Package 'multiFANOVA'

July 6, 2023

Type Package

Title Multiple Contrast Tests for Functional Data

Version 0.1.0

Date 2023-06-28

Description The provided package implements multiple contrast tests for functional data (Munko et al., 2023, <arXiv:2306.15259>). These procedures enable us to evaluate the overall hypothesis regarding equality, as well as specific hypotheses defined by contrasts. In particular, we can perform post hoc tests to examine particular comparisons of interest. Different experimental designs are supported, e.g., one-way and multi-way analysis of variance for functional data.

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

Imports doParallel, MASS, foreach, Matrix, GFDmcv, fda

RoxygenNote 7.2.3

Encoding UTF-8

NeedsCompilation no

Author Marc Ditzhaus [aut],
Merle Munko [aut],
Markus Pauly [aut],
Lukasz Smaga [aut, cre],
Jin-Ting Zhang [aut]

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

Repository CRAN

Date/Publication 2023-07-06 13:40:02 UTC

R topics documented:

---

multiFANOVA   *Multiple contrast tests for functional data*

Description

The function `multiFANOVA()` calculates the globalizing pointwise Hotelling's T^2-test (GPH) statistic for the global null hypothesis and multiple local null hypotheses. Respective p-values are obtained by a parametric bootstrap strategy.

Usage

```
multiFANOVA(
  x,
  gr_label,
  h,
  n_boot = 1000,
  alpha = 0.05,
  parallel = FALSE,
  n_cores = NULL
)
```

Arguments

x	matrix of observations $n \times j$ ($n = n_1 + ... + n_k$, j is a number of design time points).
gr_label	a vector with group labels; the integer labels (from 1 to a number of groups) should be used.
h	contrast matrix. For Dunnett's and Tukey's contrasts, it can be created by the `contr_mat()` function from the package GFDmcv (see examples).
n_boot	number of bootstrap samples.
alpha	significance level.
parallel	a logical indicating whether to use parallelization.
n_cores	if `parallel = TRUE`, a number of processes used in parallel computation. Its default value means that it will be equal to the number of cores of a computer used.

Details

The function `multiFANOVA()` concerns the tests for the heteroscedastic contrast testing problem for functional data. The details are presented in Munko et al. (2023), but here we present some summary of the problem and its solutions implemented in the package.

Suppose we have k independent functional samples $x_{i1}, \ldots, x_{in_i}$, which consist of independent and identically distributed stochastic processes defined on interval $[a, b]$ with mean function η_i and covariance function γ_i for each $i \in \{1, \ldots, k\}$. Note that the covariance functions of the different groups may differ from each other, i.e., heteroscedasticity is explicitly allowed.

We consider the null and alternative hypothesis

$$\mathcal{H}_0 : \mathbf{H}\boldsymbol{\eta}(t) = 0 \text{ for all } t \in [a, b] \quad \text{vs.} \quad \mathcal{H}_1 : \mathbf{H}\boldsymbol{\eta}(t) \neq 0 \text{ for some } t \in [a, b],$$

where $\mathbf{H} \in \mathbb{R}^{r \times k}$ denotes a known contrast matrix, i.e., $\mathbf{H}\mathbf{1}_k = \mathbf{0}_r$, $\boldsymbol{\eta} := (\eta_1, \ldots, \eta_k)^\top$ is the vector of the mean functions. The formulation of this testing framework is very general and contains many special cases like the analysis of variance for functional data (FANOVA) problem. In detail, we may choose $\mathbf{H} = \mathbf{P}_k$ for the one-way FANOVA problem to test the null hypothesis of no main effect, where $\mathbf{P}_k := \mathbf{I}_k - \mathbf{J}_k/k$ with $\mathbf{I}_k \in \mathbb{R}^{k \times k}$ denoting the unit matrix and $\mathbf{J}_k := \mathbf{1}_k\mathbf{1}_k^\top \in \mathbb{R}^{k \times k}$ denoting the matrix of ones. However, there are different possible choices of the contrast matrix $\mathbf{H}$ which lead to this global null hypothesis. Many-to-one comparisons can be considered by choosing Dunnett's contrast matrix $\mathbf{H} = [-\mathbf{1}_{k-1}, \mathbf{I}_{k-1}]$, where the mean functions $\eta_2, \ldots, \eta_k$ are compared to the mean function η_1 of the first group regarding the different contrasts. To compare all pairs of mean functions $\eta_{i_1}, \eta_{i_2}, i_1, i_2 \in \{1, \ldots, k\}$ with $i_1 \neq i_2$, the Tukey's contrast matrix:

$$\mathbf{H} = \begin{bmatrix} -1 & 1 & 0 & 0 & \cdots & \cdots & 0 \\ -1 & 0 & 1 & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -1 & 0 & 0 & 0 & \cdots & \cdots & 1 \\ 0 & -1 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & -1 & 0 & 1 & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix} \in \mathbb{R}^{k(k-1)/2 \times k}$$

can be used.

For this testing problem, we consider the pointwise Hotelling's T^2-test statistic

$$\mathrm{TF}_{n,\mathbf{H}}(t) := n(\mathbf{H}\widehat{\boldsymbol{\eta}}(t))^\top (\mathbf{H}\widehat{\boldsymbol{\Sigma}}(t,t)\mathbf{H}^\top)^+ \mathbf{H}\widehat{\boldsymbol{\eta}}(t)$$

for all $t \in [a, b]$, where $\widehat{\boldsymbol{\eta}} := (\widehat{\eta}_1, \ldots, \widehat{\eta}_k)^\top$ denotes the vector of all mean function estimators, $\mathbf{A}^+$ denotes the Moore-Penrose inverse of the matrix $\mathbf{A}$, and

$$\widehat{\boldsymbol{\Sigma}}(t, s) := \mathrm{diag}\left(\frac{n}{n_1}\widehat{\gamma}_1(t, s), \ldots, \frac{n}{n_k}\widehat{\gamma}_k(t, s) \right),$$

$n = n_1 + \cdots + n_k$, $\widehat{\gamma}_i(t, s)$ is the sample covariance function for the i-th group, $i = 1, \ldots, k$. Based on this pointwise Hotelling's T^2-test statistic, we construct the globalizing pointwise Hotelling's T^2-test (GPH) statistic by integrating over the pointwise Hotelling's T^2-test statistic, that is

$$T_n(\mathbf{H}) := \int_a^b \mathrm{TF}_{n,\mathbf{H}}(t) \, \mathrm{d}t.$$

We consider the parametric bootstrap test based on this test statistic. However, for better post hoc analysis, we also consider the multiple contrast testing procedures. The main idea of multiple contrast tests is to split up the global null hypothesis with matrix $\mathbf{H} = [\mathbf{H}_1^\top, \ldots, \mathbf{H}_r^\top]^\top$ into r single contrast tests with contrast vectors $\mathbf{H}_1, \ldots, \mathbf{H}_r \in \mathbb{R}^k$. This leads to the multiple testing problem with null hypotheses

$$\mathcal{H}_{0,\ell} : \mathbf{H}_\ell \boldsymbol{\eta}(t) = 0 \text{ for all } t \in [a, b], \text{ for } \ell \in \{1, \ldots, r\}.$$

To verify this family of null hypotheses, we adopt two approaches. First, we simply apply the above test to each hypothesis $\mathcal{H}_{0,\ell}$, and the resulting p-values are then corrected by the Bonferroni's method. However, this approach, denoted in the package as GPH, may give conservative test and loss of power. Thus, we also consider the test adopting the idea for the construction of simultaneous confidence bands proposed by Buhlmann (1998). This test is denoted by mGPH in the package and is a more powerful solution than the GPH procedure, which was shown in Munko et al. (2023).

Note that the value of the test statistic for the mGPH test for global hypotheses is equals to

$$\max_{\ell \in \{1,\ldots,r\}} \frac{T_n(\mathbf{H}_\ell)}{q_{l,\widetilde{\beta}}^{\mathcal{P}}},$$

where $q_{l,\widetilde{\beta}}^{\mathcal{P}}$ are the quantiles calculated using the adaptation of the method by Buhlmann (1998). The critical value for it is always 1.

Please have a look at a summary function designed for this package. It can be used to simplify the output of `multiFANOVA()` function.

Value

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

res_global	a data frame containing the results for testing the global null hypothesis, i.e., test statistics and p-values.
res_multi	all results of multiple contrasts tests for particular hypothesis in a contrast matrix h, i.e., test statistics, critical values and p-values.
k	a number of groups.
j	a number of design time points.
n	a vector of sample sizes.
h	an argument h.
h_boot	an argument `n_boot`.
alpha	an argument `alpha`.

References

Buhlmann P. (1998) Sieve bootstrap for smoothing in nonstationary time series. Annals of Statistics 26, 48-83.

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

Munko M., Ditzhaus M., Pauly M., Smaga L., Zhang J.T. (2023) General multiple tests for functional data. Preprint https://arxiv.org/abs/2306.15259

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

Examples

```
# Some of the examples may run some time.

# Canadian weather data set
# There are three samples of mean temperatures for
# fifteen weather stations in Eastern Canada,
# another fifteen in Western Canada, and
# the remaining five in Northern Canada.
library(fda)
data_set <- t(CanadianWeather$dailyAv[,, "Temperature.C"])
k <- 3
gr_label <- rep(c(1, 2, 3), c(15, 15, 5))
# trajectories of mean temperatures
matplot(t(data_set), type = "l", col = gr_label, lty = 1,
        xlab = "Day", ylab = "Temperature (C)",
        main = "Canadian weather data set")
legend("bottom", legend = c("Eastern Canada", "Western Canada", "Northern Canada"),
        col = 1:3, lty = 1)

# Tukey's contrast matrix
h_tukey <- GFDmcv::contr_mat(k, type = "Tukey")
# testing without parallel computing
res <- multiFANOVA(data_set, gr_label, h_tukey)
summary(res, digits = 3)
# plots for pointwise Hotelling's T^2-test statistics
oldpar <- par(mfrow = c(2, 2), mar = c(2, 2, 2, 0.1))
plot(ph_test_statistic(data_set, gr_label, h_tukey), type = "l",
     ylim = c(0, max(ph_test_statistic(data_set, gr_label, h_tukey))),
     main = "Global hypothesis")
plot(ph_test_statistic(data_set, gr_label, matrix(h_tukey[1, ], 1)), type = "l",
     ylim = c(0, max(ph_test_statistic(data_set, gr_label, h_tukey))),
     main = "Contrast 1")
plot(ph_test_statistic(data_set, gr_label, matrix(h_tukey[2, ], 1)), type = "l",
     ylim = c(0, max(ph_test_statistic(data_set, gr_label, h_tukey))),
     main = "Contrast 2")
plot(ph_test_statistic(data_set, gr_label, matrix(h_tukey[3, ], 1)), type = "l",
     ylim = c(0, max(ph_test_statistic(data_set, gr_label, h_tukey))),
     main = "Contrast 3")
par(oldpar)


# testing with parallel computing
library(doParallel)
res <- multiFANOVA(data_set, gr_label, h_tukey, parallel = TRUE, n_cores = 2)
summary(res, digits = 3)


# Dunnett's contrast matrix
h_dunnett <- GFDmcv::contr_mat(k, type = "Dunnett")
res <- multiFANOVA(data_set, gr_label, h_dunnett)
summary(res, digits = 3)
# plots for pointwise Hotelling's T^2-test statistics
```

```
oldpar <- par(mfrow = c(3, 1), mar = c(2, 2, 2, 0.1))
plot(ph_test_statistic(data_set, gr_label, h_dunnett), type = "l",
     ylim = c(0, max(ph_test_statistic(data_set, gr_label, h_dunnett))),
     main = "Global hypothesis")
plot(ph_test_statistic(data_set, gr_label, matrix(h_dunnett[1, ], 1)), type = "l",
     ylim = c(0, max(ph_test_statistic(data_set, gr_label, h_dunnett))),
     main = "Contrast 1")
plot(ph_test_statistic(data_set, gr_label, matrix(h_dunnett[2, ], 1)), type = "l",
     ylim = c(0, max(ph_test_statistic(data_set, gr_label, h_dunnett))),
     main = "Contrast 2")
par(oldpar)
```

---

ph_test_statistic   *Pointwise Hotelling's $T\hat{}2$-test statistic*

Description

The function `ph_test_statistic()` calculates the pointwise Hotelling's T^2-test statistic.

Usage

```
ph_test_statistic(x, gr_label, h)
```

Arguments

x	matrix of observations $n \times j$ ($n = n_1 + ... + n_k$, j is a number of design time points).
gr_label	a vector with group labels; the integer labels (from 1 to a number of groups) should be used.
h	contrast matrix. For Dunnett's and Tukey's contrasts, it can be created by the `contr_mat()` function from the package GFDmcv (see examples).

Details

For details, see the documentation of the `multiFANOVA()` function or the paper Munko et al. (2023).

Value

A vector of values of the pointwise Hotelling's T^2-test statistic.

References

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

Munko M., Ditzhaus M., Pauly M., Smaga L., Zhang J.T. (2023) General multiple tests for functional data. Preprint https://arxiv.org/abs/2306.15259

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

Examples

```
# Some of the examples may run some time.

# Canadian weather data set
# There are three samples of mean temperatures for
# fifteen weather stations in Eastern Canada,
# another fifteen in Western Canada, and
# the remaining five in Northern Canada.
library(fda)
data_set <- t(CanadianWeather$dailyAv[,, "Temperature.C"])
k <- 3
gr_label <- rep(c(1, 2, 3), c(15, 15, 5))
# trajectories of mean temperatures
matplot(t(data_set), type = "l", col = gr_label, lty = 1,
        xlab = "Day", ylab = "Temperature (C)",
        main = "Canadian weather data set")
legend("bottom", legend = c("Eastern Canada", "Western Canada", "Northern Canada"),
       col = 1:3, lty = 1)

# Tukey's contrast matrix
h_tukey <- GFDmcv::contr_mat(k, type = "Tukey")
# testing without parallel computing
res <- multiFANOVA(data_set, gr_label, h_tukey)
summary(res, digits = 3)
# plots for pointwise Hotelling's T^2-test statistics
oldpar <- par(mfrow = c(2, 2), mar = c(2, 2, 2, 0.1))
plot(ph_test_statistic(data_set, gr_label, h_tukey), type = "l",
     ylim = c(0, max(ph_test_statistic(data_set, gr_label, h_tukey))),
     main = "Global hypothesis")
plot(ph_test_statistic(data_set, gr_label, matrix(h_tukey[1, ], 1)), type = "l",
     ylim = c(0, max(ph_test_statistic(data_set, gr_label, h_tukey))),
     main = "Contrast 1")
plot(ph_test_statistic(data_set, gr_label, matrix(h_tukey[2, ], 1)), type = "l",
     ylim = c(0, max(ph_test_statistic(data_set, gr_label, h_tukey))),
     main = "Contrast 2")
plot(ph_test_statistic(data_set, gr_label, matrix(h_tukey[3, ], 1)), type = "l",
     ylim = c(0, max(ph_test_statistic(data_set, gr_label, h_tukey))),
     main = "Contrast 3")
par(oldpar)
```

---

summary.multifanova   *Print "multifanova" object*

Description

Prints the summary of the global and multiple contrasts testing for functional data.

Usage

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

Arguments

object   a "multifanova" object.

...   integer indicating the number of decimal places to be used to present the numer-
                ical results. It can be named digits as in the round() function (see examples).

Details

The function prints out the information about the number of samples, number of observations in
each sample, number of design time points, contrasts used, test statistics, critical values, p-values
of tests performed by the multiFANOVA() function. It also gives the decisions.

Value

No return value, called for side effects.

Examples

```
# Some of the examples may run some time.

# Canadian weather data set
# There are three samples of mean temperatures for
# fifteen weather stations in Eastern Canada,
# another fifteen in Western Canada, and
# the remaining five in Northern Canada.
library(fda)
data_set <- t(CanadianWeather$dailyAv[,, "Temperature.C"])
k <- 3
gr_label <- rep(c(1, 2, 3), c(15, 15, 5))

# Tukey's contrast matrix
h_tukey <- GFDmcv::contr_mat(k, type = "Tukey")
# testing without parallel computing
res <- multiFANOVA(data_set, gr_label, h_tukey)
summary(res, digits = 3)
```

Index