

Package ‘BaHZING’

February 17, 2025

Title Bayesian Hierarchical Zero-Inflated Negative Binomial Regression with G-Computation

Version 1.0.0

Description A Bayesian model for examining the association between environmental mixtures and all Taxa measured in a hierarchical microbiome dataset in a single integrated analysis. Compared with analyzing the associations of environmental mixtures with each Taxa individually, 'BaHZING' controls Type 1 error rates and provides more stable effect estimates when dealing with small sample sizes.

License GPL (>= 3)

Depends R (>= 4.1.0), rjags (>= 4.0.0)

Imports bayestestR, dplyr, magrittr, phyloseq, pscl, R2jags, stats, stringr, tidyr

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

Config/testthat/parallel true

Encoding UTF-8

LazyData true

RoxygenNote 7.3.2

SystemRequirements JAGS 4.x.y (<http://mcmc-jags.sourceforge.net>)

NeedsCompilation no

Author Hailey Hampson [aut],
Jesse Goodrich [aut, cre] (<<https://orcid.org/0000-0001-6615-0472>>),
Hongxu Wang [aut],
Tanya Alderete [ctb],
Shardul Nazirkar [ctb],
David Conti [aut]

Maintainer Jesse Goodrich <jagoodri@usc.edu>

Repository CRAN

Date/Publication 2025-02-17 11:20:07 UTC

Contents

BaHZING_Model	2
Format_BaHZING	4
iHMP	5
iHMP_Reduced	5
Index	6

BaHZING_Model	<i>BaHZING_Model Function This function implements the BaHZING model for microbiome data analysis.</i>
---------------	--

Description

BaHZING_Model Function This function implements the BaHZING model for microbiome data analysis.

Arguments

formatted_data	An object containing formatted microbiome data.
x	A vector of column names of the exposures.
covar	An optional vector of the column names of covariates.
n.chains	An optional integer specifying the number of parallel chains for the model in jags.model function. Default is 3.
n.adapt	An optional integer specifying the number of iterations for adaptation in jags.model function. Default is 5000.
n.iter.burnin	An optional integer specifying number of iterations in update function. Default is 10000.
n.iter.sample	An optional integer specifying the number of iterations in coda.samples function. Default is 10000.
exposure_standardization	Method for standardizing the exposures. Should be one of "standard_normal" (the default), "quantile", or "none". If "none", exposures are not standardized before analysis, and counterfactual profiles must be specified by the user.
counterfactual_profiles	A 2xP matrix or a vector with length of 2; P is the number of exposures in x. If a 2xP matrix is provided, the effect estimates for the mixture are interpreted as the estimated change in the outcome when changing each exposure p in 1:P is changed from counterfactual_profiles[1,p] to counterfactual_profiles[2,p]. If a vector of length 2 is provided, the effect estimates for the mixture are interpreted as the estimated change in the outcome when changing each exposure from counterfactual_profiles[1] to counterfactual_profiles[2]. If exposure_standardization = "standard_normal", then the default is c(-0.5, 0.5), and the effect estimate is calculated based on increasing all exposures in the mixture by one standard deviation. If exposure_standardization = "quantile", then

	the default is $c(0,1)$, and the effect estimate is calculated based on increasing all exposures in the mixture by one quantile (where the number of quantiles is based on the parameter q).
<code>q</code>	An integer specifying the number of quantiles. Only required if <code>exposure_standardization = "quantile"</code> . If <code>exposure_standardization = "quantile"</code> and <code>q</code> is not specified, then a default of <code>q = 4</code> is used.
<code>verbose</code>	If TRUE (default), function returns information a data quality check.
<code>return_all_estimates</code>	If FALSE (default), results do not include the dispersion and omega estimates from the BaHZING model.
<code>ROPE_range</code>	Region of practical equivalence (ROPE) for calculating <code>p_ropes</code> . Default is $c(-0.1, 0.1)$.

Value

A data frame containing results of the Bayesian analysis, with the following columns:

- `taxa_full`: Full Taxa information, including all levels of the taxonomy. Taxonomic levels are split by two underscores (`'__'`).
- `taxa_name`: Taxa name, which is the last level of the taxonomy.
- `domain`: domain of the taxa.
- `exposure`: Exposure name (either one of the individual exposures, or the mixture).
- `component`: Zero inflated model estimate or the Count model estimate.
- `estimate`: Point estimate of the posterior distributions.
- `bci_lcl`: 95% Bayesian Credible Interval Lower Limit. Calculated as the equal tailed interval of posterior distributions using the quantiles method.
- `bci_ucl`: 95% Bayesian Credible Interval Upper Limit. Calculated as the equal tailed interval of posterior distributions using the quantiles method.
- `p_direction`: The Probability of Direction, calculated with `bayestestR`. A higher value suggests a higher probability that the estimate is strictly positive or negative. In other words, the closer the value to 1, the higher the probability that the estimate is non-zero. Values can not be less than 50%. From `bayestestR`: also known as the Maximum Probability of Effect (MPE). This can be interpreted as the probability that a parameter (described by its posterior distribution) is strictly positive or negative (whichever is the most probable). Although differently expressed, this index is fairly similar (i.e., is strongly correlated) to the frequentist p-value.
- `p_ropes`: The probability that the estimate is not within the Region of practical equivalence (ROPE), calculated with `bayestestR`. The proportion of the whole posterior distribution that doesn't lie within the `ROPE_range`.
- `p_map`: Bayesian equivalent of the p-value, calculated with `bayestestR`. From `bayestestR`: `p_map` is related to the odds that a parameter (described by its posterior distribution) has against the null hypothesis (h_0) using Mills' (2014, 2017) Objective Bayesian Hypothesis Testing framework. It corresponds to the density value at the null (e.g., 0) divided by the density at the Maximum A Posteriori (MAP).

Format_BaHZING

Format_BaHZING Function

Description

This function takes a phyloseq object and performs formatting operations on it, including modifying the taxonomic table, uniting taxonomic levels, and creating matrices based on taxonomic information.

Arguments

phyloseq.object

A phyloseq object.

Details

The Format_BaHZING function is the core function of the Format_BaHZING package. It takes a phyloseq object as input and performs various formatting operations to prepare the data for analysis. The function modifies the taxonomic table to add taxonomic prefixes (e.g., "d__" for Kingdom), unites taxonomic levels, and creates matrices based on taxonomic information. The formatted data is then returned as a list containing different data frames for further analysis.

The package relies on the phyloseq, dplyr, and stringr packages for data manipulation, and also uses functions from tidyr to unite taxonomic levels.

The main function Format_BaHZING is exported and can be accessed by other packages or scripts that depend on the functionalities provided by this package.

The column names 'Kingdom', 'Phylum', 'Class', 'Order', 'Family', 'Genus', and 'Species' in the tax_table of the phyloseq object should be user-defined and assigned in this function. The function will use these column names to perform various operations.

Value

A list with the following elements:

- `Table`: Formatted microbiome data as a data frame.
- `Species.Genus.Matrix`: Data frame for species-genus relationships (optional).
- `Genus.Family.Matrix`: Data frame for genus-family relationships (optional).
- `Family.Order.Matrix`: Data frame for family-order relationships (optional).
- `Order.Class.Matrix`: Data frame for order-class relationships (optional).
- `Class.Phylum.Matrix`: Data frame for class-phylum relationships (optional).

Note

The column names 'Kingdom', 'Phylum', 'Class', 'Order', 'Family', 'Genus', and 'Species' in the tax_table are expected to be user-defined and assigned within the function.

*iHMP**iHMP data*

Description

A subset of data from the Integrative Human Microbiome Project

Usage

iHMP

Format

iHMP:

A phyloseq object with microbiome data for 105 participants at their first visit.

Source

<https://www.nature.com/articles/s41586-019-1237-9>>

*iHMP_Reduced**iHMP_Reduced data*

Description

A subset of data from the Integrative Human Microbiome Project

Usage

iHMP_Reduced

Format

iHMP_Reduced:

A phyloseq object with subset microbiome data for 105 participants at their first visit.

Source

<https://www.nature.com/articles/s41586-019-1237-9>>

Index

* datasets

iHMP, [5](#)

iHMP_Reduced, [5](#)

BaHZING_Model, [2](#)

Format_BaHZING, [4](#)

iHMP, [5](#)

iHMP_Reduced, [5](#)