Package 'hypergea'

January 26, 2018

Type Package

Title Hypergeometric Tests

Version 1.3.6

Date 2018-01-26

Author Markus Boenn

Maintainer Markus Boenn <markus.boenn.sf@googlemail.com>

Description Performs (exact) hypergeometric tests on IxJ and 2x2x2 contingency tables using parallelized C code.

License GPL (>= 2)

Depends R (>= 3.0.0)

Imports

Suggests RUnit

SystemRequirements OpenMP (>=3.0)

NeedsCompilation yes

Repository CRAN

Date/Publication 2018-01-26 13:24:17 UTC

R topics documented:

getExpectedValues *Expected values of multidimensional contingency tables*

Description

Calculates expected values of a multidimensional contingency table

Usage

```
getExpectedValues(x)
```

Arguments

x	table (matrix or array)

Value

an array containing the expected values

Examples

```
x <- array(c(1:8), dim=c(2,2,2))
getExpectedValues(x)
```

getMargins *Marginal sums of multidimensional contingency tables*

Description

Calculates marginal sums for $I \times J$ and $I \times J \times K$ tables

Usage

```
getMargins(x)
```

Arguments

x	a table (matrix, array) without missing values

Value

A list of `length(dim(x))` containing the marginal sums of each dimension

Examples

```
x <- matrix(c(1:4), nrow=2, ncol=2)
getMargins(x)

x <- matrix(c(1:6), nrow=2, ncol=3)
getMargins(x)

x <- array(c(1:8), dim=c(2,2,2))
getMargins(x)
```

`getOddsRatio` *Odds ratio*

Description

This function calculates odds ratios for 2×2 and $2 \times 2 \times 2$ contingency tables.

Usage

```
getOddsRatio(x, za=TRUE)
```

Arguments

`x`	a contingency table (array or matrix)
`za`	zero adjustment, i.e. how to behave in case of zeros (logical). If `FALSE` than odds ratios as calculated. If `TRUE`, each cell is adjusted by addition of a small number (0.5).

Details

If `x` is a 2×2 table then the odds ratio is returned, defined as

$$OR = \frac{x_{00}x_{11}}{x_{10}x_{01}}$$

If `x` is a $2 \times 2 \times 2$ table then Bartlett's ratio of odds ratios is calculated, defined as

$$OR = \frac{x_{000}x_{111}x_{011}x_{101}}{x_{101}x_{001}x_{100}x_{111}}$$

To be able to calculate odds ratios even for huge numbers, the log of each cell count is taken.

Calculation of an odds ratio becomes difficult in presence of zeros. Depending on the position(s) of zero(s) in the table, the ratio becomes zero, `NaN` or `Inf`. This corresponds to the behaviour of the `fisher.test` in such a case. However, another strategy is (Haldane) to add a small number (0.5) to each cell when the table contains at least one zero.

Note that this routine calculates the sample odds ratio (also: unconditional Maximum Likelihood estimate). This is different to other routines like `fisher.test`, which calculate the conditional Maximum Likelihood Estimate (MLE).

Value

the odds ratio (numeric) if x is a 2×2 or $2 \times 2 \times 2$ table, NA otherwise

References

MARTIN A. HAMILTON (1979) CHOOSING THE PARAMETER FOR A 2 x 2 TABLE OR A 2 x 2 x 2 TABLE ANALYSIS. Am. J. Epidemiol. (1979) 109 (3): 362-375

See Also

fisher.test, vcd package for visualisation of contingency tables

Examples

```
CT <- matrix(c(1:4), nrow=2)
getOddsRatio(CT)

CT <- array(c(1:8), dim=c(2,2,2))
getOddsRatio(CT)
```

---

hypergeom.test *Hypergeometric test*

Description

Performs exact hypergeometric tests for $I \times J$ and $2 \times 2 \times 2$ contingency tables

Usage

```
hypergeom.test(x, alternative = "two.sided",
pval.method = "fisheR", nthreads = 2, ...)
```

Arguments

x	an integer array (or matrix) without missing values
alternative	a character string specifying the alternative hypothesis, must be one of "two.sided" (default), "greater" or "less"
pval.method	a character vector specifying which approach should be used to determine the p-value in a two-sided test. Can be "fisheR", "minimum.likelihood" or "double". Ignored if 'alternative' is "less" or "greater"
nthreads	the number of threads that should be used
...	additional arguments. Currently not used

Details

There is a lot of discussion about the correct way to calculate a p-value in a two sided test. Thus, several alternatives are possible (see Agresti (1992) for a short discussion and further references).

Three types of two-sided p-values are implemented. The first one (default) is the approach used by the Fisher's test implementation In R. It is currently used only for 2×2 tables. It is similar to the minimum-likelihood approach, which is also usable. The third one is the so-called doubling approach. In case of tables with I!=2 and J=!2 or $2 \times 2 \times 2$ tables instead of the fisheR implementation the minimum-likelihood implementation is applied.

If pval.method is longer than one, all corresponding p-values are provided in the output, but only the first one is shown when calling output-object. If pval.method is `NULL`, all p-values are provided in the output, but only the first one (currently "fisheR") is shown when calling output-object.

Value

A list with class '"htest"' containing the following components

`statistic`	the value observed if the cell of interest
`p.value`	the p-value for the selected test
`estimate`	the sample odds ratio calculated from 'x'
`alternative`	a character string describing the alternative hypothesis
`method`	a character string indicating what type of hypergeometric test was performed
`conf.int`	the 95% confidence interval of the odds ratio

Source

The code for 2x2x2 hypergeometric test is a C translation of the Fortran algorithm of Zelterman (1999).

References

Agresti, Alan (1992) A survey of exact inference for contingency tables. Statistical Science Vol. 7, No. 1 (Feb., 1992), pp. 131-153

Zelterman, Daniel (1999) Models for Discrete Data. Oxford University Press

See Also

`fisher.test` to calculate p-values for 2×2 tables

Examples

```
##########
## example from Zelterman (1999), p. 119, Table 4.1
## mice were exposed to a fungicide (or not), some have a tumor
##########
dmn <- list(Tumor=c('y', 'n'), Exposition=c('y', 'n'))
CT <- matrix(c(4,12,5,74), nrow=2, dimnames=dmn)
## Not run: CT
```

```
hypergeom.test(CT)

## Not run:
##########
## example from Agresti (1992), Table 2
## Fisher's tea tasting experiment
##########
dmn <- list('poured first'=c('m', 't'), guess=c('m', 't'))
CT <- matrix(c(3,1,1,3), nrow=2, dimnames=dmn)
hypergeom.test(CT, alternative="two.sided")
hypergeom.test(CT, alternative="two.sided", pval.method="double")
out <- hypergeom.test(CT, alternative="two.sided", pval.method=NULL)
hypergeom.test(CT, alternative="less")
hypergeom.test(CT, alternative="greater")

## End(Not run)


##########
## example from Zelterman (1999), p. 44, Table 1.2
## Response to three drugs
##########
dmn <- list(A=c('F', 'U'), B=c('F', 'U'), C=c('F', 'U'))
CT <- array(c(6,16,2,4,2,4,6,6), dim=c(2,2,2), dimnames=dmn)
## Not run: CT
hypergeom.test(CT)
```

Index