# Package 'DescriptiveStats.OBeu'

November 14, 2016

**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 <katerinachat@yahoo.com>

**Maintainer** Kleanthis Koupidis <koupidis.okfgr@gmail.com>

**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

## R topics documented:

`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 generilized linear model.

## Usage

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

## Arguments

| | |
|---|---|
| `data` | The input data |
| `c.out` | Determines the length of the "whiskers" plot. If it is equal to zero no outliers will be returned. |
| `box.width` | The width level is determined 0.15 times the square root of the size of the input data. |
| `outliers` | If TRUE the outliers will be computed at the selected "c.out" level (default is 1.5 times the Interquartile Range). |
| `corr.method` | The correlation coefficient method to compute: "pearson" (default), "kendall" or "spearman". |
| `fr.select` | 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 *Calculation of Statistics for the Boxplot*

## Description

This function calculates the statistics for the boxplot of the numeric vector for visualization purposes.

## Usage

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

## Arguments

| | |
|---|---|
| x | The input matrix or data frame |
| c | Determines the length of the "whiskers" plot. If it is equal to zero 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 the statistics needed to visualize 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 *Calculation of Statistics for the Boxplot*

### Description

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

### Usage

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

### Arguments

| | |
|---|---|
| data | The input numeric vector, matrix or data frame. |
| out.level | Determines the length of the "whiskers" plot. If it is equal to zero 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 "c" level (default is 1.5 times the Interquartile Range). |

### Details

This function returns the statistics needed to visualize boxplots.

### Value

Returns a list with the extracted components of ds.box for each variable of the input data.

### Author(s)

Aikaterini Chatzopoulou, Kleanthis Koupidis

### See Also

ds.box, ds.analysis, open_spending.ds

## ds.correlation *Correlation Coefficient*

### 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 *Frequencies of nominal variables*

---

### Description

This function calculates 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 | One or more nominal variables to calculate their corresponding frequencies. |

### Details

This function returns a data frame with the frequencies of factors/characters of the input dataset.

### Author(s)

Kleanthis Koupidis

### See Also

ds.analysis, open_spending.ds

ds.glm *Generalized Linear Models*

## Description

ds.glm is used to fit generalized linear models through glm from stats package and return some results in json format.

## Usage

```
ds.glm(x, dependent=NULL, independent=NULL, distr.family = "gaussian")
```

## Arguments

| | |
|---|---|
| x | The input matrix or data frame |
| dependent | The dependent variables of the model |
| independent | The independent variables of the model |
| distr.family | A character string naming a the error distribution and link function to be used in the model(See family for details of family functions.) |

## Details

Generalized linear models are used to fit data, using glm stats package, by specifying the dependent and independent variables and the description of the distribution error (default is gaussian). If user provides only the dependent variables, the rest variables are selected as independent. If user provides the independent variables, the rest variables are selected as dependent. If user provides the independent variables, the two variables with the highest correlation are selected.

## Value

coefficients residuals fitted residuals.degfred qq.plot linear.predictors

## 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 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 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 min, max, range, mean, median, 0%,25%,50%,75%,100% quantiles variance, standartdeviation, skewness and kurtosis of the input data.

## 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 0%,25%,50%,75%,100% percentiles
- Variance: The variance of the input data
- StandardDeviation: The standard deviation of the input data
- Skewness: The Skewness of the input data
- Kurtosis: The Kurtosis of the input data

## Author(s)

Katerina Chatzopoulou, Kleanthis Koupidis

## See Also

`open_spending.ds`

## nums — *Select the numeric columns of a given dataset*

### 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 — *Read and Calculate the Basic Information for Basic Descriptive Tasks from Open Spending API*

### Description

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

### Usage

```
open_spending.ds(json_data,  what=NULL, to.what=NULL,
coef.outl=1.5, box.outliers=T, box.wdth=0.15,
cor.method= "pearson", select=NULL)
```

### Arguments

| | |
|---|---|
| json_data | The json string, URL or file from Open Spending API |
| what | ... |
| to.what | ... |
| 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.wdth | 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". |
| `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_df_rudolf` *Sample data from Rudolf*

---

## Description

Sample data of Revised Budget phase amounts of Municipality of Athens

- 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 json format file

## Source

Rudolf

## 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*

### Description

Sample data of Revised Budget phase amounts of Municipality of Athens

- 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 to json format file

### Source

Rudolf

`sample_json_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 json format file

## Source

OpenSpending

# Index