fisher.test {stats}

Fisher's Exact Test for Count Data

Description

Performs Fisher's exact test for testing the null of independence of rows and columns in a contingency table with fixed marginals.

Usage

`fisher.test(x, y = NULL, workspace = 200000, hybrid = FALSE, control = list(), or = 1, alternative = "two.sided", conf.int = TRUE, conf.level = 0.95)`

Arguments

- `x`: either a two-dimensional contingency table in matrix form, or a factor object.
- `y`: a factor object; ignored if `x` is a matrix.
- `workspace`: an integer specifying the size of the workspace used in the network algorithm. In units of 4 bytes.
- `hybrid`: a logical. Only used for larger than 2 by 2 tables, in which cases it indicated whether the exact probabilities (default) or a hybrid approximation thereof should be computed. See Details.
- `control`: a list with named components for low level algorithm control. At present the only one used is "mult", a positive integer >= 2 with default 30. This says how many times as much space should be allocated to paths as to keys: see file ‘fexact.c’ in the sources of this package.
- `or`: the hypothesized odds ratio. Only used in the 2 by 2 case.
- `alternative`: indicates the alternative hypothesis and must be one of "two.sided", "greater" or "less". You can specify just the initial letter. Only used in the 2 by 2 case.
- `conf.int`: logical indicating if a confidence interval should be computed (and returned).
- `conf.level`: confidence level for the returned confidence interval. Only used in the 2 by 2 case.

Details

If `x` is a matrix, it is taken as a two-dimensional contingency table, and hence its entries should be nonnegative integers. Otherwise, both `x` and `y` must be vectors of the same length. Incomplete cases are removed, the vectors are coerced into factor objects, and the contingency table is computed from these.

In the one-sided 2 by 2 cases (and where `or` is specified), p-values are obtained directly using the hypergeometric distribution. Otherwise, computations are based on a C version of the FORTRAN subroutine FEXACT which implements the network developed by Mehta and Patel (1986) and improved by Clarkson, Fan & Joe (1993). The FORTRAN code can be obtained from `http://www.netlib.org/toms/643`. Note this fails (with an error message) when the entries of the table are too large. (It transposes the table if necessary so it has no more rows than columns. One constraint is that the product of the row marginals be less than $2^{31} - 1$.)

In the 2 by 2 case, the null of conditional independence is equivalent to the hypothesis that the odds ratio equals one. ‘Exact’ inference can be based on observing that in general, given all marginal totals fixed, the first element of the contingency table has a non-central hypergeometric distribution with non-centrality parameter given by the odds ratio (Fisher, 1935). The alternative for a one-sided test is based on the odds ratio, so alternative = "greater" is a test of the odds ratio being bigger than or.

Two-sided tests are based on the probabilities of the tables, and take as ‘more extreme’ all tables with probabilities less than or equal to that of the observed table, the p-value being the sum of such probabilities.

For larger then 2 by 2 tables and hybrid = TRUE, asymptotic chi-squared probabilities are only used if the “Cochran conditions” are satisfied, that is if no cell has count zero, and more than 80% of the cells have counts at least 5.

Value

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

- **p.value**: the p-value of the test.
- **conf.int**: a confidence interval for the odds ratio. Only present in the 2 by 2 case.
- **estimate**: an estimate of the odds ratio. Note that the conditional Maximum Likelihood Estimate (MLE) rather than the unconditional MLE (the sample odds ratio) is used. Only present in the 2 by 2 case.
- **null.value**: the odds ratio under the null, or. Only present in the 2 by 2 case.
- **alternative**: a character string describing the alternative hypothesis.
- **method**: the character string "Fisher's Exact Test for Count Data".
- **data.name**: a character string giving the names of the data.

References


See Also
Examples

## Agresti (1990), p. 61f, Fisher's Tea Drinker
## A British woman claimed to be able to distinguish whether milk or
## tea was added to the cup first. To test, she was given 8 cups of
## tea, in four of which milk was added first. The null hypothesis
## is that there is no association between the true order of pouring
## and the women's guess, the alternative that there is a positive
## association (that the odds ratio is greater than 1).
TeaTasting <-
  matrix(c(3, 1, 1, 3),
         nr = 2,
         dimnames = list(Guess = c("Milk", "Tea"),
                         Truth = c("Milk", "Tea")))
fisher.test(TeaTasting, alternative = "greater")
## => p=0.2429, association could not be established

## Fisher (1950, 1962), Convictions of like-sex twins in criminals
Convictions <-
  matrix(c(2, 10, 15, 3),
         nr = 2,
         dimnames =
                    list(c("Dizygotic", "Monozygotic"),
                         c("Convicted", "Not convicted")))
Convoctions
fisher.test(Convictions, alternative = "less")
fisher.test(Convictions, conf.int = FALSE)
fisher.test(Convictions, conf.level = 0.95)$conf.int
fisher.test(Convictions, conf.level = 0.99)$conf.int

[Package stats version 2.2.1 Index]

Last updated with Webcuts support: Sun Mar 12 18:48:14 MET 2006