Package 'Metabol'

May 16, 2019

Type Package

Title The statistical analysis of metabolomic data

Version 1.0.0

Maintainer Alzbeta Gardlo <alzbeta.gardlo@gmail.com>

Description
The package of preprocessing (quality control treatment, zero imputation) and statistical analysis of metabolomic data sets made by logratio methodology. Also comparison with the other transformations/scalings (log, ln, PQN, no transformation, ln(PQN),Pareto scaling) is possible. In graphical part multivariate (PCA, PLS-DA, VIP plots, OPLS-DA, cluster analysis) and univariate (boxplots, ROC curves, volcano plots) analyses are done.

Depends openxlsx, sm, ropls

Imports compositions,robCompositions,car,pls,dendextend,gplots,rgl, pROC, stats, graphics, grDevices, rrcov, utils

Author
Alzbeta Gardlo, David Friedecky, Karel Hron, Lukas Najdekr, Radana Karlikova, Tomas Adam

License GPL-2

LazyData TRUE

RoxygenNote 6.1.1

BuildManual yes

NeedsCompilation no

R topics documented:

`batchSort` *A toy example of metabolomic data set sorted according to batch order*

Description

A dataset containing the information about 100 metabolites in samples from 14 patients with one specific disease and 10 healthy controls. Quality control samples (QCs) are also included. All samples are sorted according to batch order.

Usage

```
batchSort
```

Format

A data frame with 100 rows (metabolites) and 32 variables (samples of patients, controls and QCs sorted according to batch order)

`bigQC` *Quality control samples (QCs) checking*

Description

Quality control samples (QCs) are checked to data irregularities. It is used for data from untargeted metabolomic analysis.

Usage

```
bigQC(data, name, groupnames)
```

Arguments

`data`	Data table with variables (metabolites) in columns. Samples in rows are sorted according to specific groups.
`name`	A character string or expression indicating a name of data set. It occurs in names of every output.
`groupnames`	A character vector defining specific groups in data. Every string must be specific for each group and they must not overlap.

Details

Values of QCs are evaluated and questionable values for particular variables are denoted. There are two steps of evaluation: 1. QCs with completely higher values than the maximum of data, 2. QCs higher than majority of data.

Up to twenty different groups can be distinguished in data (including QCs).

Value

Boxplots of QCs and the other data groups.

Excel file with the list of questionable variables from two steps of evaluation.

Examples

```
data=metabol
name="Metabolomics"    #name of the project
groupnames=c("Con","Pat","QC")
bigQC(data,name,groupnames)
```

GraphsBox *Boxplots*

Description

Display 5 variants of boxplots in PDF files for every variable (metabolite) in data table separately.

Usage

```
GraphsBox(data, name, groupnames, tsf = "clr", QCs = FALSE,
  newlabs = FALSE, pair = FALSE, long = FALSE)
```

Arguments

data	Data table with variables (metabolites) in columns. Samples in rows are sorted according to specific groups.
name	A character string or expression indicating a name of data set. It occurs in names of every output.
groupnames	A character vector defining specific groups in data. Every string must be specific for each group and they must not overlap.
tsf	Data transformation must be defined by "clr" (default), "log", "log10", "PQN", "lnPQN", "pareto" or "none". See Details.
QCs	logical. If FALSE (default) quality control samples (QCs) are automatically distinguished and skipped.
newlabs	logical. If FALSE (default) standard names are used. If vector of characters then the new names of points in boxplots are drawn.
pair	logical. If TRUE then the paired boxplots are drawn. Deaful is FALSE.
long	logical. Used for long group names. If TRUE, labels of x-axis are written vertically. Deaful is FALSE.

Details

Data transformation: with "clr" clr trasformation is used (see References), with "log" natural logarithm is used, with "log10" decadic logarithm is used, with "pareto" data are only scaled by Pareto scaling, with "PQN" probabilistic quotient normalization is done, with "lnPQN" natural logarithm of PQN transformed data is done, with "none" no tranformation is done.

Up to twenty different groups can be distinguished in data (including QCs).

Value

Boxplot graphs.

Excel file with medians and differences of these medians of all groups in data.

References

Aitchison, J. (1986) The Statistical Analysis of Compositional Data Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). p. 416.

Examples

```
data=metabol
name="Metabolomics"    #name of the project
groupnames=c("Con","Pat","QC")
GraphsBox(data,name,groupnames)
```

GraphsBoxPvalue *Boxplots with p-values*

Description

Display 4 variants of boxplots with p-values from t-test or Wilcoxon test in PDF files for every variable (metabolite) in data table separately.

Usage

```
GraphsBoxPvalue(data, name, groupnames, type = "par", tsf = "clr",
  QCs = FALSE)
```

Arguments

data	Data table with variables (metabolites) in columns. Samples in rows are sorted according to specific groups.
name	A character string or expression indicating a name of data set. It occurs in names of every ouput.
groupnames	A character vector defining specific groups in data. Every string must be specific for each group and they must not overlap.
type	Parametric ("par" - default) or nonparametric ("nonpar") test must be chosen.
tsf	Data transformation must be defined by "clr" (default), "log", "log10", "PQN", "lnPQN", "pareto" or "none". See Details.
QCs	logical. If FALSE (default) quality control samples (QCs) are automatically distinguished and skipped.

Details

Data transformation: with "clr" clr trasformation is used (see References), with "log" natural logarithm is used, with "log10" decadic logarithm is used, with "pareto" data are only scaled by Pareto scaling, with "PQN" probabilistic quotient normalization is done, with "lnPQN" natural logarithm of PQN transformed data is done, with "none" no tranformation is done.

Up to nine different groups can be distinguished in data.

Value

Boxplot graphs with p-values and stars representing p-values.

Excel file with medians and differences of these medians of all groups in data.

References

Aitchison, J. (1986) The Statistical Analysis of Compositional Data Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). p. 416.

Examples

```
data=metabol
name="Metabolomics"    #name of the project
groupnames=c("Con","Pat","QC")
GraphsBoxPvalue(data,name,groupnames)
```

GraphsClust *Cluster analysis*

Description

Display cluster plots and heat maps.

Usage

```
GraphsClust(data, name, groupnames, tsf = "clr", QCs = TRUE)
```

Arguments

data	Data table with variables (metabolites) in columns. Samples in rows are sorted according to specific groups.
name	A character string or expression indicating a name of data set. It occurs in names of every output.
groupnames	A character vector defining specific groups in data. Every string must be specific for each group and they must not overlap.
tsf	Data transformation must be defined by "clr" (default), "log", "log10", "PQN", "lnPQN", "pareto" or "none". See Details.
QCs	logical. If TRUE (default) quality control samples (QCs) are automatically distinguished. See Details.

Details

Data transformation: with "clr" clr trasformation is used (see References), with "log" natural logarithm is used, with "log10" decadic logarithm is used, with "pareto" data are only scaled by Pareto scaling, with "PQN" probabilistic quotient normalization is done, with "lnPQN" natural logarithm of PQN transformed data is done, with "none" no tranformation is done.

Up to twenty different groups can be distinguished in data (including QCs).

If quality control samples (QCs) are present in data and QCs=TRUE, versions with QCs and without them are displayed. If QCs=TRUE and QCs are not present in data, this step is automatically skipped.

Value

Cluster plots and heat maps.

References

Aitchison, J. (1986) The Statistical Analysis of Compositional Data Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). p. 416.

Examples

```
data=metabol
name="Metabolomics"    #name of the project
groupnames=c("Con","Pat","QC")
GraphsClust(data,name,groupnames)
```

GraphsCorr *Correlation graphs*

Description

Evaluate correlations (Pearson and Spearman) between all pairs of variables, display correlation plots.

Usage

```
GraphsCorr(data, name, groupnames, tsf = "clr", QCs = FALSE)
```

Arguments

data	Data table with variables (metabolites) in columns. Samples in rows are sorted according to specific groups.
name	A character string or expression indicating a name of data set. It occurs in names of every output.
groupnames	A character vector defining specific groups in data. Every string must be specific for each group and they must not overlap.
tsf	Data transformation must be defined by "clr" (default), "log", "log10", "PQN", "lnPQN", "pareto" or "none". See Details.
QCs	logical. If FALSE (default) quality control samples (QCs) are automatically distinguished and skipped.

Details

Data transformation: with "clr" clr trasformation is used (see References), with "log" natural logarithm is used, with "log10" decadic logarithm is used, with "pareto" data are only scaled by Pareto scaling, with "PQN" probabilistic quotient normalization is done, with "lnPQN" natural logarithm of PQN transformed data is done, with "none" no tranformation is done.

Up to twenty different groups can be distinguished in data (including QCs).

Cor_S denotes Spearman correlation coefficient.

Value

Correlation graph.

Excel file with Pearson and Spearman correlations in standard tables and also sorted in one column. Significance tests are also evaluated.

References

Aitchison, J. (1986) The Statistical Analysis of Compositional Data Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). p. 416.

Examples

```
data=metabol
name="Metabolomics"    #name of the project
groupnames=c("Con","Pat","QC")
GraphsCorr(data,name,groupnames)
```

GraphsCorrDSC *Correlation graphs with class separation*

Description

Evaluate correlations between all pairs of variables and degree of class separation.

Usage

```
GraphsCorrDSC(data, name, groupnames, tsf = "clr", QCs = FALSE)
```

Arguments

data	Data table with variables (metabolites) in columns. Samples in rows are sorted according to specific groups.
name	A character string or expression indicating a name of data set. It occurs in names of every output.
groupnames	A character vector defining specific groups in data. Every string must be specific for each group and they must not overlap.
tsf	Data transformation must be defined by "clr" (default), "log", "log10", "PQN", "lnPQN", "pareto" or "none". See Details.
QCs	logical. If FALSE (default) quality control samples (QCs) are automatically distinguished and skipped.

Details

Data transformation: with "clr" clr trasformation is used (see References), with "log" natural logarithm is used, with "log10" decadic logarithm is used, with "pareto" data are only scaled by Pareto scaling, with "PQN" probabilistic quotient normalization is done, with "lnPQN" natural logarithm of PQN transformed data is done, with "none" no tranformation is done.

Up to twenty different groups can be distinguished in data (including QCs).

Value

Correlation graph with evaluated clusters.

Excel file with degree of class separation (DCS) of all possible pairs of groups in data for all pairs of comparisons of two variables.

References

Aitchison, J. (1986) The Statistical Analysis of Compositional Data Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). p. 416.

Pierce, K.M. et al. (2005) Classification of gasoline data obtained by gas chromatography using a piecewise alignment algorithm combined with feature selection and principal component analysis, J CHROMATOGR A 1096, p. 101-110.

Examples

```
data=metabol
name="Metabolomics"    #name of the project
groupnames=c("Con","Pat","QC")
GraphsCorrDSC(data,name,groupnames)
```

GraphsOPLSDA *Orthogonal partial least squares - discriminant analysis (OPLS-DA)*

Description

Makes orthogonal partial least squares - discriminant analysis (OPLS-DA), displays score plots and s-plots.

Usage

```
GraphsOPLSDA(data, name, groupnames, tsf = "clr", axx = "p1corr",
  top = 30, qu = 0.75, QCs = FALSE)
```

Arguments

data	Data table with variables (metabolites) in columns. Samples in rows are sorted according to specific groups.
name	A character string or expression indicating a name of data set. It occurs in names of every output.
groupnames	A character vector defining specific groups in data. Every string must be specific for each group and they must not overlap.

tsf	Data transformation must be defined by "clr" (default), "log", "log10", "PQN", "lnPQN", "pareto" or "none". See Details.
axx	Which axis of S-plot is used to choose most important variables: p1 or p1corr (default).
top	How many most important variables (in absolute values) should be highlighted in s-plot? The default is 30.
qu	Which quantile of the important variables (in absolute values) should be highlighted in s-plot? The default is 0.75.
QCs	logical. If FALSE (default) quality control samples (QCs) are automatically distinguished and skipped.

Details

Data transformation: with "clr" clr trasformation is used (see References), with "log" natural logarithm is used, with "log10" decadic logarithm is used, with "pareto" data are only scaled by Pareto scaling, with "PQN" probabilistic quotient normalization is done, with "lnPQN" natural logarithm of PQN transformed data is done, with "none" no tranformation is done.

S-plots can be used only for comparison of two groups. If there is more groups in data, all possible combinations of pairs are evaluated.

Up to twenty different groups can be distinguished in data (including QCs).

Value

Score plot and s-plots of OPLS-DA.

Excel file with s-plot summary for every variable.

References

Aitchison, J. (1986) The Statistical Analysis of Compositional Data Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). p. 416.

Gaude, E.et al. (2012) muma: Metabolomics Univariate and Multivariate Analysis. R package version 1.4. https://CRAN.R-project.org/package=muma

---

GraphsOutlier *Outlier analysis of PCA and PLS-DA*

Description

Outlier analysis of PCA and PLS-DA.

Usage

```
GraphsOutlier(data, name, groupnames, tsf = "clr", type = "pls",
  QCs = FALSE)
```

Arguments

data	Data table with variables (metabolites) in columns. Samples in rows are sorted according to specific groups.
name	A character string or expression indicating a name of data set. It occurs in names of every output.
groupnames	A character vector defining specific groups in data. Every string must be specific for each group and they must not overlap.
tsf	Data transformation must be defined by "clr" (default), "log", "log10", "PQN", "lnPQN", "pareto" or "none". See Details.
type	Definition which type of method do you want to use: "pls" (default) and "pca". If two groups are not nicely separated, the algorithm returns error. In this case set this parameter to "err".
QCs	logical. If FALSE (default) quality control samples (QCs) are automatically distinguished and skipped.

Details

Data transformation: with "clr" clr trasformation is used (see References), with "log" natural logarithm is used, with "log10" decadic logarithm is used, with "pareto" data are only scaled by Pareto scaling, with "PQN" probabilistic quotient normalization is done, with "lnPQN" natural logarithm of PQN transformed data is done, with "none" no tranformation is done.

Up to twenty different groups can be distinguished in data (including QCs).

Value

Plot of outliers.

References

Aitchison, J. (1986) The Statistical Analysis of Compositional Data Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). p. 416.

Thevenot E. A. et al., Analysis of the human adult urinary metabolome variations with age, body mass index and gender by implementing a comprehensive workflow for univariate and OPLS statistical analyses,Journal of Proteome Research 14, 3322-3335, 2015.

GraphsPCA *Principal component analysis (PCA)*

Description

Makes principal component analysis (PCA), displays score plots, loading plots, scree plots and biplots.

Usage

```
GraphsPCA(data, name, groupnames, type = "points", tsf = "clr",
  top = 20, QCs = TRUE, pairs = TRUE)
```

Arguments

data	Data table with variables (metabolites) in columns. Samples in rows are sorted according to specific groups.
name	A character string or expression indicating a name of data set. It occurs in names of every output.
groupnames	A character vector defining specific groups in data. Every string must be specific for each group and they must not overlap.
type	A type of plots must be defined by "points" (default), "names" or "both".
tsf	Data transformation must be defined by "clr" (default), "log", "log10", "PQN", "lnPQN", "pareto" or "none". See Details.
top	How many rays with highest lengths should be visualised in biplot? The default is 20.
QCs	logical. If TRUE (default) quality control samples (QCs) are automatically distinguished. See Details.
pairs	logical. If TRUE (default) PCAs of all combinations of pairs of specific groups are displayed.

Details

Data transformation: with "clr" clr trasformation is used (see References), with "log" natural logarithm is used, with "log10" decadic logarithm is used, with "pareto" data are only scaled by Pareto scaling, with "PQN" probabilistic quotient normalization is done, with "lnPQN" natural logarithm of PQN transformed data is done, with "none" no tranformation is done.

Up to twenty different groups can be distinguished in data (including QCs).

If quality control samples (QCs) are present in data and QCs=TRUE, versions with QCs and without them are displayed. If QCs=TRUE and QCs are not present in data, this step is automatically skipped.

Value

Score plot, scree plot and biplot of PCA.

Excel file with lengths of rays in biplot and degree of class separation (DCS) of all possible pairs of groups in data.

References

Aitchison, J. (1986) The Statistical Analysis of Compositional Data Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). p. 416.

Pierce, K.M. et al. (2005) Classification of gasoline data obtained by gas chromatography using a piecewise alignment algorithm combined with feature selection and principal component analysis, J CHROMATOGR A 1096, p. 101-110.

Examples

```
data=metabol
name="Metabolomics"    #name of the project
groupnames=c("Con","Pat","QC")
GraphsPCA(data,name,groupnames)
```

GraphsPCA3D *3D principal component analysis (PCA)*

Description

Makes 3D principal component analysis (PCA).

Usage

```
GraphsPCA3D(data, name, groupnames, type = "points", tsf = "clr",
  QCs = TRUE)
```

Arguments

data	Data table with variables (metabolites) in columns. Samples in rows are sorted according to specific groups.
name	A character string or expression indicating a name of data set. It occurs in names of every output.
groupnames	A character vector defining specific groups in data. Every string must be specific for each group and they must not overlap.
type	A type of plots must be defined by "points" (default) or "names".
tsf	Data transformation must be defined by "clr" (default), "log", "log10", "PQN", "lnPQN", "pareto" or "none". See Details.
QCs	logical. If TRUE (default) quality control samples (QCs) are are left in the graph. If FALSE QCs are automatically distinguished and skipped.

Details

Data transformation: with "clr" clr trasformation is used (see References), with "log" natural logarithm is used, with "log10" decadic logarithm is used, with "pareto" data are only scaled by Pareto scaling, with "PQN" probabilistic quotient normalization is done, with "lnPQN" natural logarithm of PQN transformed data is done, with "none" no tranformation is done.

Up to twenty different groups can be distinguished in data (including QCs).

If quality control samples (QCs) are present in data and QCs=TRUE, versions with QCs and without them are displayed. If QCs=TRUE and QCs are not present in data, this step is automatically skipped.

Value

3D score plot of PCA.

References

Aitchison, J. (1986) The Statistical Analysis of Compositional Data Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). p. 416.

GraphsPermut *Permutation test of PLS-DA and OPLS-DA*

Description

Makes permutation test of PLS-DA and OPLS-DA, displays score plots (for OPLS-DA only), permutation plots and summary table.

Usage

```
GraphsPermut(data, name, groupnames, tsf = "clr", type = "opls",
  permut = 100, QCs = FALSE)
```

Arguments

data	Data table with variables (metabolites) in columns. Samples in rows are sorted according to specific groups.
name	A character string or expression indicating a name of data set. It occurs in names of every output.
groupnames	A character vector defining specific groups in data. Every string must be specific for each group and they must not overlap.
tsf	Data transformation must be defined by "clr" (default), "log", "log10", "PQN", "lnPQN", "pareto" or "none". See Details.
type	Definition which type of method do you want to use. If "opls" is set (default), OPLS-DA is performed and the number of orthogonal components is automatically computed by using the cross-validation (with a maximum of 9 orthogonal components). If two groups are not nicely separated, the algorithm returns error. In this case set this parameter to "err". When "pls" is set, PLS-DA is done.
permut	Number of random permutations of response labels to estimate R2Y and Q2Y significance by permutation testing. Default is 100.
QCs	logical. If FALSE (default) quality control samples (QCs) are automatically distinguished and skipped.

Details

Data transformation: with "clr" clr trasformation is used (see References), with "log" natural logarithm is used, with "log10" decadic logarithm is used, with "pareto" data are only scaled by Pareto scaling, with "PQN" probabilistic quotient normalization is done, with "lnPQN" natural logarithm of PQN transformed data is done, with "none" no tranformation is done.

Up to twenty different groups can be distinguished in data (including QCs).

Value

Score plots and permutation plots of OPLS-DA.

Excel file with summary.

References

Aitchison, J. (1986) The Statistical Analysis of Compositional Data Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). p. 416.

Thevenot E. A. et al., Analysis of the human adult urinary metabolome variations with age, body mass index and gender by implementing a comprehensive workflow for univariate and OPLS statistical analyses,Journal of Proteome Research 14, 3322-3335, 2015.

`GraphsPLSDA` *Partial least squares - discriminant analysis (PLS-DA)*

Description

Makes partial least squares - discriminant analysis (PLS-DA), displays score plots, loading plots and biplots.

Usage

```
GraphsPLSDA(data, name, groupnames, type = "points", tsf = "clr",
  top = 20, QCs = TRUE)
```

Arguments

`data`	Data table with variables (metabolites) in columns. Samples in rows are sorted according to specific groups.
`name`	A character string or expression indicating a name of data set. It occurs in names of every output.
`groupnames`	A character vector defining specific groups in data. Every string must be specific for each group and they must not overlap.
`type`	A type of plots must be defined by "points" (default), "names" or "both".
`tsf`	Data transformation must be defined by "clr" (default), "log", "log10", "PQN", "lnPQN", "pareto" or "none". See Details.
`top`	How many rays with highest lengths should be in biplot? The default is 20.
`QCs`	logical. If TRUE (default) quality control samples (QCs) are automatically distinguished. See Details.

Details

Data transformation: with "clr" clr trasformation is used (see References), with "log" natural logarithm is used, with "log10" decadic logarithm is used, with "pareto" data are only scaled by Pareto scaling, with "PQN" probabilistic quotient normalization is done, with "lnPQN" natural logarithm of PQN transformed data is done, with "none" no tranformation is done.

Up to twenty different groups can be distinguished in data (including QCs).

If quality control samples (QCs) are present in data and QCs=TRUE, versions with QCs and without them are displayed. If QCs=TRUE and QCs are not present in data, this step is automatically skipped.

Value

Score plot and biplot of PLS-DA. Cross validation plot.

Excel file with lengths of rays in biplot and mean square error of prediction (MSEP) of data.

References

Aitchison, J. (1986) The Statistical Analysis of Compositional Data Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). p. 416.

Examples

```
data=metabol
name="Metabolomics"    #name of the project
groupnames=c("Con","Pat","QC")
GraphsPLSDA(data,name,groupnames)
```

`GraphsROC` *ROC curves*

Description

Display ROC curves with confidence intervals for every variable.

Usage

```
GraphsROC(data, name, groupnames, tsf = "clr", QCs = FALSE)
```

Arguments

`data`	Data table with variables (metabolites) in columns. Samples in rows are sorted according to specific groups.
`name`	A character string or expression indicating a name of data set. It occurs in names of every output.
`groupnames`	A character vector defining specific groups in data. Every string must be specific for each group and they must not overlap.
`tsf`	Data transformation must be defined by "clr" (default), "log", "log10", "PQN", "lnPQN", "pareto" or "none". See Details.
`QCs`	logical. If FALSE (default) quality control samples (QCs) are automatically distinguished and skipped.

Details

Data transformation: with "clr" clr trasformation is used (see References), with "log" natural logarithm is used, with "log10" decadic logarithm is used, with "pareto" data are only scaled by Pareto scaling, with "PQN" probabilistic quotient normalization is done, with "lnPQN" natural logarithm of PQN transformed data is done, with "none" no tranformation is done.

ROC curves can be used only for comparison of two groups. If there is more groups in data, all possible combinations of pairs are evaluated.

Value

ROC curves with confidence intervals of area under the curve (AUC).

Excel sheet with summary of AUCs.

References

Aitchison, J. (1986) The Statistical Analysis of Compositional Data Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). p. 416.

Examples

```
data=metabol
name="Metabolomics"    #name of the project
groupnames=c("Con","Pat","QC")
GraphsROC(data,name,groupnames)
```

`GraphsROCcomb` *ROC curves for combination of variables*

Description

Display ROC curves with confidence intervals for chosen combination of variables. Variables are combined by logistic regression.

Usage

```
GraphsROCcomb(data, name, groupnames, selm, selg = c(1, 2),
  tsf = "clr", QCs = FALSE)
```

Arguments

`data`	Data table with variables (metabolites) in columns. Samples in rows are sorted according to specific groups.
`name`	A character string or expression indicating a name of data set. It occurs in names of every output.
`groupnames`	A character vector defining specific groups in data. Every string must be specific for each group and they must not overlap.
`selm`	Select sequence number of metabolites, their combination by logistic regression will be plotted in ROC curve. Up to six metabolites can be chosen.
`selg`	Select sequence number of groups of data for which ROC curve will be plotted. The dafult is c(1,2).
`tsf`	Data transformation must be defined by "clr" (default), "log", "log10", "PQN", "lnPQN", "pareto" or "none". See Details.
`QCs`	logical. If FALSE (default) quality control samples (QCs) are automatically distinguished and skipped.

Details

Data transformation: with "clr" clr trasformation is used (see References), with "log" natural logarithm is used, with "log10" decadic logarithm is used, with "pareto" data are only scaled by Pareto scaling, with "PQN" probabilistic quotient normalization is done, with "lnPQN" natural logarithm of PQN transformed data is done, with "none" no tranformation is done.

Value

ROC curve of chosen combination of variables with confidence intervals of area under the curve (AUC).

References

Aitchison, J. (1986) The Statistical Analysis of Compositional Data Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). p. 416.

Examples

```
data=metabol
name="Metabolomics"    #name of the project
groupnames=c("Con","Pat","QC")
GraphsROCcomb(data,name,groupnames,selm=c(2,3))
```

`GraphsViol` *Violin plots*

Description

Display violin plots for all specific groups of data.

Usage

```
GraphsViol(data, name, groupnames, tsf = "clr", QCs = FALSE)
```

Arguments

`data`	Data table with variables (metabolites) in columns. Samples in rows are sorted according to specific groups.
`name`	A character string or expression indicating a name of data set. It occurs in names of every output.
`groupnames`	A character vector defining specific groups in data. Every string must be specific for each group and they must not overlap.
`tsf`	Data transformation must be defined by "clr" (default), "log", "log10", "PQN", "lnPQN", "pareto" or "none". See Details.
`QCs`	logical. If FALSE (default) quality control samples (QCs) are automatically distinguished and skipped.

Details

Data transformation: with "clr" clr trasformation is used (see References), with "log" natural logarithm is used, with "log10" decadic logarithm is used, with "pareto" data are only scaled by Pareto scaling, with "PQN" probabilistic quotient normalization is done, with "lnPQN" natural logarithm of PQN transformed data is done, with "none" no tranformation is done.

Up to twenty different groups can be distinguished in data (including QCs).

Value

Violin plots.

References

Aitchison, J. (1986) The Statistical Analysis of Compositional Data Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). p. 416.

Examples

```
data=metabol
name="Metabolomics"    #name of the project
groupnames=c("Con","Pat","QC")
GraphsViol(data,name,groupnames)
```

GraphsVIP *Variable importance plots (VIP)*

Description

Makes variable importance plots (VIP) of partial least squares - discriminant analysis (PLS-DA), also displays score plots and biplots.

Usage

```
GraphsVIP(data, name, groupnames, tsf = "clr", top = 30, QCs = FALSE)
```

Arguments

data	Data table with variables (metabolites) in columns. Samples in rows are sorted according to specific groups.
name	A character string or expression indicating a name of data set. It occurs in names of every output.
groupnames	A character vector defining specific groups in data. Every string must be specific for each group and they must not overlap.
tsf	Data transformation must be defined by "clr" (default), "log", "log10", "PQN", "lnPQN", "pareto" or "none". See Details.
top	How many most important variables should be in zoomed VIP plot and biplot? The default is 30.
QCs	logical. If FALSE (default) quality control samples (QCs) are automatically distinguished and skipped.

Details

Data transformation: with "clr" clr trasformation is used (see References), with "log" natural logarithm is used, with "log10" decadic logarithm is used, with "pareto" data are only scaled by Pareto scaling, with "PQN" probabilistic quotient normalization is done, with "lnPQN" natural logarithm of PQN transformed data is done, with "none" no tranformation is done.

VIP plot can be used only for comparison of two groups. If there is more groups in data, all possible combinations of pairs are evaluated.

Up to twenty different groups can be distinguished in data (including QCs).

Value

VIP plots, score plots and biplots of PLS-DA.

Excel file with mean, sd and angles of rays from group center for every variable.

References

Aitchison, J. (1986) The Statistical Analysis of Compositional Data Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). p. 416.

Wold, S. et al. (2001) PLS-regression: a basic tool of chemometrics, CHEMOLAB 58 (2), p. 109-130.

Chong, I.G., Jun, C.H. (2005) Performance of some variable selection methods when multicollinearity is present, CHEMOLAB 78, p. 103-112.

Examples

```
data=metabol
name="Metabolomics"    #name of the project
groupnames=c("Con","Pat","QC")
GraphsVIP(data,name,groupnames)
```

GraphsVIPOPLSDA *Orthogonal partial least squares - discriminant analysis (OPLS-DA)*

Description

Makes orthogonal partial least squares - discriminant analysis (OPLS-DA), displays score plots and s-plots.

Usage

```
GraphsVIPOPLSDA(data, name, groupnames, tsf = "clr", top = 30,
  qu = 0.75, QCs = FALSE)
```

Arguments

data	Data table with variables (metabolites) in columns. Samples in rows are sorted according to specific groups.
name	A character string or expression indicating a name of data set. It occurs in names of every output.

groupnames	A character vector defining specific groups in data. Every string must be specific for each group and they must not overlap.
tsf	Data transformation must be defined by "clr" (default), "log", "log10", "PQN", "lnPQN", "pareto" or "none". See Details.
top	How many most important variables (in absolute values) should be highlighted in s-plot? The default is 30.
qu	Which quantile of the important variables (in absolute values) should be highlighted in s-plot? The default is 0.75.
QCs	logical. If FALSE (default) quality control samples (QCs) are automatically distinguished and skipped.

Details

Data transformation: with "clr" clr trasformation is used (see References), with "log" natural logarithm is used, with "log10" decadic logarithm is used, with "pareto" data are only scaled by Pareto scaling, with "PQN" probabilistic quotient normalization is done, with "lnPQN" natural logarithm of PQN transformed data is done, with "none" no tranformation is done.

S-plots can be used only for comparison of two groups. If there is more groups in data, all possible combinations of pairs are evaluated.

Up to twenty different groups can be distinguished in data (including QCs).

Value

Score plot and s-plots of OPLS-DA.

Excel file with s-plot summary for every variable.

References

Aitchison, J. (1986) The Statistical Analysis of Compositional Data Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). p. 416.

Gaude, E.et al. (2012) muma: Metabolomics Univariate and Multivariate Analysis. R package version 1.4. https://CRAN.R-project.org/package=muma

---

GraphsVolcano *Logratio volcano plots*

Description

Display logratio volcano plot comparing two specific groups.

Usage

```
GraphsVolcano(data, name, groupnames, type = "par", tsf = "clr",
  pair = FALSE, QCs = FALSE, hy = (0.05/length(colnames(data))),
  hx = FALSE, hhx = 1.5)
```

Arguments

data	Data table with variables (metabolites) in columns. Samples in rows are sorted according to specific groups.
name	A character string or expression indicating a name of data set. It occurs in names of every output.
groupnames	A character vector defining specific groups in data. Every string must be specific for each group and they must not overlap.
type	Parametric ("par" - default) or nonparametric ("nonpar") test must be chosen.
tsf	Data transformation must be defined by "clr" (default), "log", "log10", "PQN", "lnPQN", "pareto" or "none". See Details.
pair	logical. If TRUE then the paired tests are done. Default is FALSE.
QCs	logical. If FALSE (default) quality control samples (QCs) are automatically distinguished and skipped.
hy	Important points are evaluted on y-axis with this limit. The default is alpha=0.05 with Bonferroni correction.
hx	logical. If TRUE important points are evaluated from both, x-axis and y-axis. Default is FALSE.
hhx	If hx=TRUE, set the absolute value of the limit for most important differences of medians (deafult is 1.5).

Details

Data transformation: with "clr" clr trasformation is used (see References), with "log" natural logarithm is used, with "log10" decadic logarithm is used, with "pareto" data are only scaled by Pareto scaling, with "PQN" probabilistic quotient normalization is done, with "lnPQN" natural logarithm of PQN transformed data is done, with "none" no tranformation is done.

Logratio volcano plot can be used only for comparison of two groups. If there is more groups in data, all possible combinations of pairs are evaluated.

Logratio volcano plot is a version of standard volcano plot with difference of group medians on x-axis (only for tsf=c("clr","log")) and -log10 of p-value from t-test or Wilcoxon test on y-axis. For "pareto" transformation standard volcano plot is done with log2(fold change) on x-axis.

Value

Volcano plots.

Excel sheet with differences of medians and p-values from chosen test.

References

Aitchison, J. (1986) The Statistical Analysis of Compositional Data Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). p. 416.

Examples

```
data=metabol
name="Metabolomics"    #name of the project
groupnames=c("Con","Pat","QC")
GraphsVolcano(data,name,groupnames)
```

metabol *A toy example of metabolomic data set*

Description

A small dataset containing the information about 50 metabolites in samples from 5 patients with one specific disease and 5 healthy controls. Quality control samples (QCs) are also included.

Usage

```
metabol
```

Format

A data frame with 13 rows (samples) and 50 variables (metabolites)

Preproc *Data preprocessing*

Description

Counting zeros, interpolation by LOESS, checking of coefficient of variation (CV) of quality control samples (QCs), zero imputatuion, summary of data preprocessing.

Usage

```
Preproc(data, name, groupnames, CVlim = 30)
```

Arguments

data	Data table sorted according to batch order (in columns) with variables (metabolites) in rows.
name	A character string or expression indicating a name of data set. It occurs in names of every output.
groupnames	A character vector defining specific groups in data. Every string must be specific for each group and they must not overlap.
CVlim	Limit of coefficient of variation (CV) of quality control samples (QCs). All metabolites with CV of QCs higher than this limit will be omitted from the analysis. The default value is 30.

Details

Up to fifteen different groups can be distinguished in data (including QCs).

Value

LOESS curves.

Excel file with preprocessed data.

Examples

```
data=batchSort
name="Preprocessing"    #name of the project
groupnames=c("Con","Pat","QC")
Preproc(data,name,groupnames)
```

Tests *Standard tests*

Description

Makes Shapiro-Wilk test, ANOVA, Kruskal-Wallis test, t-test for every variable and Wilcoxon test for every variable.

Usage

```
Tests(data, name, groupnames, tsf = "clr", pair = FALSE, QCs = FALSE)
```

Arguments

data	Data table with variables (metabolites) in columns. Samples in rows are sorted according to specific groups.
name	A character string or expression indicating a name of data set. It occurs in names of every output.
groupnames	A character vector defining specific groups in data. Every string must be specific for each group and they must not overlap.
tsf	Data transformation must be defined by "clr" (default), "log", "log10", "PQN", "lnPQN", "pareto" or "none". See Details.
pair	logical. If TRUE then the paired tests are done. Default is FALSE.
QCs	logical. If FALSE (default) quality control samples (QCs) are automatically distinguished and skipped.

Details

Data transformation: with "clr" clr trasformation is used (see References), with "log" natural logarithm is used, with "log10" decadic logarithm is used, with "pareto" data are only scaled by Pareto scaling, with "PQN" probabilistic quotient normalization is done, with "lnPQN" natural logarithm of PQN transformed data is done, with "none" no tranformation is done.

Value

Excel file with p-values from Shapiro-Wilk test, ANOVA, Kruskal-Wallis test, t-test for every variable and Wilcoxon test for every variable and their summary.

References

Aitchison, J. (1986) The Statistical Analysis of Compositional Data Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). p. 416.

Examples

```
data=metabol
name="Metabolomics"    #name of the project
groupnames=c("Con","Pat","QC")
Tests(data,name,groupnames)
```

`transf` *Transformation*

Description

Transform data according to chosen function.

Usage

```
transf(data, name, tsf = "clr")
```

Arguments

`data`	Data table with variables (metabolites) in columns. Samples in rows are sorted according to specific groups.
`name`	A character string or expression indicating a name of data set. It occurs in names of every output.
`tsf`	Data transformation must be defined by "clr" (default), "log", "log10", "PQN", "lnPQN", "pareto" or "none". See Details.

Details

Data transformation: with "clr" clr trasformation is used (see References), with "log" natural logarithm is used, with "log10" decadic logarithm is used, with "pareto" data are only scaled by Pareto scaling, with "PQN" probabilistic quotient normalization is done, with "lnPQN" natural logarithm of PQN transformed data is done, with "none" no tranformation is done.

Up to twenty different groups can be distinguished in data (including QCs).

Value

Excel file with data transformed according to chosen function.

References

Aitchison, J. (1986) The Statistical Analysis of Compositional Data Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). p. 416.

Examples

```
data=metabol
name="Metabolomics"    #name of the project
groupnames=c("Con","Pat","QC")
transf(data,name)
```

Index