Package 'coda4microbiome'

July 17, 2024

Title Compositional Data Analysis for Microbiome Studies

Version 0.2.4

Description Functions for microbiome data analysis that take into account its compositional nature. Performs variable selection through penalized regression for both, cross-sectional and longitudinal studies, and for binary and continuous outcomes.

License MIT + file LICENSE

URL https://malucalle.github.io/coda4microbiome/

BugReports https://github.com/malucalle/coda4microbiome/issues

Encoding UTF-8

RoxygenNote 7.2.3

LazyData false

Imports corrplot, glmnet, plyr, pROC, ggpubr, ggplot2, ComplexHeatmap, circlize, survival, survminer

Suggests rmarkdown

NeedsCompilation no

Author Malu Calle [aut] (<https://orcid.org/0000-0001-9334-415X>), Toni Susin [aut, cre] (<https://orcid.org/0000-0002-0874-2784>), Meritxell Pujolassos [aut] (<https://orcid.org/0000-0003-0313-3506>)

Maintainer Toni Susin <toni.susin@upc.edu>

Depends R (>= 3.5.0)

Repository CRAN

Date/Publication 2024-07-17 15:50:02 UTC

Contents

`coda4microbiome` *coda4microbiome: Compositional Data Analysis for Microbiome Studies*

Description

This package provides a set of functions to explore and study microbiome data within the CoDA framework, with a special focus on identification of microbial signatures (variable selection).

coda_coxnet *coda_coxnet*

Description

Microbial signatures in survival studies The algorithm performs variable selection through an elastic-net penalized Cox regression conveniently adapted to CoDA. The result is expressed as the (weighted) balance between two groups of taxa. It allows the use of non-compositional covariates.

Usage

```
coda_coxnet(
  x,
  time,
  status,
  covar = NULL,
  lambda = "lambda.1se",
  nvar = NULL,
  alpha = 0.9,
  nfolds = 10,
  showPlots = TRUE,
  coef_threshold = 0
)
```

Arguments

x	abundance matrix or data frame (rows are samples, columns are variables (taxa))
time	time to event or follow up time for right censored data. Must be a numericvector.
status	event occurrence. Vector (type: numeric or logical) specifying 0, or FALSE, for no event occurrence, and 1, or TRUE, for event occurrence.
covar	data frame with covariates (default = NULL)
lambda	penalization parameter (default = "lambda.1se")
nvar	number of variables to use in the glmnet.fit function (default = NULL)
alpha	elastic net parameter (default = 0.9)
nfolds	number of folds
showPlots	if TRUE, shows the plots (default = TRUE)
coef_threshold	coefficient threshold, minimum absolute value of the coefficient for a variable to be included in the model (default =0)

Value

list with "taxa.num","taxa.name","log-contrast coefficients","risk.score","apparent Cindex","mean cv-Cindex","sd cv-Cindex","risk score plot","signature plot".

Author(s)

M. Calle, M. Pujolassos, T. Susin

Examples

```
data(data_survival, package = "coda4microbiome")
time <- Event_time
status <- Event
set.seed(12345)
coda_coxnet(x = x,
           time = Event_time,
           status = Event,
           covar = NULL,
           lambda = "lambda.1se", nvar = NULL,
           alpha = 0.9, nfolds = 10, showPlots = TRUE, coef_threshold = 0)
```

`coda_glmnet` *coda_glmnet*

Description

Microbial signatures in cross-sectional studies. The algorithm performs variable selection through penalized regression on the set of all pairwise log-ratios. The result is expressed as the (weighted) balance between two groups of taxa. It allows the use of non-compositional covariates.

Usage

```
coda_glmnet(
  x,
  y,
  covar = NULL,
  lambda = "lambda.1se",
  nvar = NULL,
  alpha = 0.9,
  nfolds = 10,
  showPlots = TRUE,
  coef_threshold = 0
)
```

Arguments

`x`	abundance matrix or data frame (rows are samples, columns are variables (taxa))
`y`	outcome (binary or continuous); data type: numeric, character or factor vector
`covar`	data frame with covariates (default = NULL)

lambda	penalization parameter (default = "lambda.1se")
nvar	number of variables to use in the glmnet.fit function (default = NULL)
alpha	elastic net parameter (default = 0.9)
nfolds	number of folds
showPlots	if TRUE, shows the plots (default = TRUE)
coef_threshold	coefficient threshold, minimum absolute value of the coefficient for a variable to be included in the model (default =0)

Value

if y is binary: list with "taxa.num","taxa.name","log-contrast coefficients","predictions","apparent AUC","mean cv-AUC","sd cv-AUC","predictions plot","signature plot" if not:list with "taxa.num","taxa.name","log-contrast coefficients","predictions","apparent Rsq","mean cv-MSE","sd cv-MSE","predictions plot","signature plot"

Author(s)

M. Calle - T. Susin

Examples

```
data(Crohn, package = "coda4microbiome")

set.seed(123)

coda_glmnet(x_Crohn[,(1:10)],y_Crohn,showPlots= FALSE)
```

`coda_glmnet0` *coda_glmnet0*

Description

Internal function for the permutational test

Usage

```
coda_glmnet0(
  x,
  lrX,
  idlrX,
  nameslrX,
  y,
  covar = NULL,
  lambda = "lambda.1se",
  alpha = 0.9
)
```

Arguments

x	.
lrX	.
idlrX	.
nameslrX	.
y	.
covar	.
lambda	.
alpha	.

Value

.

Author(s)

M. Calle - T. Susin

---

coda_glmnet_longitudinal

coda_glmnet_longitudinal

Description

Microbial signatures in longitudinal studies. Identification of a set of microbial taxa whose joint dynamics is associated with the phenotype of interest. The algorithm performs variable selection through penalized regression over the summary of the log-ratio trajectories (AUC). The result is expressed as the (weighted) balance between two groups of taxa.

Usage

```
coda_glmnet_longitudinal(
  x,
  y,
  x_time,
  subject_id,
  ini_time,
  end_time,
  covar = NULL,
  lambda = "lambda.1se",
  nvar = NULL,
  alpha = 0.9,
  nfolds = 10,
  showPlots = TRUE,
  coef_threshold = 0
)
```

Arguments

x	abundance matrix or data frame in long format (several rows per individual)
y	outcome (binary); data type: numeric, character or factor vector
x_time	observation times
subject_id	subject id
ini_time	initial time to be analyzed
end_time	end time to be analyzed
covar	data frame with covariates (default = NULL)
lambda	penalization parameter (default = "lambda.1se")
nvar	number of variables to use in the glmnet.fit function (default = NULL)
alpha	elastic net parameter (default = 0.9)
nfolds	number of folds (default = 10)
showPlots	if TRUE, shows the plots (default = FALSE)
coef_threshold	coefficient threshold, minimum absolute value of the coefficient for a variable to be included in the model (default =0)

Value

in case of binary outcome: list with "taxa.num","taxa.name","log-contrast coefficients","predictions","apparent AUC","mean cv-AUC","sd cv-AUC","predictions plot","signature plot","trajectories plot"

Author(s)

M. Calle - T. Susin

Examples

```
data(ecam_filtered, package = "coda4microbiome")   # load the data

ecam_results<-coda_glmnet_longitudinal (x=x_ecam[,(1:4)],y= metadata$diet,
x_time= metadata$day_of_life, subject_id = metadata$studyid, ini_time=0,
end_time=60,lambda="lambda.min",nfolds=4, showPlots=FALSE)

ecam_results$taxa.num
```

coda_glmnet_longitudinal0 *coda_glmnet_longitudinal0*

Description

internal function

Usage

```
coda_glmnet_longitudinal0(
  x,
  lrX,
  idlrX,
  nameslrX,
  y,
  x_time,
  subject_id,
  ini_time,
  end_time,
  covar = NULL,
  ktop = NULL,
  lambda = "lambda.1se",
  alpha = 0.9,
  nfolds = 10
)
```

Arguments

x	abundance matrix or data frame in long format (several rows per individual)
lrX	log-ratio matrix
idlrX	indices table in the log-ratio matrix
nameslrX	colnames of the log-ratio matrix
y	outcome (binary); data type: numeric, character or factor vector
x_time	observation times
subject_id	subject id
ini_time	initial time to be analyzed
end_time	end time to be analyzed
covar	data frame with covariates (default = NULL)
ktop	given number of selected taxa or compute the best number in case it is NULL (default = NULL)
lambda	penalization parameter (default = "lambda.1se")
alpha	elastic net parameter (default = 0.9)
nfolds	number of folds

Value

.

Author(s)

M. Calle - T. Susin

`coda_glmnet_longitudinal_null` *coda_glmnet_longitudinal_null*

Description

Performs a permutational test for the coda_glmnet_longitudinal() algorithm: It provides the distribution of results under the null hypothesis by implementing the coda_glmnet_longitudinal() on different rearrangements of the response variable.

Usage

```
coda_glmnet_longitudinal_null(
  x,
  y,
  x_time,
  subject_id,
  ini_time,
  end_time,
  niter = 100,
  covar = NULL,
  alpha = 0.9,
  lambda = "lambda.1se",
  nfolds = 10,
  sig = 0.05
)
```

Arguments

`x`	abundance matrix or data frame in long format (several rows per individual)
`y`	outcome (binary); data type: numeric, character or factor vector
`x_time`	observation times
`subject_id`	subject id
`ini_time`	initial time to be analyzed
`end_time`	end time to be analyzed
`niter`	number of sample iterations
`covar`	data frame with covariates (default = NULL)

alpha	elastic net parameter (default = 0.9)
lambda	penalization parameter (default = "lambda.1se")
nfolds	number of folds
sig	significance value (default = 0.05)

Value

list with "accuracy" and "confidence interval"

Author(s)

M. Calle - T. Susin

Examples

```
set.seed(123) # to reproduce the results

data(ecam_filtered, package = "coda4microbiome")   # load the data

x=x_ecam # microbiome abundance
x_time = metadata$day_of_life    # observation times
subject_id = metadata$studyid   # subject id
y= metadata$diet           # diet ("bd"= breast diet, "fd"=formula diet)
ini_time = 0
end_time = 90

 coda_glmnet_longitudinal_null (x,y, x_time, subject_id, ini_time, end_time,
                                    lambda="lambda.min",nfolds=4, niter=3)
```

---

coda_glmnet_null *coda_glmnet_null*

Description

Performs a permutational test for the coda_glmnet() algorithm: It provides the distribution of results under the null hypothesis by implementing the coda_glmnet() on different rearrangements of the response variable.

Usage

```
coda_glmnet_null(
  x,
  y,
  niter = 100,
  covar = NULL,
  lambda = "lambda.1se",
```

```
  alpha = 0.9,
  sig = 0.05
)
```

Arguments

x	abundance matrix or data frame (rows are samples, columns are variables (taxa))
y	outcome (binary or continuous); data type: numeric, character or factor vector
niter	number of iterations (default = 100)
covar	data frame with covariates (default = NULL)
lambda	penalization parameter (default = "lambda.1se")
alpha	elastic net parameter (default = 0.9)
sig	significance level for the confidence interval (default = 0.05)

Value

a list with "accuracy" and "confidence interval"

Author(s)

M. Calle - T. Susin

Examples

```
data(Crohn, package = "coda4microbiome")

coda_glmnet_null(x=x_Crohn[,(1:10)], y=y_Crohn, niter=2,covar=NULL,lambda="lambda.1se",
                                                alpha=0.9,sig=0.05)
```

Crohn *Crohn*

Description

Microbiome composition at genus level from a Crohn's disease study: 48 taxa and 975 individuals (662 patients with Crohn's disease and 313 controls)

Format

The dataset contains two objects:

x_Crohn microbiome abundance matrix for 975 individuals (rows) and 48 genera (columns)

y_Crohn a `factor`, indicating if the sample corresponds to a case (*CD*) or a control (*no*).

References

doi:10.1016/j.chom.2014.02.005

ecam_filtered *ecam_filtered*

Description

Microbiome composition at genus level from Early childhood and the microbiome (ECAM) study (Bokulich et al. 2016). Metadata and microbiome data were downloaded from https://github.com/caporaso-lab/longitudinal-notebooks. Filtered data contains information on 42 children and 37 taxa.

Format

The dataset contains three objects:

x_ecam microbiome abundance matrix in long format (873 rows) and 37 genera (columns)

taxanames vector containing the taxonomy of the 37 taxa

metadata matrix with information on the individuals at the observation time

References

Bokulich et al. (2016). Antibiotics, birth mode, and diet shape microbiome maturation during early life. Sci Transl Med 8:343

Event *data_survival*

Description

Survival Data simulated from the Crohn's disease original study: 48 taxa and 150 individuals

Format

The dataset contains three objects:

x microbiome abundance matrix for 150 individuals (rows) and 48 genera (columns)

Event a `numeric`, event occurrence. Vector (type: numeric or logical) specifying 0 or FALSE for no event occurrence, and 1 or TRUE for event occurrence.

Event_time a `numeric`, time to event or follow up time for right censored data. Must be a vector (type:numeric) specifying time to event for each sample for right censored data.

References

doi:10.1016/j.chom.2014.02.005

Event_time *data_survival*

Description

Survival Data simulated from the Crohn's disease original study: 48 taxa and 150 individuals

Format

The dataset contains three objects:

x microbiome abundance matrix for 150 individuals (rows) and 48 genera (columns)

Event a `numeric`, event occurrence. Vector (type: numeric or logical) specifying 0 or FALSE for no event occurrence, and 1 or TRUE for event occurrence.

Event_time a `numeric`, time to event or follow up time for right censored data. Must be a vector (type:numeric) specifying time to event for each sample for right censored data.

References

doi:10.1016/j.chom.2014.02.005

explore_logratios *explore_logratios*

Description

Explores the association of each log-ratio with the outcome. Summarizes the importance of each variable (taxa) as the aggregation of the association measures of those log-ratios involving the variable. The output includes a plot of the association of the log-ratio with the outcome where the variables (taxa) are ranked by importance

Usage

```
explore_logratios(
  x,
  y,
  decreasing = TRUE,
  measure = "AUC",
  covar = NULL,
  shownames = FALSE,
  maxrow = 15,
  maxcol = 15,
  showtitle = TRUE,
  mar = c(0, 0, 1, 0)
)
```

Arguments

x	abundance matrix or data frame (rows are samples, columns are variables (taxa))
y	outcome (binary or continuous); data type: numeric, character or factor vector
decreasing	order of importance (default = TRUE)
measure	association measures "AUC", "Pearson","Spearman", "glm" (default = "AUC")
covar	data frame with covariates (default = NULL)
shownames	logical, if TRUE, shows the names of the variables in the rows of the plot (default = FALSE)
maxrow	maximum number of rows to display in the plot (default = 15)
maxcol	maximum number of columns to display in the plot (default = 15)
showtitle	logical, if TRUE, shows the title of the plot (default = TRUE)
mar	mar numerical vector of the form c(bottom, left, top, right) which gives the number of lines of margin to be specified on the four sides of the plot (default mar=c(0,0,1,0))

Value

list with "max log-ratio","names max log-ratio", "order of importance", "name of most important variables", "association log-ratio with y" and "top log-ratios plot"

Author(s)

M. Calle - T. Susin

Examples

```
data(HIV, package = "coda4microbiome")

explore_logratios(x_HIV,y_HIV)
```

---

explore_lr_longitudinal

explore_lr_longitudinal

Description

Explores the association of summary (integral) of each log-ratio trajectory with the outcome. Summarizes the importance of each variable (taxa) as the aggregation of the association measures of those log-ratios involving the variable. The output includes a plot of the association of the log-ratio with the outcome where the variables (taxa) are ranked by importance

Usage

```
explore_lr_longitudinal(
  x,
  y,
  x_time,
  subject_id,
  ini_time,
  end_time,
  showPlots = FALSE,
  decreasing = TRUE,
  covar = NULL,
  shownames = FALSE,
  maxrow = 15,
  maxcol = 15,
  showtitle = TRUE,
  mar = c(0, 0, 1, 0)
)
```

Arguments

x	abundance matrix or data frame in long format (several rows per individual)
y	outcome (binary); data type: numeric, character or factor vector
x_time	observation times
subject_id	subject id
ini_time	initial time to be analyzed
end_time	end time to be analyzed
showPlots	if TRUE, shows the plot (default = FALSE)
decreasing	order of importance (default = TRUE)
covar	data frame with covariates (default = NULL)
shownames	if TRUE, shows the names of the variables in the rows of the plot (default = FALSE)
maxrow	maximum number of rows to display in the plot (default = 15)
maxcol	maximum number of columns to display in the plot (default = 15)
showtitle	logical, if TRUE, shows the title of the plot (default = TRUE)
mar	mar numerical vector of the form c(bottom, left, top, right) which gives the number of lines of margin to be specified on the four sides of the plot (default mar=c(0,0,1,0))

Value

list with "max log-ratio","names max log-ratio","order of importance","name of most important variables","association log-ratio with y","top log-ratios plot"

Author(s)

M. Calle - T. Susin

Examples

```
set.seed(123) # to reproduce the results

data(ecam_filtered, package = "coda4microbiome")   # load the data

x=x_ecam # microbiome abundance
x_time = metadata$day_of_life    # observation times
subject_id = metadata$studyid   # subject id
y= metadata$diet           # diet ("bd"= breast diet, "fd"=formula diet)
ini_time = 0
end_time = 90

ecam_logratios<-explore_lr_longitudinal(x,y,x_time,subject_id,ini_time,end_time)
```

`explore_zeros` *explore_zeros*

Description

Provides the proportion of zeros for a pair of variables (taxa) in table x and the proportion of samples with zero in both variables. A bar plot with this information is also provided. Results can be stratified by a categorical variable.

Usage

```
explore_zeros(x, id1, id2, strata = NULL)
```

Arguments

`x`	abundance matrix or data frame (rows are samples, columns are variables (taxa))
`id1`	column number in x for the first taxa
`id2`	column number in x for the second taxa
`strata`	stratification variable (default = NULL)

Value

a list with the frequency table and the associated bar plot

Author(s)

M. Calle - T. Susin

Examples

```
data(HIV, package = "coda4microbiome")

explore_zeros(x_HIV,5,6)

explore_zeros(x_HIV,5,6, strata=y_HIV)
```

`filter_longitudinal` *filter_longitudinal*

Description

Filters those individuals and taxa with enough longitudinal information

Usage

```
filter_longitudinal(
  x,
  taxanames = NULL,
  x_time,
  subject_id,
  metadata,
  ini_time = min(x_time),
  end_time = max(x_time),
  percent_indv = 0.7,
  min_obs = 3
)
```

Arguments

`x`	abundance matrix or data frame in long format (several rows per individual)
`taxanames`	names of different taxa
`x_time`	observation times
`subject_id`	subject id
`metadata`	matrix sample data
`ini_time`	initial time to be analyzed
`end_time`	end time to be analyzed
`percent_indv`	percentage of individuals with more than min_obs observations
`min_obs`	required minimum number of observations per individual

Value

list with filtered abundance table, taxanames and metadata

Author(s)

M. Calle - T. Susin

Examples

```
data(ecam_filtered, package = "coda4microbiome")   # load the data

x=x_ecam # microbiome abundance
x_time = metadata$day_of_life    # observation times
subject_id = metadata$studyid   # subject id
ini_time = 0
end_time = 360

data_filtered<-filter_longitudinal(x,taxanames,x_time, subject_id, metadata,
                                                ini_time, end_time, min_obs=4)
```

HIV *HIV*

Description

Microbiome abundances (60 taxa and 155 individuals) from an HIV study (Noguera-Julian et al. 2016).

Format

The dataset contains three objects:

x_HIV microbiome abundance matrix for 155 individuals (rows) and 60 genera (columns)

y_HIV a factor, specifying if the individual is HIV positive or (Pos) or negative (Neg).

MSM_HIV a factor, indicating sexual preferences: MSM (*Men who have Sex with Men*) or not (nonMSM).

References

doi:10.1016/j.ebiom.2016.01.032

impute_zeros *impute_zeros*

Description

Simple imputation: When the abundance table contains zeros, a positive value is added to all the values in the table. It adds 1 when the minimum of table is larger than 1 (i.e. tables of counts) or it adds half of the minimum value of the table, otherwise.

Usage

```
impute_zeros(x)
```

Arguments

x	abundance matrix or data frame (rows are samples, columns are variables (taxa))

Value

x abundance matrix or data frame with zeros substituted by imputed positive values

Author(s)

M. Calle - T. Susin

Examples

```
data(HIV, package = "coda4microbiome")

x<-impute_zeros(x_HIV)
```

integralFun *integralFun*

Description

Integral of the curve trajectory of subject_id in the interval a,b

Usage

```
integralFun(x, y, id, a, b)
```

Arguments

x	abundance matrix or data frame in long format (several rows per individual)
y	outcome (binary); data type: numeric, character or factor vector
id	subjects-ids
a	interval initial time
b	interval final time

Value

matrix with integrals for each individual (rows) and each taxa (columns)

Author(s)

M. Calle - T. Susin

---

logratios_matrix *logratios_matrix*

Description

Computes a large matrix with all the log-ratios between pairs of taxa (columns) in the abundance table

Usage

```
logratios_matrix(x)
```

Arguments

x	abundance matrix or data frame (rows are samples, columns are variables (taxa))

Value

list with matrix of log-ratios, matrix indicating the pairs of variables involved in each log-ratio, and a matrix indicating the names of the variables involved in each log-ratio.

Author(s)

M. Calle - T. Susin

Examples

```
data(HIV, package = "coda4microbiome")

lrHIV<-logratios_matrix(x_HIV[,(1:4)])

# matrix of log-ratios (first 6 rows and 6 columns):

lrHIV[[1]][1:6,1:6]

# variables involved in each log-ratio

head(lrHIV[[2]])

# names of the variables involved in each log-ratio

head(lrHIV[[3]])
```

`metadata` *ecam_filtered*

Description

Microbiome composition at genus level from Early childhood and the microbiome (ECAM) study (Bokulich et al. 2016). Metadata and microbiome data were downloaded from https://github.com/caporaso-lab/longitudinal-notebooks. Filtered data contains information on 42 children and 37 taxa.

Format

The dataset contains three objects:

x_ecam microbiome abundance matrix in long format (873 rows) and 37 genera (columns)

taxanames vector containing the taxonomy of the 37 taxa

metadata matrix with information on the individuals at the observation time

References

Bokulich et al. (2016). Antibiotics, birth mode, and diet shape microbiome maturation during early life. Sci Transl Med 8:343

MSM_HIV *HIV*

Description

Microbiome abundances (60 taxa and 155 individuals) from an HIV study (Noguera-Julian et al. 2016).

Format

The dataset contains three objects:

x_HIV microbiome abundance matrix for 155 individuals (rows) and 60 genera (columns)

y_HIV a factor, specifying if the individual is HIV positive or (`Pos`) or negative (`Neg`).

MSM_HIV a factor, indicating sexual preferences: `MSM` (*Men who have Sex with Men*) or not (`nonMSM`).

References

doi:10.1016/j.ebiom.2016.01.032

plotMedianCurve *plotMedianCurve*

Description

Internal plot function

Usage

```
plotMedianCurve(iNum, iDen, X, Y, x_time, subject_id, ini_time, end_time)
```

Arguments

`iNum`	.
`iDen`	.
`X`	.
`Y`	.
`x_time`	.
`subject_id`	.
`ini_time`	.
`end_time`	.

Value

.

Author(s)

M. Calle - T. Susin

`plot_prediction` *plot_prediction*

Description

Plot of the predictions of a fitted model: Multiple box-plot and density plots for binary outcomes and Regression plot for continuous outcome

Usage

```
plot_prediction(prediction, y, strata = NULL, showPlots = TRUE)
```

Arguments

`prediction`	the fitted values of predictions for the model
`y`	outcome (binary or continuous); data type: numeric, character or factor vector
`strata`	stratification variable (default = NULL)
`showPlots`	if TRUE, shows the plots (default = TRUE)

Value

prediction plot

Author(s)

M. Calle - T. Susin

Examples

```
# prediction plot for the log-ratio between columns 3 and 32 on HIV status

data(HIV, package = "coda4microbiome")

x<-impute_zeros(x_HIV)

lr<-log(x[,3])-log(x[,32])

plot_prediction(lr, y_HIV)
```

plot_riskscore *plot_riskscore*

Description

Plots samples ordered by microbial risk score values along time to event.

Usage

```
plot_riskscore(risk.score, x, time, status, showPlots = TRUE)
```

Arguments

risk.score	microbial risk score values resulting from the coda_coxnet model
x	original survival data
time	time to event or follow up time for right censored data. Must be a vector (type:numeric) specifying time to event for each sample for right censored data.
status	event occurrence. Vector (numeric or logical) specifying 0 (or FALSE) for no event occurrence, and 1 (or TRUE) for event occurrence.
showPlots	(default: TRUE)

Value

returns an object of class HeatmapList.

Author(s)

M. Calle, M. Pujolassos, T. Susin

Examples

```
set.seed(12345)

data(data_survival, package = "coda4microbiome")
time <- Event_time
status <- Event
crohn_cox <- coda_coxnet(x = x,
                         time = Event_time,
                         status = Event,
                         covar = NULL,
                         lambda = "lambda.1se", nvar = NULL,
                         alpha = 0.9, nfolds = 10, showPlots = TRUE, coef_threshold = 0)
plot_riskscore(risk.score = crohn_cox$risk.score,
                    x = x,
                    time = Event_time,
                    status =  Event,
                    showPlots = TRUE)
```

```
#-----------------------------------------------------------------------------
```

---

plot_signature *plot_signature*

Description

Graphical representation of the variables selected and their coefficients

Usage

```
plot_signature(vars, coeff, showPlots = TRUE, varnames = NULL)
```

Arguments

vars	variables selected
coeff	associated coefficients
showPlots	if TRUE, shows the plots (default = TRUE)
varnames	if TRUE, shows the names of the variables

Value

bar plot

Author(s)

M. Calle - T. Susin

Examples

```
plot_signature(c(2,10, 3, 15, 4), c(0.8, -0.1, 0.2, -0.6, -0.3))
```

plot_signature_curves *plot_signature_curves*

Description

Graphical representation of the signature trajectories

Usage

```
plot_signature_curves(
  varNum,
  coeff,
  x,
  y,
  x_time,
  subject_id,
  ini_time,
  end_time,
  color = c("chocolate1", "slateblue2"),
  showLabel = TRUE,
  location = "bottomright",
  inset = c(0.01, 0.02),
  cex = 0.8,
  y.intersp = 0.7,
  main_title = NULL
)
```

Arguments

varNum	column number of the variables (taxa)
coeff	coefficients (coefficients must sum-up zero)
x	microbiome abundance matrix in long format
y	binary outcome; data type: numeric, character or factor vector
x_time	observation times
subject_id	subject id
ini_time	initial time to be analyzed
end_time	end time to be analyzed
color	color graphical parameter
showLabel	graphical parameter (see help(label))
location	graphical parameter (see help(label))
inset	graphical parameter (see help(label))
cex	graphical parameter (see help(label))
y.intersp	graphical parameter (see help(label))
main_title	title plot

Value

trajectories plot

Author(s)

M. Calle - T. Susin

Examples

```
x=matrix(c(2, 3, 4, 1, 2, 5, 10, 20, 15, 30, 40, 12), ncol=2)
x_time = c(0,10,20,1,15, 25)
subject_id = c(1,1,1,2,2,2)
y=c(0,0,0,1,1,1)
plot_signature_curves(varNum=c(1,2), coeff=c(1,-1), x, y,x_time, subject_id,
                      ini_time=0, end_time=25)
```

`plot_survcurves` *plot_survcurves*

Description

Plots survival curves stratifying samples based on their signature value. Signature value for stratification can be set by the user.

Usage

```
plot_survcurves(risk.score, time, status, strata.quantile = 0.5)
```

Arguments

`risk.score`	microbial risk score values resulting from the coda_coxnet model
`time`	time to event or follow up time for right censored data. Must be a vector (type:numeric) specifying time to event for each sample for right censored data (what about interval data?).
`status`	event occurrence. Vector (type: numeric or logical) specifying 0 or FALSE for no event occurrence, and 1 or TRUE for event occurrence.
`strata.quantile`	cut-off quantile (values 0, 1) for sample stratification based on signature value. Default is set to 0.5 quantile of the signature.

Value

return an object of class ggsurvplot which is list containing the following: plot: the survival plot (ggplot object). table: the number of subjects at risk table per time (ggplot object). data.survplot: data used to plot the survival curves (data.frame). data.survtable: data used to plot the tables under the main survival curves (data.frame).

Author(s)

M. Calle, M. Pujolassos, T. Susin

Examples

```
set.seed(12345)

data(data_survival, package = "coda4microbiome")
time <- Event_time
status <- Event
crohn_cox <- coda_coxnet(x = x,
                         time = Event_time,
                         status = Event,
                         covar = NULL,
                         lambda = "lambda.1se", nvar = NULL,
                         alpha = 0.9, nfolds = 10, showPlots = TRUE, coef_threshold = 0)
plot_survcurves(risk.score = crohn_cox$risk.score,
                time,
                status,
                strata.quantile = 0.5)


#------------------------------------------------------------------------------
```

---

sCD14 *sCD14*

Description

Microbiome composition (60 taxa and 151 individuals) and inflammatory parameter sCD14 from an HIV study (Noguera-Julian et al. 2016). A dataset containing the number of counts of 60 different genera in a group of 151 samples (including HIV - infected and non - infected patients).

Format

The dataset contains two objects:

x_sCD14 microbiome abundance matrix for 151 individuals (rows) and 60 genera (columns)

y_sCD14 a `numeric` variable with the value of the inflammation parameter sCD14 for each sample

References

Rivera-Pinto et al. (2018) Balances: a new perspective for microbiome analysis. mSystems 3 (4)

shannon *shannon*

Description

Shannon information

Usage

```
shannon(x)
```

Arguments

x	abundance composition (vector)

Value

shannon information

Author(s)

M. Calle - T. Susin

Examples

```
data(HIV, package = "coda4microbiome")

shannon(x_HIV[1,])
```

shannon_effnum *shannon_effnum*

Description

Shannon effective number of variables in a composition

Usage

```
shannon_effnum(x)
```

Arguments

x	abundance composition (vector)

Value

shannon information

Author(s)

M. Calle - T. Susin

Examples

```
data(HIV, package = "coda4microbiome")

shannon_effnum(x_HIV[1,])
```

`shannon_sim` *shannon_sim*

Description

Shannon similarity between two compositions

Usage

```
shannon_sim(x, y)
```

Arguments

x	abundance composition (vector)
y	abundance composition (vector)

Value

shannon similarity (value between 0 and 1)

Author(s)

M. Calle - T. Susin

Examples

```
data(HIV, package = "coda4microbiome")

shannon_sim(x_HIV[1,],x_HIV[2,])
```

taxanames *ecam_filtered*

Description

Microbiome composition at genus level from Early childhood and the microbiome (ECAM) study (Bokulich et al. 2016). Metadata and microbiome data were downloaded from https://github.com/caporaso-lab/longitudinal-notebooks. Filtered data contains information on 42 children and 37 taxa.

Format

The dataset contains three objects:

x_ecam microbiome abundance matrix in long format (873 rows) and 37 genera (columns)

taxanames vector containing the taxonomy of the 37 taxa

metadata matrix with information on the individuals at the observation time

References

Bokulich et al. (2016). Antibiotics, birth mode, and diet shape microbiome maturation during early life. Sci Transl Med 8:343

x *data_survival*

Description

Survival Data simulated from the Crohn's disease original study: 48 taxa and 150 individuals

Format

The dataset contains three objects:

x microbiome abundance matrix for 150 individuals (rows) and 48 genera (columns)

Event a `numeric`, event occurrence. Vector (type: numeric or logical) specifying 0 or FALSE for no event occurrence, and 1 or TRUE for event occurrence.

Event_time a `numeric`, time to event or follow up time for right censored data. Must be a vector (type:numeric) specifying time to event for each sample for right censored data.

References

doi:10.1016/j.chom.2014.02.005

x_Crohn *Crohn*

Description

Microbiome composition at genus level from a Crohn's disease study: 48 taxa and 975 individuals (662 patients with Crohn's disease and 313 controls)

Format

The dataset contains two objects:

x_Crohn microbiome abundance matrix for 975 individuals (rows) and 48 genera (columns)

y_Crohn a `factor`, indicating if the sample corresponds to a case (*CD*) or a control (*no*).

References

doi:10.1016/j.chom.2014.02.005

x_ecam *ecam_filtered*

Description

Microbiome composition at genus level from Early childhood and the microbiome (ECAM) study (Bokulich et al. 2016). Metadata and microbiome data were downloaded from https://github.com/caporaso-lab/longitudinal-notebooks. Filtered data contains information on 42 children and 37 taxa.

Format

The dataset contains three objects:

x_ecam microbiome abundance matrix in long format (873 rows) and 37 genera (columns)

taxanames vector containing the taxonomy of the 37 taxa

metadata matrix with information on the individuals at the observation time

References

Bokulich et al. (2016). Antibiotics, birth mode, and diet shape microbiome maturation during early life. Sci Transl Med 8:343

x_HIV *HIV*

Description

Microbiome abundances (60 taxa and 155 individuals) from an HIV study (Noguera-Julian et al. 2016).

Format

The dataset contains three objects:

x_HIV microbiome abundance matrix for 155 individuals (rows) and 60 genera (columns)

y_HIV a factor, specifying if the individual is HIV positive or (`Pos`) or negative (`Neg`).

MSM_HIV a factor, indicating sexual preferences: `MSM` (*Men who have Sex with Men*) or not (`nonMSM`).

References

doi:10.1016/j.ebiom.2016.01.032

x_sCD14 *sCD14*

Description

Microbiome composition (60 taxa and 151 individuals) and inflammatory parameter sCD14 from an HIV study (Noguera-Julian et al. 2016). A dataset containing the number of counts of 60 different genera in a group of 151 samples (including HIV - infected and non - infected patients).

Format

The dataset contains two objects:

x_sCD14 microbiome abundance matrix for 151 individuals (rows) and 60 genera (columns)

y_sCD14 a `numeric` variable with the value of the inflammation parameter sCD14 for each sample

References

Rivera-Pinto et al. (2018) Balances: a new perspective for microbiome analysis. mSystems 3 (4)

---

y_Crohn *Crohn*

Description

Microbiome composition at genus level from a Crohn's disease study: 48 taxa and 975 individuals (662 patients with Crohn's disease and 313 controls)

Format

The dataset contains two objects:

x_Crohn microbiome abundance matrix for 975 individuals (rows) and 48 genera (columns)

y_Crohn a `factor`, indicating if the sample corresponds to a case (*CD*) or a control (*no*).

References

doi:10.1016/j.chom.2014.02.005

---

y_HIV *HIV*

Description

Microbiome abundances (60 taxa and 155 individuals) from an HIV study (Noguera-Julian et al. 2016).

Format

The dataset contains three objects:

x_HIV microbiome abundance matrix for 155 individuals (rows) and 60 genera (columns)

y_HIV a factor, specifying if the individual is HIV positive or (`Pos`) or negative (`Neg`).

MSM_HIV a factor, indicating sexual preferences: `MSM` (*Men who have Sex with Men*) or not (`nonMSM`).

References

doi:10.1016/j.ebiom.2016.01.032

`y_sCD14` *sCD14*

Description

Microbiome composition (60 taxa and 151 individuals) and inflammatory parameter sCD14 from an HIV study (Noguera-Julian et al. 2016). A dataset containing the number of counts of 60 different genera in a group of 151 samples (including HIV - infected and non - infected patients).

Format

The dataset contains two objects:

x_sCD14 microbiome abundance matrix for 151 individuals (rows) and 60 genera (columns)

y_sCD14 a `numeric` variable with the value of the inflammation parameter sCD14 for each sample

References

Rivera-Pinto et al. (2018) Balances: a new perspective for microbiome analysis. mSystems 3 (4)

Index