

Package ‘SurvBal’

September 12, 2024

Type Package

Title Selection of Compositional Microbiome Balances for Survival Outcomes

Version 1.2.0

Description Enables the selection of microbiome balances in relation to censored survival or time-to-event outcomes, which are of considerable interest in many biomedical studies. The most commonly used survival models – the Cox proportional hazards and parametric survival (including accelerated failure time) models are included in the package, which are used in combination with step-wise selection procedures to identify the optimal associated balance of microbiome, i.e., the ratio of the geometric means of two groups of taxa's relative abundances.

Author Ying Li [aut],
Wodan Ling [aut, cre, cph]

Maintainer Ying Li <yil4013@med.cornell.edu>

URL <https://github.com/yinglia/SurvBal>

BugReports <https://github.com/yinglia/SurvBal/issues>

Depends R (>= 3.5.0)

Imports boot,
ggplot2,
nonnestcox,
survival (>= 3.5.7),
survminer,
zCompositions (<= 1.4.0.1),
vegan,
MiRKAT

Suggests knitr,
rmarkdown

VignetteBuilder knitr

License GPL (>= 3)

Encoding UTF-8

LazyData true

RoxygenNote 7.3.2

R topics documented:

bacteria	2
balance_selection	2
gvhd	4

Index	6
--------------	----------

bacteria	<i>Raw Taxon Count Table</i>
----------	------------------------------

Description

The data is a raw taxon count table consisting of 63 samples and 139 taxa

Usage

```
bacteria
```

Format

An object of class `data.frame` with 63 rows and 139 columns.

balance_selection	<i>Selection of the optimal balance of microbiome associated with survival outcomes</i>
-------------------	---

Description

`balances_selection()` identifies the optimal balance, i.e., the log-ratio of the geometric means of two sets of taxa that is most associated with the survival outcome using a greedy forward step-wise selection approach

Usage

```
balance_selection(
  Surv_obj,
  data,
  covariates = NULL,
  min_prevalence = 0.1,
  mult_repl = TRUE,
  model = "coxph",
  dist = "weibull",
  stopping_pvalue = 0.15,
  sequential_test = FALSE,
  sequential_alpha = 0.25,
  selection_criterion = "min_decrement_pvalue",
  selection_threshold = 0.15,
  quantile_plotted = c(0.25, 0.75),
  alpha = 0.05
)
```

Arguments

Surv_obj	An object of class Surv generated by <code>survival::Surv</code>
data	A $n \times p$ matrix which is the raw taxon count table, where rows are the samples, columns are the taxa
covariates	A $n \times q$ data frame containing important covariates for adjustment, which could be a mixture of q continuous and discrete (pre-defined as factor) variables. The default value is NULL
min_prevalence	The minimum prevalence of taxa that could be considered in the selection procedure. The default value is 0.1
mult_repl	A logical variable indicating which algorithm is used to process the raw taxon count table. If TRUE then the geometric Bayesian multiplicative replacement (<code>zCompositions::cmultRepl</code>) is used to impute the inflated zeros. Otherwise, a small pseudo-count (0.5) will be added to all raw counts. Finally, either processed data will be converted to relative abundance, then log of relative abundance will be used in the selection procedure. The default value is TRUE
model	Specifies which kind of survival regression model is built. The options are "coxph" (<code>survival::coxph</code>) and "parametric" (<code>survival::survreg</code>). Make sure there is no zero survival time if choose "parametric". The default value is "coxph"
dist	Specifies which kind of parametric distribution is used if model is "parametric". The options include "weibull", "exponential", "gaussian", "logistic", "lognormal" and "loglogistic". For more details, please see <code>survival::survreg</code> . The default value is "weibull"
stopping_pvalue	The threshold of p-value to stop the forward search. The forward search where, at each step, a new taxon is added to the existing balance, will be stopped if the resulting p-value is larger than stopping_pvalue. The default value is 0.15
sequential_test	A logical variable indicating whether sequential testing, in addition to stopping_pvalue, is used to stop the forward search. If TRUE, the forward search where, at each step, a new taxon is added to the existing balance, will be stopped if the new balance is not significantly different from the current one in the survival regression model. The default value is FALSE
sequential_alpha	The level of significance for the sequential testing if sequential_test is TRUE. The default value is 0.25
selection_criterion	The criterion to select the global balance in the final model. The options include "min_pvalue" and "min_decrement_pvalue". If "min_pvalue" is used, the balance that has the smallest p-value along the forward selection path will be selected. If "min_decrement_pvalue" is used, the decrement of p-value along the forward selection path will be calculated, the balance before the first decrement that is smaller than selection_threshold will be selected. The default value is "min_decrement_pvalue"
selection_threshold	A threshold of p-value decrement used for balance selection in the final model if selection_criterion is "min_decrement_pvalue". The default value is 0.15
quantile_plotted	A vector specifies the quantiles of the final selected balance that will be shown in the survival plot. The default value is <code>c(0.25, 0.75)</code>

alpha The level of significance for survival plot of the final selected balance. The corresponding confidence intervals will be shown in the survival plot. The default value is 0.05

Value

`balance_selection` return a list containing the following components:

<code>global_p</code>	Omnibus p-value from the global community-level association test, MiRKAT-S, which examines whether there is an overall shift in the microbiome composition (presence-absence status and abundance, encoded by Bray-Curtis and Jaccard distances) regarding the survival outcome
<code>selection_path</code>	A matrix showing the forward selection path. If <code>sign</code> is 1, the corresponding taxon is in the numerator of the balance; If <code>sign</code> is 0, the corresponding taxon is in the denominator of the balance. <code>p_value</code> is the p-value of the balance consisting of taxa up to the current taxon
<code>survival_model</code>	An object of class <code>coxph</code> or <code>survreg</code> depending on the user's input, which is the final model with the final selected balance of microbiome
<code>balance_name</code>	A list containing names of the taxa in the numerator and the denominator of the final selected balance
<code>balance</code>	The value of the final selected balance of microbiome for each sample
<code>survival_plot</code>	A plot of survival curves with confidence intervals, stratified by chosen quantiles of the final selected balance of microbiome

References

Rivera-Pinto, J., Egozcue, J. J., Pawlowsky-Glahn, V., Paredes, R., Noguera-Julian, M., & Calle, M. L. "Balances: a new perspective for microbiome analysis". *MSystems* 3.4 (2018): 10-1128

Fine, J. P. "Comparing nonnested Cox models". *Biometrika* 89.3 (2002): 635-648

Wahrendorf, J., Becher, H., & Brown, C. C. "Bootstrap comparison of non-nested generalized linear models: applications in survival analysis and epidemiology". *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 36.1 (1987): 72-81

Examples

```
balance_selection(Surv_obj = gvhd, data = bacteria, mult_repl = TRUE,
model = "coxph", selection_criterion = "min_pvalue")

balance_selection(Surv_obj = gvhd, data = bacteria, mult_repl = FALSE,
model = "parametric", dist = "weibull", selection_criterion = "min_decrement_pvalue")
```

gvhd

Survival Data

Description

The data is a survival object containing survival time and censoring/event information for 63 samples

gvhd

5

Usage

gvhd

Format

An object of class Surv with 63 rows and 2 columns.

Index

* datasets

bacteria, [2](#)

gvhd, [4](#)

bacteria, [2](#)

balance_selection, [2](#)

gvhd, [4](#)