Package ‘GFD’

October 12, 2022

Type Package

Title Tests for General Factorial Designs

Version 0.3.3

Date 2022-01-17

Author Sarah Friedrich, Frank Konietschke, Markus Pauly, Marc Ditzhaus, Philipp Steinhauer

Maintainer Sarah Friedrich <sarah.friedrich@math.uni-augsburg.de>

Depends R (>= 3.3)

Description Implemented are the Wald-type statistic, a permuted version thereof as well as the ANOVA-type statistic for general factorial designs, even with non-normal error terms and/or heteroscedastic variances, for crossed designs with an arbitrary number of factors and nested designs with up to three factors. Friedrich et al. (2017) <doi:10.18637/jss.v079.c01>.

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), methods, shiny (>= 1.4), shinyjs, shinyWidgets, shinythemes, tippy

LazyData TRUE

Suggests RGtk2 (>= 2.20.31), knitr, rmarkdown, HSAUR

RoxygenNote 6.1.1

VignetteBuilder knitr, rmarkdown

Encoding UTF-8

NeedsCompilation no

Repository CRAN

Date/Publication 2022-01-18 07:52:51 UTC

R topics documented:

calculateGUI .......................................................... 2
curdies ................................................................. 2
calculateGUI

A graphical user interface for the package GFD

Description

This function provides a graphical user interface for calculating statistical tests in general factorial designs.

Usage

calculateGUI()

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 permutations (default: 10,000) and the significance level alpha (default: 0.05) need to be specified. If the plot option is chosen, an additional window opens containing information on the plots.

curdies

Curdies river data set

Description

A dataset containing the number of flatworms (dugesia) sampled in two seasons at different sites in the Curdies River in Western Victoria.

Usage

data(curdies)

Format

A data frame with 36 rows and 3 variables:

- season  a factor with levels "SUMMER" and "WINTER"
- site    a factor with levels 1 to 6, nested within "season"
- dugesia number of flatworms counted on a particular stone (in no./dm^2)
**GFD**

*Tests for General Factorial Designs*

---

**Description**

The GFD function calculates the Wald-type statistic (WTS), the ANOVA-type statistic (ATS) as well as a permutation version of the WTS for general factorial designs.

**Usage**

```r
GFD(formula, data = NULL, nperm = 10000, alpha = 0.05, nested.levels.unique = FALSE, CI.method = "t-quantile")
```

**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`. The default option is `NULL`.
- `nperm`: The number of permutations used for calculating the permuted Wald-type statistic. The default option is `10000`.
- `alpha`: A number specifying the significance level; the default is `0.05`.
- `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.
- `CI.method`: Method for calculating the confidence intervals. Default is `"t-quantile"` for CIs based on the corresponding t-quantile. Additionally, the quantile of the permutation distribution can be used (`"perm"`).

**Details**

The package provides the Wald-type statistic, a permuted version thereof as well as the ANOVA-type statistic for general factorial designs, even with non-normal error terms and/or heteroscedastic variances. It is implemented for both crossed and hierarchically nested designs and allows for an arbitrary number of factor combinations as well as different sample sizes in the crossed design. The GFD function returns three p-values: One for the ATS based on an F-quantile and two for the WTS, one based on the $\chi^2$ distribution and one based on the permutation procedure. Since the ATS is only an approximation and the WTS based on the $\chi^2$ distribution is known to be very liberal for small sample sizes, we recommend to use the WTPS in these situations.
Value

A GFD 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, variance and 100*(1-alpha)% confidence intervals.

WTS  The value of the WTS along with degrees of freedom of the central chi-square distribution and p-value, as well as the p-value of the permutation procedure.

ATS  The value of the ATS, degrees of freedom of the central F distribution and the corresponding p-value.

References


Examples

data(startup)
model <- GFD(Costs ~ company, data = startup, CI.method = "perm")
summary(model)

---

GFD_GUI

A shiny app for the package GFD

Description

This function provides a shiny app for calculating GFD and QANOVA test statistics and respective p-values.

Usage

GFD_GUI()

Author(s)

Philipp Steinhauer
pizza

Pizza delivery times

Description

A dataset containing the delivery times for pizza (in minutes) under different conditions.

Usage

data(pizza)

Format

A data frame with 16 rows and 6 variables:

- **Crust** a factor with levels "thick" and "thin"
- **Coke** whether or not Coke was ordered with the pizza ("yes" or "no")
- **Bread** whether or not garlic bread was ordered with the pizza ("yes" or "no")
- **Driver** the sex of the driver, a factor with levels "M" and "F"
- **Hour** time of order in hours after midnight
- **Delivery** Delivery time in minutes

Source

http://www.statsci.org/data/oz/pizza.html

QANOVA

QANOVA: Quantile-based analysis-of-variance

Description

The function qanova calculates the Wald-type statistic based on the quantiles and/or their linear combinations, e.g. the interquartile range. Respective p-values are obtained by a $\chi^2$-approximation and a permutation approach, respectively.

Usage

QANOVA(formula, data = NULL, quantiles = c(0.5), lin_mat = NULL, var_method = "interval", nperm = 1999, var_level = 0.95, 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 and the censoring status indicator. Default option is NULL.

**quantiles**
A vector of probabilities corresponding to the quantiles of interest. By default is c(0.5), i.e. just the median is included.

**lin_mat**
A matrix specifying which linear combination of the quantiles should be included for the analysis. By default (NULL) the identity matrix is chosen, i.e. all chosen quantiles are considered simulatenously.

**var_method**
Method for the variance estimation of the sample quantiles. The default ("interval") is the interval-based estimator of Price and Bonett (2001). Additionally, the bootstrap method ("boot") of Efron (1987) or a kernel density approach ("kernel") can be chosen.

**nperm**
The number of permutations used for calculating the permuted p-value. The default option is 1999.

**var_level**
A number between 0 and 1 specifying the confidence level for the interval variance estimation method; the default value is 0.95.

**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.

Details

The qanova function calculates the Wald-type statistic based on quantiles and linear combinations of them for general factorial designs. The procedure is fully nonparametric and no specific assumption of the underlying distribution is required. In particular, heteroscedastic settings can be studied. The analysis can be based on a single quantile (e.g. the median, default choice), a linear combination of quantiles (e.g. the interquartile range, set quantiles=c(0.25,0.75) and lin_mat = matrix(c(-1,1),ncol=2)) or on several (combinations of) quantiles simulatenously.

The qanova function returns the test statistic as well as two corresponding p-values: the first is based on a \( \chi^2 \) approximation and the second one is based on a permutation procedure.

@return A qanova object containing the following components:

- **pvalues_stat** The p-values obtained by \( \chi^2 \)-approximation
- **pvalues_per** The p-values of the permutation approach
- **statistics** The value of the qanova along with degrees of freedom of the central chi-square distribution and p-value, as well as the p-value of the permutation procedure.
- **nperm** The number of permutations used for calculating the permuted p-value.

Author(s)

Philipp Steinhauer
References

Examples
QANOVA(weightgain ~ source*type, data = HSAUR::weightgain, var_method = "interval", nperm =199)

startup

Startup Costs of five different companies

Description
A dataset containing the startup costs (in thousands of dollars) of five companies.

Usage
data(startup)

Format
A data frame with 60 rows and 2 variables:

Costs price, in thousands of dollars
company company, a factor with levels "pets", "pizza", "gifts", "shoes" and "bakery"

Source
https://college.cengage.com/mathematics/brase/understandable_statistics/7e/students/datasets/owan/frames/frame.html
Index

* datasets
  curdies, 2
  pizza, 5
  startup, 7
calculateGUI, 2
curdies, 2
formula, 3
GFD, 3
GFD_GUI, 4
GFDsurvGUI (GFD_GUI), 4
GUI (calculateGUI), 2

pizza, 5
QANOVA, 5
startup, 7