Package 'ecocbo'

August 18, 2023

Title Calculating Optimum Sampling Effort in Community Ecology

Version 0.10.2

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 four functions (1) sim_beta() estimates statistical power and type 2 error by using Permutational Multivariate Analysis of Variance, (2) plot_power() represents the results of the previous function, (3) scompvar() calculates the variation components necessary for (4) 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.

License GPL (>= 3)

Encoding UTF-8

RoxygenNote 7.2.3

Imports ggplot2, ggpubr, sampling, stats, foreach, parallel, doParallel, vegan

Depends R (>= 2.10)

LazyData true

Suggests knitr,rmarkdown, testthat (>= 3.0.0), SSP

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 2023-08-18 06:32:33 UTC

R topics documented:

ecocbo-package 2
epiBetaR 4
epiDat 5
plot_power 6
prints 7
scompvar 8
simH0Dat 9
simHaDat 10
sim_beta 11
sim_cbo 13

Index **14**

---

ecocbo-package *ecocbo: Calculating Optimum Sampling Effort in Community Ecology*

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 four functions (1) sim_beta() estimates statistical power and type 2 error by using Permutational Multivariate Analysis of Variance, (2) plot_power() represents the results of the previous function, (3) scompvar() calculates the variation components necessary for (4) 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.

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 he appropriate sampling effort in community ecology studies.

ecocbo is composed by four functions: `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. `scompvar` calculates the components of variation for the analized 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.

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

```
library(SSP)


# Load and adjust data.
data(epiDat)

epiH0 <- epiDat
epiH0[,"site"] <- as.factor("T0")
epiHa <- epiDat
epiHa[,"site"] <- as.factor(epiHa[,"site"])

# Calculate simulation parameters.
parH0 <- SSP::assempar(data = epiH0, type = "counts", Sest.method = "average")
parHa <- SSP::assempar(data = epiHa, type = "counts", Sest.method = "average")

# Simulation.
simH0Dat <- SSP::simdata(parH0, cases = 3, N = 1000, sites = 1)
simHaDat <- SSP::simdata(parHa, cases = 3, N = 100, sites = 10)

# Determination of power
epiBetaR <- sim_beta(simH0Dat, simHaDat,
                     n = 5, m = 4, k = 30,
                     alpha = 0.05,
                     transformation = "square root", method = "bray",
                     dummy = FALSE,
                     useParallel = FALSE)

epiBetaR

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

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

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

`epiBetaR` *Dataset containing an example of the results of applying ecocbo::sim_beta().*

Description

The dataset contains the results of applying ecocbo::sim_beta() to epiDat. The result is a list with three levels. $Power A data frame with the results of statistical power according to different sampling efforts. $Results 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. $alpha The value of Type I Error as chosen by the user.

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("epiBetaR")
```

Format

An object of class "ecocbo_beta", also a list of two data frames and a numeric value. The format is:

$Power		a data frame that contains power levels at different experimental designs.
	m	number of sites.
	n	number of replicates within each site.
	Power	estimation of power for the experimental design.
	Beta	estimation of error type II for the experimental design.
	fCrit	value of observed F according to the required alpha.
$Results		a data frame that contains the results of the evaluation of sim_beta.
	dat.sim	simulation from which the results are obtained.
	k	number of resample for the result.
	m	number of sites considered for the result.
	n	number of replicates within each site for the result.
	pseudoFH0	observed F value for the experimental design, when all observations belong to one site.
	pseudoFHa	observed F value for the experimental design, when observations belong to different sites.
	AMSHa	calculated mean squares among sites in the experiment.
	RMSHa	calculated mean squares for the residuals in the experiment.
$alpha		a numeric value for the Type I Error as chosen by the user.

Details

This dataset comes from applying ecocbo::sim_beta() 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(epiBetaR)

plot_power(epiBetaR, n = 3, m = 4, 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, method = "both")
```

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) that results in a sampling effort close to 95% in power. If provided, said number of samples will be used.
`m`	Site label to be used as basis for the plot.
`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 a power curve in which 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 a power curve 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()`

Examples

```
plot_power(data = epiBetaR, n = 4, m = 2, method = "both")
plot_power(data = epiBetaR, n = NULL, m = 3, method = "power")
plot_power(data = epiBetaR, n = NULL, m = 3, method = "density")
```

`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_beta" that results from sim_beta().
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()

Examples

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

simH0Dat *Dataset for using in the example of 'sim_beta'*

Description

This dataset contains half of the necessary data to test `sim_beta()`, the other half being `simHaDat`. These datasets are the result of applying the preparation steps mentioned in the vignette "ecocbo-guide": (1) subset 'epibionts', (2) adjust the 'site' labels, (3) apply `assempar()`, and (4) apply `simdata()`.

Usage

```
data("simH0Dat")
```

Format

A list formed by 3 data frames that contain 1000 rows (observations) and 95 columns (93 simulated species and 2 labels) each. The last column `site` is a factor with one level, to indicate how all observations belong to the same site.

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("simH0Dat")
data("simHaDat")

epiBetaR <- sim_beta(simH0Dat, simHaDat,
                     n = 5, m = 4, k = 50,
                     alpha = 0.05,
                     transformation = "square root", method = "bray",
                     dummy = FALSE,
                     useParallel = FALSE)

epiBetaR
```

simHaDat *Dataset for using in the example of 'sim_beta'*

Description

This dataset contains half of the necessary data to test `sim_beta()`, the other half being `simH0Dat`. These datasets are the result of applying the preparation steps mentioned in the vignette "ecocbo-guide": (1) subset 'epibionts', (2) adjust the 'site' labels, (3) apply `assempar()`, and (4) apply `simdata()`.

Usage

```
data("simHaDat")
```

Format

A list formed by 3 data frames that contain 1000 rows (observations) and 95 columns (93 simulated species and 2 labels) each. The last column `site` is a factor with 10 levels to indicate how the observations belong to several different sites.

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("simH0Dat")
data("simHaDat")

epiBetaR <- sim_beta(simH0Dat, simHaDat,
                     n = 5, m = 4, k = 50,
                     alpha = 0.05,
                     transformation = "square root", method = "bray",
                     dummy = FALSE,
                     useParallel = FALSE)

epiBetaR
```

`sim_beta` *Calculate beta and power out of simulated samples*

Description

`sim_beta()` can be used to assess the power of a study by comparing the variation when one can assume wether 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(
  simH0,
  simHa,
  n,
  m,
  k = 50,
  alpha = 0.05,
  transformation = "none",
  method = "bray",
  dummy = FALSE,
  useParallel = FALSE
)
```

Arguments

`simH0`	Simulated community from `SSP::simdata()` in which H0 is true.
`simHa`	Simulated community from `SSP::simdata()` in which H0 is false.
`n`	Maximum number of samples to consider.
`m`	Maximum number of sites.
`k`	Number of resamples the process will take. Defaults to 50.
`alpha`	Level of significance for Type I error. Defaults to 0.05.
`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 `vegan::vegdist()` 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 FALSE.

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()` `SSP::assempar()` `SSP::simdata()`

Examples

```
sim_beta(simH0Dat, simHaDat, n = 5, m = 4, k = 30, alpha = 0.05,
         transformation = "square root", method = "bray", dummy = FALSE,
         useParallel = FALSE)
```

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)
```

Arguments

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

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**
epiBetaR, 4
epiDat, 5
simH0Dat, 9
simHaDat, 10
∗ **package**
ecocbo-package, 2

compvar *(*scompvar*)*, 8

ecocbo *(*ecocbo-package*)*, 2
ecocbo-package, 2
epiBetaR, 4
epiDat, 5

plot_power, *2*, 6
plot_power(), *8*, *12, 13*
plotpower *(*plot_power*)*, 6
print.ecocbo_beta *(*prints*)*, 7
prints, 7

scompvar, *2*, 8
scompvar(), *7*, *12, 13*
sim_beta, *2*, 11
sim_beta(), *6–8*, *13*
sim_cbo, *2*, 13
sim_cbo(), *7, 8*, *12*
simbeta *(*sim_beta*)*, 11
simcbo *(*sim_cbo*)*, 13
simH0Dat, 9, *10*
simHaDat, *9*, 10
SSP::assempar(), *12*
SSP::simdata(), *12*

vegan::vegdist(), *11*