Package ‘CBCgrps’

July 27, 2018

Type Package

Title Compare Baseline Characteristics Between Groups

Version 2.3

Date 2018-07-27

Author Zhongheng Zhang, 
Sir Run-Run Shaw hospital, 
Zhejiang university school of medicine

Maintainer Zhongheng Zhang <zh_zhang1984@zju.edu.cn>

Depends R (>= 3.2.0), nortest (>= 1.0-4)

Description Compare baseline characteristics between two or more groups. The variables being compared can be factor and numeric variables. The function will automatically judge the type and distribution of the variables, and make statistical description and bivariate analysis.

License GPL-2

NeedsCompilation no

Repository CRAN

Date/Publication 2018-07-27 08:20:02 UTC

R topics documented:

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBCgrps2.0-package</td>
<td>2</td>
</tr>
<tr>
<td>df</td>
<td>2</td>
</tr>
<tr>
<td>multigrps</td>
<td>3</td>
</tr>
<tr>
<td>twogrps</td>
<td>5</td>
</tr>
</tbody>
</table>

Index 8
Description

The package aims to automate the process of comparing Baseline Characteristics between groups.

Details

The DESCRIPTION file: In clinical studies employing electronic medical records, the variables to be investigated are usually large in number. It is sometimes cumbersome to compare these variables between two or more groups one by one. I design this package to automate the process of statistical description and bivariate statistical inference.

Author(s)

Zhongheng Zhang Department of emergency medicine, Sir Run-Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, 310016, China. <zh_zhang1984@zju.edu.cn> Maintainer: Zhongheng Zhang

References


See Also

No other reference.

Examples

data(df)
a<-twogrps(df,"mort")

Description

A data frame with 1000 observations on the following 7 variables.

Usage

data("df")
**Format**

A data frame with 1000 observations on the following 7 variables.

- **crp** a numeric vector, C-reactive protein measured in mg/l
- **hb** a numeric vector, hemoglobin
- **ddimer** a numeric vector
- **wbc** a numeric vector, white blood cell
- **comorbid** a factor with levels cirrhosis COPD diabetes heartfailure hypertension renalfailure stroke
- **sex** a factor with levels female male
- **mort** a factor with levels alive dead

**Details**

The dataset is generated as a working example without clinical relevance.

**Source**

Simulated dataset without sources.

**References**

simulated dataset without reference.

**Examples**

```r
data(df)
## maybe str(df); plot(df) ... 
```

---

**multigrps**  
*Compare Baseline Characteristics between three or more groups*

**Description**

The main function of the CBCgrps package.

**Usage**

```r
multigrps(df, gvar, p.rd = 3, normtest = "yes", norm.rd = 2, sk.rd = 2, tabNA = "no", cat.rd = 0, maxfactorlevels = 30, minfactorlevels = 10, sim = FALSE, workspace = 2e+05)
```
Arguments

- **df**: The data frame on which statistical description and inference are performed.
- **gvar**: The group variable.
- **p.rd**: Decimal space of p value to be displayed.
- **normtest**: Whether test for normal distribution is performed or not. "yes" for turning on the normality test and "no" for not testing normality.
- **norm.rd**: Decimal space of normally distributed numeric variables to be displayed.
- **sk.rd**: Decimal space of skewed numeric variables to be displayed.
- **tabNA**: Whether categorical variables with NA be displayed or not. "no" to be omitted, "ifany" to be displayed. The default value is "no".
- **cat.rd**: Decimal space of categorical variables (the proportion) to be displayed.
- **maxfactorlevels**: The maximum levels for factor variables, the default is 30. The argument is used to avoid treating date or time variables as factor variables.
- **minfactorlevels**: If a numeric variable has only several values, it is treated as categorical variable. The default value is 10.
- **sim**: a logical indicating whether to compute p-values by Monte Carlo simulation, in larger than 2 by 2 tables. The default is FALSE.
- **workspace**: If the fisher.test() function requires more workspace, it can be defined here. The default is workspace equals to 2e+05.

Details

The function compares differences in categorical and continuous variables between three or more groups. The function automatically judges the distribution of the continuous variable and uses appropriate description for them. Chi-square test is used for categorical data. Analysis of variance is used for normally distributed numeric data. Kruskal-Wallis rank sum test is used for non-normally distributed data. It is common that some categorical variables contain numeric or integer values. For example, the gender variable may contain values 1 and 2, representing male and female respectively. Such a variable can be identified by counting the number of integer values. Thus, the minfactorlevels argument is used to define the minimum value for a variable to be considered as numeric variable. For some large dataset, the normality test is extremely sensitive that a small deviation from normal distribution will lead to the rejection of the null hypothesis of normality. In such a circumstance, users may opt to switch off the normality test. Setting "yes" in the normtest argument for turning on the normality test and "no" for not testing normality.

Value

- **table**: The table with string values. The mean and standard error are put in a single cell, and connected by plus and minus symbol.
- **table.norm**: The table contains only normally distributed numeric variables. Mean and standard deviation are put in separate cells.
- **table.skew**: The table contains only skewed numeric variables. Median and interquartile range are put in separate cells.
The table contains only factor variables. Number and proportion are put in separate cells.

g1     level of group 1

g2     level of group 2

Note
No further notes

Author(s)
Zhongheng Zhang Department of emergency medicine, Sir Run-Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, 310016, China. <zh_zhang1984@zju.edu.cn>

References

See Also
No other reference

Examples

data(df)
b<-multigrps(df,"comorbid")

twogrps  Compare Baseline Characteristics between two groups

Description
The main function of the CBCgrps package. The function compares differences in categorical and continuous variables between two groups. The function automatically judges the distribution of the continuous variable and use appropriate description for them. Wilcoxon rank sum test is employed for non-normal data.

Usage

twogrps(df, gvar, p.rd = 3, normtest = "yes", norm.rd = 2, sk.rd = 2, tabNA = "no", cat.rd = 0, maxfactorlevels = 30, minfactorlevels = 10, sim = FALSE, workspace = 2e+05)
Arguments

df The data frame on which statistical description and inference are performed.
gvar The group variable.
p.rd Decimal space of p value to be displayed.
normtest Whether test for normal distribution is performed or not. For some large dataset, the normality test is extremely sensitive that a small deviation from normal distribution will lead to the rejection of the null hypothesis of normality. In such a circumstance, users may opt to switch off the normality test. "yes" for turning on the normality test and "no" for not testing normality.
norm.rd Decimal space of normally distributed numeric variables to be displayed.
sk.rd Decimal space of skewed numeric variables to be displayed.
tabNA Weather categorical variables with NA be displayed or not. "no" to be omitted, "ifany" to be displayed. The default value is "no".
cat.rd Decimal space of categorical variables (the proportion) to be displayed.
maxfactorlevels The maximum levels for factor variables, the default is 30. The argument is used to avoid treating date or time variables as factor variables.
minfactorlevels If a numeric variable has only several values, it is treated as categorical variable. The default value is 10.
sim A logical indicating whether to compute p-values by Monte Carlo simulation, in larger than 2 by 2 tables. The default is FALSE.
workspace If the fisher.test() function requires more workspace, it can be defined here. The default is workspace=2e+05.

Details

It is common that some categorical variables contain numeric or integer values. For example, the gender variable may contain values 1 and 2, representing male and female respectively. Such a variable can be identified by counting the number of integer values. Thus, the minfactorlevels argument is used to define the minimum value for a variable to be considered as numeric variable.

Value

table The table with string values. The mean and standard error are put in a single cell, and connected by plus and minus symbol.
table.norm The table contains only normally distributed numeric variables. Mean and standard deviation are put in separate cells.
table.skew The table contains only skewed numeric variables. Median and interquartile range are put in separate cells.
table.cat The table contains only factor variables. Number and proportion are put in separate cells.
g1 level of group 1
g2 level of group 2
Note
No further notes

Author(s)
Zhongheng Zhang Department of emergency medicine, Sir Run-Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, 310016, China. <zh_zhang1984@zju.edu.cn>

References

See Also
No other reference

Examples
```r
data(df)
a<-twogrps(df,"mort")```
Index

*Topic Compare
twogrps, 5
*Topic baseline
 multigrps, 3
twogrps, 5
*Topic bivariate analysis; statistical description
 CBCgrps2.0-package, 2
*Topic compare
 multigrps, 3
*Topic datasets
df, 2

CBCgrps (CBCgrps2.0-package), 2
CBCgrps2.0-package, 2
df, 2
multigrps, 3
twogrps, 5