Package 'DescriptiveWH'

January 20, 2025

Title Descriptive Statistics

Depends R(>= 3.6)

Version 1.0.3

Description Exploratory analysis of a data base. Using the functions of this package is possible to filter the data set detecting atypical values (outliers) and to perform exploratory analysis through visual inspection or dispersion measures. With this package you can explore the structure of your data using several parameters at the same time joining statistical parameters with different graphics. Finally, this package aid to confirm or reject the hypothesis that your data structure presents a normal distribution. Therefore this package is useful to get a previous insight of your data before to carry out statistical analysis.

Imports stats (>= 3.6.3), graphics (>= 3.6.3)

License GPL-3

URL https://github.com/William-HC/DescriptiveWH

Encoding UTF-8

RoxygenNote 7.1.1

NeedsCompilation no

Author William Herrera-Caceres [aut, cre]

Maintainer William Herrera-Caceres <williamandresh@hotmail.com>

Repository CRAN

Date/Publication 2021-06-17 05:50:16 UTC

Contents

---

destatWH	*Descriptive statistics*

Description

This function is useful to see the basic statistical information of a variable.

Usage

```
destatWH(x, dec = NULL, op = FALSE)
```

Arguments

x	Numeric. A vector containing the data set.
dec	Numeric. A value defined to set the number of decimals in the returns.
op	Logical. If TRUE, returns plot, histogram, boxplot and qqnorm of the data set.

Details

This function shows statistical values including several visual tools.

Value

If op=FALSE (defined by default), the function returns a matrix (1 row, 9 columns) with values of the mean, the standart deviadtion, the minumun, the median, the maximum, the variance, the variation coefficient, the first quartil value and the third quartil value of the data set. Whereas if op=TRUE , the function returns four graphics: 1) A plot with X axis the number of samples and Y axis the amount of values included in the data set, 2) The density histogram of data set with a red line representing the distribution shape of the data, 3) A classical boxplot of the data, and 4) the qqnorm plot including the qqline used to check the normal disribution in a data set through visual inspection.

Examples

```
#Example 1
data=rnorm(1000,30,10)
destatWH( x=data ,dec=2 )
#Example 2
destatWH( x=data , op=TRUE )
```

---

distrWH   *Distribution of records*

Description

This function is useful to see the frequency of data and descriptive statistics

Usage

```
distrWH(x, dec = FALSE)
```

Arguments

x   Numeric. A vector containing the data set.

dec   Numeric. A value defined to set the number of decimals in the returns.

Details

This function shows statistical descriptive values of a vector.

Value

List including the table with the absolute frequency of data, also the mean, median, mode and range of data.

Examples

```
#Example 1
dat=c(1,5,2,4,5,1,3,5,3,5)
distrWH(x=dat,dec=3)
```

---

freqWH   *Distribution of frequency*

Description

Function used to compute the absolute and relative frecuency of a categorical variable.

Usage

```
freqWH(x, op = FALSE)
```

Arguments

x   Numeric or Character. A vector of the data set.

op   Logical. Optional argument to see the data structure in graphical representation.

Details

This function shows a table including the frequency of values from a categorical vector. This vector can be formed by numeric, character or mixed values, that is, numerical and character values. The table also includes the number of categories in which data is distributed. The absolute and relative frecuencies can be showed using histogram.

Value

A data frame is returned by default (op=FALSE), including the absolute and relative frequeincies of the data set. In addition, cumulative fequencies are included in this table. When "op=TRUE", the function returns only the absolute and relative frequencies in a graphical representation through histogram.

Examples

```
#Example 1
data1=c("A","A","A","A",2,3,2,3,"B","B","B")
freqWH(data1)
#Example 2
data2=c("A","A","B","A","C")
freqWH(data2)
#Example 3
data3=c(5,5,3,1,5,2,3,6,9,6,6)
freqWH(data3)
#Example 4
data4=c("A1","A2","B3","A2","C")
freqWH(data4)
```

---

meanWH   *Arithmetic, Harmonic, Geometric and Weighted Mean*

Description

Function used to compute different kind of mean in variables.

Usage

```
meanWH(x, meth = NULL, narm = FALSE, dec = NULL)
```

Arguments

x	Numeric. A vector of the data set.
meth	Character. To compute the arithmetic (meth="a"), harmonic ("h"), geometric ("g") or weighted ("w") mean. Only for weighted mean option, an additional column including the weight of values is required.
narm	Logical. Optional argument used to compute the mean removing NA values.
dec	Numeric. A value to define the number of decimals in the returns.

Details

This function shows the mean as a numerical value from the data set. For the case of weighted mean, an additional column must be included in the dataset containing the assigned weigths of each value. The weighted mean is computed taking into account the total percentage of weights, that is, for cases that the sumed of wights is 100% or different (i.e. 90%, 120%, etc).

Value

This function returns a numerical value representing the mean depending on used method.

Examples

```
#Example 1
data1=c(7,6,NA,8,NA)
meanWH(x=data1,meth="a",narm=TRUE)
#Example 2
data2=rnorm(10,50,3)
meanWH(x=data2,meth="a",dec=2)
#Example 3
data3=rnorm(100,50,11)
meanWH(x=data3,meth="g")
#Example 4
data4=rnorm(10,20,4)
meanWH(data4,meth="h")
#Example 5  (total weight is closed to 100%)
df1=data.frame(value=c(9,8,7,8),weight=c(0.5,0.2,0.2,0.1))
meanWH(df1,meth="w")
#Example 6  (total weight is not closed 100%)
df2=data.frame(value=c(9,8,7,8),weight=c(0.5,0.2,0.2,0.2))
meanWH(df2,meth="w")
```

---

outlierWH   *Outliers detection*

Description

The function used to detect atypical values in a data set

Usage

```
outlierWH(x, lower = NULL, upper = NULL, dec = NULL)
```

Arguments

x	Numeric. A vector of the data set.
lower	Numeric. A value representing the lower limit of outliers.
upper	Numeric. A value representing the upper limit of outliers.
dec	Numeric. A value to define the decimal numbers

Details

This function shows a list with the putative outliers and other without those. If you do not type lower and upper values, by default the value is assumed as the standard deviation of "x" multipied two.

Value

A list including three elements: 1) A data frame including the mean, the standart deviation, and the upper and lower limits of outliers, 2) A numeric vector with data considered no outliers, and 2) A numeric vector including the putative outliers.

Examples

```
#Example 1
data=rnorm(100,30,15)
outlierWH( x=data ,lower=15 ,upper=35 )
#Example 2
outlierWH( x=data )
#Example 3
outlierWH( x=data ,lower=15 )
```

Index