Package ‘MVN’

June 30, 2021

Type Package
Title Multivariate Normality Tests
Version 5.9
Date 2021-06-29
Maintainer Selcuk Korkmaz <selcukkorkmaz@gmail.com>
Depends R (>= 3.5.0)
Suggests BiocStyle, knitr, rmarkdown
VignetteBuilder knitr
Imports methods, nortest, moments, MASS, plyr, psych, boot, energy, car
Collate mvn.R
Encoding UTF-8
License GPL (>= 2)
RoxygenNote 7.1.1
NeedsCompilation no
Author Selcuk Korkmaz [aut, cre] (<https://orcid.org/0000-0003-4632-6850>), Dincer Goksuluk [aut], Gokmen Zararsiz [aut]
Repository CRAN
Date/Publication 2021-06-30 08:40:12 UTC

R topics documented:

mvn ................................................................. 2
Index 6

1
Description

Performs multivariate normality tests, including Mardia, 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

```r
mvn(
  data,
  subset = NULL,
  mvnTest = "hz",
  covariance = TRUE,
  tol = 1e-25,
  alpha = 0.5,
  scale = FALSE,
  desc = TRUE,
  transform = "none",
  R = 1000,
  univariateTest = "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 is used 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 columns of the 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 (default) 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 calculates 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 calculates 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 calculates 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 calculates the Doornik-Hansen’s multivariate normality test. The code is adapted from asbio package (Aho, 2017).

If `mvnTest = "energy"`, it calculates the Energy multivariate normality test. The code is adapted from energy package (Rizzo and Szekely, 2017).

**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, <selcukkorkmaz@gmail.com>

**References**


Examples

```r
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