Package 'MVET'

July 21, 2025

Type Package

Title Multivariate Estimates and Tests

Version 0.1.0

Author Yeonseok Choi [aut, cre],
Yong-Seok Choi [ctb]

Maintainer Yeonseok Choi <yeonseok3488@gmail.com>

Description Multivariate estimation and testing, currently a package for testing parametric data. To deal with parametric data, various multivariate normality tests and outlier detection are performed and visualized using the 'ggplot2' package. Homogeneity tests for covariance matrices are also possible, as well as the Hotelling's T-square test and the multivariate analysis of variance test. We are exploring additional tests and visualization techniques, such as profile analysis and randomized complete block design, to be made available in the future and making them easily accessible to users.

License MIT + file LICENSE

URL https://github.com/YeonSeok-Choi/MVET

Encoding UTF-8

RoxygenNote 7.3.1

Depends R (>= 3.5.0)

Imports stats, ggplot2 (>= 3.4.0), gridExtra (>= 2.3)

LazyData true

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

NeedsCompilation no

Repository CRAN

Date/Publication 2024-06-06 16:40:02 UTC

Contents

`.mean_parallel_plot` *Mean Value Parallel Coordinates Plot (Use to HT2test & VManova)*

Description

Mean Value Parallel Coordinates Plot (Use to HT2test & VManova)

Usage

```
.mean_parallel_plot(data, grp.name, scale = FALSE)
```

Arguments

`data`	A numeric matrix or data frame. If data frame, group(class) column can be a factor or a string.
`grp.name`	The name of a column of `string` representing groups(classes) in the input data.
`scale`	If `TRUE`, the data will be scaled before calculating mean values and used in the plot. (default scale = `FALSE`)

Value

Mean Value Parallel Coordinates Plot

`boxMtest` *Box's M-test*

Description

Performs Box's M-test for homogeneity of covariance matrices derived from multivariate normality data according to a single classification factor. This test is based on the chi-square approximation.

Usage

```
boxMtest(data,
         group)
```

Arguments

data	A numeric matrix or data frame.
group	In either vector or factor form, the length of the group must correspond to the number of observations n in the data.

Value

M.stat	Box's M-test statistic approximates the chi-square distribution.
df	The degree of freedom is related to the test statistic.
p.value	The p-value of the test statistic.

See Also

mardiatest

Examples

```
data(wine)
class <- wine$class
winedata <- subset(wine, select = -class)
boxMtest(winedata, class)
```

HT2test *Hotelling T Square Test*

Description

The mean vector test (Hotelling T square test) to compare one sample or two samples that satisfy the multivariate normality test and the homogeneity of covariance matrices test.

Usage

```
HT2test(data1,
        data2,
        mu0 = NULL,
        sample = "two",
        plot.scale = FALSE)
```

Arguments

data1	The data frame or matrix must consist of only numbers, and the data must consist of only a single group or class. It should not contain columns that separate groups or classes.

data2	The data frame or matrix must consist of only numbers, and the data must consist of only a single group or class. It should not contain columns that separate groups or classes. The data2 is for comparison with data1 and is not used to compare one sample to another.
mu0	The mu0 is used to test the mean vector hypothesis of data1. It is only used to compare one-sample.
sample	The options for specifying the number of groups for group comparisons are one and two, where one is used to compare one-sample and two is used to compare two-samples. (default sample = two)
plot.scale	If TRUE, the data will be scaled before calculating mean values and used in the plot. It has no direct effect on the data. It only applies to two samples. (default plot.scale = FALSE)

Value

One.HT2	The Hotelling T square test in one-sample, showing the degrees of freedom required for the F test, the Hotelling t square statistic, the F test statistic, and the probability of significance.
Mean.val.plot	Plot the mean value parallel coordinates, representing the two samples using the mean values for each variable.
Two.HT2	The Hotelling T square test in two-sample, showing the degrees of freedom required for the F test, the Hotelling t square statistic, the F test statistic, and the probability of significance.

References

Johnson, R. A., & Wichern, D. W. (2007). Applied Multivariate Statistical Analysis (6th ed.). Pearson Prentice Hall.

See Also

mardiatest for multivariate normality (Includes outlier remove)

PPCCtest for multivariate normality

SPCCtest for multivariate normality

boxMtest for homogeneity of covariance matrices

Examples

```
data(wine)
class1.wine <- subset(wine, class == 1)[, -1]
class2.wine <- subset(wine, class == 2)[, -1]
modified.class2.wine <- outlier(class2.wine, lim = 0, level = 0.05, option = "all")$modified.data

## one sample
value <- 0
p <- ncol(class1.wine)
mu0 <- matrix(rep(value, p), nrow = p, ncol = 1)
HT2test(data1 = class1.wine, mu0 = mu0, sample = "one")
```

```
## two sample
HT2test(data1 = class1.wine, data2 = modified.class2.wine, sample = "two", plot.scale = TRUE)
```

---

mardiatest *Mardia Test for Multivariate Normality Test*

Description

Performs a multivariate normality test by conducting a mardia test using skewness and kurtosis. If both skewness and kurtosis are satisfied, multivariate normality is satisfied.

Usage

```
mardiatest(data,
           level = 0.05,
           showplot = FALSE,
           showoutlier = FALSE,
           outlieropt = "all",
           shownewdata = FALSE)
```

Arguments

data	A numeric matrix or data frame.
level	The significance level of the skewness and kurtosis statistics. (default = 0.05)
showplot	If TRUE, show a chi-square Q-Q plot using ggplot2. If 'showoutlier' is TRUE, outliers are also displayed. (default = FALSE)
showoutlier	If TRUE, show the outliers number and count. (default = FALSE)
outlieropt	An "option" in the outlier function. (default = "all")
shownewdata	If TRUE Shows the new data with outliers removed. (default = FALSE)

Value

mult.nomality	Calculate statistics and p-values for skewness and kurtosis to ultimately determine whether multivariate normality is satisfied.
QQPlot	Shows Chi-Square Q-Q plot.
...	Same as the result of outlier

References

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

Mardia, K. V. (1974), Applications of Some Measures of Multivariate Skewness and Kurtosis in Testing Normality and Robustness Studies. Sankhya, 36, 115-128.

See Also

outlier

Examples

```
## Simple Mardia Test
data(wine)
class2.wine <- subset(wine, class == 2)[, -1]
mardiatest(class2.wine, level = 0.05, showplot = TRUE)

## Mardia Test and Outlier Detection
data(wine)
class2.wine <- subset(wine, class == 2)[, -1]
mardiatest(class2.wine, level = 0.05, showplot = TRUE,
           showoutlier = TRUE, outlieropt = "all", shownewdata = TRUE)
```

`outlier` *Outliers Detection*

Description

Using the mardia test, outliers are detected based on skewness and kurtosis. However, outliers don't detect more than half of the total observation data.(Can be modified with the `lim` option.)

Usage

```
outlier(data,
        lim = 0,
        level = 0.05,
        option = "all")
```

Arguments

`data`	A numeric matrix or data frame.
`lim`	The number of outliers detected can be limited. If 0 is entered, detection is possible up to half of the data. (default = `0`)
`level`	The significance level of the skewness and kurtosis statistics of the "`madiatest`" function. (default = `0.05`)
`option`	"`skew`" refers to skewness, "`kurt`" refers to kurtosis, "`all`" refers to skewness and kurtosis. Outliers are detected until the corresponding option in the `mardiatest` is "Accept". (default = "`all`")

Value

modified.data	The modified data without outliers.
modified.mvn	The modified Mardia test result without outliers.
outlier.num	The position of outliers.
outlier.cnt	Total number of outliers.

References

Jobson, J. D.(1992). Applied Multivariate Data Analysis, Springer-Verlag, New York.

See Also

mardiatest

Examples

```
data(wine)
class2.wine <- subset(wine, class == 2)[, -1]
outlier(class2.wine, lim = 0, level = 0.05, option = "all")
```

PPCCtest *Probability Plot Correlation Coefficient(PPCC) Test for Multivariate Normality Test*

Description

The correlation coefficient of the quantiles and mahalanobis square are tested by using the critical value table by Filliben (1975) for the multivariate normality test.

Usage

```
PPCCtest(data,
         level = 0.05)
```

Arguments

data	A numeric matrix or data frame.
level	At the 0.01 or 0.05 significance level, the critical value. (default = 0.05)

Value

data.cnt	Observation n data count.
PPCC.value	Correlation coefficient value.
critical.value	Critical value proposed by Filliben (1975), corresponding to data.cnt and PPCC.value.
test.res	Final result of multivariate normality.
QQPlot	Shows Chi-Square Q-Q plot.

References

Filliben, J. J. (1975), The Probability Plot Correlation Coefficient Test for Normality, *Technometrics* 17, 111-117.

Examples

```
data(wine)
class1.wine <- subset(wine, class == 1)[, -1]
PPCCtest(class1.wine, level = 0.05)
```

`SPCCtest` *Srivastava Plot Correlation Coefficient(SPCC) Test for Multivariate Normality Test*

Description

Using principal component analysis, the number of eigenvalues is selected such that the ratio of eigenvalues exceeds 70%. The principal component score vectors corresponding to these selected eigenvalues are used, and testing is conducted using the threshold defined by Filliben (1975). Users have the option to select the number of eigenvalues for the analysis based on their requirements.

Usage

```
SPCCtest(data,
         k = 0,
         level = 0.05)
```

Arguments

`data`	A numeric matrix or data frame.
`k`	The number of principal components can be manually selected. If 0 is entered, it automatically finds k components such that the explained variance ratio is at least 70%. (default = `0`)
`level`	At the `0.01` or `0.05` significance level, the critical value. (default = `0.05`)

Value

`Srivastava.QQplot`	Shows a chi-Square Q-Q plot for each PCs using ggplot2.
`data.cnt`	Observation n data count.
`explain.ratio`	Displays all explained variance ratios.
`critical.value`	Critical value proposed by Filliben (1975), corresponding to `data.cnt` and `PPCC.value`.
`result`	Final result of multivariate normality.

References

Srivastava, M. S. (1984), A measure of skewness and kurtosis and a graphical method for assessing multivariate normality. Statistics & Probability Letters, 2(5), 263-267.

Filliben, J. J. (1975), The Probability Plot Correlation Coefficient Test for Normality, *Technometrics* 17, 111-117.

Examples

```
data(wine)
class1.wine <- subset(wine, class == 1)[, -1]
SPCCtest(class1.wine, k = 5, level = 0.05)
```

`VManova` *Various Multivariate Anova(VManova)*

Description

Perform various types of multivariate analysis of variance (MANOVA) that satisfy tests of multivariate normality and homogeneity of covariance matrices.

Usage

```
VManova(data,
        grp1.name,
        grp2.name,
        way = "one",
        method = "all",
        plot.scale = FALSE)
```

Arguments

`data`	A numeric matrix or data frame. If data frames, group(class) column can be a factor or a string.
`grp1.name`	The name of the first group (or class) column in the input data, specified as a `string`.
`grp2.name`	The name of the second group (or class) column in the input data, specified as a `string`. Used to represent the second group(class) in a two-way MANOVA.
`way`	The type of MANOVA to perform (`"one"` for one-way or `"two"` for two-way). (default = `"one"`)
`method`	The method for MANOVA analysis. `"Wilks"` represents Wilks' lambda, `"LH"` represents Lawley-Hotelling trace, `"Pillai"` represents Pillai-Bartlett trace, `"Roy"` represents Roy's largest root, and `"all"` represents all methods. (default is `"all"`)
`plot.scale`	If `TRUE`, the data will be scaled before calculating mean values and used in the plot. It has no direct effect on the MANOVA analysis itself. (default plot.scale = `FALSE`)

Value

`Mean.val.plot`	Plot the mean value parallel coordinates, representing the two samples using the mean values for each variable.
`One.all`	Outputs the results of a one-way MANOVA test. It displays the degrees of freedom (Df1, Df2) of the F-distribution, statistics for Wilks, Lawley-Hotelling, Pillai, and Roy, the F-distribution test statistic, and the significance level in that order.
`Two.all`	Outputs the results of a two-way MANOVA test. It displays the degrees of freedom (Df1, Df2) of the F-distribution, statistics for Wilks, Lawley-Hotelling, Pillai, and Roy, the F-distribution test statistic, and the significance level in that order.

References

Rencher, A. C., & Christensen, W. F. (2002). Methods of Multivariate Analysis. John Wiley & Sons, Inc., New York.

See Also

`mardiatest` for multivariate normality (Includes outlier remove)

`PPCCtest` for multivariate normality

`SPCCtest` for multivariate normality

`boxMtest` for homogeneity of covariance matrices

Examples

```
data(wine)

## one way
VManova(wine, grp1.name = "class", way = "one", method = "all", plot.scale = TRUE)

## two way
newwine <- wine
# (1: low, 2: medium, 3: high)
newwine$v4 <- ifelse(wine$v4 <= 17, 1,
                     ifelse(wine$v4 <= 22, 2, 3))
VManova(newwine, grp1.name = "class", grp2.name = "v4",
        way = "two", method = "all", plot.scale = TRUE)
```

`wine` *Wine Dataset*

Description

These data are the results of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivars. The analysis determined the quantities of 13 constituents found in each of the three types of wines.

Usage

```
wine
```

Format

A data frame with 178 observations on the following 14 variables:

class The class vector, the three different cultivars of wine are reprensented by the three integers(1 to 3).

v1 Alcohol

v2 Malic acid

v3 Ash

v4 Alcalinity of ash

v5 Magnesium

v6 Total phenols

v7 Flavanoids

v8 Nonflavanoid phenols

v9 Proanthocyanins

v10 Color intensity

v11 Hue

v12 OD280/OD315 of diluted wines

v13 Proline

Source

http://archive.ics.uci.edu/ml/datasets/Wine.

Index