Package ‘MVN’

January 30, 2019

Type Package

Title Multivariate Normality Tests

Version 5.6

Date 2019-01-30

Author Selcuk Korkmaz [aut, cre],
Dincer Goksuluk [aut],
Gokmen Zararsiz [aut]

Maintainer Selcuk Korkmaz <selcukorkmaz@gmail.com>

Depends R (>= 3.5.0)

VignetteBuilder knitr

Suggests knitr

Imports methods, nortest, moments, MASS, robustbase, mvoutlier, plyr, psych, boot, energy, kableExtra, magrittr, car

Collate mvn.R hzTest.R mardiaTest.R roystonTest.R

Description Performs multivariate normality tests and graphical approaches and implements multivariate outlier detection and univariate normality of marginal distributions through plots and tests, and performs multivariate Box-Cox transformation (Korkmaz et al, (2014), <https://journal.r-project.org/archive/2014-2/korkmaz-goksuluk-zararsiz.pdf>).

Encoding UTF-8

RoxygenNote 6.1.1

NeedsCompilation no

Repository CRAN

Date/Publication 2019-01-30 15:20:07 UTC

R topics documented:

`hzTest` *Deprecated Function for Henze-Zirklers's Multivariate Normality Test*

Description

Please use 'mvn' function instead

Usage

```
hzTest(data, cov = TRUE, qqplot = FALSE)
```

Arguments

`data`	a numeric matrix or data frame
`cov`	If `TRUE` covariance matrix is normalized by `n`, if `FALSE` it is normalized by `n-1`
`qqplot`	`TRUE` it creates chi-square Q-Q plot

`mardiaTest` *Deprecated Function for Mardia's Multivariate Normality Test*

Description

Please use 'mvn' function instead

Usage

```
mardiaTest(data, cov = TRUE, qqplot = FALSE)
```

Arguments

`data`	a numeric matrix or data frame
`cov`	If `TRUE` covariance matrix is normalized by `n`, if `FALSE` it is normalized by `n-1`
`qqplot`	`TRUE` it creates chi-square Q-Q plot

mvn *Multivariate Normality Tests*

Description

Performs multivariate normality tests, including Marida, Royston, Henze-Zirkler, Dornik-Haansen, E-Statistics, and graphical approaches and implements multivariate outlier detection and univariate normality of marginal distributions through plots and tests, and performs multivariate Box-Cox transformation.

Usage

```
mvn(data, subset = NULL, mvnTest = c("mardia", "hz", "royston", "dh",
  "energy"), covariance = TRUE, tol = 1e-25, alpha = 0.5,
  scale = FALSE, desc = TRUE, transform = "none", R = 1000,
  univariateTest = c("SW", "CVM", "Lillie", "SF", "AD"),
  univariatePlot = "none", multivariatePlot = "none",
  multivariateOutlierMethod = "none", bc = FALSE, bcType = "rounded",
  showOutliers = FALSE, showNewData = FALSE)
```

Arguments

`data`	a numeric matrix or data frame
`subset`	define a variable name if subset analysis is required
`mvnTest`	select one of the MVN tests. Type `"mardia"` for Mardia's test, `"hz"` for Henze-Zirkler's test, `"royston"` for Royston's test, `"dh"` for Doornik-Hansen's test and `energy` for E-statistic. Default is Henze-Zirkler's test `"hz"`. See details for further information.
`covariance`	this option works for `"mardia"` and `"royston"`. If `TRUE` covariance matrix is normalized by `n`, if `FALSE` it is normalized by `n-1`
`tol`	a numeric tolerance value which isused for inversion of the covariance matrix (`default = 1e-25`
`alpha`	a numeric parameter controlling the size of the subsets over which the determinant is minimized. Allowed values for the alpha are between 0.5 and 1 and the default is 0.5.
`scale`	if `TRUE` scales the colums of data
`desc`	a logical argument. If `TRUE` calculates descriptive statistics
`transform`	select a transformation method to transform univariate marginal via logarithm (`"log"`), square root (`"sqrt"`) and square (`"square"`).
`R`	number of bootstrap replicates for Energy test, default is 1000.
`univariateTest`	select one of the univariate normality tests, Shapiro-Wilk (`"SW"`), Cramer-von Mises (`"CVM"`), Lilliefors (`"Lillie"`), Shapiro-Francia (`"SF"`), Anderson-Darling (`"AD"`). Default is Anderson-Darling (`"AD"`). Do not apply Shapiro-Wilk's test, if dataset includes more than 5000 cases or less than 3 cases.

`univariatePlot`	select one of the univariate normality plots, Q-Q plot (`"qq"`), histogram (`"histogram"`), box plot (`"box"`), scatter (`"scatter"`)
`multivariatePlot`	`"qq"` for chi-square Q-Q plot, `"persp"` for perspective plot, `"contour"` for contour plot
`multivariateOutlierMethod`	select multivariate outlier detection method, `"quan"` quantile method based on Mahalanobis distance and `"adj"` adjusted quantile method based on Mahalanobis distance
`bc`	if `TRUE` it applies Box-Cox power transformation
`bcType`	select `"optimal"` or `"rounded"` type of Box-Cox power transformation, only applicable if `bc = TRUE`, default is `"rounded"`
`showOutliers`	if `TRUE` prints multivariate outliers
`showNewData`	if `TRUE` prints new data without outliers

Details

If `mvnTest = "mardia"`, it calculate the Mardia's multivariate skewness and kurtosis coefficients as well as their corresponding statistical significance. It can also calculate corrected version of skewness coefficient for small sample size (n< 20). For multivariate normality, both p-values of skewness and kurtosis statistics should be greater than 0.05. If sample size less than 20 then p.value.small should be used as significance value of skewness instead of p.value.skew. If there are missing values in the data, a listwise deletion will be applied and a complete-case analysis will be performed.

If `mvnTest = "hz"`, it calculate the Henze-Zirkler's multivariate normality test. The Henze-Zirkler test is based on a non-negative functional distance that measures the distance between two distribution functions. If the data is multivariate normal, the test statistic HZ is approximately lognormally distributed. It proceeds to calculate the mean, variance and smoothness parameter. Then, mean and variance are lognormalized and the p-value is estimated. If there are missing values in the data, a listwise deletion will be applied and a complete-case analysis will be performed.

If `mvnTest = "royston"`, it calculate the Royston's multivariate normality test. A function to generate the Shapiro-Wilk's W statistic needed to feed the Royston's H test for multivariate normality However, if kurtosis of the data greater than 3 then Shapiro-Francia test is used for leptokurtic samples else Shapiro-Wilk test is used for platykurtic samples. If there are missing values in the data, a listwise deletion will be applied and a complete-case analysis will be performed. Do not apply Royston's test, if dataset includes more than 5000 cases or less than 3 cases, since it depends on Shapiro-Wilk's test.

If `mvnTest = "dh"`, it calculate the Doornik-Hansen's multivariate normality test. The code is adapted from asbio package (Aho, 2017).

#'If `mvnTest = "energy"`, it calculate the Doornik-Hansen's multivariate normality test. The code is adapted from energy package (Rizzo and Szekely, 2017)i

Value

`multivariateNormality` corresponding multivariate normality test statistics and p-value

`univariateNormality` corresponding univariate normality test statistics and p-value

`Descriptives` Descriptive statistics

`multivariateOutliers` multivariate outliers

`newData` new data without multivariate outliers

multivariate normality plots, Q-Q, perspective or contour

chi-square Q-Q plot for multivariate outliers

univariate normality plots, Q-Q plot, histogram, box plot, scatter

Author(s)

Selcuk Korkmaz, <selcukorkmaz@gmail.com>

References

Korkmaz S, Goksuluk D, Zararsiz G. MVN: An R Package for Assessing Multivariate Normality. The R Journal. 2014 6(2):151-162. URL https://journal.r-project.org/archive/2014-2/korkmaz-goksuluk-zararsiz.pdf

Mardia, K. V. (1970), Measures of multivariate skewnees and kurtosis with applications. Biometrika, 57(3):519-530.

Mardia, K. V. (1974), Applications of some measures of multivariate skewness and kurtosis for testing normality and robustness studies. Sankhy A, 36:115-128.

Henze, N. and Zirkler, B. (1990), A Class of Invariant Consistent Tests for Multivariate Normality. Commun. Statist.-Theor. Meth., 19(10): 35953618.

Henze, N. and Wagner, Th. (1997), A New Approach to the BHEP tests for multivariate normality. Journal of Multivariate Analysis, 62:1-23.

Royston, J.P. (1982). An Extension of Shapiro and Wilks W Test for Normality to Large Samples. Applied Statistics, 31(2):115124.

Royston, J.P. (1983). Some Techniques for Assessing Multivariate Normality Based on the Shapiro-Wilk W. Applied Statistics, 32(2).

Royston, J.P. (1992). Approximating the Shapiro-Wilk W-Test for non-normality. Statistics and Computing, 2:117-119.121133.

Royston, J.P. (1995). Remark AS R94: A remark on Algorithm AS 181: The W test for normality. Applied Statistics, 44:547-551.

Shapiro, S. and Wilk, M. (1965). An analysis of variance test for normality. Biometrika, 52:591611.

Doornik, J.A. and Hansen, H. (2008). An Omnibus test for univariate and multivariate normality. Oxford Bulletin of Economics and Statistics 70, 927-939.

G. J. Szekely and M. L. Rizzo (2013). Energy statistics: A class of statistics based on distances, Journal of Statistical Planning and Inference, http://dx.doi.org/10.1016/j.jspi.2013.03.018

M. L. Rizzo and G. J. Szekely (2016). Energy Distance, WIRES Computational Statistics, Wiley, Volume 8 Issue 1, 27-38. Available online Dec., 2015, http://dx.doi.org/10.1002/wics.1375.

G. J. Szekely and M. L. Rizzo (2017). The Energy of Data. The Annual Review of Statistics and Its Application 4:447-79. 10.1146/annurev-statistics-060116-054026

Examples

```
result = mvn(data = iris[-4], subset = "Species", mvnTest = "hz",
             univariateTest = "AD", univariatePlot = "histogram",
             multivariatePlot = "qq", multivariateOutlierMethod = "adj",
             showOutliers = TRUE, showNewData = TRUE)

#### Multivariate Normality Result
result$multivariateNormality

### Univariate Normality Result
result$univariateNormality

### Descriptives
result$Descriptives

### Multivariate Outliers
result$multivariateOutliers

### New data without multivariate outliers
result$newData

# Note that this function also creates univariate histograms,
# multivariate Q-Q plots for multivariate normality assessment
# and multivariate outlier detection.
```

---

roystonTest *Deprecated Function for Royston's Multivariate Normality Test*

Description

Please use 'mvn' function instead

Usage

```
roystonTest(data, qqplot = FALSE)
```

Arguments

`data`	a numeric matrix or data frame
`qqplot`	`TRUE` it creates chi-square Q-Q plot

Index