Package 'mdapack'

May 21, 2020

Title Medical Data Analysis Pack

Version 0.0.2

Depends R (>= 2.10)

Description An implementation of functions for medical data analysis which perform basic data pre processing,
univariate and bivariate analysis.Kirkwood et al. (2003) <doi:10.1002/sim.1961>.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

RoxygenNote 7.1.0

Imports stats, ggplot2, reshape2, devtools, roxygen2, knitr, testthat, rmarkdown, spelling, httr, covr, gh, rlang, openssl, git2r, pkgbuild, utils, curl, VIM, FactoMineR, mice, grDevices, graphics

VignetteBuilder knitr

NeedsCompilation no

Author Henry Nanji [aut, cre], Saisakul Chernbumroong [aut]

Maintainer Henry Nanji <henry.nanji@nottingham.ac.uk>

Repository CRAN

Date/Publication 2020-05-20 22:30:02 UTC

R topics documented:

`checkValidity` *A checkValidity Function*

Description

'checkValidity' checks the class of the different variables in the data frame and returns a list of numeric and factor or character variables. It also returns basic summary statistics(mean, median, min, max) for numeric and categorical/character variables It also returns the number of colunms and rows in the data frame.

Usage

```
checkValidity(dat, numeric = FALSE, cat = FALSE)
```

Arguments

`dat`	Data in data frame format.
`numeric`	Numeric data type.
`cat`	Categorical or factor data type.

Value

Returns a list of numeric and character(factor) variables, basic summary statistics and number of rows and columns in the data frame

Author(s)

Henry Nanji, Saisakul Chernbumroong

Examples

```
library(FactoMineR)
data(wine)
checkValidity(wine, numeric=TRUE)
```

checkVariableNames *A checkVariableNames Function*

Description

'checkVariableNames'checks if the given parameters' names and the columns' names of the data are the same.

Usage

```
checkVariableNames(raw, parameterName)
```

Arguments

raw	Data in data.frame format.
parameterName	Factor or characters of parameters' names.

Value

Returns a list of the correct variable names

Author(s)

Henry Nanji,Saisakul Chernbumroong

Examples

```
data(iris)
checkVariableNames(iris, c("Sepal.Length","Sepal.Width","Petal.Length","Petal.Width","Species"))
```

correctType *A correctType Function*

Description

'correctType' reads data from .csv file into the correct data type. This function processes the raw data into the correct data type.

Usage

```
correctType(raw, datatype, cols = NULL)
```

Arguments

raw	Data in data frame format.
datatype	The data type for each raw data column. Accepted data types are factor, numeric, integer, date, ordered, and character.
cols	If not all columns need correction, specify the column number that need correction.

Value

Returns the data set with the corrected data type.

Author(s)

Henry Nanji, Saisakul Chernbumroong

Examples

```
library(FactoMineR)
data(wine)
correctType(wine, "factor", 1)
```

---

discretize *A discretize Function*

Description

'discretize' function performs a basic discretization using binning.

Usage

```
discretize(x, no.bin = 5)
```

Arguments

x	Data frame.
no.bin	Number of bin. 'FD' to use the Freedman-Diaconis rule for identifying number of bins. The number of bin is 'max-min/2*IQR*n^-(1/3)'. 'ThreeStage' to discretize data into three stages. 1 if data is more than mu+sigma/2; 0=-1 if data is less than mu - sigma/2; otherwise 0.

Value

Return the discretized data

Author(s)

Saisakul Chernbumroong

Examples

```
data(iris)
iris.bin = discretize(iris, no.bin = 5)
```

`missingAnalysis` *A missingAnalysis Function*

Description

'missingAnalysis' function reveals the variables with missing values, the number of missing values for each variable, and in what combinations It also produces a plot for visualizing pattern of missing values and returns a data frame showing correlation between pairs of variables that has missing values It also returns a data frame with complete cases

Usage

```
missingAnalysis(
  dat,
  plot = FALSE,
  miss_pattern = FALSE,
  complete_dat = FALSE,
  miss_cor_pattern = FALSE
)
```

Arguments

`dat`	Data in data frame format.
`plot`	A plot of the missing data pattern.
`miss_pattern`	A data frame with missing data pattern for each variable.
`complete_dat`	A data frame with complete cases.
`miss_cor_pattern`	Correlation between variables with missing values.

Value

Returns a complete data frame, a tabulation of missing data pattern and missing data plot for combination of variables

Author(s)

Henry Nanji

Examples

```
library(VIM)
data(sleep)
missingAnalysis(sleep, complete_dat  = TRUE)
```

mutualInfo *A mutualInfo Function*

Description

'mutualInfo' function calculates mutual information of the two variables.

Usage

```
mutualInfo(x, y)
```

Arguments

x	Numerical or factor data.
y	Numerical or factor data.

Value

Returns the mutual information value.

Author(s)

Henry Nanji, Saisakul Chernbumroong

Examples

```
data(iris)
mutualInfo(iris[, 1], iris[, 5])
```

outlier_detection *Outlier detection function*

Description

'outlier_detection'visually detect and highlights outliers in a univariate continuous variable. The function fetches the values of data points that lie beyond the extremes of the whiskers(observations that lie outside of 1.5 * IQR.

Usage

```
outlier_detection(dat, ...)
```

Arguments

dat	A data frame for data values.
...	Other arguments.

Value

Returns a box plot showing the outliers for each variable.

Author(s)

Henry Nanji

Examples

```
library(FactoMineR)
data(wine)
outlier_detection(wine)
```

`plotBoxplotGroup` *A plotBoxplotGroup Function*

Description

'plotBoxplotGroup' function plots boxplots by group.

Usage

```
plotBoxplotGroup(dat, x, ...)
```

Arguments

`dat`	Data in frame format.
`x`	A grouping variable.
`...`	Other parameters.

Value

Returns a box plot for each grouping variable

Author(s)

Henry Nanji

Examples

```
data(iris)
plotBoxplotGroup(iris, 'Species')
```

plotHeatmapGroup *A plotHeatmapGroup Function*

Description

This function plots a heatmap of aall numerical variables in data frame to show the correlation.

Usage

```
plotHeatmapGroup(dat)
```

Arguments

dat	data in data.frame format.

Author(s)

Saisakul Chernbumroong, Henry Nanji

Examples

```
data(iris)
plotHeatmapGroup(iris[,1:4])
```

univariateAnalysis *A univariateAnalysis Function*

Description

This function produces univariate plots including histogram with density plot, box plot, q-q plot, and calculate the Shapiro-Wilk statistic for numeric data.

Usage

```
univariateAnalysis(
  dat,
  hist = FALSE,
  boxplot = FALSE,
  qqnorm = FALSE,
  shapiro = FALSE
)
```

Arguments

dat	data in data.frame format
hist	a plot of histogram
boxplot	a box plot showing distribution of the variable
qqnorm	a qnantile quantile plot
shapiro	a Shapiro-Wilk normality test.

Value

Return plots from univariate analysis.

Author(s)

Saisakul Chernbumroong, Henry Nanji

Examples

```
data(iris)
univariateAnalysis (iris, qqnorm  = TRUE)
```

Index