Package 'EasyDescribe'

April 16, 2023

Type Package

Title A Convenient Way of Descriptive Statistics

Version 0.1.2

Date 2023-04-13

Author Xiuquan Nie [aut, cph, cre],
Mengyi Wang [ctb, rev],
Weihong Chen [fnd]

Maintainer Xiuquan Nie <niexiuquan1995@foxmail.com>

Depends R (>= 3.5.0)

Imports multiCA, CATT, gmodels, psych, rcompanion, FSA, fitdistrplus, nortest, clinfun, car

Suggests R.rsp

VignetteBuilder R.rsp

Description Descriptive Statistics is essential for publishing articles. This package can perform descriptive statistics according to different data types. If the data is a continuous variable, the mean and standard deviation or median and quartiles are automatically output; if the data is a categorical variable, the number and percentage are automatically output. In addition, if you enter two variables in this package, the two variables will be described and their relationships will be tested automatically according to their data types. For example, if one of the two input variables is a categorical variable, another variable will be described hierarchically based on the categorical variable and the statistical differences between different groups will be compared using appropriate statistical methods. And for groups of more than two, the post hoc test will be applied. For more information on the methods we used, please see the following references:
Libiseller, C. and Grimvall, A. (2002) <doi:10.1002/env.507>,
Patefield, W. M. (1981) <doi:10.2307/2346669>,
Hope, A. C. A. (1968) <doi:10.1111/J.2517-6161.1968.TB00759.X>,
Mehta, C. R. and Patel, N. R. (1983) <doi:10.1080/01621459.1983.10477989>,
Mehta, C. R. and Patel, N. R. (1986) <doi:10.1145/6497.214326>,
Clarkson, D. B., Fan, Y. and Joe, H. (1993) <doi:10.1145/168173.168412>,
Cochran, W. G. (1954) <doi:10.2307/3001616>,
Armitage, P. (1955) <doi:10.2307/3001775>,
Szabo, A. (2016) <doi:10.1080/00031305.2017.1407823>,

David, F. B. (1972) <doi:10.1080/01621459.1972.10481279>,
Joanes, D. N. and Gill, C. A. (1998) <doi:10.1111/1467-9884.00122>,
Dunn, O. J. (1964) <doi:10.1080/00401706.1964.10490181>,
Copenhaver, M. D. and Holland, B. S. (1988) <doi:10.1080/00949658808811082>,
Chambers, J. M., Freeny, A. and Heiberger, R. M. (1992) <doi:10.1201/9780203738535-5>,
Shaffer, J. P. (1995) <doi:10.1146/annurev.ps.46.020195.003021>,
Myles, H. and Douglas, A. W. (1973) <doi:10.2307/2063815>,
Rahman, M. and Tiwari, R. (2012) <doi:10.4236/health.2012.410139>,
Thode, H. J. (2002) <doi:10.1201/9780203910894>,
Jonckheere, A. R. (1954) <doi:10.2307/2333011>,
Terpstra, T. J. (1952) <doi:10.1016/S1385-7258(52)50043-X>.

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 7.2.3

NeedsCompilation no

Repository CRAN

Date/Publication 2023-04-16 07:50:09 UTC

R topics documented:

EasyDescribe-package 2
fundescribe 2
T2D 5

Index **6**

`EasyDescribe-package` *A Convenient Way of Descriptive Statistics*

Description

EasyDescribe provide a convenient way of descriptive statistics.

`fundescribe` *A Convenient Way of Descriptive Statistics*

Description

This function can perform descriptive statistics according to different data types.

Usage

```
fundescribe(x, y, data = NULL, na.rm = TRUE, norm.t = NULL)
```

Arguments

x	A vector or a factor. A continuous variable or a categorical variable.
y	A vector or a factor. A continuous variable or a categorical variable.
data	An optional parameter, the name of the data containing x and y.
na.rm	An optional parameter, if FALSE, the information of NA will be given.
norm.t	An optional parameter, there are seven normal test methods available: c("ks.test", "shapiro.test", "cvm.test", "lillie.test", "pearson.test", "sf.test", "ad.test").

Details

This function can perform descriptive statistics according to different data types. If the data is a continuous variable, the mean and standard deviation or median and quartiles are automatically output; if the data is a categorical variable, the number and percentage are automatically output. In addition, if you enter two variables in this function, the two variables will be described and their relationships will be tested automatically according to their data types. For example, if one of the two input variables is a categorical variable, another variable will be described hierarchically based on the categorical variable and the statistical differences between different groups will be compared using appropriate statistical methods. And for groups of more than two, the post hoc test will be applied.

Value

No return value, called for side effects.

Author(s)

Xiuquan Nie, niexiuquan1995@foxmail.com

References

Libiseller, C. and Grimvall, A. (2002) Performance of partial Mann-Kendall tests for trend detection in the presence of covariates. *Environmetrics*, 13, 71-84. doi:10.1002/env.507.

Patefield, W. M. (1981) An efficient method of generating r x c tables with given row and column totals. *Applied Statistics*, 30, 91-97. doi:10.2307/2346669

Hope, A. C. A. (1968) A simplified Monte Carlo significance test procedure. *Journal of the Royal Statistical Society Series B*, 30, 582-598. doi:10.1111/J.25176161.1968.TB00759.X

Mehta, C. R. and Patel, N. R. (1983) A network algorithm for performing Fisher's exact test in r x c contingency tables. *Journal of the American Statistical Association*, 78, 427-434. doi:10.1080/01621459.1983.10477989

Mehta, C. R. and Patel, N. R. (1986) Algorithm 643: FEXACT, a FORTRAN subroutine for Fisher's exact test on unordered r x c contingency tables. *ACM Transactions on Mathematical Software*, 12, 154-161. doi:10.1145/6497.214326

Clarkson, D. B., Fan, Y. and Joe, H. (1993) A Remark on Algorithm 643: FEXACT: An Algorithm for Performing Fisher's Exact Test in r x c Contingency Tables. *ACM Transactions on Mathematical Software*, 19, 484-488. doi:10.1145/168173.168412

Cochran, W. G. (1954) Some methods for strengthening the common chi-squared tests. *International Biometric Society*, 10 (4), 417-451. doi:10.2307/3001616

Armitage, P. (1955) Tests for Linear Trends in Proportions and Frequencies. *International Biometric Society*, 11 (3), 375-386. doi:10.2307/3001775

Szabo, A. (2016) Test for trend with a multinomial outcome. *American Statistician*, 73 (4), 313-320. doi:10.1080/00031305.2017.1407823

David, F. B. (1972) Constructing confidence sets using rank statistics. *Journal of the American Statistical Association*, 67, 687-690. doi:10.1080/01621459.1972.10481279

Joanes, D. N. and Gill, C. A. (1998) Comparing measures of sample skewness and kurtosis. *The Statistician*, 47, 183-189. doi:10.1111/14679884.00122

Dunn, O. J. (1964) Multiple comparisons using rank sums. *Technometrics*, 6, 241-252. doi:10.1080/00401706.1964.10490181

Copenhaver, M. D. and Holland, B. S. (1988) Computation of the distribution of the maximum studentized range statistic with application to multiple significance testing of simple effects. *Journal of Statistical Computation and Simulation*, 30, 1-15. doi:10.1080/00949658808811082

Chambers, J. M., Freeny, A. and Heiberger, R. M. (1992) *Statistical Models in S.* 49-52. doi:10.1201/97802037385355

Shaffer, J. P. (1995) Multiple hypothesis testing. *Annual Review of Psychology*, 46, 561-584. doi:10.1146/annurev.ps.46.020195.003021

Myles, H. and Douglas, A. W. (1973) *Nonparametric Statistical Methods.* 115-120. doi:10.2307/2063815

Rahman, M. and Tiwari, R. (2012) Pairwise comparisons in the analysis of carcinogenicity data. *Health*, 4, 910-918. doi:10.4236/health.2012.410139

Thode, H. J. (2002) *Testing for Normality.* doi:10.1201/9780203910894

Jonckheere, A. R. (1954) A distribution-free k-sample test again ordered alternatives. *Biometrika*, 41, 133-145. doi:10.2307/2333011

Terpstra, T. J. (1952) The asymptotic normality and consistency of Kendall's test against trend, when ties are present in one ranking. *Indagationes Mathematicae*, 14, 327-333. doi:10.1016/S1385-7258(52)50043X

Examples

```
data(T2D)
fundescribe(T2D$age, norm.t = c("lillie.test"))
fundescribe(gender, data = T2D)
fundescribe(education, diabetes, data = T2D)
fundescribe(glucose, age, data = T2D)
fundescribe(T2D$glucose, T2D$diabetes)
```

T2D *A data for 20 diabetes patients*

Description

A data for 20 diabetes patients. The data were fabricated.

Usage

```
T2D
```

Format

A data.frame contains 20 obs of 8 variables. The variables are:

ID The ID of these 20 people.

gender A character ("F" and "M").

age A numeric.

education A ordered factor ("1"<"2"<"3"<"4").

marriage A logical.

smoke A factor (0: "never smoker", 1: "current smoker", 2: "ex-smoker").

glucose A numeric.

diabetes A factor (0: "normal people", 1: "patient").

Index

∗ **datasets**
T2D, 5

EasyDescribe *(*EasyDescribe-package*)*, 2
EasyDescribe-package, 2

fundescribe, 2

T2D, 5