

Package ‘DescriptiveStats.OBeu’

February 2, 2017

Type Package

Title Descriptive Statistics OpenBudgets.eu

Version 1.1.5

Date 2016-09-18

Description Descriptive Statistics and other related analysis for OBeu datasets.

Author Kleanthis Koupidis

Contributor Aikaterini Chatzopoulou <kchatzopoul@okfn.gr>, Charalampos Bratsas <charalampos.bratsas@okfn.org>

Maintainer Kleanthis Koupidis <koupidis@okfn.gr>

URL <https://github.com/okgreece/DescriptiveStats.OBeu>

BugReports <https://github.com/okgreece/DescriptiveStats.OBeu/issues>

License GPL-2 | file LICENSE

LazyData true

Imports grDevices,jsonlite,reshape,stats

RoxygenNote 5.0.1

Suggests knitr, rmarkdown

VignetteBuilder knitr

R topics documented:

ds.analysis	2
ds.box	3
ds.boxplot	4
ds.correlation	4
ds.frequency	5
ds.hist	6
ds.kurtosis	7
ds.skewness	7
ds.statistics	8
nums	9
open_spending.ds	9
sample_json_link_openspending	10
sample_json_link_rudolf	11
Index	12

`ds.analysis`*Calculation of some Descriptive Tasks*

Description

The input of this function can be a matrix or data frame. The function calculates the basic descriptive measures, the correlation and the boxplot parameters of all the numerical variables, the frequencies of all the nominal variables and the necessary components of the selected generalized linear model.

Usage

```
ds.analysis(data, c.out=1.5, box.width=0.15, outliers=T, hist.class="Sturges",  
corr.method= "pearson", fr.select=NULL)
```

Arguments

<code>data</code>	The input data
<code>c.out</code>	Determines the length of the "whiskers" plot. If it is equal to zero no outliers will be returned.
<code>box.width</code>	The width level is determined 0.15 times the square root of the size of the input data.
<code>outliers</code>	If TRUE the outliers will be computed at the selected "c.out" level (default is 1.5 times the Interquartile Range).
<code>hist.class</code>	The method or the number of classes for the histogram.
<code>corr.method</code>	The correlation coefficient method to compute: "pearson" (default), "kendall" or "spearman".
<code>fr.select</code>	One or more nominal variables to calculate their corresponding frequencies.

Details

...

Value

A list with...

Author(s)

Kleanthis Koupidis

See Also

[open_spending.ds](#)

`ds.box`*Boxplot Parameters of a numeric vector*

Description

This function calculates the statistical measures needed to visualize the boxplot of a numeric vector.

Usage

```
ds.box(x, c=1.5, c.width = 0.15 , out = T)
```

Arguments

x	The input numeric vector
c	Determines the length of the "whiskers" plot. If it is equal to zero or out=F, no outliers will be returned.
c.width	The width level is determined 0.15 times the square root of the size of the input vector
out	If TRUE the outliers will be computed at the selected "c" level (default is 1.5 times the Interquartile Range).

Details

This function returns a list with the parameters needed to visualize a boxplot.

Value

Returns a list with the following components:

- lo.whisker The extreme of the lower whisker
- lo.hinge The lower "hinge"
- median The median
- up.hinge The upper "hinge"
- up.whisker The extreme of the upper whisker
- box.width The width of the box (default is 0.15 times the square root of the size of the vector)
- lo.out The values of any data points which lie below the extreme of the lower whisker
- up.out The values of any data points which lie above the extreme of the upper whisker
- n The non-NA observations of the vector

Author(s)

Kleanthis Koupidis

See Also

[ds.analysis](#), [open_spending.ds](#)

ds.boxplot

Boxplot Parameters of a matrix or data frame

Description

This function calculates the statistics of the boxplot for the input matrix or data frame.

Usage

```
ds.boxplot(data, out.level=1.5, width = 0.15 , outl = T)
```

Arguments

data	The input numeric matrix or data frame.
out.level	Determines the length of the "whiskers" plot. If it is equal to zero or "outl" is set to F, no outliers will be returned.
width	The width level is determined 0.15 times the square root of the size of the input data.
outl	If TRUE the outliers will be computed at the selected "out.level" level (default is 1.5 times the Interquartile Range).

Details

This function returns as a list object the statistical parameters needed to visualize boxplot.

Value

Returns a list with the extracted components of [ds.box](#) for each variable/column of the input data.

Author(s)

Aikaterini Chatzopoulou, Kleanthis Koupidis

See Also

[ds.box](#), [ds.analysis](#), [open_spending.ds](#)

ds.correlation

Correlation Coefficient of a dataframe

Description

This functions calculates the correlation coefficient of the input vectors, matrix or data frame. By default, the correlation coefficient of pearson is computed.

Usage

```
ds.correlation(x, y=NULL, cor.method="pearson")
```

Arguments

x	A vector, matrix or data frame
y	A vector, matrix or data frame
cor.method	The correlation coefficient method to compute: "pearson" (default), "kendall" or "spearman".

Details

This function returns an upper triangle matrix with the correlation coefficients of the input data. The correlation coefficient of pearson is computed, by default. Other options are "kendall" or "spearman".

Author(s)

Aikaterini Chatzopoulou, Kleanthis Koupidis

See Also

[ds.analysis](#), [open_spending.ds](#)

ds.frequency

Barplot parameters

Description

This function calculates the frequencies and the relative frequencies of factors/characters of the input dataset.

Usage

```
ds.frequency(data, select=NULL)
```

Arguments

data	A vector, matrix or data frame which includes at least one factor/character.
select	Select one or more specific nominal variables to calculate their corresponding frequencies, if it's not specified the result corresponds to frequencies of every factor variable in the data.

Details

This function returns a list with the frequencies and relative frequencies of factors/characters of the input dataset.

Author(s)

Kleanthis Koupidis

See Also

[ds.analysis](#), [open_spending.ds](#)

`ds.hist`*Histogram breaks and frequencies*

Description

This function computes the histogram parameters of the numeric input vector. The default for breaks is the value resulted from "Sturges" algorithm.

Usage

```
ds.hist(x, breaks="Sturges")
```

Arguments

x	The input numeric vector.
breaks	The method or the number of classes for the histogram.

Details

The possible values for breaks are "Sturges"(see `nclass.Sturges`), "Scott"(see `nclass.scott`), "FD" / "Freedman-Diaconis" (see `nclass.FD`) and other integer value.

Value

A list with the following components:

- cuts The boundaries of the histogram classes
- counts The frequency of each histogram class
- normal.curve The normal curve
- mean The average value of the input vector
- median The median value of the input data

Author(s)

Kleanthis Koupidis

See Also

[ds.analysis,open_spending.ds](#)

`ds.kurtosis`*Calculation of Kurtosis*

Description

This function calculates kurtosis of the input vector, matrix or data frame.

Usage

```
ds.kurtosis(x)
```

Arguments

`x` A vector, matrix or data frame.

Details

This function returns the kurtosis, based on a scaled version of the fourth moment, of numbers of the input data.

Author(s)

Aikaterini Chatzopoulou

See Also

[ds.skewness](#), [ds.statistics](#), [ds.analysis](#), [open_spending.ds](#)

`ds.skewness`*Calculation of Skewness*

Description

This function calculates skewness of the input vector, matrix or data frame.

Usage

```
ds.skewness(x)
```

Arguments

`x` A vector, matrix or data frame.

Details

This function returns the skewness, also known as Pearson's moment coefficient of skewness, of numbers of the input data.

Author(s)

Aikaterini Chatzopoulou

See Also

[ds.kurtosis](#), [ds.statistics](#), [ds.analysis](#), [open_spending.ds](#)

ds.statistics

Calculation of the Statistic Measures

Description

This function calculates the basic descriptive measures of the input dataset.

Usage

```
ds.statistics(data)
```

Arguments

data A numeric vector, matrix or data frame

Details

This function returns the following values of the input data: minimum, maximum, range, mean, median, first and third quantiles, variance, standart deviation, skewness and kurtosis.

Value

A json file with the following components:

- Min The minimum observed value of the input data
- Max The maximum observed value of the input data
- Range The range, defined as the difference of the maximum and the minimum value.
- Mean The average value of the input data
- Median The median value of the input data
- Quantiles The 25% and 75% percentiles
- Variance The variance of the input data
- Standard Deviation The standard deviation of the input data
- Skewness The Skewness of the input data
- Kurtosis The Kurtosis of the input data

Author(s)

Aikaterini Chatzopoulou, Kleanthis Koupidis

See Also

[open_spending.ds](#)

nums	<i>Select the numeric columns of a given dataset</i>
------	--

Description

Extract and return a data frame with the columns that include only numeric values

Usage

```
nums(data)
```

Arguments

data The input data frame

Value

This function returns a data frame with the numeric columns of the input dataset.

Author(s)

Kleanthis Koupidis

open_spending.ds	<i>Read and Calculate the Basic Information for Basic Descriptive Tasks from Open Spending API</i>
------------------	--

Description

Extract and analyze the input data provided from Open Spending API, using the ds.analysis function.

Usage

```
open_spending.ds(json_data, dimensions=NULL, amounts=NULL, measured.dimensions=NULL,
coef.outl=1.5, box.outliers=T, box.width=0.15,
cor.method= "pearson", freq.select=NULL)
```

Arguments

json_data	The json string, URL or file from Open Spending API
dimensions	The dimensions of the input data
amounts	The measures of the input data
measured.dimensions	The dimensions to which correspond amount/numeric variables
coef.outl	Determines the length of the "whiskers" plot. If it is equal to zero no outliers will be returned.
box.outliers	If TRUE the outliers will be computed at the selected "coef.outl" level (default is 1.5 times the Interquartile Range).

box.width	The width level is determined 0.15 times the square root of the size of the input data.
cor.method	The correlation coefficient method to compute: "pearson" (default), "kendall" or "spearman".
freq.select	One or more nominal variables to calculate their corresponding frequencies.

Details

This function is used to read data in json format from Open Spending API, in order to implement some basic descriptive tasks through [ds.analysis](#) function.

Value

A json string with the resulted parameters of the [ds.analysis](#) function.

Author(s)

Kleanthis Koupidis

See Also

[ds.analysis](#)

sample_json_link_openspending

Sample data from Open Spending

Description

Sample data of Revised Budget phase amounts

- The year (2016) of the recorded approved budget phase amounts
- The revised budget phase amounts of 2016
- The original amounts of this year
- The functional classification description
- The functional classification code

Format

A link with the json format data

Source

OpenSpending

`sample_json_link_rudolf`*Sample data from Rudolf API*

Description

Sample data from Rudolf API

- The year (2016) of the recorded approved budget phase amounts
- The revised budget phase amounts of 2016
- The original amounts of this year
- The functional classification description
- The functional classification code

Format

A link with the json format data

Source

Rudolf

Index

`ds.analysis`, [2](#), [3–8](#), [10](#)
`ds.box`, [3](#), [4](#)
`ds.boxplot`, [4](#)
`ds.correlation`, [4](#)
`ds.frequency`, [5](#)
`ds.hist`, [6](#)
`ds.kurtosis`, [7](#), [8](#)
`ds.skewness`, [7](#), [7](#)
`ds.statistics`, [7](#), [8](#), [8](#)

`nums`, [9](#)

`open_spending.ds`, [2–8](#), [9](#)

`sample_json_link_openspending`, [10](#)
`sample_json_link_rudolf`, [11](#)