Package ‘MANOVA.RM’

August 21, 2017

Type Package
Title Analysis of Multivariate Data and Repeated Measures Designs
Version 0.1.2
Date 2017-08-21
Author Sarah Friedrich, Frank Konietschke, Markus Pauly
Maintainer Sarah Friedrich <sarah.friedrich@uni-ulm.de>
Depends R (>= 3.2.3)
Description Implemented are various tests for semi-parametric repeated measures
and general MANOVA designs that do neither assume multivariate normality nor
covariance homogeneity, i.e., the procedures are applicable for a wide range
of general multivariate factorial designs.
License GPL-2 | GPL-3
Imports plyr (>= 1.8.3), MASS (>= 7.3-43), Matrix (>= 1.2-2), magic
(>= 1.5-6), plotrix (>= 3.5-12), parallel, methods, ellipse
LazyData TRUE
Suggests RGtk2 (>= 2.20.31), knitr, rmarkdown, HSAUR, tidyR, GFD
RoxygenNote 5.0.1
VignetteBuilder knitr, rmarkdown
NeedsCompilation no
Repository CRAN
Date/Publication 2017-08-21 09:33:08 UTC

R topics documented:

   MANOVA.RM-package .................................................. 2
   conf.reg .......................................................... 3
   EEG .............................................................. 4
   EEGwide ......................................................... 5
   GUI.MANOVA ...................................................... 6
   GUI.RM ......................................................... 6
The MANOVA.RM package provides two important functions: MANOVA() and RM() which will be explained in detail below.

**MANOVA function**

The MANOVA() function provides the Wald-type statistic (WTS), the ANOVA-type statistic (ATS) as well as a modified ANOVA-type statistic (MATS) as in Friedrich and Pauly (2017) for multivariate designs with metric data as described in Konietschke et al. (2015). These are even applicable for non-normal error terms, different sample sizes and/or heteroscedastic variances. The MATS can even handle designs involving singular covariance matrices. The tests are implemented for designs with an arbitrary number of crossed factors or for nested designs. In addition to the asymptotic p-values, they also provide p-values based on resampling approaches (parametric or wild bootstrap). For further details, see MANOVA.

**RM function**

The RM() function provides the Wald-type statistic as well as the ANOVA-type statistic for repeated measures designs with metric data as described in Friedrich et al. (2017). These are even applicable for non-normal error terms and/or heteroscedastic variances. It is implemented for designs with an arbitrary number of whole-plot and sub-plot factors and allows for different sample sizes. In addition to the asymptotic p-values, it also provides p-values based on resampling approaches (Permutation, parametric bootstrap, Wild bootstrap). For further details, see RM.

**References**


The `conf.reg()` function calculates confidence regions for contrasts in multivariate factorial designs. In the two-dimensional case, confidence ellipsoids can be plotted via the generic `plot()` function.

### Description

The `conf.reg()` function calculates confidence regions for contrasts in multivariate factorial designs. In the two-dimensional case, confidence ellipsoids can be plotted via the generic `plot()` function.

### Usage

```r
conf.reg(object, nullhypo)
```

### Arguments

- **object**: A `manova` object.
- **nullhypo**: In scenarios with more than one factor, the null hypothesis, i.e., the contrast of interest must be specified.

### Value

A `confreg` object containing the following components:

- **center**: The center of the confidence ellipsoid.
- **scale**: The scaling factors for the axis of the confidence ellipsoid calculated as $\sqrt{\lambda \cdot c/N}$, where $\lambda$ are the eigenvalues, $c$ denotes the bootstrap quantile and $N$ is the total sample size. See Friedrich and Pauly (2017) for details.
- **eigenvectors**: The corresponding eigenvectors, which determine the axes of the ellipsoid.

### References


### Examples

```r
data(EEG)
EEG_mod <- MANOVA(resp ~ sex * diagnosis, 
  data = EEG, subject = "id", resampling = "paramBS", 
  alpha = 0.05, iter = 100, CPU = 1)
conf.reg(EEG_mod, nullhypo = "sex")
```
Description

At the Department of Neurology, University Clinic of Salzburg, 160 patients were diagnosed with either AD, MCI, or SCC, based on neuropsychological diagnostics. This data set contains z-scores for brain rate and Hjorth complexity, each measured at frontal, temporal and central electrode positions and averaged across hemispheres. In addition to standardization, complexity values were multiplied by -1 in order to make them more easily comparable to brain rate values: For brain rate we know that the values decrease with age and pathology, while Hjorth complexity values are known to increase with age and pathology. The three between-subjects factors considered were sex (men vs. women), diagnosis (AD vs. MCI vs. SCC), and age (< 70 vs. 70 years). Additionally, the within-subjects factors region (frontal, temporal, central) and feature (brain rate, complexity) structure the response vector.

Usage
data(EEG)

Format

A data frame with 960 rows and 7 variables:

- **resp**  EEG measurements
- **sex**  sex of the patient
- **age**  age of the patient, coded as 0 for less than 70 years and 1 for >= 70 years
- **diagnosis**  neuropsychological diagnosis, AD for Alzheimer’s Disease, MCI for mild cognitive impairment or SCC for subjective cognitive complaints without clinically significant deficits
- **region**  brain region of the EEG measurements, one of "temporal", "frontal" and "central"
- **feature**  feature of the EEG measurements, either "brainrate" or "complexity"
- **id**  Subject id

Source

**Description**

At the Department of Neurology, University Clinic of Salzburg, 160 patients were diagnosed with either AD, MCI, or SCC, based on neuropsychological diagnostics. This data set contains z-scores for brain rate and Hjorth complexity, each measured at frontal, temporal and central electrode positions and averaged across hemispheres. In addition to standardization, complexity values were multiplied by -1 in order to make them more easily comparable to brain rate values: For brain rate we know that the values decrease with age and pathology, while Hjorth complexity values are known to increase with age and pathology. The three between-subjects factors considered were sex (men vs. women), diagnosis (AD vs. MCI vs. SCC), and age (< 70 vs. >= 70 years). Additionally, the within-subjects factors region (frontal, temporal, central) and feature (brain rate, complexity) structure the response vector.

**Usage**

`data(EEGwide)`

**Format**

A data frame with 160 rows and 9 variables:

- `brainrate_temporal` EEG measurements for brainrate in temporal regions
- `brainrate_frontal` EEG measurements for brainrate in frontal regions
- `brainrate_central` EEG measurements for brainrate in central regions
- `complexity_temporal` EEG measurements for complexity in temporal regions
- `complexity_frontal` EEG measurements for complexity in frontal regions
- `complexity_central` EEG measurements for complexity in central regions
- `sex` sex of the patient
- `age` age of the patient, coded as 0 for less than 70 years and 1 for >= 70 years
- `diagnosis` neuropsychological diagnosis, AD for Alzheimer’s Disease, MCI for mild cognitive impairment or SCC for subjective cognitive complaints without clinically significant deficits

**Source**

GUI.RM

Description
This function provides a graphical user interface for calculating statistical tests for multivariate data.

Usage
GUI.MANOVA()

Details
The function produces a GUI for the calculation of the test statistics. Data can be loaded via the "load data" button. The formula, number of resampling iterations (default: 10,000) and the significance level alpha (default: 0.05) need to be specified. Furthermore, the column name specifying the subjects in the data needs to be provided. For the resampling methods, the user can choose between a parametric bootstrap approach (see e.g. Konietschke et al. (2015)) and a Wild bootstrap using Rademacher weights (see e.g. Bathke et al. (2016)).

GUI.RM

Description
This function provides a graphical user interface for calculating statistical tests in repeated measures designs.

Usage
GUI.RM()

Details
The function produces a GUI for the calculation of the test statistics and for plotting. Data can be loaded via the "load data" button. The formula, number of resampling iterations (default: 10,000) and the significance level alpha (default: 0.05) need to be specified. Furthermore, the number of sub-plot factors and the column name specifying the subjects in the data need to be provided. For the resampling methods, the user can choose between a permutation approach (Friedrich et al. (2017)), a parametric bootstrap approach (Konietschke et al. (2015)) and a Wild bootstrap using Rademacher weights (Bathke et al. (2016)). If the plot option is chosen, an additional window opens containing information on the plots.
Description

The MANOVA function calculates the Wald-type statistic (WTS), the ANOVA-type statistic (ATS) and a modified ATS (MATS) as well as resampling versions of these test statistics for semi-parametric multivariate data.

Usage

`MANOVA(formula, data, subject, iter = 10000, alpha = 0.05, resampling = "paramBS", CPU, seed, nested.levels.unique = FALSE)`

Arguments

- `formula`: A model `formula` object. The left hand side contains the response variable and the right hand side contains the factor variables of interest. An interaction term must be specified.
- `data`: A data.frame, list or environment containing the variables in `formula`. Data must be in long format.
- `subject`: The column name of the subjects in the data.
- `iter`: The number of iterations used for calculating the resampled statistic. The default option is 10,000.
- `alpha`: A number specifying the significance level; the default is 0.05.
- `resampling`: The resampling method to be used, one of "paramBS" (parametric bootstrap approach) and "WildBS" (wild bootstrap approach with Rademacher weights). The Wild Bootstrap is calculated for all test statistics.
- `CPU`: The number of cores used for parallel computing. If omitted, cores are detected via `detectCores`.
- `seed`: A random seed for the resampling procedure. If omitted, no reproducible seed is set.
- `nested.levels.unique`: A logical specifying whether the levels of the nested factor(s) are labeled uniquely or not. Default is FALSE, i.e., the levels of the nested factor are the same for each level of the main factor. For an example and more explanations see the GFD package and the corresponding vignette.

Details

The MANOVA() function provides the Wald-type statistic (WTS) as well as the ANOVA-type statistic (ATS) for multivariate designs with metric data as described in Konietschke et al. (2015). Furthermore, it contains a modified ATS (MATS), which is invariant under scale transformations of the components and applicable to designs with singular covariance matrices, see Friedrich and Pauly (2017) for details. These tests are even applicable for non-normal error terms, different sample
sizes and/or heteroscedastic variances. They are implemented for designs with an arbitrary number of crossed factors or for nested designs. In addition to the asymptotic p-values, the function also provides p-values based on resampling approaches.

Value

A MANOVA object containing the following components:

- **Descriptive**: Some descriptive statistics of the data for all factor level combinations. Displayed are the number of individuals per factor level combination and the vector of means (one column per dimension).
- **Covariance**: The estimated covariance matrix.
- **WTS**: The value of the WTS along with degrees of freedom of the central chi-square distribution and p-value.
- **ATS**: The value of the ATS, degrees of freedom of the central F distribution and the corresponding p-value.
- **MATS**: The value of the MATS.
- **resampling**: p-values for the test statistic based on the chosen resampling approach.

NOTE

The number of resampling iterations has been set to 100 in the examples due to run time restrictions on CRAN. Usually it is recommended to use at least 1000 iterations. For more information and detailed examples also refer to the package vignette.

References


See Also

RM
Examples

data(EEG)
EEG_mod <- MANOVA(formula = resp ~ sex * diagnosis,
                   data = EEG, subject = "id", resampling = "paramBS",
                   alpha = 0.05, iter = 100, CPU = 1)

summary(EEG_mod)

MANOVA.wide

Tests for Multivariate Data in Semi-Parametric Factorial Designs

Description

The MANOVA.wide function calculates the Wald-type statistic (WTS), the ANOVA-type statistic (ATS) and a modified ATS (MATS) as well as resampling versions of these test statistics for semi-parametric multivariate data provided in wide format.

Usage

MANOVA.wide(formula, data, iter = 10000, alpha = 0.05,
             resampling = "paramBS", CPU, seed, nested.levels.unique = FALSE)

Arguments

formula A model formula object. The left hand side contains the matrix of response variables and the right hand side contains the factor variables of interest. An interaction term must be specified.
data A data.frame, list or environment containing the variables in formula. Data must be in wide format.
iter The number of iterations used for calculating the resampled statistic. The default option is 10,000.
alpha A number specifying the significance level; the default is 0.05.
resampling The resampling method to be used, one of "paramBS" (parametric bootstrap approach) and "WildBS" (wild bootstrap approach with Rademacher weights). The Wild Bootstrap is calculated for all test statistics.
CPU The number of cores used for parallel computing. If omitted, cores are detected via detectCores.
seed A random seed for the resampling procedure. If omitted, no reproducible seed is set.
nested.levels.unique A logical specifying whether the levels of the nested factor(s) are labeled uniquely or not. Default is FALSE, i.e., the levels of the nested factor are the same for each level of the main factor. For an example and more explanations see the GFD package and the corresponding vignette.
Value

See MANOVA

NOTE

The number of resampling iterations has been set to 100 in the examples due to run time restrictions on CRAN. Usually it is recommended to use at least 1000 iterations.

See Also

MANOVA

Examples

```r
# Example on producing plastic film from Krzanowski (1998, p. 381), see \code{\link{manova.summary}}
tear <- c(6.5, 6.2, 5.8, 6.5, 6.5, 6.9, 7.2, 6.9, 6.1, 6.3, 6.7, 6.6, 7.2, 7.1, 6.8, 7.1, 7.0, 7.2, 7.5, 7.6)
gloss <- c(9.5, 9.9, 9.6, 9.6, 9.2, 9.1, 10.0, 9.9, 9.5, 9.4, 9.1, 9.3, 8.3, 8.4, 8.5, 9.2, 8.8, 9.7, 10.1, 9.2)
opacity <- c(4.4, 6.4, 3.0, 4.1, 0.8, 5.7, 2.0, 3.9, 1.9, 5.7, 2.8, 4.1, 3.8, 1.6, 3.4, 8.4, 5.2, 6.9, 2.7, 1.9)
rate <- gl(2,10, labels = c("Low", "High"))
additive <- gl(2, 5, length = 20, labels = c("Low", "High"))
example <- data.frame(tear, gloss, opacity, rate, additive)
fit <- MANOVA.wide(cbind(tear, gloss, opacity) ~ rate * additive, data = example, iter = 100, CPU = 1)
summary(fit)
```

### o2cons

Oxygen Consumption of Leukocytes

Description

A dataset containing measurements on the oxygen consumption of leukocytes in the presence and absence of inactivated staphylococci.

Usage

```r
data(o2cons)
```

Format

A data frame with 144 rows and 5 variables:

- **O2** oxygen consumption of leukocytes in $\mu$l
- **Staphylococci**: whether or not inactivated staphylococci were added, 1 denotes yes, 0 no
- **Time**: the measurements were taken after 6, 12 and 18 minutes
- **Group**: the treatment group, either P for Placebo or V for Verum
- **Subject**: the subject id
Source

Description
The RM() function calculates the Wald-type statistic (WTS), the ANOVA-type statistic (ATS) as well as resampling versions of these test statistics for semi-parametric repeated measures designs.

Usage
RM(formula, data, subject, no.subf = 1, iter = 10000, alpha = 0.05, resampling = "Perm", CPU, seed)

Arguments
- formula: A model formula object. The left hand side contains the response variable and the right hand side contains the factor variables of interest. An interaction term must be specified. The time variable must be the last factor in the formula.
- data: A data.frame, list or environment containing the variables in formula. Data must be in long format.
- subject: The column name of the subjects in the data.
- no.subf: The number of sub-plot factors in the data, default is 1.
- iter: The number of iterations used for calculating the resampled statistic. The default option is 10,000.
- alpha: A number specifying the significance level; the default is 0.05.
- resampling: The resampling method to be used, one of "Perm" (randomly permute all observations), "paramBS" (parametric bootstrap approach) and "WildBS" (wild bootstrap approach with Rademacher weights). Except for the Wild Bootstrap, all methods are applied to the WTS only.
- CPU: The number of cores used for parallel computing. If omitted, cores are detected via detectCores.
- seed: A random seed for the resampling procedure. If omitted, no reproducible seed is set.

Details
The RM() function provides the Wald-type statistic as well as the ANOVA-type statistic for repeated measures designs with metric data as described in Friedrich et al. (2017). These are even applicable for non-normal error terms and/or heteroscedastic variances. It is implemented for designs with an arbitrary number of whole-plot and sub-plot factors and allows for different sample sizes. In addition to the asymptotic p-values, it also provides p-values based on resampling approaches.
Value

An RM object containing the following components:

- **Descriptive**: Some descriptive statistics of the data for all factor level combinations. Displayed are the number of individuals per factor level combination, the mean and 100*(1-alpha)% confidence intervals (based on t-quantiles).
- **Covariance**: The estimated covariance matrix.
- **WTS**: The value of the WTS along with degrees of freedom of the central chi-square distribution and corresponding p-value.
- **ATS**: The value of the ATS, degrees of freedom of the central F distribution and the corresponding p-value.
- **Resampling**: p-values for the test statistics based on the chosen resampling approach.

References


See Also

GFD, nparLD, MANOVA

Examples

data(o2cons)
oxy <- RM(O2 ~ Group * Staphylococci * Time, data = o2cons, 
  subject = "Subject", no.subf = 2, iter = 100, resampling = "Perm", CPU = 1)
summary(oxy)
plot(oxy, factor = "Group")

# For more details including the output of the examples also refer to the package vignette.

# using the EEG data, consider additional within-subjects factors 'brain region' and 'feature' data(EEG)
## Not run:
set.seed(987)
EEG_model <- RM(resp ~ sex * diagnosis * feature * region, 
  data = EEG, subject = "id", no.subf = 2, resampling = "WildBS", 
  iter = 1000, alpha = 0.01, CPU = 1)
summary(EEG_model)

## End(Not run)
## Index

*Topic **datasets**
- EEG, 4
- EEGwide, 5
- o2cons, 10

- conf.reg, 3
- detectCores, 7, 9, 11

- EEG, 4
- EEGwide, 5
- formula, 7, 9, 11

- GFD, 12
- GUI.MANOVA, 6
- GUI.RM, 6

- MANOVA, 7, 10, 12
- MANOVA.RM-package, 2
- MANOVA.wide, 9

- nparLD, 12
- o2cons, 10

- RM, 8, 11