

Package ‘MVN’

February 24, 2018

Type Package

Title Multivariate Normality Tests

Version 5.0

Date 2018-02-23

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

Maintainer Selcuk Korkmaz <selcukorkmaz@gmail.com>

URL <http://www.biosoft.hacettepe.edu.tr/MVN/>

Depends R (>= 3.3.0)

VignetteBuilder knitr

Suggests knitr

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

Collate mvn.R

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

License GPL (>= 2)

RoxygenNote 6.0.1

NeedsCompilation no

Repository CRAN

Date/Publication 2018-02-24 08:35:02 UTC

R topics documented:

mvn	2
Index	5

Description

Performs multivariate normality tests, including Marida, Royston, Henze-Zirkler, Dornik-Haansen, E-Statistics

Usage

```
mvn(data, subset = NULL, mvnTest = c("mardia", "hz", "royston", "dh",
  "energy"), covariance = TRUE, scale = FALSE, desc = TRUE,
  transform = "none", R = 1000, univariateTest = c("SW", "CVM", "Lillie",
  "SF", "AD"), univariatePlot = "none", multivariatePlot = "none",
  multivariateOutlierMethod = "none", 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. 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
scale	if TRUE scales the columns 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")
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
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.

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>

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.
```

Index

mvn, [2](#)