Package 'exstat'

December 13, 2022

Title Designed for the statistical description of data

Version 1.0.0

Description It is designed for the statistical description of data.
This module provides functions for creating the basic descriptive
statistics table, testing normality distribution of variable,
screening extreme values, comparing the size between/among groups,
and calculating correlation coefficient between variables.
The visualization of the statistical results are also provided.

License GPL (>= 3)

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.2.2

Imports httr,vroom,ggplot2,readxl,gridExtra,gt

NeedsCompilation no

Author Yanqiu Feng [aut, cre],
bin Wang [aut, cph]

Maintainer Yanqiu Feng <1610306103@pku.edu.cn>

R topics documented:

InitStat *Initialize ExpoStat Module*

Description

The first step to start ExpoStat Module

Usage

```
InitStat()
```

Value

An R6 class object.

Author(s)

Yanqiu Feng, Bin Wang (corresponding author)

Examples

```
res = InitStat()
```

LoadStat *Upload data file for ExpoStat Module*

Description

Upload data file for ExpoStat Module

Usage

```
LoadStat(PID, UseExample = "default", DataPath = NULL, VocaPath = NULL)
```

Arguments

PID	chr. Program ID. It must be the same with the PID generated by InitStat.
UseExample	chr. Whether uses example data for analyses, available option include "example#1" for using example data1 and "default" for using data.
DataPath	chr. Input file directory, e.g. "D:/test/expostat_data.xlsx". It should be noted that the slash symbol is "/", not "\".
VocaPath	chr. Input file directory, e.g. "D:/test/expostat_voca.xlsx". It should be noted that the slash symbol is "/", not "\".

Value

A list object containing imported data.

Author(s)

Yanqiu Feng, Bin Wang (corresponding author)

Examples

```
res = InitStat()
   res1 = LoadStat(PID = res$PID, UseExample = "example#1")
```

`StatComp` *Size comparison between groups*

Description

Size comparison between groups

Usage

```
StatComp(PID, OutPath="default", Group, Task = "mean",
   Vars, VarsBy, Method = "wilcox", Layout = "density", Brightness = "dark" ,
   Palette = "default1")
```

Arguments

`PID`	chr. Program ID. It must be the same with the PID generated by InitStat.
`OutPath`	chr. Output file directory, e.g. "D:/ExpoStat/StatComp". It should be noted that the slash symbol is "/", not "\".If "default", the current working directory will be set.
`Group`	lgl. Whether to separate dataset into train and test data for normality test. The default is "TRUE".
`Task`	chr. Comparison task. At present, only the mean comparison is available.
`Vars`	chr. Target variables used for modelling. It should be noted that there is fixed format for the entering characters separated with "," and without space. The default values is "all" (all variables are included).
`VarsBy`	chr. Variable used to group the observation for size comparison.
`Method`	chr. Comparison method. At present, only "wilcox" (Wilcoxon rank sum test) is available.
`Layout`	chr. Visualization layout. Available values include "column.points", "density".
`Brightness`	chr. Visualization brightness. Available values include "light" and "dark".
`Palette`	chr. Visualization palette. Available values include "default1", "default2" , "default3" and 5 journal option including "cell", "nature", "science", "lancet", "nejm".

Value

A list object containing the results of size comparisons between groups for variables and visualization of the results.

Author(s)

Yanqiu Feng, Bin Wang (corresponding author)

Examples

```
res = InitStat()
   res1 = LoadStat(PID = res$PID, UseExample = "example#1")
   res2 = StatComp(PID=res$PID, Group = T, Task = "mean", Vars = "X5,X6,X7,X8,X9",
   VarsBy = "Y1", Method = "wilcox", Layout = "density", Brightness = "dark",
   Palette = "default1")
```

StatCorr *Correlation analysis between variables*

Description

Correlation analysis between variables

Usage

```
StatCorr(PID, OutPath="default", Group, VarsX, VarsY, VarsBy,
  Method = "spearman", Layout= "bubble", Brightness = "dark" , Palette ="default1")
```

Arguments

PID	chr. Program ID. It must be the same with the PID generated by InitStat.
OutPath	chr. Output file directory, e.g. "D:/ExpoStat/StatCorr". It should be noted that the slash symbol is "/", not "\".If "default", the current working directory will be set.
Group	lgl. Whether to separate dataset into train and test data for normality test. The default is "TRUE".
VarsX	chr. Target variables used for modelling. It should be noted that there is fixed format for the entering characters separated with "," and without space. The default values is "all.x" (all variables are included).
VarsY	chr. Target outcome variables used for correlation analysis.
VarsBy	chr. Variable used to group the observation for correlation analysis.
Method	chr. Method for orrelation analysis. Available values include "spearman" (Spearman's rank correlation analysis) and "pearson" (Pearson correlation analysis).
Layout	chr. Visualization layout. Available values include "heatmap", "bubble", "matrix".

Brightness	chr. Visualization brightness. Available values include "light" and "dark".
Palette	chr. Visualization palette. Available values include "default1", "default2" , "default3" and 5 journal option including "cell", "nature", "science", "lancet", "nejm".

Value

A list object containing the results of correlation analysis between variables and visualization of the results.

Author(s)

Yanqiu Feng, Bin Wang (corresponding author)

Examples

```
res = InitStat()
   res1 = LoadStat(PID = res$PID, UseExample = "example#1")
   res2 = StatCorr(PID = res$PID, Group = T, VarsX = "X5,X6,X7,X8,X9", VarsY = "Y1",
   VarsBy = "Y1", Method = "pearson", Layout = "bubble", Brightness = "dark",
   Palette = "nature")
```

StatDesc *Variable description*

Description

Variable description

Usage

```
StatDesc(PID, OutPath="default", Group, Vars, VarsBy, Layout="box",
   Brightness="light", Palette="default1")
```

Arguments

PID	chr. Program ID. It must be the same with the PID generated by InitStat.
OutPath	chr. Output file directory, e.g. "D:/ExpoStat/StatDesc". It should be noted that the slash symbol is "/", not "\".If "default", the current working directory will be set.
Group	lgl. Whether to separate dataset into train and test data for normality test. The default is "TRUE".
Vars	chr. Target variables used for modelling. It should be noted that there is fixed format for the entering characters separated with "," and without space. The default values is "all" (all variables are included).
VarsBy	chr. Variable used to group the observation for size description.

Layout	chr. Visualization layout. Available values include "box", "violin".
Brightness	chr. Visualization brightness. Available values include "light" and "dark".
Palette	chr. Visualization palette. Available values include "default1", "default2" and 5 journal option including "cell", "nature", "science", "lancet", "nejm".

Value

A list object containing the results of variable description for continuous and discrete variables respectively and visualization of the continuous variables.

Author(s)

Yanqiu Feng, Bin Wang (corresponding author)

Examples

```
res = InitStat()
   res1 = LoadStat(PID = res$PID, UseExample = "example#1")
   res2 = StatDesc(PID = res$PID, Group = T, Vars = "C1,C2,X5,X6,X7,X8,X9",
   VarsBy = NULL, Layout = "box", Brightness = "dark", Palette = "default2")
```

StatExtre *Extreme value calculation*

Description

Extreme value calculation

Usage

```
StatExtre(PID, OutPath="default", Group, Vars, LimitLow = 0.025,
   LimitUpper = 0.975, Layout = "column.points", Brightness = "light",
   Palette = "default2")
```

Arguments

PID	chr. Program ID. It must be the same with the PID generated by InitStat.
OutPath	chr. Output file directory, e.g. "D:/ExpoStat/StatExtre". It should be noted that the slash symbol is "/", not "\".If "default", the current working directory will be set.
Group	lgl. Whether to separate dataset into train and test data for normality test. The default is "TRUE".
Vars	chr. Target variables used for modelling. It should be noted that there is fixed format for the entering characters separated with "," and without space. The default values is "all" (all variables are included).
LimitLow	num. Lower limit ratio to screen the small extreme values located from 0 to this lower limit of the target variables.

LimitUpper	num. Upper limit ratio to screen the large extreme values located from this lower limit to 1 of the target variables.
Layout	chr. Visualization layout . Available values include "column.points", "heatmap".
Brightness	chr. Visualization brightness . Available values include "light" and "dark".
Palette	chr. Visualization palette . Available values include "default1", "default2", "default3" and 5 journal option including "cell", "nature", "science", "lancet", "nejm".

Value

A list object containing the results of extremum for variables and visualization of the results.

Author(s)

Yanqiu Feng, Bin Wang (corresponding author)

Examples

```
res = InitStat()
   res1 = LoadStat(PID = res$PID, UseExample = "example#1")
   res2 = StatExtre(PID = res$PID, Group = T, Vars = "X5,X6,X7,X8,X9",
   LimitLow = 0.025, LimitUpper = 0.975, Layout = "column.points",
   Brightness = "dark", Palette = "default2")
```

StatNorm *Normality test for numeric variables*

Description

Normality test for numeric variables

Usage

```
StatNorm(PID, OutPath="default", Group, Vars, Method = "shapiro.test", Layout = "rose.chart",
   Brightness = "light", Palette = "default1")
```

Arguments

PID	chr. Program ID. It must be the same with the PID generated by InitStat.
OutPath	chr. Output file directory, e.g. "D:/ExpoStat/StatNorm". It should be noted that the slash symbol is "/", not "\".If "default", the current working directory will be set.
Group	lgl. Whether to separate dataset into train and test data for normality test. The default is "TRUE".
Vars	chr. Target variables used for modelling. It should be noted that there is fixed format for the entering characters separated with "," and without space. The default values is "all" (all variables are included).

Method	chr. Normality test method. Only "shapiro.test" method is available at present.
Layout	chr. Visualization layout. Available values include "column", "column.points", "rose.chart", and "density".
Brightness	chr. Visualization brightness. Available values include "light" and "dark".
Palette	chr. Visualization palette. Available values include "default1", "default2", "default3" and 5 journal option including "cell", "nature", "science", "lancet", "nejm".

Value

A list object containing the results of normality test for variables and visualization of the results.

Author(s)

Yanqiu Feng, Bin Wang (corresponding author)

Examples

```
res = InitStat()
   res1 = LoadStat(PID = res$PID, UseExample = "example#1")
   res2 = StatNorm(PID=res$PID, Group = T, Vars = 'X5,X6,X7,X8,X9',
   Method = "shapiro.test",Layout = "rose.chart" , Brightness = "dark",
   Palette = "default3")
```

StatTable1 *Create Table 1 for for different epidemilogical study designs*

Description

Create Table 1 for different epidemilogical study designs

Usage

```
StatTable1(PID, OutPath="default", EpiDesign = "cohort",
   Group, VarsY, VarsC, Missing = "ifany")
```

Arguments

PID	chr. Program ID. It must be the same with the PID generated by InitStat.
OutPath	chr. Output file directory, e.g. "D:/ExpoStat/StatTable1". It should be noted that the slash symbol is "/", not "\".If "default", the current working directory will be set.
EpiDesign	chr. Research types provided for users, include "cohort" , "case-control", "cross-section".
Group	lgl. Whether to separate dataset into train and test data for creating Table 1. The default is "TRUE".
VarsY	chr. Outcome variable used for modelling. Only one variable can be entered.

VarsC	chr. Covariate variables needing further statistical test. It should be noted that there is fixed format for the entering characters separated with "," and without space. The defaults value is all covariate variables listed in the data file, which can be entered with "all.c".
Missing	chr. Counts of missing values in the table, available options include are "no" (never display missing values), "ifany" (only display if any missing values), and "always" (includes missing count row for all variables). Default is "ifany".

Value

A list object containing standardized table 1.

Author(s)

Yanqiu Feng, Bin Wang (corresponding author)

Examples

```
res = InitStat()
   res1 = LoadStat(PID = res$PID, UseExample = "example#1")
   res2 = StatTable1(PID = res$PID, EpiDesign = "cohort" ,
   Group = 'T', VarsY = "Y1", VarsC = "C1,C2,C3,C4,C5,C6",
   Missing = "ifany")
```

Index