Package 'KbMvtSkew'

January 20, 2025

Type Package

Title Khattree-Bahuguna's Univariate and Multivariate Skewness

Version 1.0.2

Maintainer Zhixin Lun <zlun@oakland.edu>

Description Computes Khattree-Bahuguna's univariate and multivariate skewness, principal-component-based Khattree-Bahuguna's multivariate skewness. It also provides several measures of univariate or multivariate skewnesses including, Pearson's coefficient of skewness, Bowley's univariate skewness and Mardia's multivariate skewness. See Khattree, R. and Bahuguna, M. (2019) <doi:10.1007/s41060-018-0106-1>.

License GPL-3

Encoding UTF-8

LazyData true

Imports stats

Depends R (>= 3.6.0)

RoxygenNote 7.0.2

NeedsCompilation no

Author Zhixin Lun [aut, cre] (<https://orcid.org/0000-0002-8980-1554>),
Ravindra Khattree [aut] (<https://orcid.org/0000-0002-9305-2365>)

Repository CRAN

Date/Publication 2020-03-23 15:40:05 UTC

Contents

BowleySkew   *Bowley's Univariate Skewness*

Description

Compute Bowley's Univariate Skewness.

Usage

```
BowleySkew(x)
```

Arguments

x   a vector of original observations.

Details

Bowley's skewness is defined in terms of quantiles as

$$\hat{\gamma} = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$$

where Q_i is the ith quartile $i = 1, 2, 3$ of the data.

Value

BowleySkew gives the Bowley's univariate skewness of the data.

References

Bowley, A. L. (1920). *Elements of Statistics*. London : P.S. King & Son, Ltd.

Examples

```
# Compute Bowley's univariate skewness

set.seed(2019)
x <- rnorm(1000) # Normal Distribution
BowleySkew(x)

set.seed(2019)
y <- rlnorm(1000, meanlog = 1, sdlog = 0.25) # Log-normal Distribution
BowleySkew(y)
```

---

data-Chatterjee *Chatterjee, Hadi and Price Data*

Description

Chatterjee, Hadi and Price Data

Usage

```
data(Chatterjee)
```

Format

The format is a dataframe of 40 observations and 7 variables.

Source

The data come from Chatterjee, Hadi and Price (2000).

References

Chatterjee, S., Hadi, A. S., and Price, B. (2000). *Regression Analysis by Example*. Hoboken: Wiley.

Examples

```
data(Chatterjee)
```

---

data-OlymWomen *1984 Los Angeles Olympic records data of track events for women*

Description

Data are time records from 1984 Olympic track events for women from 55 countries: 100-meter, 200-meter, 400-meter, 800-meter, 1500-meter, 3000-meter, and Marathon. The corresponding variables are named as m100, m200, m400, m800, m1500, m3000, and marathon. The time measurements are recorded in seconds.

Usage

```
data(OlymWomen)
```

Format

The format is a dataframe of 55 observations and 8 variables.

Source

The data come from Khattree and Naik (2000, pp. 511-512).

References

Khattree, R. and Naik, D. (2000). *Multivariate Data Reduction and Discrimination with SAS®
Software*. Cary, NC: SAS Institute Inc.

Examples

```
data(OlymWomen)
```

---

data-SwissHead   *Measurements of Heads of Swiss Soldiers*

Description

Data are measurements, in millimeters, of the heads of 200 Swiss soldiers.

Usage

```
data(SwissHead)
```

Format

The format is a dataframe of 200 observations and 6 variables.

Source

The data come from Flurry and Riedwyl (1988).

References

Flurry, B. and Riedwyl, H. (1988). *Multivariate Statistics: A Practical Approach*. London: Chap-
man and Hall.

Examples

```
data(SwissHead)
```

kbMvtSkew	*Khattree-Bahuguna's Multivariate Skewness*

Description

Compute Khattree-Bahuguna's Multivariate Skewness.

Usage

```
kbMvtSkew(x)
```

Arguments

x   a matrix of original observations.

Details

Let $\mathbf{X} = (X_1, \ldots, X_p)'$ be the multivariate random vector and $(X_{i_1}, X_{i_2}, \ldots, X_{i_p})'$ be one of the $p!$ permutations of $(X_1, \ldots, X_p)'$. We predict X_{i_j} conditionally on subvector $(X_{i_1}, \ldots, X_{i_{j-1}})$ and compute the corresponding residual V_{i_j} through a linear regression model for $j = 2, \cdots, p$. For $j = 1$, we define $V_{i_1} = X_{i_1} - \bar{X}_{i_1}$, where $\bar{X}_{i_1}$ is the mean of X_{i_1}. For $j \geq 2$, we have

$$\hat{X}_{i_2} = \hat{\beta}_0 + \hat{\beta}_1 X_{i_1}, \quad V_{i_2} = X_{i_2} - \hat{X}_{i_2}$$

$$\hat{X}_{i_3} = \hat{\beta}_0 + \hat{\beta}_1 X_{i_1} + \hat{\beta}_2 X_{i_2}, \quad V_{i_3} = X_{i_3} - \hat{X}_{i_3}$$

$$\vdots$$

$$\hat{X}_{i_p} = \hat{\beta}_0 + \hat{\beta}_1 X_{i_1} + \hat{\beta}_2 X_{i_2} + \cdots + \hat{\beta}_{p-1} X_{i_{p-1}}, \quad V_{i_p} = X_{i_p} - \hat{X}_{i_p}.$$

We calculate the sample skewness $\hat{\delta}_{i_j}$ of V_{i_j} by the sample Khattree-Bahuguna's univariate skewness formula (see details of kbSkew that follows) respectively for $j = 1, \cdots, p$ and define $\hat{\Delta}_i = \sum_{j=1}^{p} \hat{\delta}_{i_j}, i = 1, 2, \ldots, P$ for all $P = p!$ permutations of $(X_1, \ldots, X_p)'$. The sample Khattree-Bahuguna's multivariate skewness is defined as

$$\hat{\Delta} = \frac{1}{P} \sum_{i=1}^{P} \hat{\Delta}_i.$$

Clearly, $0 \leq \hat{\Delta} \leq \frac{p}{2}$.

Value

kbMvtSkew computes the Khattree-Bahuguna's multivairate skewness for a p-dimensional data.

References

Khattree, R. and Bahuguna, M. (2019). An alternative data analytic approach to measure the univariate and multivariate skewness. *International Journal of Data Science and Analytics*, Vol. 7, No. 1, 1-16.

See Also

kbSkew for Khattree-Bahuguna's univariate skewness.

Examples

```
# Compute Khattree-Bahuguna's multivairate skewness

data(OlymWomen)
kbMvtSkew(OlymWomen[, c("m800","m1500","m3000","marathon")])
```

---

kbSkew   *Khattree-Bahuguna's Univariate Skewness*

Description

Compute Khattree-Bahuguna's Univariate Skewness.

Usage

```
kbSkew(x)
```

Arguments

x   a vector of original observations.

Details

Given a univariate random sample of size n consist of observations $x_1, x_2, \ldots, x_n$, let $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ be the order statistics of $x_1, x_2, \ldots, x_n$ after being centered by their mean. Define

$$y_i = \frac{x_{(i)} + x_{(n-i+1)}}{2}$$

and

$$w_i = \frac{x_{(i)} - x_{(n-i+1)}}{2}$$

The sample Khattree-Bahuguna's univariate skewness is defined as

$$\hat{\delta} = \frac{\sum y_i^2}{\sum y_i^2 + \sum w_i^2}.$$

It can be shown that $0 \leq \hat{\delta} \leq \frac{1}{2}$. Values close to zero indicate, low skewness while those close to $\frac{1}{2}$ indicate the presence of high degree of skewness.

Value

kbSkew gives the Khattree-Bahuguna's univariate skewness of the data.

References

Khattree, R. and Bahuguna, M. (2019). An alternative data analytic approach to measure the univariate and multivariate skewness. *International Journal of Data Science and Analytics*, Vol. 7, No. 1, 1-16.

Examples

```
# Compute Khattree-Bahuguna's univariate skewness

set.seed(2019)
x <- rnorm(1000) # Normal Distribution
kbSkew(x)

set.seed(2019)
y <- rlnorm(1000, meanlog = 1, sdlog = 0.25) # Log-normal Distribution
kbSkew(y)
```

MardiaMvtSkew   *Mardia's Multivariate Skewness*

Description

Compute Mardia's Multivariate Skewness.

Usage

```
MardiaMvtSkew(x)
```

Arguments

x   a matrix of original observations.

Details

Given a p-dimensional multivariate random vector with mean vector $\boldsymbol{\mu}$ and positive definite variance-covariance matrix $\boldsymbol{\Sigma}$, Mardia's multivariate skewness is defined as

$$\beta_{1,p} = E[(\boldsymbol{X}_1 - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_2 - \boldsymbol{\mu})]^3,$$

where $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are independently and identically distributed copies of $\boldsymbol{X}$. For a multivariate random sample of size n, $\boldsymbol{x}_1, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, its sample version is defined as

$$\hat{\beta}_{1,p} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} [(\boldsymbol{x}_i - \bar{\boldsymbol{x}})'\boldsymbol{S}^{-1}(\boldsymbol{x}_j - \bar{\boldsymbol{x}})]^3,$$

where the sample mean $\bar{\boldsymbol{x}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i$ and the sample variance-covariance matrix $\boldsymbol{S} = \frac{1}{n} \sum_{i=1}^{n} (\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})'$. It is assumed that $n \geq p$.

Value

`MardiaMvtSkew` gives the sample Mardia's multivairate skewness.

References

Mardia, K.V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57(3), 519–530.

Examples

```
# Compute Mardia's multivairate skewness

data(OlymWomen)
MardiaMvtSkew(OlymWomen[, c("m800","m1500","m3000","marathon")])
```

---

pcKbSkew	*Principal-component-based Khattree-Bahuguna's Multivariate Skewness*

Description

Compute Principal-component-based Khattree-Bahuguna's Multivariate Skewness.

Usage

```
pcKbSkew(x, cor = FALSE)
```

Arguments

x   a matrix of original scale observations.

cor   a logical value indicating whether the calculation should use the correlation matrix (cor = TRUE) or the covariance matrix (cor = FALSE). The default value is cor = FALSE.

Details

Let $\mathbf{X} = X_1, \ldots, X_p$ be a p-dimensional multivariate random vector. We compute the sample skewness for p principal components of $\mathbf{X}$ respectively by the sample Khattree-Bahuguna's univariate skewness formula (see details of [kbSkew](#) that follows). Let $\eta_1, \eta_2, \ldots, \eta_p$ be the p univariate skewnesses for p principal components. Principal-component-based Khattree-Bahuguna's multivariate skewness for a sample is then defined as

$$\eta = \sum_{i=1}^{p} \eta_i.$$

Clearly, $0 \leq \eta \leq \frac{p}{2}$.

Value

pcKbSkew gives the sample principal-component-based Khattree-Bahuguna's multivairate skewness.

References

Khattree, R. and Bahuguna, M. (2019). An alternative data analytic approach to measure the univariate and multivariate skewness. *International Journal of Data Science and Analytics*, Vol. 7, No. 1, 1-16.

See Also

kbSkew for Khattree-Bahuguna's univariate skewness.

Examples

```
# Compute principal-component-based Khattree-Bahuguna's multivairate skewness

data(OlymWomen)
pcKbSkew(OlymWomen[, c("m800","m1500","m3000","marathon")])
```

PearsonSkew	*Pearson's coefficient of skewness*

Description

Compute Pearson's coefficient of skewness.

Usage

```
PearsonSkew(x)
```

Arguments

x   a vector of original observations.

Details

Pearson's coefficient of skewness is defined as

$$\gamma_1 = \frac{E[(X - \mu)^3]}{(\sigma^3)}$$

where $\mu = E(X)$ and $\sigma^2 = E[(X - \mu)^2]$. The sample version based on a random sample $x_1, x_2, \ldots, x_n$ is defined as

$$\hat{\gamma_1} = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{ns^3}$$

where $\bar{x}$ is the sample mean and s is the sample standard deviation of the data, respectively.

Value

`PearsonSkew` gives the sample Pearson's univariate skewness.

References

Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Philos. Trans. R. Soc. Lond.* A 185, 71-110.

Pearson, K. (1895). Contributions to the mathematical theory of evolution II: skew variation in homogeneous material. *Philos. Trans. R. Soc. Lond.* A 86, 343-414.

Examples

```
# Compute Pearson's univariate skewness

set.seed(2019)
x <- rnorm(1000) # Normal Distribution
PearsonSkew(x)

set.seed(2019)
y <- rlnorm(1000, meanlog = 1, sdlog = 0.25) # Log-normal Distribution
PearsonSkew(y)
```

Index