Package 'MiMIR'

February 1, 2024

Title Metabolomics-Based Models for Imputing Risk

Version 1.5

Description Provides an intuitive framework for ad-hoc statistical analysis of 1H-NMR metabolomics by Nightingale Health. It allows to easily explore new metabolomics measurements assayed by Nightingale Health, comparing the distributions with a large Consortium (BBMRI-nl); project previously published metabolic scores [<doi:10.1016/j.ebiom.2021.103764>, <doi:10.1161/CIRCGEN.119.002610>, <doi:10.1038/s4146
019-11311-
9>, <doi:10.7554/eLife.63033>, <doi:10.1161/CIRCULATIONAHA.114.013116>, <doi:10.1007/s00125-019-05001-w>]; and calibrate the metabolic surrogate values to a desired dataset.

License GPL-3

Encoding UTF-8

RoxygenNote 7.1.2

Depends R (>= 4.1.0)

Imports caret, DT, foreach, ggplot2, heatmaply, matrixStats, plotly, pROC, purrr, shiny, shinycssloaders, shinyFiles, shinydashboard, shinyjs, shinyWidgets, stats, survival, survminer, dplyr, fs

LazyData true

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

Config/testthat/edition 3

NeedsCompilation no

Author Daniele Bizzarri [aut, cre] (<https://orcid.org/0000-0002-6881-273X>), Marcel Reinders [aut, ths] (<https://orcid.org/0000-0002-1148-1562>), Marian Beekman [aut] (<https://orcid.org/0000-0003-0585-6206>), Pieternella Eline Slagboom [aut, ths] (<https://orcid.org/0000-0002-2875-4723>), Erik van den Akker [aut, ths] (<https://orcid.org/0000-0002-7693-0728>)

Maintainer Daniele Bizzarri <d.bizzarri@lumc.nl>

Repository CRAN

Date/Publication 2024-02-01 08:50:02 UTC

R topics documented:

acc_LOBOV *acc_LOBOV*

Description

Accuracy of the Leave One Biobank Out Validation of the surrogate metabolic-modesl performed in BBMRI-nl

Usage

```
data("acc_LOBOV")
```

Format

An object of class `list` of length 20.

Details

Dataframe containing the accuracy obtained during the Leave One Biobank Out Validation of the surrogate metabolic-modesl in BBMRI-nl.

References

The method is described in: Bizzarri,D. et al. (2022) 1H-NMR metabolomics-based surrogates to impute common clinical risk factors and endpoints. EBioMedicine, 75, 103764, doi:10.1016/j.ebiom.2021.103764

Examples

```
data("acc_LOBOV")
```

Ahola_Olli_betas *T2D-score Betas*

Description

The coefficients used to compute the T2Diabetes score by Ahola Olli.

Usage

```
data("Ahola_Olli_betas")
```

Format

An object of class `data.frame` with 7 rows and 3 columns.

Details

Dataframe containing the abbreviation of the metabolites, the metabolites names and finally the Coefficients to compute the T2Diabetes score

References

Ahola-Olli,A.V. et al. (2019) Circulating metabolites and the risk of type 2 diabetes: a prospective study of 11,896 young adults from four Finnish cohorts. Diabetologia, 62, 2298-2309, doi:10.1007/s00125-019-05001-w

Examples

```
data("Ahola_Olli_betas")
```

`BBMRI_hist` *BBMRI_hist*

Description

Distributions of the Nightingale Health metabolic features in BBMRI-nl

Usage

```
data("BBMRI_hist")
```

Format

An object of class `list` of length 57.

Details

List containing the histograms of the metabolomics-features in BBMRI-nl.

Examples

```
data("BBMRI_hist")
```

BBMRI_hist_plot *multi_hist*

Description

Function to plot the ~60 metabolites used for the metabolomics-based scores and compare them to to their distributions in BBMRI-nl

Usage

```
BBMRI_hist_plot(
  dat,
  x_name,
  color = MiMIR::c21,
  scaled = FALSE,
  datatype = "metabolite",
  main = "Comparison with the metabolites measures in BBMRI"
)
```

Arguments

dat	data.frame or matrix with the metabolites
x_name	string with the name of the selected variable
color	colors selected for all the variables
scaled	logical to z-scale the variables
datatype	a character vector indicating what data type is being plotted
main	title of the plot

Details

This function plots the distribution of a metabolic feature in the uploaded dataset, compared to their distributions in BBMRI-nl. The selection of features available is done following the metabolic scores features.

Value

plotly image with the histogram of the selected variable compared to the distributions in BBMRI-nl

References

The selection of metabolic features available is the one selected by the papers: Deelen,J. et al. (2019) A metabolic profile of all-cause mortality risk identified in an observational study of 44,168 individuals. Nature Communications, 10, 1-8, doi:10.1038/s41467-019-11311-9 Ahola-Olli,A.V. et al. (2019) Circulating metabolites and the risk of type 2 diabetes: a prospective study of 11,896 young adults from four Finnish cohorts. Diabetologia, 62, 2298-2309, doi:10.1007/s00125-019-05001-w Wurtz,P. et al. (2015) Metabolite profiling and cardiovascular event risk:

a prospective study of 3 population-based cohorts. Circulation, 131, 774-785, doi:10.1161/CIRCULATIONAHA.114.013116 Bizzarri,D. et al. (2022) 1H-NMR metabolomics-based surrogates to impute common clinical risk factors and endpoints. EBioMedicine, 75, 103764, doi:10.1016/j.ebiom.2021.103764 van den Akker Erik B. et al. (2020) Metabolic Age Based on the BBMRI-NL 1H-NMR Metabolomics Repository as Biomarker of Age-related Disease. Circulation: Genomic and Precision Medicine, 13, 541-547, doi:10.1161/CIRCGEN.119.002610

Examples

```
library(plotly)
library(MiMIR)

#load the metabolites dataset
metabolic_measures <- synthetic_metabolic_dataset

BBMRI_hist_plot(metabolic_measures, x_name="alb", scaled=TRUE)
```

`BBMRI_hist_scaled` *BBMRI_hist_scaled*

Description

Z-scaled distributions of the Nightingale Health metabolic features in BBMRI-nl

Usage

```
data("BBMRI_hist_scaled")
```

Format

An object of class `list` of length 57.

Details

List containing the histograms of the scaled metabolomics-features in BBMRI-nl.

Examples

```
data("BBMRI_hist_scaled")
```

`binarize_all_pheno` *binarize_all_pheno*

Description

Helper function created to binarize the phenotypes used to calculate the metabolomics based surrogate made by Bizzarri et al.

Usage

```
binarize_all_pheno(data)
```

Arguments

`data`	phenotypes data.frame containing some of the following variables (with the same namenclature): "sex","diabetes", "lipidmed", "blood_pressure_lowering_med", "current_smoking", "metabolic_syndrome", "alcohol_consumption", "age","BMI", "ln_hscrp","waist_circumference", "weight","height", "triglycerides", "ldl_chol", "hdlchol", "totchol", "eGFR","wbc","hgb"

Details

Bizzarri et al. built multivariate models,using 56 metabolic features quantified by Nightingale, to predict the 19 binary characteristics of an individual. The binary variables are: sex, diabetes status, metabolic syndrome status, lipid medication usage, blood pressure lowering medication, current smoking, alcohol consumption, high age, middle age, low age, high hsCRP, high triglycerides, high ldl cholesterol, high total cholesterol, low hdl cholesterol, low eGFR, low white blood cells, low hemoglobin levels.

Value

The phenotypic variables binarized following the thresholds in in the metabolomics surrogates made by by Bizzarri et al.

References

This function was made to binarize the variables following the same rules indicated in the article: Bizzarri,D. et al. (2022) 1H-NMR metabolomics-based surrogates to impute common clinical risk factors and endpoints. EBioMedicine, 75, 103764, doi:10.1016/j.ebiom.2021.103764

See Also

pheno_barplots

Examples

```
library(MiMIR)

#load the phenotypes dataset
phenotypes <- synthetic_phenotypic_dataset
#Calculate BMI, LDL cholesterol and eGFR
binarized_phenotypes<-binarize_all_pheno(phenotypes)
```

`BMI_LDL_eGFR` *BMI_LDL_eGFR*

Description

#' Function created to calculate: 1) BMI using height and weight; 2) LDL cholesterol using HDL cholesterol, triglycerides, totchol; 3) eGFR creatinine levels, sex and age.

Usage

```
BMI_LDL_eGFR(phenotypes, metabo_measures)
```

Arguments

`phenotypes`	data.frame containing height and weight, HDL cholesterol, triglycerides, totchol, sex and age
`metabo_measures`	numeric data-frame with Nightingale metabolomics quantifications containing creatinine levels (crea)

Value

phenotypes data.frame with the addition of BMI, LDL cholesterol and eGFR

References

This function is constructed to calculate BMI, LDL cholesterol and eGFR as in the following papers:

BMI: Flint AJ, Rexrode KM, Hu FB, Glynn RJ, Caspard H, Manson JE et al. Body mass index, waist circumference, and risk of coronary heart disease: a prospective study among men and women. Obes Res Clin Pract 2010; 4: e171-e181, doi:10.1016/j.orcp.2010.01.001

LDL-cholesterol: Friedewald WT, Levy RI, Fredrickson DS. Estimation of the Concentration of Low-Density Lipoprotein Cholesterol in Plasma, Without Use of the Preparative Ultracentrifuge. Clin Chem 1972; 18: 499-502, <doi.org/10.1093/clinchem/18.6.499>

eGFR: Carrero Juan Jesus, Andersson Franko Mikael, Obergfell Achim, Gabrielsen Anders, Jernberg Tomas. hsCRP Level and the Risk of Death or Recurrent Cardiovascular Events in Patients With Myocardial Infarction: a Healthcare-Based Study. J Am Heart Assoc 2019; 8: e012638, <doi: 10.1161/JAHA.119.012638>

Examples

```
library(MiMIR)

#load the dataset
metabolic_measures <- synthetic_metabolic_dataset
phenotypes <- synthetic_phenotypic_dataset
#Calculate BMI, LDL cholesterol and eGFR
phenotypes<-BMI_LDL_eGFR(phenotypes, metabolic_measures)
```

---

c21 *c21*

Description

Colors attributed to each metabolomics-based model in MiMIR

Usage

```
data("c21")
```

Format

An object of class `character` of length 21.

Examples

```
data("c21")
```

---

calculate_surrogate_scores *calculate_surrogate_scores*

Description

Function to compute the surrogate scores by Bizzarri et al. from the Nightingale metabolomics matrix

Usage

```
calculate_surrogate_scores(
  met,
  pheno,
  PARAM_surrogates,
  bin_names = c("sex", "diabetes"),
  Nmax_miss = 1,
  Nmax_zero = 1,
  post = TRUE,
  roc = FALSE,
  quiet = FALSE
)
```

Arguments

met	numeric data-frame with Nightingale-metabolomics
pheno	phenotypic data.frame including this clinical variables (with the same nomenclature): "sex","diabetes", "lipidmed", "blood_pressure_lowering_med", "current_smoking", "metabolic_syndrome", "alcohol_consumption", "age","BMI", "ln_hscrp","waist_circumference", "weight","height", "triglycerides", "ldl_chol", "hdlchol", "totchol", "eGFR","wbc","hgb"
PARAM_surrogates	list containing the parameters to compute the metabolomics-based surrogates
bin_names	vector of strings containing the names of the binary variables
Nmax_miss	numeric value indicating the maximum number of missing values allowed per sample (Number suggested=1)
Nmax_zero	numeric value indicating the maximum number of zeros allowed per sample (Number suggested=1)
post	logical to indicate if the function should calculate the posterior probabilities
roc	logical to plot ROC curves for the metabolomics surrogate (available only for the phenotypes included)
quiet	logical to suppress the messages in the console

Details

Bizzarri et al. built multivariate models,using 56 metabolic features quantified by Nightingale, to predict the 19 binary characteristics of an individual. The binary variables are: sex, diabetes status, metabolic syndrome status, lipid medication usage, blood pressure lowering medication, current smoking, alcohol consumption, high age, middle age, low age, high hsCRP, high triglycerides, high ldl cholesterol, high total cholesterol, low hdl cholesterol, low eGFR, low white blood cells, low hemoglobin levels.

Value

if pheno is not available: list with the surrogates and the Nightingale metabolomics matrix after QC. if pheno is available: list with the surrogates, ROC curves, phenotypes, binarized phenotypes and the Nightingale metabolomics matrix after QC,

References

This function was made to vidualize the binarized variables calculated following the rules indicated in the article: Bizzarri,D. et al. (2022) 1H-NMR metabolomics-based surrogates to impute common clinical risk factors and endpoints. EBioMedicine, 75, 103764, doi:10.1016/j.ebiom.2021.103764

See Also

QCprep_surrogates

Examples

```
require(MiMIR)
require(foreach)
require(pROC)
require(foreach)

#load dataset
m <- synthetic_metabolic_dataset
p <- synthetic_phenotypic_dataset
#Apply the surrogates
sur<-calculate_surrogate_scores(met=m,pheno=p,MiMIR::PARAM_surrogates,bin_names=c("sex","diabetes"))
```

`comp.CVD_score` *comp.CVD_score*

Description

Function to compute CVD-score made by Peter Wurtz et al. made by Deelen et al. on Nightingale metabolomics data-set.

Usage

```
comp.CVD_score(met, phen, betas, quiet = FALSE)
```

Arguments

`met`	numeric data-frame with Nightingale-metabolomics
`phen`	data-frame containing phenotypic information of the samples (specifically: sex, systolic_blood_pressure, current_smoking, diabetes, blood_pressure_lowering_med, lipidmed, totchol, and hdlchol)
`betas`	The betas of the linear regression composing the CVD-score
`quiet`	logical to suppress the messages in the console

Value

data-frame containing the value of the CVD-score on the uploaded data-set

References

This function is constructed to be able to apply the CVD-score as described in: Wurtz,P. et al. (2015) Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. Circulation, 131, 774-785, doi:10.1161/CIRCULATIONAHA.114.013116

See Also

prep_met_for_scores, CVD_score_betas, comp.T2D_Ahola_Olli, comp.mort_score

Examples

```
library(MiMIR)

#load the dataset
met <- synthetic_metabolic_dataset
phen<-synthetic_phenotypic_dataset
#Prepare the metabolic features fo the mortality score
CVDscore<-comp.CVD_score(met= met, phen=phen, betas=MiMIR::CVD_score_betas, quiet=TRUE)
```

---

comp.mort_score *comp.mort_score*

Description

Function to compute the mortality score made by Deelen et al. on Nightingale metabolomics dataset.

Usage

```
comp.mort_score(dat, betas = mort_betas, quiet = FALSE)
```

Arguments

dat	numeric data-frame with Nightingale-metabolomics
betas	data.frame containing the coefficients used for the regression of the mortality score
quiet	logical to suppress the messages in the console

Details

This multivariate model predicts all-cause mortality at 5 or 10 years better than clinical variables normally associated with mortality. It is constituted of 14 metabolic features quantified by Nightingale Health. It was originally trained using a stepwise Cox regression analysis in a meta-analysis on 12 cohorts composed by 44,168 individuals.

Value

data-frame containing the value of the mortality score on the uploaded data-set

References

This function is constructed to be able to apply the mortality score as described in: Deelen,J. et al. (2019) A metabolic profile of all-cause mortality risk identified in an observational study of 44,168 individuals. Nature Communications, 10, 1-8, doi:10.1038/s41467-019-11311-9

See Also

prep_met_for_scores, mort_betas, comp.T2D_Ahola_Olli, comp.CVD_score

Examples

```
library(MiMIR)

#load the Nightignale metabolomics dataset
metabolic_measures <- synthetic_metabolic_dataset
#Prepare the metabolic features fo the mortality score
mortScore<-comp.mort_score(metabolic_measures,quiet=TRUE)
```

---

comp.T2D_Ahola_Olli *comp.T2D_Ahola_Olli*

Description

Function to compute the T2D score made by Ahola Olli et al. on Nightingale metabolomics data-set.

Usage

```
comp.T2D_Ahola_Olli(met, phen, betas, quiet = FALSE)
```

Arguments

met	numeric data-frame with Nightingale-metabolomics
phen	data-frame containing phenotypic information of the samples (in particular: sex, age, BMI and the clinically measured glucose)
betas	The betas of the linear regression composing the T2D-score
quiet	logical to suppress the messages in the console

Details

This metabolomics-based score is associated with incident Type 2 Diabetes, made by Ahola-Olli et al. It is constructed using phe, l_vldl_ce_percentage and l_hdl_fc quantified by Nightingale Health, and some phenotypic information: sex, age, BMI, fasting glucose. It was trained using a stepwise logistic regression on 3 cohorts.

Value

data-frame containing the value of the T2D-score on the uploaded data-set

References

This function is constructed to be able to apply the T2D-score as described in: Ahola-Olli,A.V. et al. (2019) Circulating metabolites and the risk of type 2 diabetes: a prospective study of 11,896 young adults from four Finnish cohorts. Diabetologia, 62, 2298-2309, doi:10.1007/s00125-019-05001-w

See Also

prep_met_for_scores, Ahola_Olli_betas, comp.mort_score, comp.CVD_score

Examples

```
library(MiMIR)

#load the dataset
met <- synthetic_metabolic_dataset
phen<-synthetic_phenotypic_dataset
#Prepare the metabolic features fo the mortality score
T2Dscore<-comp.T2D_Ahola_Olli(met= met, phen=phen,betas=MiMIR::Ahola_Olli_betas, quiet=TRUE)
```

---

comp_covid_score *comp_covid_score*

Description

Function to compute the COVID severity score made by Nightingale Health UK Biobank Initiative et al. on Nightingale metabolomics data-set.

Usage

```
comp_covid_score(dat, betas = MiMIR::covid_betas, quiet = FALSE)
```

Arguments

dat	numeric data-frame with Nightingale-metabolomics
betas	data.frame containing the coefficients used for the regression of the COVID-score
quiet	logical to suppress the messages in the console

Details

Multivariate model predicting the risk of severe COVID-19 infection. It is based on 37 metabolic features and trained using LASSO regression on 52,573 samples from the UK-biobanks.

Value

data-frame containing the value of the COVID-score on the uploaded data-set

References

This function is constructed to be able to apply the COVID-score as described in: Nightingale Health UK Biobank Initiative et al. (2021) Metabolic biomarker profiling for identification of susceptibility to severe pneumonia and COVID-19 in the general population. eLife, 10, e63033, doi:10.7554/eLife.63033

See Also

prep_data_COVID_score, covid_betas, comp.mort_score

Examples

```
library(MiMIR)

#load the Nightignale metabolomics dataset
metabolic_measures <- synthetic_metabolic_dataset

#Compute the mortality score
mortScore<-comp_covid_score(dat=metabolic_measures, quiet=TRUE)
```

`cor_assoc` *cor_assoc*

Description

Function to calulate the correlation between 2 matrices

Usage

```
cor_assoc(dat1, dat2, feat1, feat2, method = "pearson", quiet = FALSE)
```

Arguments

`dat1`	matrix 1
`dat2`	matrix 2
`feat1`	vector of strings with the names of the selected variables in dat
`feat2`	vector if strings with the names of the selected variables in dat2
`method`	indicates which methods of the correlation to use
`quiet`	logical to suppress the messages in the console

Value

correlations of the selected variables in the 2 martrices

See Also

plot_corply

Examples

```
library(stats)

#load the dataset
m <- as.matrix(synthetic_metabolic_dataset)

#Compute the pearson correlation of all the variables in the data.frame metabolic_measures
cors<-cor_assoc(m, m, MiMIR::metabolites_subsets$MET63,MiMIR::metabolites_subsets$MET63)
```

---

covid_betas *COVID-score betas*

Description

The coefficients used to compute the COVID score by Nightingale Health UK Biobank Initiative et al.

Usage

```
data("covid_betas")
```

Format

An object of class `data.frame` with 25 rows and 3 columns.

Details

Dataframe containing the abbreviation of the metabolites, the metabolites names and finally the Coefficients to compute the COVID score

References

Nightingale Health UK Biobank Initiative et al. (2021) Metabolic biomarker profiling for identification of susceptibility to severe pneumonia and COVID-19 in the general population. eLife, 10, e63033, doi:10.7554/eLife.63033

Examples

```
data("covid_betas")
```

CVD_score_betas *CVD-score betas*

Description

The coefficients used to compute the CVD score by Wurtz et al.

Usage

```
data("CVD_score_betas")
```

Format

An object of class `data.frame` with 12 rows and 3 columns.

Details

Dataframe containing the abbreviation of the metabolites, the metabolites names and finally the Coefficients to compute the COVID score

References

Wurtz,P. et al. (2015) Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. Circulation, 131, 774-785, doi:10.1161/CIRCULATIONAHA.114.013116

Examples

```
data("CVD_score_betas")
```

find_BBMRI_names *find_BBMRI_names*

Description

Function to translate Nightingale metabolomics alternative metabolite names to the ones used in BBMRI-nl

Usage

```
find_BBMRI_names(names)
```

Arguments

names	vector of strings with the metabolic features names to be translated

Value

data.frame with the uploaded metabolites names on the first column and the BBMRI names on the second column.

References

This is a function originally created for the package ggforestplot and modified ad hoc for our package (https://nightingalehealth.github.io/ggforestplot/articles/index.html).

Examples

```
library(MiMIR)
library(purrr)

#load the Nightignale metabolomics dataset
metabolic_measures <- synthetic_metabolic_dataset
#Find the metabolites names used in BBMRI-nl
nam<-find_BBMRI_names(colnames(metabolic_measures))
```

`hist_plots` *hist_plots*

Description

#' Function to plot the histograms for all the variables in dat

Usage

```
hist_plots(
  dat,
  x_name,
  color = MiMIR::c21,
  scaled = FALSE,
  datatype = "metabolic score",
  main = "Predictors Distributions"
)
```

Arguments

`dat`	data.frame or matrix with the variables to plot
`x_name`	string with the names of the selected variables in dat
`color`	colors selected for all the variables
`scaled`	logical to z-scale the variables
`datatype`	a character vector indicating what data type is beeing plotted
`main`	title of the plot

Value

plotly image with the histograms of the selected variables

Examples

```
require(MiMIR)
require(plotly)
require(matrixStats)
#load the metabolites dataset
m <- synthetic_metabolic_dataset

#Apply a surrogate models and plot the ROC curve
surrogates<-calculate_surrogate_scores(m, PARAM_surrogates=MiMIR::PARAM_surrogates, roc=FALSE)
#Plot the histogram of the surrogate sex values scaled
hist_plots(surrogates$surrogates, x_name="s_sex", scaled=TRUE)
```

`hist_plots_mortality` *hist_plots_mortality*

Description

#' Function to plot the histogram of the mortality score separated for different age ranges as a plotly image

Usage

```
hist_plots_mortality(mort_score, phenotypes)
```

Arguments

`mort_score`	data.frame containing the mortality score
`phenotypes`	data.frame containing age

Value

plotly image with the histogram of the mortality score separated in 3 age ranges

Examples

```
library(MiMIR)
library(plotly)
#' #load the dataset
metabolic_measures <- synthetic_metabolic_dataset
phenotypes <- synthetic_phenotypic_dataset

#Compute the mortality score
mortScore<-comp.mort_score(metabolic_measures,quiet=TRUE)
#Plot the mortality score histogram at different ages
```

```
hist_plots_mortality(mortScore, phenotypes)
```

---

kapmeier_scores *kapmeier_scores*

Description

#' Function that creates a Kaplan Meier comparing first and last tertile of a metabolic score

Usage

```
kapmeier_scores(predictors, pheno, score, Eventname = "Event")
```

Arguments

predictors	The data.frame containing the predictors
pheno	The data.frame containing the phenotypes
score	a character string indicating which predictor to use
Eventname	a character string with the name of the event to print on the plot

Value

plotly with a Kaplan Meier comparing first and last tertile of a metabolic score

Examples

```
require(MiMIR)
require(plotly)
require(survminer)
require(ggfortify)
require(ggplot2)

#load the dataset
metabolic_measures <- synthetic_metabolic_dataset
phenotypes <- synthetic_phenotypic_dataset

#Compute the mortality score
mortScore<-comp.mort_score(metabolic_measures,quiet=TRUE)

#Plot a Kaplan Meier
kapmeier_scores(predictors=mortScore, pheno=phenotypes, score="mortScore")
```

LOBOV_accuracies *LOBOV_accuracies*

Description

Function created to visualize the accuracies in the current dataset compared to the accuracies in the Leave One Biobank Out Validation in Bizzarri et al.

Usage

```
LOBOV_accuracies(surrogates, bin_phenotypes, bin_pheno_available, acc_LOBOV)
```

Arguments

surrogates	numeric data.frame containing the surrogate values by Bizzarri et al.
bin_phenotypes	numeric data.frame with the binarized phenotypes output of binarize_all_pheno
bin_pheno_available	vector of strings with the available phenotypes
acc_LOBOV	accuracy of LOBOV calculated in Bizzarri et al.

Details

Comparison of the AUCs of the surrogates in the updated dataset and the results of the Leave One Biobank Out Validation made in BBMRI-nl.

Value

Boxplot with the accuracies of the LOBOV

References

This function was made to vidualize the binarized variables calculated following the rules indicated in the article: Bizzarri,D. et al. (2022) 1H-NMR metabolomics-based surrogates to impute common clinical risk factors and endpoints. EBioMedicine, 75, 103764, doi:10.1016/j.ebiom.2021.103764

Examples

```
require(pROC)
require(plotly)
require(MiMIR)
require(foreach)
require(ggplot2)

#load the dataset
m <- synthetic_metabolic_dataset
p<- synthetic_phenotypic_dataset
```

```
#Calculating the binarized surrogates
b_p<-binarize_all_pheno(p)
#Apply a surrogate models and plot the ROC curve
sur<-calculate_surrogate_scores(m, p, MiMIR::PARAM_surrogates, bin_names=colnames(b_p))
p_avail<-colnames(b_p)[c(1:5)]
LOBOV_accuracies(sur$surrogates, b_p, p_avail, MiMIR::acc_LOBOV)
```

`metabolites_subsets` *metabolomics feature subsets*

Description

List containing all the subset of the metabolomics-based features used for our models

Usage

```
data("metabolites_subsets")
```

Format

An object of class `list` of length 8.

References

The selection of metabolic features available is the one selected by the papers: Deelen,J. et al. (2019) A metabolic profile of all-cause mortality risk identified in an observational study of 44,168 individuals. Nature Communications, 10, 1-8, doi:10.1038/s41467-019-11311-9 Ahola-Olli,A.V. et al. (2019) Circulating metabolites and the risk of type 2 diabetes: a prospective study of 11,896 young adults from four Finnish cohorts. Diabetologia, 62, 2298-2309, doi:10.1007/s00125-019-05001-w Wurtz,P. et al. (2015) Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. Circulation, 131, 774-785, doi:10.1161/CIRCULATIONAHA.114.013116 Bizzarri,D. et al. (2022) 1H-NMR metabolomics-based surrogates to impute common clinical risk factors and endpoints. EBioMedicine, 75, 103764, doi:10.1016/j.ebiom.2021.103764 van den Akker Erik B. et al. (2020) Metabolic Age Based on the BBMRI-NL 1H-NMR Metabolomics Repository as Biomarker of Age-related Disease. Circulation: Genomic and Precision Medicine, 13, 541-547, doi:10.1161/CIRCGEN.119.002610

Examples

```
data("metabolites_subsets")
```

MetaboWAS *MetaboWAS*

Description

Function to calculate a Metabolome Wide Association study

Usage

```
MetaboWAS(met, pheno, test_variable, covariates, img = TRUE, adj_method = "BH")
```

Arguments

met	numeric data.frame with the metabolomics features
pheno	data.frame containing the phenotype of interest
test_variable	string vector with the name of the phenotype of interest
covariates	string vector with the name of the variables to be added as a covariate
img	logical indicating if the function should plot a Manhattan plot
adj_method	multiple testing correction method

Details

This is a function to compute linear associations individually for each variable in the first data.frame with the test variable and corrected for the selected covariates. This function to computes linear regression modelindividually for each variable in the first data.frame with the test variable and adjusted for potential confounders. False Discovery Rate (FDR) is applied to account for multiple testing correction. The user has the faculty to select the test variable and the potential covariates within the pool of variables in the phenotypic file input. The results of the associations are reported in a Manhattan plot

The p-value of the association is then corrected using Benjamini Hochberg. Finally we use plotly to plot a Manhattan Plot, which reports on the x-axis the list of metabolites reported in the Nightingale Health, divided in groups, and on the y-axis the -log (adjusted p-value).

Value

res= the results of the MetaboWAS, manhplot= the Manhattan plot made with plotly, N_hits= the number of significant hits

References

This method is also described and used in: Bizzarri,D. et al. (2022) 1H-NMR metabolomics-based surrogates to impute common clinical risk factors and endpoints. EBioMedicine, 75, 103764, doi:10.1016/j.ebiom.2021.103764

Examples

```
require(MiMIR)
require(plotly)
require(ggplot2)

#' #load the dataset
metabolic_measures <- synthetic_metabolic_dataset
phenotypes <- synthetic_phenotypic_dataset

#Computing a MetaboWAS for age corrected by sex
MetaboWAS(met=metabolic_measures, pheno=phenotypes, test_variable="age", covariates= "sex")
```

---

metabo_names_translator

metabolomics feature nomenclatures

Description

Translator of the names of the metabolomics-features to the ones used in BBMRI-nl

Usage

```
data("metabo_names_translator")
```

Format

An object of class `data.frame` with 228 rows and 9 columns.

References

This is a list originally created for the package ggforestplot and modified ad-hoc for our package (https://nightingalehealth.github.io/ggforestplot/articles/index.html).

Examples

```
data("metabo_names_translator")
```

mort_betas *Mortality score betas*

Description

The coefficients used to compute the mortality score by Deelen et al.

Usage

```
data("mort_betas")
```

Format

An object of class `data.frame` with 14 rows and 3 columns.

Details

Dataframe containing the abbreviation of the metabolites, the metabolites names and finally the Coefficients to compute the mortality score

References

Deelen,J. et al. (2019) A metabolic profile of all-cause mortality risk identified in an observational study of 44,168 individuals. Nature Communications, 10, 1-8, doi:10.1038/s41467-019-11311-9

Examples

```
data("mort_betas")
```

multi_hist *multi_hist*

Description

#' Function to plot the histograms for all the variables in dat

Usage

```
multi_hist(dat, color = MiMIR::c21, scaled = FALSE)
```

Arguments

`dat`	data.frame or matrix with the variables to plot
`color`	colors selected for all the variables
`scaled`	logical to z-scale the variables

Value

plotly image with the histograms for all the variables in dat

Examples

```
library(plotly)
library(MiMIR)

#load the dataset
metabolic_measures <- synthetic_metabolic_dataset

multi_hist(metabolic_measures[,MiMIR::metabolites_subsets$MET14], scaled=T)
```

PARAM_metaboAge *PARAMETERS MetaboAge*

Description

The coefficients used to compute the MetaboAge by van den Akker et al.

Usage

```
data("PARAM_metaboAge")
```

Format

An object of class `list` of length 8.

Details

List containing all the information to pre-process and compute the MetaboAge.

References

van den Akker Erik B. et al. (2020) Metabolic Age Based on the BBMRI-NL 1H-NMR Metabolomics Repository as Biomarker of Age-related Disease. Circulation: Genomic and Precision Medicine, 13, 541-547, doi:10.1161/CIRCGEN.119.002610

Examples

```
data("PARAM_metaboAge")
```

PARAM_surrogates *PARAMETERS surrogates*

Description

The coefficients used to compute the metabolomics-based surrogate clinical variables by Bizzarri et al.

Usage

```
data("PARAM_surrogates")
```

Format

An object of class `list` of length 6.

Details

List containing all the information to pre-process and compute the surrogate clinical variables.

References

Bizzarri,D. et al. (2022) 1H-NMR metabolomics-based surrogates to impute common clinical risk factors and endpoints. EBioMedicine, 75, 103764, doi:10.1016/j.ebiom.2021.103764

Examples

```
data("PARAM_surrogates")
```

phenotypes_names *phenotypic features names*

Description

List containing all the subsets of phenotypics variables used in the app

Usage

```
data("phenotypes_names")
```

Format

An object of class `list` of length 5.

Examples

```
data("phenotypes_names")
```

pheno_barplots *pheno_barplots*

Description

#' Function created to binarize the phenotypes used to calculate the metabolomics based surrogate made by Bizzarri et al.

Usage

```
pheno_barplots(bin_phenotypes)
```

Arguments

bin_phenotypes	phenotypes data.frame containing some of the following variables (with the same namenclature): "sex","diabetes", "lipidmed", "blood_pressure_lowering_med", "current_smoking", "metabolic_syndrome", "alcohol_consumption", "age","BMI", "ln_hscrp","waist_circumference", "weight","height", "triglycerides", "ldl_chol", "hdlchol", "totchol", "eGFR","wbc","hgb"

Details

Bizzarri et al. built multivariate models,using 56 metabolic features quantified by Nightingale, to predict the 19 binary characteristics of an individual. The binary variables are: sex, diabetes status, metabolic syndrome status, lipid medication usage, blood pressure lowering medication, current smoking, alcohol consumption, high age, middle age, low age, high hsCRP, high triglycerides, high ldl cholesterol, high total cholesterol, low hdl cholesterol, low eGFR, low white blood cells, low hemoglobin levels.

Value

The phenotypic variables binarized following the thresholds in in the metabolomics surrogates made by by Bizzarri et al.

References

This function was made to vidualize the binarized variables calculated following the rules indicated in the article: Bizzarri,D. et al. (2022) 1H-NMR metabolomics-based surrogates to impute common clinical risk factors and endpoints. EBioMedicine, 75, 103764, doi:10.1016/j.ebiom.2021.103764

See Also

binarize_all_pheno

Examples

```
require(MiMIR)
require(foreach)

#load the phenotypes dataset
phenotypes <- synthetic_phenotypic_dataset

#Calculate BMI, LDL cholesterol and eGFR
binarized_phenotypes<-binarize_all_pheno(phenotypes)
#Plot the variables
pheno_barplots(binarized_phenotypes)
```

`plattCalibration` *plattCalibration*

Description

Function that calculates the Platt Calibrations

Usage

```
plattCalibration(r.calib, p.calib, nbins = 10, pl = FALSE)
```

Arguments

`r.calib`	observed binary phenotype
`p.calib`	predicted probabilities
`nbins`	number of bins to create the plots
`pl`	logical indicating if the function should plot the Reliability diagram and histogram of the calibrations

Details

Many popular machine learning algorithms produce inaccurate predicted probabilities, especially when applied on a dataset different than the training set. Platt (1999) proposed an adjustment, in which the original probabilities are used as a predictor in a single-variable logistic regression to produce more accurate adjusted predicted probabilities. The function will also help the evaluation of the calibration, by plotting: reliability diagrams and distributions of the calibrated and non-calibrated probabilities. The reliability diagrams plots the mean predicted value within a certain range of posterior probabilities, against the fraction of accurately predicted values. Finally, we also report accuracy measures for the calibrations: the ECE, MCE and the Log-Loss of the probabilities before and after calibration.

Value

list with samples, responses, calibrations, ECE, MCE and calibration plots if save==T

References

This is a function originally created for the package in eRic, under the name prCalibrate and modified ad hoc for our purposes (Github)

J. C. Platt, 'Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods', in Advances in Large Margin Classifiers, 1999, pp. 61-74.

Examples

```
library(stats)
library(plotly)

#load the dataset
met <- synthetic_metabolic_dataset
phen <- synthetic_phenotypic_dataset

#Calculating the binarized surrogates
b_phen<-binarize_all_pheno(phen)
#Apply a surrogate models and plot the ROC curve
surr<-calculate_surrogate_scores(met, phen,MiMIR::PARAM_surrogates, bin_names=colnames(b_phen))
#Calibration of the surrogate sex
real_data<-as.numeric(b_phen$sex)
pred_data<-surr$surrogates[,"s_sex"]
plattCalibration(r.calib=real_data, p.calib=pred_data, nbins = 10, pl=TRUE)
```

---

plot_corply *plot_corply*

Description

Function creating plottig the correlation between 2 datasets, dat1 x dat2 on basis of (partial) correlations

Usage

```
plot_corply(
  res,
  main = NULL,
  zlim = NULL,
  reorder.x = FALSE,
  reorder.y = reorder.x,
  resort_on_p = FALSE,
  abs = FALSE,
  cor.abs = FALSE,
  reorder_dend = FALSE
)
```

Arguments

res	associations obtained with cor.assoc
main	title of the plot
zlim	max association to plot
reorder.x	logical indicating if the function should reorder the x axis based on clustering
reorder.y	logical indicating if the function should reorder the y axis based on clustering
resort_on_p	logical indicating if the function should reorder x and y axis based on the pvalues of the associations
abs	logical indicating if the function should reorder based the absolute values
cor.abs	logical indicating if the function should reorder the plot base on the absolute values
reorder_dend	Tlogical indicating if the function should reorder the plot based on dendrogram

Value

heatmap with the results of cor.assoc

See Also

cor_assoc

Examples

```
library(stats)

#load the dataset
m <- as.matrix(synthetic_metabolic_dataset)

#Compute the pearson correlation of all the variables in the data.frame metabolic_measures
cors<-cor_assoc(m, m, MiMIR::metabolites_subsets$MET63,MiMIR::metabolites_subsets$MET63)
#Plot the correlations
plot_corply(cors, main="Correlations metabolites")
```

---

plot_na_heatmap *plot_na_heatmap*

Description

Function plotting information about missing & zero values on the indicated matrix.

Usage

```
plot_na_heatmap(dat)
```

Arguments

dat	The matrix or data.frame

Details

This heatmap indicates the available values in grey and missing or zeros in white. On the sides two bar plots on the sides, one showing the missingn or zero values per row and another to show the missing or zeroes per column.

Value

Plot with a central heatmap and two histogram on the sides

Examples

```
library(graphics)
library(MiMIR)

#load the metabolites dataset
metabolic_measures <- synthetic_metabolic_dataset
#Plot the missing values in the metabolomics matrix
plot_na_heatmap(metabolic_measures)
```

---

prep_data_COVID_score *prep_data_COVID_score*

Description

Helper function to pre-process the Nightingale Health metabolomics data-set before applying the COVID score.

Usage

```
prep_data_COVID_score(
  dat,
  featID = c("gp", "dha", "crea", "mufa", "apob_apoa1", "tyr", "ile", "sfa_fa", "glc",
   "lac", "faw6_faw3", "phe", "serum_c", "faw6_fa", "ala", "pufa", "glycine", "his",
    "pufa_fa", "val", "leu", "alb", "faw3", "ldl_c", "serum_tg"),
  quiet = FALSE
)
```

Arguments

dat	numeric data-frame with Nightingale-metabolomics
featID	vector of strings with the names of metabolic features included in the COVID-score
quiet	logical to suppress the messages in the console

Value

The Nightingale-metabolomics data-frame after pre-processing (checked for zeros, z-scaled and log-transformed) according to what has been done by the authors of the original papers.

References

This function is constructed to be able to follow the pre-processing steps described in: Nightingale Health UK Biobank Initiative et al. (2021) Metabolic biomarker profiling for identification of susceptibility to severe pneumonia and COVID-19 in the general population. eLife, 10, e63033, doi:10.7554/eLife.63033

See Also

prep_met_for_scores, covid_betas, comp_covid_score

Examples

```
require(MiMIR)
require(matrixStats)

#load the Nightignale metabolomics dataset
metabolic_measures <- synthetic_metabolic_dataset
#Prepare the metabolic features fo the mortality score
prepped_met <- prep_data_COVID_score(dat=metabolic_measures)
```

---

prep_met_for_scores *prep_met_for_scores*

Description

Helper function to pre-process the Nightingale Health metabolomics data-set before applying the mortality, Type-2-diabetes and CVD scores.

Usage

```
prep_met_for_scores(dat, featID, plusone = FALSE, quiet = FALSE)
```

Arguments

dat	numeric data-frame with Nightingale-metabolomics
featID	vector of strings with the names of metabolic features included in the score selected
plusone	logical to determine if a value of 1.0 should be added to all metabolic features (TRUE) or only to the ones featuring zeros before log-transforming (FALSE)
quiet	logical to suppress the messages in the console

Value

The Nightingale-metabolomics data-frame after pre-processing (checked for zeros, zscale and log-transformed) according to what has been done by the authors of the original papers.

References

This function is constructed to be able to follow the pre-processing steps described in: Deelen,J. et al. (2019) A metabolic profile of all-cause mortality risk identified in an observational study of 44,168 individuals. Nature Communications, 10, 1-8, doi:10.1038/s41467-019-11311-9.

Ahola-Olli,A.V. et al. (2019) Circulating metabolites and the risk of type 2 diabetes: a prospective study of 11,896 young adults from four Finnish cohorts. Diabetologia, 62, 2298-2309, doi:10.1007/s00125-019-05001-w

Wurtz,P. et al. (2015) Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. Circulation, 131, 774-785, doi:10.1161/CIRCULATIONAHA.114.013116

See Also

comp.mort_score, mort_betas, comp.T2D_Ahola_Olli, comp.CVD_score

Examples

```
library(MiMIR)

#load the Nightingale metabolomics dataset
metabolic_measures <- synthetic_metabolic_dataset
#Prepare the metabolic features fo the mortality score
prepped_met <- prep_met_for_scores(metabolic_measures,featID=MiMIR::mort_betas$Abbreviation)
```

`QCprep` *QCprep*

Description

Helper function to pre-process the Nightingale Health metabolomics data-set before applying the MetaboAge score by van den Akker et al.

Usage

```
QCprep(mat, PARAM_metaboAge, quiet = TRUE, Nmax_zero = 1, Nmax_miss = 1)
```

Arguments

mat	numeric data-frame NH-metabolomics matrix.
PARAM_metaboAge	list containing all the parameters to compute the metaboAge (metabolic features list,BBMRI-nl means and SDs of the metabolic features, and coefficients)
quiet	logical to suppress the messages in the console
Nmax_zero	numberic value indicating the maximum number of zeros allowed per sample (Number suggested=1)
Nmax_miss	numberic value indicating the maximum number of missing values allowed per sample (Number suggested=1)

Value

Nightingale-metabolomics data-frame after pre-processing (checked for zeros, missing values, samples>5SD from the BBMRI-mean, imputing the missing values and z-scaled)

References

This function is constructed to be able to follow the pre-processing steps described in: van den Akker Erik B. et al. (2020) Metabolic Age Based on the BBMRI-NL 1H-NMR Metabolomics Repository as Biomarker of Age-related Disease. Circulation: Genomic and Precision Medicine, 13, 541-547, doi:10.1161/CIRCULATIONAHA.114.013116

See Also

apply.fit

Examples

```
library(MiMIR)

#load the Nightignale metabolomics dataset
metabolic_measures <- synthetic_metabolic_dataset

#Pre-process the metabolic features
prepped_met<-QCprep(as.matrix(metabolic_measures[,metabolites_subsets$MET63]), PARAM_metaboAge)
```

QCprep_surrogates *QCprep_surrogates*

Description

Helper function to pre-process the Nightingale Health metabolomics data-set before applying metabolomics-based surrogates by Bizzarri et al.

Usage

```
QCprep_surrogates(
  mat,
  PARAM_surrogates,
  Nmax_miss = 1,
  Nmax_zero = 1,
  quiet = FALSE
)
```

Arguments

mat	numeric data-frame Nightingale metabolomics matrix.
PARAM_surrogates	is a list holding the parameters to compute the surrogates
Nmax_miss	numeric value indicating the maximum number of missing values allowed per sample (Number suggested=1)
Nmax_zero	numeric value indicating the maximum number of zeros allowed per sample (Number suggested=1)
quiet	logical to suppress the messages in the console

Details

Bizzarri et al. built multivariate models,using 56 metabolic features quantified by Nightingale, to predict the 19 binary characteristics of an individual. The binary variables are: sex, diabetes status, metabolic syndrome status, lipid medication usage, blood pressure lowering medication, current smoking, alcohol consumption, high age, middle age, low age, high hsCRP, high triglycerides, high ldl cholesterol, high total cholesterol, low hdl cholesterol, low eGFR, low white blood cells, low hemoglobin levels.

Value

Nightingale-metabolomics data-frame after pre-processing (checked for zeros, missing values, samples>5SD from the BBMRI-mean, imputing the missing values and z-scaled)

References

This function was made to vidualize the binarized variables calculated following the rules indicated in the article: Bizzarri,D. et al. (2022) 1H-NMR metabolomics-based surrogates to impute common clinical risk factors and endpoints. EBioMedicine, 75, 103764, doi:10.1016/j.ebiom.2021.103764

See Also

binarize_all_pheno

Examples

```
library(MiMIR)

#load the Nightignale metabolomics dataset
metabolic_measures <- synthetic_metabolic_dataset
#Pre-process the metabolic features
prepped_met<-QCprep_surrogates(as.matrix(metabolic_measures), MiMIR::PARAM_surrogates)
```

---

roc_surro *roc_surro*

Description

Function that creates a ROC curve of the selected metabolic surrogates as a plotly image

Usage

```
roc_surro(surrogates, bin_phenotypes, x_name)
```

Arguments

surrogates	numeric data.frame of metabolomics-based surrogate values by Bizzarri et al.
bin_phenotypes	logic data.frame of binarized phenotypes
x_name	vector of strings with the names of the selected binary phenotypes for the roc

Value

plotly image with the ROC curves for one or more selected variables

Examples

```
require(pROC)
require(plotly)
require(foreach)
require(MiMIR)

#load the dataset
met <- synthetic_metabolic_dataset
phen<- synthetic_phenotypic_dataset

#Calculating the binarized surrogates
b_phen<-binarize_all_pheno(phen)
#Apply a surrogate models and plot the ROC curve
surr<-calculate_surrogate_scores(met, phen, MiMIR::PARAM_surrogates, colnames(b_phen))
#Plot the ROC curves
roc_surro(surr$surrogates, b_phen, "sex")
```

roc_surro_subplots *roc_surro_subplots*

Description

Function that plots the ROCs of the surrogates of all the available surrogate models as plotly subplots

Usage

```
roc_surro_subplots(surrogates, bin_phenotypes)
```

Arguments

surrogates	numeric data.frame containing the surrogate values by Bizzarri et al.
bin_phenotypes	numeric data.frame with the binarized phenotypes output of binarize_all_pheno

Value

plotly image with all the ROCs for all the available clinical variables

Examples

```
library(pROC)
library(plotly)
library(MiMIR)

#load the dataset
met <- synthetic_metabolic_dataset
phen<- synthetic_phenotypic_dataset

#Calculating the binarized surrogates
b_phen<-binarize_all_pheno(phen)
#Apply a surrogate models and plot the ROC curve
surr<-calculate_surrogate_scores(met, phen, MiMIR::PARAM_surrogates, colnames(b_phen))

roc_surro_subplots(surr$surrogates, b_phen)
```

scatterplot_predictions *scatterplot_predictions*

Description

Function to visualize a scatter-plot comparing two variables

Usage

```
scatterplot_predictions(x, p, title, xname = "x", yname = "predicted x")
```

Arguments

x	numeric vector
p	second numeric vector
title	string vector with the title
xname	string vector with the name of the variable on the x axis
yname	string vector with the name of the variable on the y axis

Value

plotly image with the scatterplot

Examples

```
library(plotly)
#load the dataset
metabolic_measures <- synthetic_metabolic_dataset
phenotypes <- synthetic_phenotypic_dataset

#Pre-process the metabolic features
prepped_met<-QCprep(as.matrix(metabolic_measures), MiMIR::PARAM_metaboAge)
#Apply the metaboAge
metaboAge<-apply.fit(prepped_met, FIT=PARAM_metaboAge$FIT_COEF)

age<-data.frame(phenotypes$age)
rownames(age)<-rownames(phenotypes)
scatterplot_predictions(age, metaboAge, title="Chronological Age vs MetaboAge")
```

`startApp` *startMiMIR*

Description

Start the application MiMIR.

Usage

```
startApp(launch.browser = TRUE)
```

Arguments

`launch.browser`	TRUE/FALSE

Details

This function starts the R-Shiny tool called MiMIR (Metabolomics-based Models for Imputing Risk), a graphical user interface that provides an intuitive framework for ad-hoc statistical analysis of Nightingale Health's 1H-NMR metabolomics data and allows for the projection and calibration of 24 pre-trained metabolomics-based models, without any pre-required programming knowledge.

Value

Opens application. If `launch.browser=TRUE` in default web browser

References

Deelen,J. et al. (2019) A metabolic profile of all-cause mortality risk identified in an observational study of 44,168 individuals. Nature Communications, 10, 1-8, doi: 10.1038/s41467-019-11311-9. Ahola-Olli,A.V. et al. (2019) Circulating metabolites and the risk of type 2 diabetes: a prospective study of 11,896 young adults from four Finnish cohorts. Diabetologia, 62, 2298-2309, doi: 10.1007/s00125-019-05001-w Wurtz,P. et al. (2015) Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. Circulation, 131, 774-785, doi: 10.1161/CIRCULATIONAHA.114.013116 Bizzarri,D. et al. (2022) 1H-NMR metabolomics-based surrogates to impute common clinical risk factors and endpoints. EBioMedicine, 75, 103764, doi: 10.1016/j.ebiom.2021.103764 van den Akker Erik B. et al. (2020) Metabolic Age Based on the BBMRI-NL 1H-NMR Metabolomics Repository as Biomarker of Age-related Disease. Circulation: Genomic and Precision Medicine, 13, 541-547, doi:10.1161/CIRCGEN.119.002610

synthetic_metabolic_dataset

synthetic metabolomics dataset

Description

Data.frame containing a synthetic dataset of the Nightingale Metabolomics dataset created with the package synthpop from the LLS_PAROFF dataset.

Usage

```
data("synthetic_metabolic_dataset")
```

Format

An object of class `data.frame` with 500 rows and 229 columns.

References

M. Schoenmaker et al., 'Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study', Eur. J. Hum. Genet., vol. 14, no. 1, Art. no. 1, Jan. 2006, doi:10.1038/sj.ejhg.5201508 B. Nowok, G. M. Raab, and C. Dibben, 'synthpop: Bespoke Creation of Synthetic Data in R', J. Stat. Softw., vol. 74, no. 1, Art. no. 1, Oct. 2016, doi:10.18637/jss.v074.i11

Examples

```
data("synthetic_metabolic_dataset")
```

synthetic_phenotypic_dataset

synthetic metabolomics dataset

Description

Data.frame containing a synthetic dataset of phenotypic dataset created with the package synthpop from the LLS_PAROFF dataset.

Usage

```
data("synthetic_metabolic_dataset")
```

Format

An object of class `data.frame` with 500 rows and 24 columns.

References

M. Schoenmaker et al., 'Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study', Eur. J. Hum. Genet., vol. 14, no. 1, Art. no. 1, Jan. 2006, doi:10.1038/sj.ejhg.5201508 B. Nowok, G. M. Raab, and C. Dibben, 'synthpop: Bespoke Creation of Synthetic Data in R', J. Stat. Softw., vol. 74, no. 1, Art. no. 1, Oct. 2016, doi:10.18637/jss.v074.i11

Examples

```
data("synthetic_metabolic_dataset")
```

`ttest_scores` *ttest_scores*

Description

#' Function that creates a boxplot with a continuous variable split using the binary variable

Usage

```
ttest_scores(dat, pred, pheno)
```

Arguments

`dat`	The data.frame containing the 2 variables
`pred`	character indicating the y variable
`pheno`	character indicating the binary variable

Value

plotly boxplot with the continuous variable split using the binary variable

Examples

```
library(MiMIR)
library(plotly)

#load the dataset
metabolic_measures <- synthetic_metabolic_dataset
phenotypes <- synthetic_phenotypic_dataset

#Compute the mortality score
mortScore<-comp.mort_score(metabolic_measures,quiet=TRUE)
dat<-data.frame(predictor=mortScore, pheno=phenotypes$sex)
colnames(dat)<-c("predictor","pheno")
ttest_scores(dat = dat, pred= "mortScore", pheno="sex")
```

ttest_surrogates *ttest_surrogates*

Description

Function that calculates a t-test and a plotly image of the selected surrogates

Usage

```
ttest_surrogates(surrogates, bin_phenotypes)
```

Arguments

surrogates	numeric data.frame containing the surrogate values by Bizzarri et al.
bin_phenotypes	numeric data.frame with the binarized phenotypes output of binarize_all_pheno

Details

Barplot and T-test indicating if the surrogate variables could split accordingly the real value of the binary clinical variables.

Value

plotly image with all the ROCs for all the available clinical variables

Examples

```
require(pROC)
require(plotly)
require(MiMIR)
require(foreach)

#load the dataset
m <- synthetic_metabolic_dataset
p <- synthetic_phenotypic_dataset

#Calculating the binarized surrogates
b_p<-binarize_all_pheno(p)
#Apply a surrogate models and plot the ROC curve
surr<-calculate_surrogate_scores(met=m, pheno=p, MiMIR::PARAM_surrogates, bin_names=colnames(b_p))
ttest_surrogates(surr$surrogates, b_p)
```

Index