

Package ‘GFDmcv’

February 2, 2023

Type Package

Title General Hypothesis Testing Problems for Multivariate
Coefficients of Variation

Version 0.1.0

Date 2023-02-02

Description

Performs test procedures for general hypothesis testing problems for four multivariate coefficients of variation (Ditzhaus and Smaga, 2023 <[arXiv:2301.12009](#)>). We can verify the global hypothesis about equality as well as the particular hypotheses defined by contrasts, e.g., we can conduct post hoc tests. We also provide the simultaneous confidence intervals for contrasts.

License LGPL-2 | LGPL-3 | GPL-2 | GPL-3

Imports Rcpp (>= 1.0.9), mvtnorm, doParallel, MASS, foreach, Matrix,
stringr, HSAUR

LinkingTo Rcpp, RcppArmadillo

RoxygenNote 7.2.3

Encoding UTF-8

NeedsCompilation yes

Author Marc Ditzhaus [aut],
Lukasz Smaga [aut, cre]

Maintainer Lukasz Smaga <ls@amu.edu.pl>

Repository CRAN

Date/Publication 2023-02-02 17:00:02 UTC

R topics documented:

contr_mat	2
e_mcv	3
GFDmcv	5
plot.gfdmcv	8
summary.gfdmcv	9

Index	11
--------------	-----------

contr_mat	<i>Contrasts' matrices</i>
-----------	----------------------------

Description

The output are different contrast matrices, namely the centering matrix as well as the matrices of Dunnett's and Tukey's contrasts for given number of groups.

Usage

```
contr_mat(k, type = c("center", "Dunnett", "Tukey"))
```

Arguments

k an integer denoting a number of groups

type an character denoting type of contrasts. The possible values are "center" (default), "Dunnett", "Tukey".

Details

The centering matrix is $\mathbf{P}_k = \mathbf{I}_k - \mathbf{J}_k/k$, where \mathbf{I}_k is the unity matrix and $\mathbf{J}_k = \mathbf{1}\mathbf{1}^\top \in R^{k \times k}$ consisting of 1's only. The matrix of Dunnett's contrasts:

$$\begin{pmatrix} -1 & 1 & 0 & \dots & \dots & 0 \\ -1 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -1 & 0 & \dots & \dots & 0 & 1 \end{pmatrix} \in R^{(k-1) \times k}.$$

The matrix of Tukey's contrasts:

$$\begin{pmatrix} -1 & 1 & 0 & \dots & \dots & 0 & 0 \\ -1 & 0 & 1 & 0 & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -1 & 0 & 0 & 0 & \dots & \dots & 1 \\ 0 & -1 & 1 & 0 & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & \dots & \dots & 0 & -1 & 1 \end{pmatrix} \in R^{\binom{k}{2} \times k}.$$

Value

The matrix of contrasts.

References

Ditzhaus M., Smaga L. (2022) Permutation test for the multivariate coefficient of variation in factorial designs. *Journal of Multivariate Analysis* 187, 104848.

Ditzhaus M., Smaga L. (2023) Inference for all variants of the multivariate coefficient of variation in factorial designs. Preprint <https://arxiv.org/abs/2301.12009>.

Dunnett C. (1955) A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association* 50, 1096-1121.

Tukey J.W. (1953) *The problem of multiple comparisons*. Princeton University.

Examples

```
contr_mat(4, type = "center")
contr_mat(4, type = "Dunnett")
contr_mat(4, type = "Tukey")
```

e_mcv

Estimators and confidence intervals of four multivariate coefficients of variation and their reciprocals

Description

Calculates the estimators with respective $(1 - \alpha)$ -confidence intervals for the four different variants of the multivariate coefficients (MCV) and their reciprocals by Reyment (1960), Van Valen (1974), Voinov and Nikulin (1996) and Albert and Zhang (2010).

Usage

```
e_mcv(x, conf_level = 0.95)
```

Arguments

`x` a matrix of data of size $n \times d$.
`conf_level` a confidence level. By default, it is equal to 0.95.

Details

The function `e_mcv()` calculates four different variants of multivariate coefficient of variation for d -dimensional data. These variant were introduced by by Reyment (1960, RR), Van Valen (1974, VV), Voinov and Nikulin (1996, VN) and Albert and Zhang (2010, AZ):

$$\hat{C}^{RR} = \sqrt{\frac{(\det \hat{\Sigma})^{1/d}}{\hat{\boldsymbol{\mu}}^\top \hat{\boldsymbol{\mu}}}}, \quad \hat{C}^{VV} = \sqrt{\frac{\text{tr} \hat{\Sigma}}{\hat{\boldsymbol{\mu}}^\top \hat{\boldsymbol{\mu}}}}, \quad \hat{C}^{VN} = \sqrt{\frac{1}{\hat{\boldsymbol{\mu}}^\top \hat{\Sigma}^{-1} \hat{\boldsymbol{\mu}}}}, \quad \hat{C}^{AZ} = \sqrt{\frac{\hat{\boldsymbol{\mu}}^\top \hat{\Sigma} \hat{\boldsymbol{\mu}}}{(\hat{\boldsymbol{\mu}}^\top \hat{\boldsymbol{\mu}})^2}},$$

where n is the sample size, $\hat{\boldsymbol{\mu}}$ is the empirical mean vector and $\hat{\boldsymbol{\Sigma}}$ is the empirical covariance matrix:

$$\hat{\boldsymbol{\mu}}_i = \frac{1}{n} \sum_{j=1}^n \mathbf{X}_j, \quad \hat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{j=1}^n (\mathbf{X}_j - \hat{\boldsymbol{\mu}})(\mathbf{X}_j - \hat{\boldsymbol{\mu}})^\top.$$

In the univariate case ($d = 1$), all four variants reduce to coefficient of variation. Furthermore, their reciprocals, the so-called standardized means, are determined:

$$\hat{B}^{RR} = \sqrt{\frac{\hat{\boldsymbol{\mu}}^\top \hat{\boldsymbol{\mu}}}{(\det \hat{\boldsymbol{\Sigma}})^{1/d}}}, \quad \hat{B}^{VV} = \sqrt{\frac{\hat{\boldsymbol{\mu}}^\top \hat{\boldsymbol{\mu}}}{\text{tr} \hat{\boldsymbol{\Sigma}}}}, \quad \hat{B}^{VN} = \sqrt{\hat{\boldsymbol{\mu}}^\top \hat{\boldsymbol{\Sigma}}^{-1} \hat{\boldsymbol{\mu}}}, \quad \hat{B}^{AZ} = \sqrt{\frac{(\hat{\boldsymbol{\mu}}^\top \hat{\boldsymbol{\mu}})^2}{\hat{\boldsymbol{\mu}}^\top \hat{\boldsymbol{\Sigma}} \hat{\boldsymbol{\mu}}}}.$$

In addition to the estimators, the respective confidence intervals [C_lwr, C_upr] for a given confidence level $1 - \alpha$ are calculated by the `e_mcv()` function. These confidence intervals are based on an asymptotic approximation by a normal distribution, see Ditzhaus and Smaga (2023) for the technical details. These approximations do not rely on any specific (semi-)parametric assumption on the distribution and are valid nonparametrically, even for tied data.

Value

When $d > 1$ (respectively $d = 1$) a data frame with four rows (one row) corresponding to the four MCVs (the univariate CV) and six columns containing the estimators `C_est` for the MCV (CV) and the estimators `B_est` for their reciprocals as well as the upper and lower bounds of the corresponding confidence intervals [C_lwr, C_upr] and [B_lwr, B_upr].

References

- Albert A., Zhang L. (2010) A novel definition of the multivariate coefficient of variation. *Biometrical Journal* 52:667-675.
- Ditzhaus M., Smaga L. (2023) Inference for all variants of the multivariate coefficient of variation in factorial designs. Preprint <https://arxiv.org/abs/2301.12009>.
- Reyment R.A. (1960) Studies on Nigerian Upper Cretaceous and Lower Tertiary Ostracoda: part 1. Senonian and Maastrichtian Ostracoda, *Stockholm Contributions in Geology*, vol 7.
- Van Valen L. (1974) Multivariate structural statistics in natural history. *Journal of Theoretical Biology* 45:235-247.
- Voinov V., Nikulin M. (1996) *Unbiased Estimators and Their Applications, Vol. 2, Multivariate Case*. Kluwer, Dordrecht.

Examples

```
# d > 1 (MCVs)
data_set <- lapply(list(iris[iris$Species == "setosa", 1:3],
                      iris[iris$Species == "versicolor", 1:3],
                      iris[iris$Species == "virginica", 1:3]),
                 as.matrix)
lapply(data_set, e_mcv)
# d = 1 (CV)
data_set <- lapply(list(iris[iris$Species == "setosa", 1],
                      iris[iris$Species == "versicolor", 1],
                      iris[iris$Species == "virginica", 1]),
```

```

                                as.matrix)
lapply(data_set, e_mcv)

```

GFDmcv

Inference for four multivariate coefficients of variation

Description

The function `GFDmcv()` calculates the Wald-type statistic for global null hypotheses and max-type statistics for multiple local null hypotheses, both in terms of the four variants of the multivariate coefficient of variation. Respective p -values are obtained by a χ^2 -approximation, a pooled bootstrap strategy and a pooled permutation approach (only for the Wald-type statistic), respectively.

Usage

```

GFDmcv(
  x,
  h_mct,
  h_wald,
  alpha = 0.05,
  n_perm = 1000,
  n_boot = 1000,
  parallel = FALSE,
  n_cores = NULL
)

```

Arguments

<code>x</code>	a list of length k with elements being $n_i \times d$ matrices of data, $i = 1, \dots, k$.
<code>h_mct</code>	a $r \times k$ contrast matrix \mathbf{H} of full row rank for multiple contrast tests. Remember to specify it correctly taking into account the order of elements of the list <code>x</code> .
<code>h_wald</code>	a $q \times k$ contrast matrix \mathbf{H} of full row rank for the Wald-type tests. Remember to specify it correctly taking into account the order of elements of the list <code>x</code> .
<code>alpha</code>	a significance level (then $1 - \alpha$ is the confidence level).
<code>n_perm</code>	a number of permutation replicates.
<code>n_boot</code>	a number of bootstrap replicates.
<code>parallel</code>	a logical indicating whether to use parallelization.
<code>n_cores</code>	if <code>parallel = TRUE</code> , a number of processes used in parallel computation. Its default value means that it will be equal to a number of cores of a computer used.

Details

The function `GFDmcv()` calculates the Wald-type statistic for global null hypotheses of the form

$$\mathcal{H}_0 : \mathbf{H}(C_1, \dots, C_k)^\top = \mathbf{0} \text{ and } \mathcal{H}_0 : \mathbf{H}(B_1, \dots, B_k)^\top = \mathbf{0},$$

where \mathbf{H} is a contrast matrix reflecting the research question of interest and C_i (B_i) are the subgroup-specific MCVs (and their reciprocal) by Reyment (1960, RR), Van Valen (1974, VV), Voinov and Nikulin (1996, VN) or Albert and Zhang (2010, AZ), respectively. We refer to the function `e_mcv()` for the detailed definitions of the different variants. The p -value of the Wald-type statistic relies on a χ^2 -approximation, a (pooled) bootstrap or permutation approach.

Furthermore, the function `GFDmcv()` calculates a max-type test statistic for the multiple comparison of q local null hypotheses:

$$\mathcal{H}_{0,\ell} : \mathbf{h}_\ell^\top \mathbf{C} = \mathbf{0} \text{ or } \mathcal{H}_{0,\ell} : \mathbf{h}_\ell^\top \mathbf{B} = \mathbf{0}, \ell = 1, \dots, q,$$

where $\mathbf{C} = (C_1, \dots, C_k)^\top$ and $\mathbf{B} = (B_1, \dots, B_k)^\top$. The p -values are determined by a Gaussian approximation and a bootstrap approach, respectively. In addition to the local test decisions, multiple adjusted confidence intervals for the contrasts $\mathbf{h}_\ell^\top \mathbf{C}$ and $\mathbf{h}_\ell^\top \mathbf{B}$, respectively, are calculated.

Please have a look on the plot and summary functions designed for this package. They can be used to simplify the output of `GFDmcv()`.

Value

A list of class `gfdmcv` containing the following components:

<code>overall_res</code>	a list of two elements representing the results for testing the global null hypothesis. The first one is a matrix <code>test_stat</code> of the test statistics, while the second is a matrix <code>p_values</code> of the respective p -values.
<code>mct_res</code>	all results of MCT tests for particular hypothesis in <code>h_mct</code> , i.e., the estimators and simultaneous confidence intervals for $\mathbf{h}_\ell^\top \mathbf{C}$ and for $\mathbf{h}_\ell^\top \mathbf{B}$, the test statistics and critical values as well as the decisions.
<code>h_mct</code>	an argument <code>h_mct</code> .
<code>h_wald</code>	an argument <code>h_wald</code> .
<code>alpha</code>	an argument <code>alpha</code> .

References

- Albert A., Zhang L. (2010) A novel definition of the multivariate coefficient of variation. *Biometrical Journal* 52:667-675.
- Ditzhaus M., Smaga L. (2022) Permutation test for the multivariate coefficient of variation in factorial designs. *Journal of Multivariate Analysis* 187, 104848.
- Ditzhaus M., Smaga L. (2023) Inference for all variants of the multivariate coefficient of variation in factorial designs. Preprint <https://arxiv.org/abs/2301.12009>.
- Reyment R.A. (1960) Studies on Nigerian Upper Cretaceous and Lower Tertiary Ostracoda: part 1. Senonian and Maastrichtian Ostracoda, *Stockholm Contributions in Geology*, vol 7.
- Van Valen L. (1974) Multivariate structural statistics in natural history. *Journal of Theoretical Biology* 45:235-247.
- Voinov V., Nikulin M. (1996) *Unbiased Estimators and Their Applications*, Vol. 2, Multivariate Case. Kluwer, Dordrecht.

Examples

```

# Some of the examples may run some time.

# one-way analysis for MCV and CV
# d > 1 (MCV)
data_set <- lapply(list(iris[iris$Species == "setosa", 1:3],
                      iris[iris$Species == "versicolor", 1:3],
                      iris[iris$Species == "virginica", 1:3]),
                  as.matrix)
# estimators and confidence intervals of MCVs and their reciprocals
lapply(data_set, e_mcv)
# contrast matrices
k <- length(data_set)
# Tukey's contrast matrix
h_mct <- contr_mat(k, type = "Tukey")
# centering matrix P_k
h_wald <- contr_mat(k, type = "center")

# testing without parallel computing
res <- GFDmcv(data_set, h_mct, h_wald)
summary(res, digits = 3)
oldpar <- par(mar = c(4, 5, 2, 0.3))
plot(res)
par(oldpar)

# testing with parallel computing
library(doParallel)
res <- GFDmcv(data_set, h_mct, h_wald, parallel = TRUE, n_cores = 2)
summary(res, digits = 3)
oldpar <- par(mar = c(4, 5, 2, 0.3))
plot(res)
par(oldpar)

# two-way analysis for CV (based on the example in Ditzhaus and Smaga, 2022)
library(HSAUR)
data_set <- lapply(list(BtheB$bdi.pre[BtheB$drug == "No" & BtheB$length == "<6m"],
                      BtheB$bdi.pre[BtheB$drug == "No" & BtheB$length == ">6m"],
                      BtheB$bdi.pre[BtheB$drug == "Yes" & BtheB$length == "<6m"],
                      BtheB$bdi.pre[BtheB$drug == "Yes" & BtheB$length == ">6m"]),
                  as.matrix)
# estimators and confidence intervals of CV and its reciprocal
lapply(data_set, e_mcv)

# interaction
h_mct <- contr_mat(4, type = "Tukey")
h_wald <- kronecker(contr_mat(2, type = "center"),
                  contr_mat(2, type = "center"))
res <- GFDmcv(data_set, h_mct, h_wald)
summary(res, digits = 3)
oldpar <- par(mar = c(4, 6, 2, 0.1))
plot(res)
par(oldpar)

```

```

# main effect drug
h_mct <- matrix(c(1, 1, -1, -1), nrow = 1)
h_wald <- kronecker(contr_mat(2, type = "center"), 0.5 * matrix(1, 1, 2))
res <- GFDmcb(data_set, h_mct, h_wald)
summary(res, digits = 3)
oldpar <- par(mar = c(4, 6, 2, 0.1))
plot(res)
par(oldpar)

# main effect length
h_mct <- matrix(c(1, -1, 1, -1), nrow = 1)
h_wald <- kronecker(0.5 * matrix(1, 1, 2), contr_mat(2, type = "center"))
res <- GFDmcb(data_set, h_mct, h_wald)
summary(res, digits = 3)
oldpar <- par(mar = c(4, 6, 2, 0.1))
plot(res)
par(oldpar)

```

plot.gfdmcb

Plot simultaneous confidence intervals for contrasts

Description

Simultaneous confidence intervals for contrasts for CV and MCVs and their reciprocals are plotted.

Usage

```

## S3 method for class 'gfdmcb'
plot(x, ...)

```

Arguments

x	an "gfdmcb" object.
...	additional arguments not used.

Value

No return value, called for side effects.

Examples

```

# Some of the examples may run some time.
# For more examples, see the documentation of the GFDmcb() function.
data_set <- lapply(list(iris[iris$Species == "setosa", 1:3],
                      iris[iris$Species == "versicolor", 1:3],
                      iris[iris$Species == "virginica", 1:3]),
                  as.matrix)

```



```

# estimators and confidence intervals of MCVs and their reciprocals
lapply(data_set, e_mcv)
# contrast matrices
k <- length(data_set)
# Tukey's contrast matrix
h_mct <- contr_mat(k, type = "Tukey")
# centering matrix P_k
h_wald <- contr_mat(k, type = "center")

# testing without parallel computing
res <- GFDmcv(data_set, h_mct, h_wald)
oldpar <- par(mar = c(4, 5, 2, 0.3))
plot(res)
par(oldpar)

```

summary.gfdmcv

Print "gfdmcv" object

Description

Prints the summary of the inference methods for CV and MCVs.

Usage

```

## S3 method for class 'gfdmcv'
summary(object, ...)

```

Arguments

object	an "gfdmcv" object.
...	integer indicating the number of decimal places to be used to present the numerical results, It can be named digits as in the round() function (see examples).

Details

The function prints out the information about the significance level, contrast matrices, test statistics, p -values, critical values, simultaneous confidence intervals for contrasts performed by the GFDmcv() function.

Value

No return value, called for side effects.

Examples

```
# Some of the examples may run some time.
# For more examples, see the documentation of the GFDMCv() function.
data_set <- lapply(list(iris[iris$Species == "setosa", 1:3],
                      iris[iris$Species == "versicolor", 1:3],
                      iris[iris$Species == "virginica", 1:3]),
                  as.matrix)
# estimators and confidence intervals of MCVs and their reciprocals
lapply(data_set, e_mcv)
# contrast matrices
k <- length(data_set)
# Tukey's contrast matrix
h_mct <- contr_mat(k, type = "Tukey")
# centering matrix P_k
h_wald <- contr_mat(k, type = "center")

# testing without parallel computing
res <- GFDMCv(data_set, h_mct, h_wald)
summary(res, digits = 3)
```

Index

`contr_mat`, [2](#)

`e_mcv`, [3](#)

`GFDmcv`, [5](#)

`plot.gfdmcv`, [8](#)

`summary.gfdmcv`, [9](#)