Package 'nparMD'

October 27, 2025

Type Package

Title Nonparametric Analysis of Multivariate Data in Factorial Designs

Version 0.2.3

Depends R (>= 3.1.0)

Imports matrixStats, matrixcalc, MASS, gtools, Formula, methods, stats

Maintainer Maximilian Kiefel <physik210@hotmail.com>

Description Analysis of multivariate data with two-way completely randomized factorial design. The analysis is based on fully nonparametric, rank-based methods and uses test statistics based on the Dempster's ANOVA, Wilk's Lambda, Lawley-Hotelling and Bartlett-Nanda-Pillai criteria. The multivariate response is allowed to be ordinal, quantitative, binary or a mixture of the different variable types. The package offers two functions performing the analysis, one for small and the other for large sample sizes. The underlying methodology is largely described in Bathke and Harrar (2016) <doi:10.1007/978-3-319-39065-9_7> and in Munzel and Brunner (2000) <doi:10.1016/S0378-3758(99)00212-8> and in Kiefel and Bathke (2022) <doi:10.1515/stat-2022-0112>.

License GPL-2 | GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 7.3.3

NeedsCompilation no

Author Maximilian Kiefel [cre, aut],
Arne C. Bathke [ctb]

Repository CRAN

Date/Publication 2025-10-27 14:00:02 UTC

Contents

---

nparml	*Nonparametric Test For Multivariate Data With Two-Way Layout Factorial Design - Large Samples*

Description

Analysis of multivariate data with two-way completely randomized factorial design - version for large samples. The analysis is based on fully nonparametric, rank-based methods and uses an F-approximation for 'Dempster's ANOVA' and a chisquare-approximation for the criteria called 'Wilks Lambda', 'Lawley-Hotelling' and 'Bartlett-Nanda-Pillai'. These approximations are given by the asymtotic distribution of these statistics under true null-hypothesis. In contrast to the normal-approximated test (as used in the small sample version) it is designed for data with large samples (see details) while the number of factorial levels is allowed to be small. The multivariate response is allowed to be ordinal, quantitative, binary or a mixture of the different variable types. The test statistics are constructed using nonparametric relative effect estimators.

Usage

```
nparml(formula, data)
```

Arguments

formula	an object of class "formula" with two explanatory variables (factors), see examples.
data	an object of class "data.frame" containing the variables in the formula

Details

The data is analysed for main effects and interaction effect of the explanatory factors. In each case the null hypothesis "no effect" is testet. In order to obtain reliable results the considered data should include at least 7 observations per factor level combination. This method is only implemented for complete data sets without missing values.

Value

Returns a list of data frames providing the values of the test statistics, p-values, degrees of freedom, factor levels, and groupsize per factor level combination.

References

Kiefel M., Freidl J. (2024) Nonparametric Analysis of Multivariate Data in Factorial Designs with Nondetects: A Case Study with Microbiome Data In: Journal of Agricultural, Biological, and Environmental Statistics https://doi.org/10.1007/s13253-024-00671-5.

Kiefel M., Bathke A.C. (2022) Fully Nonparametric Methods for Multivariate Data in Factorial Designs. Asymptotics, Finite Sample Approximations, and Implementation in R In: Open Statistics, vol. 3, 2022, p. 74 https://doi.org/10.1515/stat-2022-0112

Kiefel M., Bathke A.C. (2020) Rank-Based Analysis of Multivariate Data in Factorial Designs and Its Implementation in R In: Nonparametric Statistics (285-294) Springer Proceedings in Mathematics & Statistics Springer International Publishing, Cham

Bathke A.C., Harrar S.W. (2016) Rank-Based Inference for Multivariate Data in Factorial Designs. In: Liu R., McKean J. (eds) Robust Rank-Based and Nonparametric Methods. Springer Proceedings in Mathematics & Statistics, vol 168. Springer, Cham

Harrar S.W., Bathke A.C. (2012) A modified two-factor multivariate analysis of variance: asymptotics and small sample approximations (and erratum). In: Annals of the Institute of Statistical Mathematics, 64(1&5):135-165&1087, 2012.

Brunner E., Dette H., Munk A. (1997) Box-Type Approximations in Nonparametric Factorial Designs In: Journal of the American Statistical Association, 92(440):1494-1502

Examples

```
data(pseudostudy1)
nparml(resp1|resp2|resp3~treatment*age, pseudostudy1)
```

---

nparms | *Nonparametric Test For Multivariate Data With Two-Way Layout Factorial Design - Small Samples*

Description

Analysis of multivariate data with two-way completely randomized factorial design - version for small samples. The analysis is based on fully nonparametric, rank-based methods and uses a N(0,1)-approximation for test statistics based on 'Dempster's ANOVA', 'Wilk's Lambda', 'Lawley-Hotelling' and 'Bartlett-Nanda-Pillai' criteria. This approximation is established by the asymptotic distribution of these four statistics under true null-hypothesis if one of the explanatory factors has a large number of levels. The multivariate response is allowed to be ordinal, quantitative, binary or a mixture of the different variable types. The test statistics are constructed using nonparametric relative effect estimators.

Usage

```
nparms(formula,data)
```

Arguments

formula	an object of class "formula" with two explanatory variables (factors), see examples.
data	an object of class "data.frame" containing the variables in the formula

Details

This method is only implemented for complete data sets without missing values. The data is analysed for main effects and interaction effect of the explanatory factors. In each case the null hypothesis "no effect" is testet. The explanatory factor that has the higher number of levels is automatically designated as factor "A". The covariance matrix estimation requires at least 4 observations (observation vectors) per factor level combination. As the estimation is very time-consuming for large groups it is performed wih a random selection of observations when a group exceeds a size of 6 observation vectors.

Value

Returns a list of data frames providing the values of the test statistics, p-values, degrees of freedom, factor levels, and groupsize per factor level combination.

References

Kiefel M., Freidl J. (2024) Nonparametric Analysis of Multivariate Data in Factorial Designs with Nondetects: A Case Study with Microbiome Data In: Journal of Agricultural, Biological, and Environmental Statistics https://doi.org/10.1007/s13253-024-00671-5.

Kiefel M., Bathke A.C. (2022) Fully Nonparametric Methods for Multivariate Data in Factorial Designs. Asymptotics, Finite Sample Approximations, and Implementation in R In: Open Statistics, vol. 3, 2022, p. 74 https://doi.org/10.1515/stat-2022-0112

Kiefel M., Bathke A.C. (2020) Rank-Based Analysis of Multivariate Data in Factorial Designs and Its Implementation in R In: Nonparametric Statistics (285-294) Springer Proceedings in Mathematics & Statistics Springer International Publishing, Cham

Bathke A.C., Harrar S.W. (2016) Rank-Based Inference for Multivariate Data in Factorial Designs. In: Liu R., McKean J. (eds) Robust Rank-Based and Nonparametric Methods. Springer Proceedings in Mathematics & Statistics, vol 168. Springer, Cham

Harrar S.W., Bathke A.C. (2012) A modified two-factor multivariate analysis of variance: asymptotics and small sample approximations (and erratum). In: Annals of the Institute of Statistical Mathematics, 64(1&5):135-165&1087, 2012.

Brunner E., Dette H., Munk A. (1997) Box-Type Approximations in Nonparametric Factorial Designs In: Journal of the American Statistical Association, 92(440):1494-1502

Examples

```
data(pseudostudy2)
nparms(resp1|resp2|resp3~treatment*age, pseudostudy2)
```

---

  pseudostudy1   *pseudostudy dataset 1*

Description

Example data set in a two-way layout without effects (in terms of disribution) by factors "treatment" and "age". In constrast to the 'pseudostudy2' data set the number of factorial levels is low while the number of observations per factor level combination is large.

Usage

```
data(pseudostudy1)
```

Format

pseudostudy consists of 56 cases (rows) and 5 variables where 'treatment'('age') is considered to be the explanatory factor 'A'('B') and the variables resp1, resp2. resp3 form the response vector.

Examples

```
data(pseudostudy1)
```

---

pseudostudy2	*pseudostudy dataset 2*

Description

Example data set in a two-way layout without effects (in terms of disribution) by factors "treatment" and "age".

Usage

```
data(pseudostudy2)
```

Format

pseudostudy consists of 107 cases (rows) and 5 variables where 'treatment'('age') is considered to be the explanatory factor 'A'('B') and the variables resp1, resp2. resp3 form the response vector.

Examples

```
data(pseudostudy2)
```

Index