Package ‘LDLcalc’

May 31, 2022

Title Calculate and Predict the Low Density Lipoprotein Values

Version 2.0

Author Petros Paplomatas [aut, cre],
Kostas Anagnostopoulos [aut]

Description
A wide variety of ways to calculate (through equations) or predict (using 9 Machine learning methods as well as a stack algorithm combination of them all) the Low Density Lipoprotein values of patients based on the values of three other metrics, namely Total Cholesterol , Triglycerides and High Density Lipoprotein. It can also calculate the variance of LDL and the Atherogenic Index of Plasma (AIP) using error propagation and bootstrapping.

License GPL-3

URL `https://sciencesandresearch.com/ldlcalc-calculate-and-predict-the-low-density-lipoprotein-values`
,`https://sciencesandresearch.com/ldlcalctutorial`

Encoding UTF-8

LazyData true

RoxygenNote 7.2.0

Imports data.table, caret, caretEnsemble, lares, corrplot,
RColorBrewer, lattice, resample, moments, ggplot2, janitor,
philentropy, sets

Depends R (>= 3.5.0)

Suggests rmarkdown, knitr, Cubist, earth, gbm, glmnet, gridExtra,
kernlab, randomForest, tidyr, testthat (>= 3.0.0)

Config/testthat/edition 3

VignetteBuilder knitr

NeedsCompilation no

Maintainer Petros Paplomatas <`P.paplomatas@hotmail.com`>

Repository CRAN

Date/Publication 2022-05-31 15:50:10 UTC

R topics documented:

AIPbootVrnc *Calculate variance of AIP using Bootstrapping.*

Description

This function can be used to calculate the variance of the Atherogenic Index of Plasma (AIP) using Bootstrapping.

Usage

```
AIPbootVrnc(
  TG,
  HDL,
  sampleSize = length(TG),
  SI = TRUE,
  noOfReps = 1000,
  pb = F
)
```

Arguments

`TG`	A vector or data frame column containing the triglyceride (TG) values to be used for the calculation of the variance of the Atherogenic Index of Plasma (AIP) using bootstrapping. TG and HDL must be of the same length.
`HDL`	A vector or data frame column containing the high density lipoprotein (HDL) values to be used for the calculation of the variance of the Atherogenic Index of Plasma (AIP) using bootstrapping. TG and HDL must be of the same length.
`sampleSize`	(default = length of TG or HDL) The sample size that will be generated at each bootstrapping sampling round. Size of bootstrapped samples cannot be larger than the original.
`SI`	Boolean (default=TRUE). AIP is by definition calculated using SI units for TG and HDL (mmol/L). If mg/dl units are provided instead, SI must be set to FALSE.
`noOfReps`	(default = 1000) Number of repetitions of the bootstrapping.
`pb`	Display a progress bar (default = FALSE)

Value

It returns a data table with four columns. The first column contains the mean of the AIP values for each iteration. The second column contains the median of each iteration.The third column contains the variance and the fourth column contains the CV of each iteration. It also returns the median of the "Mean", "Var" and "CV" columns of the data table.

Examples

```
## Not run:
AIPbootstrVar = AIPbootVrnc(sampleA$TG,sampleA$HDL)

## End(Not run)
```

AIPcalc *Calculation of Atherogenic Index of Plasma (AIP)*

Description

Calculates the Atherogenic Index of Plasma (AIP) from triglyceride and HDL values.

Usage

```
AIPcalc(TG, HDL, SI ,roundDigit)
```

Arguments

`TG`	A vector or data frame column containing the triglyceride (TG) values to be used for the calculation of the atherogenic index of plasma (AIP). TG and HDL must be of the same length.
`HDL`	A vector or data frame column containing the high density lipoprotein (HDL) values to be used for the calculation of the atherogenic index of plasma (AIP). TG and HDL must be of the same length.
`SI`	Boolean (default=TRUE). AIP is by definition calculated using SI units for TG and HDL (mmol/L). If mg/dl units are provided instead, SI must be set to FALSE.
`roundDigit`	Decimal digits to round the result to (default = 4).

Value

A vector of AIP values of length equal to the length of TG and HDL.

References

Dobiasova M, Frohlich J. The plasma parameter log (TG/HDL-C) as an atherogenic index: correlation with lipoprotein particle size and esterification rate in apo B- lipoprotein-depleted plasma (FERHDL). Clin Biochem. 2001, 34:583-88.

Tan MH, Johns D, Glazer NB. Pioglitazone reduces atherogenic index of plasma in patients with type 2 diabetes. Clin Chem. 2004, 50:1184-88.

Nwagha UI, Ingweh JC. A significant indicator for the onset of atherosclerosis during menopause in hypertensive females of South East Nigeria. J Coll Med. 2005, 10(2):67-71.

Daniels LB, Laughlin G, Sarno MJ. Lp- PLA2 is an independent predictor of incidence of coronary heart disease in apparently healthy older population. J Am Col Cardiol. 2008,51:913-19.

Xiaowei Z, Lugang Y, Hui Z, Qinhua M, Xiaohua Z, Ting L, et al. Atherogenic index of plasma is a novel and better biomarker associated with obesity: A population- based cross-sectional study in China. Lipids Health Dis. 2018,17(1):37.

Examples

```
## Not run:
AIP = AIPcalc(sampleA$TG, sampleA$HDL)

## End(Not run)
```

`AIPErrPrp` *Variance of Atherogenic Index of Plasma (AIP) using Error Propagation.*

Description

Calculate the variance of AIP using Error Propagation (The Delta Method)

Usage

```
AIPErrPrp(TG, HDL, SI = TRUE, roundDigit = 5)
```

Arguments

`TG` A vector or data frame column containing the triglyceride (TG) values to be used for the calculation of the variance of Atherogenic Index of Plasma (AIP). TG and HDL must be of the same length.

`HDL` A vector or data frame column containing the high density lipoprotein (HDL) values to be used for the calculation of variance of the Atherogenic Index of Plasma (AIP). TG and HDL must be of the same length.

`SI` Boolean (default=TRUE). AIP is by definition calculated using SI units for TG and HDL (mmol/L). If mg/dl units are provided instead, SI must be set to FALSE.

`roundDigit` Decimal digits to round the result to (default = 5).

Value

The variance of AIP using error propagation theory.

References

Casella G, Berger RL. Statistical Inference. 2nd ed. Duxbury Thomson Learning, 2002, Pages 240-245

Joint Commitee for Guides in Metrology. Evaluation of measurement data- Guide to the expression of uncertainty in measurement. 2008.

Joint Commitee for Guides in Metrology. International Vocabulary of Metrology (VIM)- Basic and General Concepts and Associated Terms. 2012.

Examples

```
## Not run:
AIPpropagationVar = AIPErrPrp(sampleA$TG,sampleA$HDL)

## End(Not run)
```

`AIPErrPrp2Ord` *Second Order Taylor expansion AIP Error propagation variance*

Description

Calculate the second order Taylor Expansion error propagation (delta method) variance of the Atherogenic Index of Plasma.

Usage

```
AIPErrPrp2Ord(TG, HDL, SI=TRUE, roundDigit=5)
```

Arguments

`TG` A vector or data frame column containing the triglyceride (TG) values to be used for the calculation of the variance of the Atherogenic Index of Plasma (AIP) using second order Taylor expansion. TG and HDL must be of the same length.

`HDL` A vector or data frame column containing the high density lipoprotein (HDL) values to be used for the calculation of the variance of the Atherogenic Index of Plasma (AIP) using second order Taylor expansion. TG and HDL must be of the same length.

`SI` Boolean (default=TRUE). AIP is by definition calculated using SI units for TG and HDL (mmol/L). If mg/dl units are provided instead, SI must be set to FALSE.

`roundDigit` Decimal digits to round the result to (default = 5).

Value

The variance of AIP using error propagation theory.

References

Casella G, Berger RL. Statistical Inference. 2nd ed. Duxbury Thomson Learning 2002, Pages 240-245

Joint Commitee for Guides in Metrology. Evaluation of measurement data- Guide to the expression of uncertainty in measurement. 2008.

Joint Commitee for Guides in Metrology. International Vocabulary of Metrology (VIM)- Basic and General Concepts and Associated Terms. 2012.

Examples

```
## Not run:
AIPpropagationVarTaylor = AIPErrPrp2Ord(sampleA$TG,sampleA$HDL)

## End(Not run)
```

AIP_DensityPlotOfbootst *Plot density of bootstrap and empirical variance for AIP*

Description

Plot density of the bootstrap variance for AIP along with the median, 2.5 and 97.5 percentile as vertical lines. Also plots the empirical variance as a segment from the bottom to the middle of the graph and the error propagation variance as a line segment from the top to the middle of the graph. DF must the dataframe of the list returned from the respective (LDL or AIP) Bootstrap variance function.

Usage

```
AIP_DensityPlotOfbootst(
  DF,
  title = "",
  empirVrnc,
  errPropVrnc,
  errPropVrnc2Ord
)
```

Arguments

`DF`	A data frame containing the bootstrap variances of AIP.
`title`	Title of the plot (default="").
`empirVrnc`	Value of the empirical (experimental) variance of AIP.
`errPropVrnc`	Value of the first order error propagation variance of AIP.
`errPropVrnc2Ord`	Value of the second order error propagation variance of AIP.

Value

Plots the respective plot.

Examples

```
## Not run:
sampleA$AIP = AIPcalc(sampleA$TG,sampleA$HDL, SI=FALSE)
AIP_empirVrnc=var(sampleA$AIP)
AIP_errPropVrnc=AIPErrPrp(sampleA$TG,sampleA$HDL, SI=FALSE)
AIP_errPropVrnc2Ord=AIPErrPrp2Ord(sampleA$TG,sampleA$HDL, SI=FALSE)
DfAIPboost=as.data.frame(AIPbootVrnc(sampleA$TG,sampleA$HDL, SI=FALSE))
AIP_DensityPlotOfbootst(DfAIPboost,"Title",AIP_empirVrnc, AIP_errPropVrnc, AIP_errPropVrnc2Ord)

## End(Not run)
```

AIP_HDLVrnc *Calculate the variance of AIP when HDL changes.*

Description

This function calculates the variance of the Atherogenic Index of Plasma (AIP) using two methods: Error Propagation (first and second order) and Bootstrap when the HDL distribution changes. It uses a data frame where each column contains a different set of HDL values of increasing variance as the column index increases.

Usage

```
AIP_HDLVrnc(dfHDL, TG, SI = TRUE, bootStrpReps = 2000)
```

Arguments

dfHDL	A data frame where each column contains a different set of HDL values of increasing variance as the column index increases.
TG	A vector or data frame column containing the triglyceride (TG) values to be used for the calculation of the variance of the Atherogenic Index of Plasma (AIP).
SI	Boolean (default=TRUE). AIP is by definition calculated using SI units for TG and HDL (mmol/L). If mg/dl units are provided instead, SI must be set to FALSE.
bootStrpReps	(Default=2000) Number of bootstrap iterations for bootstrap variance calculation.

Value

It returns a list with the first order Error Propagation variance (ErrPropVrnc), the second order Error Propagation variance (ErrPropVrnc2Ord) and the bootstrap variance (BootVrnc). Each list element is a vector of length equal to the number of columns of the HDL data frame supplied as argument (dfHDL) and each vector value corresponds to the respective variance of the corresponding data frame columns.

Examples

```
## Not run:
dfHDL = CV_Range(sampleB$HDL,0,100,maxRandIter = 1000, plot=FALSE)
AIP_HDLVrnc=AIP_HDLVrnc(dfHDL,sampleA$TG, bootStrpReps=2000)

## End(Not run)
```

`AIP_TGVrnc` *Calculate the variance of AIP when TG changes.*

Description

Calculate the variance of the Atherogenic Index of Plasma (AIP) using two methods: Error Propagation (first and second order) and Bootstrap when the triglyceride distribution changes. It uses a data frame where each column contains a different set of TG values of increasing variance as the column index increases.

Usage

```
AIP_TGVrnc(dfTG, HDL, SI = TRUE, bootStrpReps = 2000)
```

Arguments

`dfTG`	A data frame where each column contains a different set of TG values of increasing variance as the column index increases.
`HDL`	A vector or data frame column containing the HDL values to be used for the calculation of the variance of the Atherogenic Index of Plasma (AIP).
`SI`	Boolean (default=TRUE). AIP is by definition calculated using SI units for TG and HDL (mmol/L). If mg/dl units are provided instead, SI must be set to FALSE.
`bootStrpReps`	(Default=2000) Number of bootstrap iterations for bootstrap variance calculation.

Value

It returns a list with the first order Error Propagation variance (ErrPropVrnc), the second order Error Propagation variance (ErrPropVrnc2Ord) and the bootstrap variance (BootVrnc). Each list element is a vector of length equal to the number of columns of the TG data frame supplied as argument (dfTG) and each vector value corresponds to the respective variance of the corresponding data frame columns.

Examples

```
## Not run:
dfTG = CV_Range(sampleA$TG, 0, 100, maxRandIter = 100, plot=FALSE)
AIP_TGVrnc=AIP_TGVrnc(dfTG,sampleA$HDL, bootStrpReps=2000)

## End(Not run)
```

`chebyshev` *Chebysev's inequality*

Description

Function to calculate lower and upper bounds for which at least 75% of data points of a distribution lie within ±2SD of the mean using Chebyshev's inequality.

Usage

```
chebyshev(vec)
```

Arguments

`vec`	The vector containing the values of the data set.

Value

It outputs the lower and upper bound of Chebyshevs 75% range and the lower and upper value of the range of the observed (measured) distribution.

References

Bienayme I. Considerations a l appui de la de couverte d laplace. Comptes Rendus de l Acade mie des Sciences 1853; 37: 309–324.

. Chebyshev P. Des valeurs moyennes. Journal de Mathematiques Pures . Appliqee 1867; 2(12): 177–184s

Examples

```
## Not run:
chebysevBounds = chebyshev(sampleA$LDL)

## End(Not run)
```

`Corr_Plot` *Correlation plot*

Description

Plots a correlation plot to see the correlation between different columns of your data, for example LDL,HDL relative to age.

Usage

```
Corr_Plot(listDaten)
```

Arguments

listDaten — This is a list of the data between which one can see the correlation. One needs to provide at least two columns of equal length in order to see the correlation between them

Value

No return value, it prints the requested diagram.

Examples

```
Corr_Plot(SampleData[1:5])
```

CV — *Calculate the CV of a set of values.*

Description

This function calculates the coefficient of variation (CV) of the values supplied.

Usage

```
CV(Vec, roundDigits = 2)
```

Arguments

Vec — The vector of values for which to calculate the CV.

roundDigits — Number of digits to round the result to (default=2).

Value

Returns the value of the CV.

Examples

```
CV = CV(sampleA$LDL)
```

CV_Range *Change variance of distribution while keeping the mean constant*

Description

Takes as input a vector of values that have a distribution with a variance expressed as coefficient of variation (CV), It then modifies the distribution, so that the CV changes from a lower bound (lower_CV_bound) to an upper bound (upper_CV_bound), while the mean stays constant. Optionally, it can also plot the resulting variances.

Usage

```
CV_Range(
  sampleVector,
  lower_CV_Bound,
  upper_CV_Bound,
  maxRandIter = 10000,
  plot = F
)
```

Arguments

`sampleVector`	The vector containing the initial values with initial distribution and CV
`lower_CV_Bound`	The lower CV value we wish the distribution to achieve.
`upper_CV_Bound`	The upper CV value we wish the distribution to achieve.
`maxRandIter`	The function uses a stochastic algorithm to change the CV values from the lower to the upper value desired. If these bounds have not been achieved after maxRandIter iterations (default=10000), the function will exit, in order to avoid a possible infinite loop.
`plot`	If set to TRUE, will plot the resulting variances (default=FALSE)

Value

A data frame where each column contains a distribution of values with increasing CV, from the lower bound to upper bound.

Examples

```
## Not run:
DataFrame=CV_Range(sampleA$LDL,0,100,maxRandIter = 100, plot=TRUE)

## End(Not run)
```

DensityPlotOfVar *Plot variance density*

Description

Plot density of the Variance along with the median, 2.5 and 97.5 percentile as vertical lines

Usage

```
DensityPlotOfVar(Vector, title = "")
```

Arguments

Vector	A vector or data frame columns with the values whose
title	Title of the plot (default=""). distribution density is to be plotted.

Value

The function returns no values but plots the density plot.

Examples

```
## Not run:
LDLbootstrVar=as.data.frame(LDLbootVrnc(sampleA$CHOL,sampleA$HDL,sampleA$TG))
DensityPlotOfVar(LDLbootstrVar$dataTable.CV)

## End(Not run)
```

ErrorCorCov *Calculate correlation and covariance of residuals*

Description

This function calculates the residuals (the squared difference of each value from the mean) of two groups. Then it calculates the correlation and covariance between the residuals of the two groups and plots them.

Usage

```
ErrorCorCov(dataset1, dataset2, plot = F)
```

Arguments

dataset1	A vector containing the values of the first group.
dataset2	A vector containing the values of the second group.
plot	(default=TRUE). Plot the errors of the first group versus the second one.

Value

Returns a list with the residuals of the two groups (distr1Error, distr2Error), the correlation coefficient between the two groups (Correlation) and the covariance of the two groups (Covariance).

Examples

```
## Not run:
ErrorOFCorCov=ErrorCorCov(sampleA$HDL[1:20],sampleA$CHOL[1:20],plot = FALSE)

## End(Not run)
```

JSD *Calculate the Jensen-Shannon divergence (JSD)*

Description

Calculate the Jensen-Shannon divergence between the same parameter of different data sets. This works only for two distributions.

Usage

```
JSD(vec1, vec2)
```

Arguments

vec1	The vector containing the values of the first data set.
vec2	The vector containing the values of the second data set.

Value

The function returns the value of the Jensen-Shannon divergence (JSD), the frequencies of the two datasets (freq1, freq2, dfFreqs) and the probabilities of the frequencies of the two groups, which should sum up to 1.

References

D.M. Endres, J.E. Schindelin, A new metric for probability distributions, IEEE Trans. Inf. Theory (2003), https://doi.org/10.1109/TIT.2003.813506.

F. Oesterreicher, I. Vajda, A new class of metric divergences on probability spaces and its applicability in statistics, Ann. Inst. Stat. Math. (2003), https://doi.org/ 10.1007/BF02517812.

Examples

```
## Not run:
JSDCalc = JSD(model[["trainData"]]$LDLd,model[["testData"]]$LDLd)

## End(Not run)
```

JSDNormal — *Calculate the Jensen-Shannon divergence (JSD) between a discrete empirical distribution and the normal distribution.*

Description

Calculates the Jensen-Shannon divergence between a discrete distribution and the corresponding normal distribution with mean and standard deviation the same as these of the discrete one.

Usage

```
JSDNormal(dfSmpl, param)
```

Arguments

`dfSmpl` A data frame containing the values of the discrete distribution. The data frame may contain more that one column with discrete distribution values. The argument "param" specified next will determine which column will be used

`param` The name of the column to be used.

Value

The function returns the Jensen-Shannon divergence between the discrete and corresponding normal distribution. It also returns a data frame with the empirical probability of the values supplied in the column as well as the empirical probabilies one of the normal discrete distribution. distribution.

References

D.M. Endres, J.E. Schindelin, A new metric for probability distributions, IEEE Trans. Inf. Theory (2003), https://doi.org/10.1109/TIT.2003.813506.

F. Oesterreicher, I. Vajda, A new class of metric divergences on probability spaces and its applicability in statistics, Ann. Inst. Stat. Math. (2003), https://doi.org/ 10.1007/BF02517812.

Examples

```
## Not run:
JSD.between.empirical.Normal =JSDNormal(sampleA,"LDL")

## End(Not run)
```

LDLallEq *Calculates and returns the LDL values using all available equations*

Description

This function calculates and returns the LDL values computed with all of the 12 named equations.

Usage

```
LDLallEq(TC, HDL, TG)
```

Arguments

TC	The TC (Total Cholesterol) value.
HDL	The HDL (High-density lipoprotein-cholesterol) value.
TG	The TG (Triglyceride) value.

Value

The calculated LDL values, according to all the equations.

Examples

```
LDLallEq(170,35,174)
```

LDLbootVrnc *Calculate variance of LDL using bootstrapping.*

Description

Function to calculate the variance of LDL using Bootstrapping.

Usage

```
LDLbootVrnc(CHOL, HDL, TG, sampleSize = length(CHOL), noOfReps = 1000, pb = F)
```

Arguments

CHOL	A vector containing the cholesterol values to be used for LDL calculation.
HDL	A vector containing the HDL values to be used for LDL calculation.
TG	A vector containing the triglyceride values to be used for LDL calculation. The three vectors (vecCHOL, vecHDL and vecG) must contain the same number of values
sampleSize	Number of samples drawn uniformly and with replacement from the three vectors. It cannot be larger than the sample size (default=sample size).
noOfReps	Number of bootstrap iterations.
pb	Draw a text progress bar (defaut=FALSE)

Value

It returns a data table with four columns. The first column contains the mean of the LDL values for each iteration. The second column contains the median of each iteration.The third column contains the variance and the fourth column contains the CV of each iteration. It also returns the median of the "Mean", "Var" and "CV" columns of the data table.

Examples

```
## Not run:
LDLboostrpVar = LDLbootVrnc(sampleA$CHOL, sampleA$HDL, sampleA$TG)

## End(Not run)
```

LDLErrPrp *Calculate LDL variance using error propagation*

Description

Function to calculate the LDL Variance according to error propagation (delta) method.

Usage

```
LDLErrPrp(CHOL, HDL, TG, divFactor = 5)
```

Arguments

CHOL	A vector containing the cholesterol values to be used for LDL calculation.
HDL	A vector containing the HDL values to be used for LDL calculation.
TG	A vector containing the triglyceride values to be used for LDL calculation.
divFactor	The factor by which to divide the triglyceride values so as to approximate the VLDL. Default is 5, according to the Friedewald equation.

Value

The function returns the error propagation variance of LDL calculated from the cholesterol, HDL and tryglyceride values passed as arguments.

References

Casella G, Berger RL. Statistical Inference. 2nd ed. Duxbury Thomson Learning; 2002.

Examples

```
## Not run:
LDLerrorPrp = LDLErrPrp(sampleA$CHOL, sampleA$HDL, sampleA$TG)

## End(Not run)
```

LDL_CHOLVrnc *Calculate LDL variance when cholesterol variance changes*

Description

This function calculates the variance of LDL using two methods: Error Propagation and Bootstrap when the cholesterol distribution changes. It uses a data frame where each column contains a different set of cholesterol values of increasing variance as the column index increases.

Usage

```
LDL_CHOLVrnc(dfCHOL, HDL, TG, bootStrpReps = 2000)
```

Arguments

`dfCHOL`	A data frame where each column contains a different set of cholesterol values of increasing variance as the column index increases.
`HDL`	A vector or data frame column containing the HDL values to be used for the calculation of the variance of LDL.
`TG`	A vector or data frame column containing the triglyceride values to be used for the calculation of the variance of LDL.
`bootStrpReps`	(Default=2000) Number of bootstrap iterations for bootstrap variance calculation.

Value

The function returns a list with the Error Propagation variance (ErrPropVrnc) and the bootstrap variance (BootVrnc). Each list element is a vector of length equal to the number of columns of the cholesterol data frame supplied as argument (dfCHOL) and each vector value corresponds to the respective variance of the corresponding data frame columns.

Examples

```
## Not run:
dfCHOL = CV_Range(sampleA$CHOL, 0, 100, maxRandIter = 3000, plot=FALSE)
LDLCHOLVar = LDL_CHOLVrnc(dfCHOL, sampleA$HDL, sampleA$TG, bootStrpReps=200)

## End(Not run)
```

LDL_DensityPlotOfbootst *Plot density of bootstrap and empirical variance for LDL*

Description

Plot density of the bootstrap variance for LDL along with the median, 2.5 and 97.5 percentile as vertical lines. Also plots the empirical variance as a segment from the bottom to the middle of the graph and the error propagation variance as a line segment from the top to the middle of the graph. DF must the dataframe of the list returned from the respective Bootstrap variance function.

Usage

```
LDL_DensityPlotOfbootst(DF, title = "", empirVrnc, errPropVrnc)
```

Arguments

`DF`	The data frame containing the bootstrap variances of LDL.
`title`	Title of the plot (default="")
`empirVrnc`	The value of of the empirical (experimental) variance.
`errPropVrnc`	The value of the error propagation variance.

Value

Plots the respective plot.

Examples

```
## Not run:
LDL_empirVrnc = var(sampleA$LDL)
LDL_errPropVrnc = LDLErrPrp(sampleA$CHOL,sampleA$HDL,sampleA$TG)
LDLbootStrp=as.data.frame(LDLbootVrnc(sampleA$CHOL,sampleA$HDL,sampleA$TG))
LDL_DensityPlotOfbootst(LDLbootStrp,"Title",LDL_empirVrnc,LDL_errPropVrnc)

## End(Not run)
```

LDL_eq *Calculates and returns the LDL Value for any of the 12 different equations*

Description

This function calculates and returns the LDL value computed from any of the 12 named equations.

Usage

```
LDL_eq(TC, HDL, TG, EqMethod)
```

Arguments

`TC`	The TC (Total Cholesterol) value.
`HDL`	The HDL (High-density lipoprotein- cholesterol) value.
`TG`	The TG (Triglyceride) value.
`EqMethod`	The type of equation to be used to calculate the LDL value.The type of equation to be used to calculate the LDL value. EqMethod could be:("Friedewald","Ahmadi","Chen","Anandaraja" or "Rao").

Value

The calculated LDL value, according to the equation of choice or a printed error message and 404, if the equation type does not exist.

Examples

```
LDL_eq(170.5,35.12,230,"Martin360")
```

`LDL_HDLVrnc` *Calculate LDL variance when HDL variance changes*

Description

This function calculates the variance of LDL using two methods: Error Propagation and Bootstrap when the HDL distribution changes. It uses a data frame where each column contains a different set of HDL values of increasing variance as the column index increases.

Usage

```
LDL_HDLVrnc(dfHDL, CHOL, TG, bootStrpReps = 2000)
```

Arguments

`dfHDL`	A data frame where each column contains a different set of HDL values of increasing variance as the column index increases.
`CHOL`	A vector or data frame column containing the cholesterol values to be used for the calculation of the variance of LDL.
`TG`	A vector or data frame column containing the triglyceride values to be used for the calculation of the variance of LDL.
`bootStrpReps`	(Default=2000) Number of bootstrap iterations for bootstrap variance calculation.

Value

The function returns a list with the Error Propagation variance (ErrPropVrnc) and the bootstrap variance (BootVrnc). Each list element is a vector of length equal to the number of columns of the HDL data frame supplied as argument (dfHDL) and each vector value corresponds to the respective variance of the corresponding data frame columns.

Examples

```
## Not run:
dfHDL = CV_Range(sampleA$HDL, 0, 100, maxRandIter = 4000, plot=FALSE)
LDLHDLVar = LDL_HDLVrnc(dfHDL,sampleA$CHOL,sampleA$TG,bootStrpReps=2000)

## End(Not run)
```

`LDL_ML_Main` *Create, train, assess and return an ML prediction model.*

Description

This function reads data from a DATACSV.csv, or a data table file. It partitions them according to the partition parameter and labels them, trains the model (according to the ML method chosen and the first set of the partitioned data), assesses the model using the second set of the partition data and returns it.

Usage

```
LDL_ML_Main(DataCSV, partition, MLmethod, ReportMultiPlot = TRUE)
```

Arguments

`DataCSV` The .csv or a data table file, path containing the data with which the model will be trained and assessed. Must contain at least 4 columns, named "CHOL", "HDL", "TG" and "LDLd", through which the train data and the validation data will be extracted.

`partition` A value in the range (0,1) that stipulates what percentage of the input data will be used for training the model, while the remainder will be used to assess it.

`MLmethod` A string that stipulates the Machine Learning method ("lm","rlm","glmnet","earth","svmRadial","knn","g or "rf") that is to be used to train the prediction model with.

`ReportMultiPlot` A boolean that allows the user to select whether the LDL_ML_Main function will plot a diagram with 5 plots, relating different stats on the newly created model. Preset to TRUE.

Value

It initializes and returns the ML prediction model. In case of bad input, it will return either -2 (illegitimate partition input) or -3 (illegitimate ML method input).

Examples

```
model = LDL_ML_Main(SampleData,0.7,"lm",ReportMultiPlot=FALSE)
```

LDL_ML_Main_All_Models

Create, train, assess and return all ML prediction models

Description

This function reads data from a DATACSV.csv or a data.table file.It partitions them according to the partition parameter and labels them, trains all of the models, assesses them using the second set of the partition data, optionally plots some info relating the accuracy of the models and returns them for further use.

Usage

```
LDL_ML_Main_All_Models(
  DataCSV,
  partition,
  ReportMultiPlot = TRUE,
  ComparisonPlot = TRUE
)
```

Arguments

`DataCSV` The .csv or a data table file, path containing the data with which the model will be trained and assessed. Must contain at least 4 columns, named "CHOL", "HDL", "TG" and "LDLd", through which the train data and the validation data will be extracted.

`partition` A value in the range (0,1) that stipulates what percentage of the input data will be used for training the models, while the remainder will be used to assess them.

`ReportMultiPlot` A boolean that allows the user to select whether the LDL_ML_Main function will plot a diagram with 5 plots, relating different stats on the newly created model. Preset to TRUE.

`ComparisonPlot` A boolean that allows the user to select whether the LDL_ML_Main_All_Models function will plot a comparison plot, relating different stats on the newly created models. Preset to TRUE.

Value

It initializes and returns all ML prediction models. In case of bad input, it will return -2 (illegitimate partition input).

Examples

```
allModels = LDL_ML_Main_All_Models(SampleData, 0.8,ReportMultiPlot = FALSE,ComparisonPlot=FALSE)
```

`LDL_ML_Main_StackingAlgorithm`

Create, train, assess and return a Stacking Algorithm Machine Learning prediction model

---

Description

This function reads data from a DATACSV.csv or data table file. It partitions them according to the partition parameter and labels them, trains all of the models and 'stacks' them into one, assesses them using the second set of the partition data, optionally plots some info relating the accuracy of the models and returns them for further use.

Usage

```
LDL_ML_Main_StackingAlgorithm(
  DataCSV,
  partition,
  ReportMultiPlot = TRUE,
  ComparisonPlot = TRUE
)
```

Arguments

`DataCSV` The .csv or data table file, path containing the data with which the model will be trained and assessed. Must contain at least 4 columns, named "CHOL", "HDL", "TG" and "LDLd", through which the train data and the validation data will be extracted.

`partition` A value in the range (0,1) that stipulates what percentage of the input data will be used for training the model, while the remainder will be used to assess it.

`ReportMultiPlot` A boolean that allows the user to select whether the LDL_ML_Main function will plot a diagram with 5 plots, relating different stats on the newly created model. Preset to TRUE.

`ComparisonPlot` A boolean that allows the user to select whether the LDL_ML_Main_All_Models function will plot a comparison plot, relating different stats on the newly created models. Preset to TRUE.

Value

It initializes and returns the stacked algorithm prediction model. In case of bad input, it will return -2 (illegitimate partition input)

Examples

```
stackModel = LDL_ML_Main_StackingAlgorithm(SampleData,0.8,ReportMultiPlot=TRUE,ComparisonPlot=TRUE)
```

LDL_ML_predict *Predict LDL value(s)*

Description

This function predicts and returns predictions, based on the model previously trained.

Usage

```
LDL_ML_predict(model, data)
```

Arguments

model	The model with which the predictions will be made.
data	The data with which the predictions will be made, can either be a single set of (CHOL,HDL,TG) values or a data table of sets of said values.

Value

The predicted LDL value(s).

Examples

```
modelPrediction = LDL_ML_predict(model$model,data.table::data.table(CHOL=170.5,HDL=35.12,TG=175))
```

LDL_TGVrnc *Calculate LDL variance when triglyceride variance changes*

Description

This function calculates the variance of LDL using two methods: Error Propagation and Bootstrap when the triglyceride distribution changes. It uses a data frame where each column contains a different set of triglyceride values of increasing variance as the column index increases.

Usage

```
LDL_TGVrnc(dfTG, CHOL, HDL, bootStrpReps = 2000)
```

Arguments

`dfTG` A data frame where each column contains a different set of triglyceride values of increasing variance as the column index increases.

`CHOL` A vector or data frame column containing the cholesterol values to be used for the calculation of the variance of LDL.

`HDL` A vector or data frame column containing the HDL values to be used for the calculation of the variance of LDL.

`bootStrpReps` (Default=2000) Number of bootstrap iterations for bootstrap variance calculation.

Value

The function returns a list with the Error Propagation variance (ErrPropVrnc) and the bootstrap variance (BootVrnc). Each list element is a vector of length equal to the number of columns of the triglyceride data frame supplied as argument (dfTG) and each vector value corresponds to the respective variance of the corresponding data frame columns.

Examples

```
## Not run:
dfTG = CV_Range(sampleA$TG, 0, 100, maxRandIter = 3000, plot=FALSE)
LDLTGVar=LDL_TGVrnc( sampleA$CHOL,  sampleA$HDL, dfTG, bootStrpReps=2000)

## End(Not run)
```

`model` *A trained model*

Description

An already trained ML model (including trainData & testData) with the "lm" method using the default data from the package, to be used in the predict example, and potentially by the user for further predictions.

Usage

```
model
```

Format

An object of class `list` of length 3.

plotAIP_HDLVrnc *Plots AIP variance versus increasing HDL variance*

Description

Plots the variance of AIP (both Error Propagation first and second order, as well as Bootstrap variance) versus increasing HDL variance.

Usage

```
plotAIP_HDLVrnc(dfHDLVars, AIPVrncErrProp, AIPVrncErrProp2Ord, AIPbootVrnc)
```

Arguments

dfHDLVars	A data frame with increasing variances of HDL.
AIPVrncErrProp	A vector with the first order error propagation variances of AIP when the HDL variance increases.
AIPVrncErrProp2Ord	A vector with the second order error propagation variances of AIP when the HDL variance increases.
AIPbootVrnc	A vector with the bootstrap variances of AIP when the triglyceride variance increases.

Value

The function creates a plot with the AIP error propagation and bootstrap variances versus increasing triglyceride variance.

Examples

```
HDLVariances = CV_Range(sampleA$HDL,15,25,plot=FALSE)
AIPVrncChngHDLVrnc = AIP_HDLVrnc(HDLVariances,sampleA$TG, bootStrpReps=2000)
HDLErrPropVrnc = AIPVrncChngHDLVrnc$ErrPropVrnc
HDLErrPropVrnc2Ord = AIPVrncChngHDLVrnc$ErrPropVrnc2Ord
HDLBootVrnc = AIPVrncChngHDLVrnc$BootVrnc
plotAIP_HDLVrnc(HDLVariances,HDLErrPropVrnc,HDLErrPropVrnc2Ord,HDLBootVrnc)
```

plotAIP_TGVrnc *Plots AIP variance versus increasing triglyceride variance*

Description

Plots the variance of AIP (both Error Propagation first and second order, as well as Bootstrap variance) versus increasing triglyceride variance.

Usage

```
plotAIP_TGVrnc(TGVars, AIPVrncErrProp, AIPVrncErrProp2Ord, AIPbootVrnc)
```

Arguments

TGVars	A data frame with increasing variances of triglycerides.
AIPVrncErrProp	A vector with the first order error propagation variances of AIP when the triglyceride variance increases.
AIPVrncErrProp2Ord	A vector with the second order error propagation variances of AIP when the triglyceride variance increases.
AIPbootVrnc	A vector with the bootstrap variances of AIP when the triglyceride variance increases.

Value

The function creates a plot with the AIP error propagation and bootstrap variances versus increasing triglyceride variance.

Examples

```
# AIP - TG Variance
TGVariances = CV_Range(sampleA$TG,15,16,plot=FALSE)
AIPVrncChngTGVrnc = AIP_TGVrnc(TGVariances,sampleA$HDL,bootStrpReps = 2000)
TGErrPropVrnc = AIPVrncChngTGVrnc$ErrPropVrnc
TGErrPropVrnc2Ord = AIPVrncChngTGVrnc$ErrPropVrnc2Ord
TGBootVrnc = AIPVrncChngTGVrnc$BootVrnc
plotAIP_TGVrnc(TGVariances,TGErrPropVrnc,TGErrPropVrnc2Ord,TGBootVrnc)
```

plotCholVrncToLDL *Plots LDL variance versus increasing cholesterol variance*

Description

Plots the variance of LDL (both Error Propagation and Bootstrap variance) versus increasing cholesterol variance.

Usage

```
plotCholVrncToLDL(dfCholVars, LDLCHOLVrncErrProp, LDLCHOLVrncBoot)
```

Arguments

`dfCholVars` A data frame with increasing variances of cholesterol.

`LDLCHOLVrncErrProp` A vector with the error propagation variances of LDL when the cholesterol variance increases.

`LDLCHOLVrncBoot` A vector with the bootstrap variances of LDL when the cholesterol variance increases.

Value

The function creates a plot with the LDL error propagation and bootstrap variances versus increasing cholesterol variances.

Examples

```
# For HDL - CHOL:
# Make the DF of ascending variances for CHOL of sample
CHOLVariances = CV_Range(sampleA$CHOL,1,10,plot=FALSE)
LDLCHOLDependance = LDL_CHOLVrnc(CHOLVariances, sampleA$HDL, sampleA$TG, bootStrpReps=2000)
plotCholVrncToLDL(CHOLVariances,LDLCHOLDependance$ErrPropVrnc,LDLCHOLDependance$BootVrnc)
```

PlotCorrWithRegrLine *Scatterplot of pairs of parameters*

Description

Scatterplot of pairs of parameters with the corresponding regression line.

Usage

```
PlotCorrWithRegrLine(df, xParam, yParam)
```

Arguments

`df`	The data frame with the parameterrs to be plotted
`xParam`	The parameter (column name) to be plotted in the abscissa (x axis).
`yParam`	The prameter (column name) to be plotted in the ordinate (y axis).

Value

The function returns a ggplot2 object.

Examples

```
## Not run:
PlotCorrWithRegrLine(sampleA,"CHOL", "HDL")

## End(Not run)
```

`PlotDiscrHist` *Plot discrete histogram*

Description

Plot discrete histogram (barplot) of the data frame column named in the param argument. It also plots the mean as a vertical continuous line, the mean plus/minus 2 standard deviations as veritcal dotted lines and overlays a density plot of the normal distribution with mean and standard deviation corresponding to those of the data.

Usage

```
PlotDiscrHist(DF, param, title = "")
```

Arguments

`DF`	A data frame containing columns whose discrete histogram is to be plotted.
`param`	The column name of the columns for which to plot the discrete histogram.
`title`	Title of the plot (default="").

Value

It returns a ggplot object.

Examples

```
## Not run:
PlotDiscrHist(sampleA,"LDL")

## End(Not run)
```

`plotHDLVrncToLDL` *Plots LDL variance versus increasing HDL variance*

Description

Plots the variance of LDL (both Error Propagation and Bootstrap variance) versus increasing HDL variance.

Usage

```
plotHDLVrncToLDL(dfHDLVars, LDLHDLVrncErrProp, LDLHDLVrncBoot)
```

Arguments

`dfHDLVars` A data frame with increasing variances of HDL.

`LDLHDLVrncErrProp` A vector with the error propagation variances of LDL when the HDL variance increases.

`LDLHDLVrncBoot` A vector with the bootstrap variances of LDL when the HDL variance increases.

Value

The function creates a plot with the LDL error propagation and bootstrap variances versus increasing HDL variances.

Examples

```
## For LDL - HDL:
# Make the DF of ascending variances for HDL of sample
HDLVariances = CV_Range(sampleA$HDL,15,25,plot=FALSE)
# Get the Error Propagation and the Bootstrap variance of LDL relative to HDL
LDLHDLSampleDependance = LDL_HDLVrnc(HDLVariances,sampleA$CHOL, sampleA$TG, bootStrpReps=2000)
plotHDLVrncToLDL(HDLVariances,LDLHDLSampleDependance$ErrPropVrnc,LDLHDLSampleDependance$BootVrnc)
```

`plotTGVrncToLDL` *Plots LDL variance versus increasing triglyceride variance*

Description

Plots the variance of LDL (both Error Propagation and Bootstrap variance) versus increasing triglyceride variance.

Usage

```
plotTGVrncToLDL(dfTGVars, LDLTGVrncErrProp, LDLTGVrncBoot)
```

Arguments

`dfTGVars` A data frame with increasing variances of triglycerides.

`LDLTGVrncErrProp` A vector with the error propagation variances of LDL when the triglyceride variance increases.

`LDLTGVrncBoot` A vector with the bootstrap variances of LDL when the triglycerides variance increases.

Value

The function creates a plot with the LDL error propagation and bootstrap variances versus increasing triglyceride variances.

Examples

```
## For LDL - TG:
# Make the DF of ascending variances for TG of sample
TGVariances = CV_Range(sampleA$TG,15,16,plot=FALSE)
# Get the Error Propagation and the Bootstrap variance of LDL relative to TG
LDLTGSampleDependance = LDL_TGVrnc(TGVariances,sampleA$CHOL, sampleA$HDL, bootStrpReps =2000)
plotTGVrncToLDL(TGVariances,LDLTGSampleDependance$ErrPropVrnc,LDLTGSampleDependance$BootVrnc)
```

---

sampleA *Anonymous real data*

Description

A data set of repeated measurements of cholesterol, LDL and triglycerides. It is used as data for the examples.

Usage

```
sampleA
```

Format

A `data frame` with 10 elements, which are:

day Day number of each measurement

CHOLrep1, CHOLrep2, CHOL Replicate measurements of cholesterol for each day and their mean.

HDLrep1, HDLrep2, HDL eplicate measurements of HDL for each day and their mean.

LDLrep1, LDLrep2, LDL LDL calculated according to the Friedewald formula mentioned above and its mean. All measurement units are in mg/dl and not mmol/L.

sampleB *Anonymous real data*

Description

A data set of repeated measurements of cholesterol, LDL and triglycerides. It is used as data for the examples.

Usage

```
sampleB
```

Format

A `data frame` with 10 elements, which are:

day Day number of each measurement

CHOLrep1, CHOLrep2, CHOL Replicate measurements of cholesterol for each day and their mean.

HDLrep1, HDLrep2, HDL eplicate measurements of HDL for each day and their mean.

LDLrep1, LDLrep2, LDL LDL calculated according to the Friedewald formula mentioned above and its mean. All measurement units are in mg/dl and not mmol/L.

SampleData *Anonymous real data*

Description

A data set for the test of your model, containing values of 2000 cases. It is used as data for the examples.

Usage

```
SampleData
```

Format

A `data frame` with 5 elements, which are:

AGE The Age of cases

CHOL The Cholesterol of cases

TG The Triglyceride of cases

HDL The High Density Lipoprotein of cases

LDLd The direct Low Density Lipoprotein of cases

Index