Package 'ecocbo'

August 23, 2025

Title Calculating Optimum Sampling Effort in Community Ecology

Version 0.13.0

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) prep_data() takes the original dataset and creates simulated sets that can be used as a basis for estimating statistical power and type II error. (2) sim_beta() is used to estimate the statistical power for the different sampling efforts specified by the user. (3) sim_cbo() calculates then the optimal sampling effort, based on the statistical power and the sampling costs. Additionally, (4) scompvar() calculates the variation components necessary for (5) Underwood_cbo() to calculate the optimal combination of number of sites and samples depending on either an economic budget or on a desired statistical accuracy. Lastly, (6) plot_power() helps the user visualize the results of sim_beta().

License GPL (>= 3)

URL https://github.com/arturoSP/ecocbo

BugReports https://github.com/arturoSP/ecocbo/issues

Encoding UTF-8

RoxygenNote 7.3.2

Imports ggplot2, ggpubr, sampling, stats, rlang, dplyr, tidyr, tidyselect, parabar, parallelly, vegan, SSP, plotly

Depends R (>= 4.1.0)

LazyData true

Suggests knitr,rmarkdown, testthat (>= 3.0.0)

Config/testthat/edition 3

VignetteBuilder knitr

NeedsCompilation no

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

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

Repository CRAN

Date/Publication 2025-08-23 16:50:02 UTC

Contents

`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 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 he 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 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. 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, M = 10,
                        n = 5, m = 6, k = 20,
                        transformation = "none", method = "bray",
                        dummy = TRUE, useParallel = FALSE,
                        model = "single.factor")

simResults

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


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

```
# Cost-benefit optimization
cboResult <- sim_cbo(epiBetaR, cn = 75)
cboResult

# Visualization of statistical power
plot_power(data = epiBetaR, method = "power")
```

`betaNested` *Data set containing the results of applying ecocbo::sim_beta() to a nested factors experiment.*

Description

The dataset contains the results of applying ecocbo::sim_beta() to the dataset from PAPIIT experiment. The result is a list with 4 components.

Usage

```
betaNested
```

Format

An object of class "ecocbo_beta", also a list containing four components. The format is:

$Power **m** number of sites considered for the result.

n number of replicates within each site for the result.

Power estimated statistical power.

Beta estimated type II error.

fCrit estimated pseudoF value that corresponds to the 1-alpha quartile of the distribution of pseudoF.

$Results **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.

pseduoFHa observed F value for the experimental design, when observations belong to different sites.

MSB(A) calculated mean squares among sites in the experiment.

MSR calculated mean squares for the residuals in the experiment.

$alpha usually 0.05

$model "nested.symmetric"

attribute "ecocbo.beta"

Details

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.

Source

Data available from the GitHub Digital Repository: https://github.com/edlinguerra/IA206320_publico/tree/main/datos (Guerra-Castro et al. 2022).

`epiBetaR` *Data set containing the results of applying ecocbo::sim_beta() to a single factor experiment.*

Description

The dataset contains the results of applying ecocbo::sim_beta() to an excerpt from the dataset epibionts from the package SSP. The result is a list with 4 components.

Usage

```
epiBetaR
```

Format

An object of class "ecocbo_beta", also a list containing four components. The format is:

$Power **m** number of sites considered for the result.

n number of replicates within each site for the result.

Power estimated statistical power.

Beta estimated type II error.

fCrit estimated pseudoF value that corresponds to the 1-alpha quartile of the distribution of pseudoF.

$Results **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.

pseduoFHa observed F value for the experimental design, when observations belong to different sites.

MSB(A) calculated mean squares among sites in the experiment.

MSR calculated mean squares for the residuals in the experiment.

$alpha usually 0.05

$model nested.symmetric

attribute ecocbo.beta

Details

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.

Source

Data available from the GitHub Digital Repository: https://github.com/edlinguerra/SSP/tree/master/data (Guerra-Castro et al. 2022).

---

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

```
epiDat
```

Format

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

Source

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

---

macrofDat *Dataset on species count of coastal macrofauna.*

Description

This is a dataset containing a subset from the macrofauna recorded in the PAPIIT experiment.

Usage

```
macrofDat
```

Format

A dataframe with counts of individuals for 43 observations on 34 species.

Source

Data available from the GitHub Digital Repository: https://github.com/edlinguerra/IA206320_publico/tree/main/datos (Guerra-Castro et al. 2022).

`plot_power` *Plot Statistical Power and Pseudo-F Distributions*

Description

Visualizes the statistical power of a study as a function of the sampling effort. The power curve plot illustrates how power increases with sample size, while the density plot highlights overlapping areas where α and β are significant.

Usage

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

Arguments

`data`	Object of class `"ecocbo_beta"` obtained from `sim_beta()`.
`n`	Optional. Integer. Number of samples `n` within the selected `m`. Defaults to `NULL`, and the function selects the number of samples yielding a power close to $1 - \alpha$.
`m`	Optional. Integer. Number of replicates `m` to use for power computation. Defaults to `NULL`, in which case the function selects the number of sites that result in a sampling effort that is close to $1 - \alpha$.
`method`	Character. Type of plot to generate: • "power": Plots the power curve. • "density": Plots the density distribution of pseudo-F values. • "both": Displays both plots side by side. • "surface": Displays a 3d surface plot of the power curves for nested factors experiments.
`completePlot`	Logical. Is the plot to be drawn complete? If TRUE the plot will be trimmed to present a better distribution of the density plot.

Value

A plot displaying:

- If `method = "power"`, power curves for different values of `m`, with the selected `n` highlighted in red.
- If `method = "density"`: a density plot of observed pseudo-F values with a vertical line indicating significance from `sim_beta()`.
- If `method = "both"`: a composite figure with both the power curve and the density plot.
- If `method = "surface"`: a surface plot for the statistical power in different sampling designs.

The selected values of `m`, `n`, and the corresponding component of variation are displayed in all cases.

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

```
# Power curve visualization
plot_power(data = epiBetaR, method = "power")

# Density plot of pseudo-F values
plot_power(data = betaNested, method = "density")

# Composite plot with both power curve and density plot
plot_power(data = betaNested, method = "both")
```

---

prep_data *Prepare Data for Evaluation*

Description

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,
  M = 3,
```

```
  n,
  m,
  k = 50,
  transformation = "none",
  method = "bray",
  dummy = FALSE,
  useParallel = TRUE,
  model = "single.factor"
)
```

Arguments

`data`	Data frame where columns represent species names and rows correspond to samples. • For `"single.factor"` analysis: The first column should indicate the replicate to which the sample belongs. • For `"nested.symmetric"` analysis: The first column should indicate the treatment, and the second column should indicate the replicate.
`type`	Character. Nature of the data to be processed. It may be presence / absence ("P/A"), counts of individuals ("counts"), or coverage ("cover").
`Sest.method`	Character Method for estimating species richness using `vegan::specpool()`. Available methods are the incidence-based Chao ("chao"), first order jackknife ("jack1"), second order jackknife ("jack2") and Bootstrap ("boot"). By default, the average ("average") of the four estimates is used.
`cases`	Integer. Number of simulated datasets.
`N`	Integer. Total number of samples simulated per site.
`M`	Integer. Total number of replicates simulated per dataset.
`n`	Integer. Maximum number of samples to consider (must be `<= N`).
`m`	Integer. Number of replicates to consider. (must be `<=M`)
`k`	Integer. Number of resampling iterations. Defaults to 50.
`transformation`	Character. Transformation applied to reduce the weight of dominant species: "square root", "fourth root", "Log (X+1)", "P/A", "none".
`method`	Character. Dissimilarity metric used `vegan::vegdist()`. Common options include: "Gower", "Bray–Curtis", "Jaccard", etc.
`dummy`	Logical. If `TRUE`, adds a small constant to empty observations.
`useParallel`	Logical. If `TRUE`, enables parallel computation. Defaults to `TRUE`.
`model`	Character. Select the model to use. Options, so far, are `"single.factor"` and `"nested.symmetric"`.

Details

The input dataset should have:

- One or two leading columns for treatment/replicate labels.
- Subsequent columns representing species presence/absence, counts, or coverage.
- `"single.factor"` requires a single column for replicates.
- `"nested.symmetric"` requires two columns: treatment and replicate in that order.

Value

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

An object of class "ecocbo_data" is a list containing:

- `$Results`, a data frame that lists the estimates of pseudoF for `simH0` and `simHa`, useful for statistical power analysis. It also includes mean squares for variance component estimation.
- `$model`, a label for keeping track of the model that is being used in the analysis.
- `$a`, an integer for the number of treatments recorded from the original data.

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, M = 10,
                        n = 5, m = 5, k = 30,
                        transformation = "none", method = "bray",
                        dummy = FALSE, useParallel = FALSE,
                        model = "single.factor")

simResults
```

`scompvar` *Simulated Components of Variation*

Description

Computes the average components of variation among sampling units and within samples in relation to sampling effort.

Usage

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

Arguments

data	Object of class "ecocbo_data" obtained from prep_data().
n	Optional. Integer. Number of samples to consider.
m	Optional. Integer. Number of replicates to consider.

Details

If m or n are set to NULL, the function automatically uses the largest available values from the experimental design set in sim_beta().

Value

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

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.

Usage

```
simResults
```

Format

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

$Results dat.sim simulation from which the results are obtained.

k number of resample 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.

MSR calculated mean squares for the residuals in the experiment.

$model "single.factor"

attribute class: ecocbo_data

Details

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.

Source

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

---

simResultsNested *Data set containing the results of applying ecocbo::prep_data() to a nested factors experiment.*

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.

Usage

```
simResultsNested
```

Format

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

$Results **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.

MSB(A) calculated mean squares among sites in the experiment.

MSR calculated mean squares for the residuals in the experiment.

$model "single.factor"

attribute class: ecocbo_data

Details

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.

Source

Source data is available from https://github.com/edlinguerra/IA206320_publico/tree/main/datos (Guerra-Castro et al. 2020).

---

sim_beta *Calculate Beta Error and Statistical Power from Simulated Samples*

Description

Estimates the statistical power of a study by comparing variation under null and alternative hypotheses. For instance, if the beta error is 0.25, there is a 25% chance of failing to detect a real difference, and the power of the study is $1 - \beta$, meaning 0.75 in this case.

Usage

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

Arguments

`data`	An object of class `"ecocbo_data"` that results from applying `prep_data()` to a community dataset.
`alpha`	Numeric. Significance level for Type I error. Defaults to 0.05.

Details

The function displays a summary matrix with estimated power values for various sampling efforts.

Value

A list of class "ecocbo_beta", containing:

- `$Power`: a data frame with power and beta estimates across different sampling efforts (`m` sites and `n` samples).
- `$Results`: a data frame with pseudo-F estimates for `simH0` and `simHa`.
- `$alpha`: significance level 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::assempar()` `SSP::simdata()`

Examples

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

`sim_cbo` *Cost-Benefit Optimization for Sampling Effort*

Description

Given a table of statistical power estimates produced by `sim_beta`, `sim_cbo` finds the sampling design (number of replicates/site and sites) that minimizes total cost while achieving a user-specified power threshold.

Usage

```
sim_cbo(data, cn, cm = NULL)
```

Arguments

data	Object of class "ecocbo_beta", as returned by sim_beta.
cn	Numeric. Cost per sampling unit.
cm	Numeric. Fixed cost per replicate.

Value

A data frame with one row per candidate design. In the single factor case, the results include the available n values, their statistical power and cost. For the nested symmetric experiments, the results include all the available values for m, the optimal n, according to the power, and the associated cost. The results also mark a suggested sampling effort, based on the cost and power range as selected by the user.

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() Underwood_cbo()

Examples

```
# Optimization of single factor experiment
sim_cbo(data = epiBetaR, cn = 80)

# Optimization of a nested factor experiment
sim_cbo(data = betaNested, cn = 80, cm = 180)
```

Underwood_cbo *Cost-Benefit Optimization after Underwood's equations*

Description

Applies a cost-benefit optimization model based on either a desired level of precision or a predefined budget, following the approach of Underwood (1997).

Usage

```
Underwood_cbo(
  comp.var,
  multSE = NULL,
  budget = NULL,
  a = NULL,
  ca = NULL,
  cm = NULL,
  cn
)
```

Arguments

comp.var	Data frame as obtained from scompvar(), containing variance component estimates
multSE	Optional. Numeric. Required multivariate standard error for the sampling experiment.
budget	Optional. Numeric. Total budget available for the sampling experiment.
a	Numeric. Number of treatments to consider.
ca	Numeric. Cost per treatment.
cm	Numeric. Cost per replicate.
cn	Numeric. Cost per sampling unit.

Value

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

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()` `sim_cbo()`

Examples

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

# Optimization based on budget constraint
Underwood_cbo(comp.var = compVar, multSE = NULL, budget = 20000, a = 3, ca = 2500, cn = 100)

# Optimization based on precision constraint
Underwood_cbo(comp.var = compVar, multSE = 0.15, cn = 150)
```

Index