epiDat")
str(epiDat)
```

plot_power

Power curves for different sampling efforts

Description

plot_power() can be used to visualize the power of a study as a function of the sampling effort. The power curve plot shows that the power of the study increases as the sample size increases, and the density plot shows the overlapping areas where α and β are significant.

Usage

```
plot_power(data, n = NULL, m = NULL, method = "power")
```

Arguments

data	Object of class "ecocbo_beta" that results from sim_beta() .
n	Defaults to NULL, and then the function computes the number of samples 'n', within the selected 'm', that result in a sampling effort close to (1 - alpha) in power. If provided, said number of samples will be used.
m	Defaults to NULL, and then the function computes the number of sites 'm' that result in a sampling effort that is close to (1 - alpha) in power. If provided, said number of site will be used.
method	The desired plot. Options are "power", "density" or "both". "power" plots the power curve, "density" plots the density distribution of pseudoF, and "both" draws both plots one next to the other.

Value

If the method is "power", then the power curves for the different values of 'm'. The selected, or computed, 'n' is marked in red. If the method is "density", then a density plot for the observed pseudoF values and a line marking the value of pseudoF that marks the significance level indicated in [sim_beta\(\)](#). If the method is "both", then a composite with power curves and a density plot side by side.

The value of the selected 'm', 'n' and the corresponding component of variation are presented in all methods.

Author(s)

Edlin Guerra-Castro (<edlinguerra@gmail.com>), Arturo Sanchez-Porras

References

Underwood, A. J. (1997). Experiments in ecology: their logical design and interpretation using analysis of variance. Cambridge university press.

Underwood, A. J., & Chapman, M. G. (2003). Power, precaution, Type II error and sampling design in assessment of environmental impacts. *Journal of Experimental Marine Biology and Ecology*, 296(1), 49-70.

See Also

[sim_beta\(\)](#) [scompvar\(\)](#) [sim_cbo\(\)](#) [prep_data\(\)](#)

Examples

```
epiBetaR <- sim_beta(simResults, alpha = 0.05)

plot_power(data = epiBetaR, n = NULL, m = 3, method = "power")
plot_power(data = epiBetaR, n = NULL, m = 3, method = "density")
plot_power(data = epiBetaR, n = 4, m = 3, method = "both")
```

prep_data

Prepare data for evaluation

Description

`prep_data()` formats and arranges the initial data so that it can be readily used by the other functions in the package. The function first gets the species names and the number of samples for each species from the input data frame. Then, it permutes the sampling efforts and calculates the pseudo-F statistic and the mean squares for each permutation. Finally, it returns a data frame with the permutations, pseudo-F statistic, and mean squares.

Usage

```

prep_data(
  data,
  type = "counts",
  Sest.method = "average",
  cases = 5,
  N = 100,
  sites = 10,
  n,
  m,
  k = 50,
  transformation = "none",
  method = "bray",
  dummy = FALSE,
  useParallel = TRUE,
  model = "single.factor"
)

```

Arguments

data	Data frame with species names (columns) and samples (rows) information. The first column should indicate the site to which the sample belongs, regardless of whether a single site has been sampled.
type	Nature of the data to be processed. It may be presence / absence ("P/A"), counts of individuals ("counts"), or coverage ("cover")
Sest.method	Method for estimating species richness. The function specpool is used for this. Available methods are the incidence-based Chao "chao", first order jackknife "jack1", second order jackknife "jack2" and Bootstrap "boot". By default, the "average" of the four estimates is used.
cases	Number of data sets to be simulated.
N	Total number of samples to be simulated in each site.
sites	Total number of sites to be simulated in each data set.
n	Maximum number of samples to consider.
m	Maximum number of sites.
k	Number of resamples the process will take. Defaults to 50.
transformation	Mathematical function to reduce the weight of very dominant species: 'square root', 'fourth root', 'Log (X+1)', 'P/A', 'none'
method	The appropriate distance/dissimilarity metric (e.g. Gower, Bray–Curtis, Jaccard, etc). The function <code>vegan::vegdist()</code> is called for that purpose.
dummy	Logical. It is recommended to use TRUE in cases where there are observations that are empty.
useParallel	Logical. Perform the analysis in parallel? Defaults to TRUE.
model	Select the model to use. Options, so far, are 'single.factor' and 'nested.symmetric'.

Value

prep_data() returns an object of class "ecocbo_data".

An object of class "ecocbo_data" is a list containing: \$Results, a data frame that lists the estimates of pseudoF for simH0 and simHa that can be used to compute the statistical power for different sampling efforts, as well as the square means necessary for calculating the variation components.

Author(s)

Edlin Guerra-Castro (<edlinguerra@gmail.com>), Arturo Sanchez-Porras

References

Underwood, A. J. (1997). Experiments in ecology: their logical design and interpretation using analysis of variance. Cambridge university press.

Underwood, A. J., & Chapman, M. G. (2003). Power, precaution, Type II error and sampling design in assessment of environmental impacts. Journal of Experimental Marine Biology and Ecology, 296(1), 49-70.

See Also

[sim_beta\(\)](#) [plot_power\(\)](#) [sim_cbo\(\)](#) [scompvar\(\)](#)

Examples

```
simResults <- prep_data(data = epiDat, type = "counts", Sest.method = "average",
                        cases = 5, N = 100, sites = 10,
                        n = 5, m = 5, k = 30,
                        transformation = "none", method = "bray",
                        dummy = FALSE, useParallel = FALSE,
                        model = "single.factor")

simResults
```

prints

S3Methods for Printing

Description

prints for ecocbo::sim_beta() objects.

Usage

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

Arguments

x Object from `ecocbo::sim_beta()` function.
 ... Additional arguments

Value

Prints the result of `ecocbo::sim_beta()` function, showing in an ordered matrix the estimated power for the different experimental designs that were considered.

 scompvar

Simulated components of variation

Description

`scompvar` can be used to calculate the average component of variation among units and the average component of variation within samples in terms of sampling effort.

Usage

```
scompvar(data, n = NULL, m = NULL)
```

Arguments

data Object of class "ecocbo_data" that results from `prep_data()`.
 n Number of samples to be considered. Defaults to NULL.
 m Site label to be used as basis for the computation. Defaults to NULL.

Value

A data frame containing the values for the variation component among sites `compVarA` and in the residuals `compVarR`.

Note

If `m` or `n` are left as NULL, the function will calculate the components of variation using the largest available values as set in the experimental design in `sim_beta()`.

Author(s)

Edlin Guerra-Castro (<edlinguerra@gmail.com>), Arturo Sanchez-Porras

References

Underwood, A. J. (1997). Experiments in ecology: their logical design and interpretation using analysis of variance. Cambridge university press.
 Underwood, A. J., & Chapman, M. G. (2003). Power, precaution, Type II error and sampling design in assessment of environmental impacts. *Journal of Experimental Marine Biology and Ecology*, 296(1), 49-70.

See Also

[sim_beta\(\)](#) [plot_power\(\)](#) [sim_cbo\(\)](#) [prep_data\(\)](#)

Examples

```
scompvar(data = simResults)
scompvar(data = simResults, n = 5, m = 2)
```

simResults

Data set containing the results of applying ecocbo::prep_data().

Description

The dataset contains the results of applying `ecocbo::prep_data()` to `epiDat`. The result is a list with one level: `$Results` is a data frame with the results of applying PERMANOVA to `epiDat` a number of times, it contains the values of `pseudoF` and the mean squares for different repeated sampling efforts.

This dataset can be used to study the variability of the `pseudoF`-statistic, `beta` and the power when an experiment is applied to a varying number of samples, sampling units, or sampling sites.

Usage

```
data("simResults")
```

Format

An object of class "ecocbo_data", also a list containing one data frame. The format is:

<code>\$Results</code>	a data frame that contains the results of the evaluation of <code>sim_beta</code> .
<code>dat.sim</code>	simulation from which the results are obtained.
<code>k</code>	number of resample for the result.
<code>m</code>	number of sites considered for the result.
<code>n</code>	number of replicates within each site for the result.
<code>pseudoFH0</code>	observed F value for the experimental design, when all observations belong to one site.
<code>pseudoFHa</code>	observed F value for the experimental design, when observations belong to different sites.
<code>AMSHa</code>	calculated mean squares among sites in the experiment.
<code>RMSHa</code>	calculated mean squares for the residuals in the experiment.

Details

This dataset comes from applying `ecocbo::prep_data()` to the basic data from `ecocbo::epiDat`.

Source

Data available from the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.3bk3j9kj5> (Guerra-Castro et al. 2020).

References

Guerra-Castro, E. J. et al. 2016. Scales of spatial variation in tropical benthic assemblages and their ecological relevance: epibionts on Caribbean mangrove roots as a model system. – Mar. Ecol. Prog. Ser. 548: 97–110.

Examples

```
data(simResults)

sim_beta(simResults, alpha = 0.05)
```

sim_beta	<i>Calculate beta and power out of simulated samples</i>
----------	--

Description

sim_beta() can be used to assess the power of a study by comparing the variation when one can assume whether an ecological community does not have composition differences (H0 true) or it does (H0 false). For example, if the beta error is 0.25, then there is a 25% chance of failing to detect a difference even if the difference is real. The power of the study is $1 - \beta$, so in this example, the power of the study is 0.75.

Usage

```
sim_beta(data, alpha = 0.05)
```

Arguments

data	An object of class "ecocbo_data" that results from applying prep_data() to a community data frame.
alpha	Level of significance for Type I error. Defaults to 0.05.

Value

sim_data() returns an object of class "ecocbo_beta".

The function print() is used to present a matrix that summarizes the results by showing the estimate power according to different sampling efforts.

An object of class "ecocbo_beta" is a list containing the following components:

- \$Power a data frame containing the estimation of power and beta for several combination of sampling efforts (m sites and n samples).
- \$Results a data frame containing the estimates of pseudoF for simH0 and simHa.
- \$alpha level of significance for Type I error.

Author(s)

Edlin Guerra-Castro (<edlinguerra@gmail.com>), Arturo Sanchez-Porras

References

Underwood, A. J. (1997). Experiments in ecology: their logical design and interpretation using analysis of variance. Cambridge university press.

Underwood, A. J., & Chapman, M. G. (2003). Power, precaution, Type II error and sampling design in assessment of environmental impacts. *Journal of Experimental Marine Biology and Ecology*, 296(1), 49-70.

Anderson, M. J. (2014). Permutational multivariate analysis of variance (PERMANOVA). *Wiley statsref: statistics reference online*, 1-15.

Guerra-Castro, E. J., Cajas, J. C., Simões, N., Cruz-Motta, J. J., & Mascaró, M. (2021). SSP: an R package to estimate sampling effort in studies of ecological communities. *Ecography*, 44(4), 561-573.

See Also

[plot_power\(\)](#) [scompvar\(\)](#) [sim_cbo\(\)](#) [prep_data\(\)](#) [SSP::asempar\(\)](#) [SSP::simdata\(\)](#)

Examples

```
sim_beta(data = simResults, alpha = 0.05)
```

sim_cbo

Simulated cost-benefit optimization

Description

`sim_cbo()` can be used to apply a cost-benefit optimization model that depends either on a desired level of precision or on a budgeted total cost, as proposed by Underwood (1997).

Usage

```
sim_cbo(comp.var, multSE = NULL, ct = NULL, ck, cj = NULL)
```

Arguments

<code>comp.var</code>	Data frame as obtained from scompvar() .
<code>multSE</code>	Optional. Required multivariate standard error for the sampling experiment.
<code>ct</code>	Optional. Total cost for the sampling experiment.
<code>ck</code>	Cost per replicate.
<code>cj</code>	Cost per unit.

Value

A data frame containing the optimized values for m number of sites and n number of samples to consider.

Author(s)

Edlin Guerra-Castro (<edlinguerra@gmail.com>), Arturo Sanchez-Porras

References

Underwood, A. J. (1997). Experiments in ecology: their logical design and interpretation using analysis of variance. Cambridge university press.

Underwood, A. J., & Chapman, M. G. (2003). Power, precaution, Type II error and sampling design in assessment of environmental impacts. *Journal of Experimental Marine Biology and Ecology*, 296(1), 49-70.

See Also

[sim_beta\(\)](#) [plot_power\(\)](#) [scompvar\(\)](#)

Examples

```
compVar <- scompvar(data = simResults)
```

```
sim_cbo(comp.var = compVar, multSE = NULL, ct = 20000, ck = 100, cj = 2500)
```

```
sim_cbo(comp.var = compVar, multSE = 0.15, ct = NULL, ck = 100, cj = 2500)
```

Index

- * **datasets**
 - epiDat, 3
 - simResults, 9
- * **package**
 - ecocbo-package, 2
- compvar (scompvar), 8
- ecocbo (ecocbo-package), 2
- ecocbo-package, 2
- epiDat, 3
- plot_power, 2, 4
- plot_power(), 7, 9, 11, 12
- plotpower (plot_power), 4
- prep_data, 2, 5
- prep_data(), 5, 8–11
- prepdata (prep_data), 5
- print.ecocbo_beta (prints), 7
- prints, 7
- scompvar, 2, 8
- scompvar(), 5, 7, 11, 12
- sim_beta, 2, 10
- sim_beta(), 4, 5, 7–9, 12
- sim_cbo, 2, 11
- sim_cbo(), 5, 7, 9, 11
- simbeta (sim_beta), 10
- simcbo (sim_cbo), 11
- simResults, 9
- SSP::assempar(), 11
- SSP::simdata(), 11
- vegan::vegdist(), 6