

Package ‘stratastats’

July 23, 2025

Title Stratified Analysis of 2x2 Contingency Tables

Version 0.2

Description Offers a comprehensive approach for analysing stratified 2x2 contingency tables. It facilitates the calculation of odds ratios, 95% confidence intervals, and conducts chi-squared, Cochran-Mantel-Haenszel, Mantel-Haenszel, and Breslow-Day-Tarone tests. The package is particularly useful in fields like epidemiology and social sciences where stratified analysis is essential. The package also provides interpretative insights into the results, aiding in the understanding of statistical outcomes.

Depends R ($\geq 4.0.0$)

Imports abind ($\geq 1.4.0$), gt ($\geq 0.8.0$), stats ($\geq 4.0.0$)

License GPL (≥ 2)

Encoding UTF-8

RoxygenNote 7.2.3

NeedsCompilation no

Author Gianmarco Alberti [aut, cre]

Maintainer Gianmarco Alberti <gianmarcoalberti@gmail.com>

Repository CRAN

Date/Publication 2024-01-08 14:00:02 UTC

Contents

stratastats	2
Index	6

Description

This function performs a comprehensive stratified analysis of 2x2 contingency tables. It calculates odds ratios, 95 percent confidence intervals, conducts chi-squared tests, Cochran-Mantel-Haenszel tests for conditional independence, Mantel-Haenszel and Breslow-Day-Tarone tests for homogeneity of odds ratios across strata. Additionally, it produces a nicely-formatted table using the `gt` package, which includes both the results and their interpretation, facilitating easier understanding and presentation of the analysis. The function is designed to work with either a list of 2x2 tables or a 3-dimensional array representing stratified tables.

Visit this [LINK](#) to access the package's vignette.

Usage

```
stratastats(  
  input.data,  
  variable.names = c("Var1", "Var2"),  
  stratifying.variable.name = "Var3",  
  flip.or = FALSE,  
  table.font.size = 13,  
  source.note.font.size = 11  
)
```

Arguments

<code>input.data</code>	A list of 2x2 contingency tables or a 3-dimensional array where each 2x2 slice along the third dimension represents a stratum.
<code>variable.names</code>	A character vector containing the names of the two variables being cross-tabulated. Default is "Var1" and "Var2". If not provided for a 3D array, they are extracted from the array's dimnames.
<code>stratifying.variable.name</code>	A character vector containing the name of the stratifying variable. Default is "Var3". If not provided for a 3D array, it is extracted from the array's dimnames.
<code>flip.or</code>	Optional. Logical value indicating whether to flip the odds ratio (default is FALSE). If TRUE, the reciprocal of the OR is calculated.
<code>table.font.size</code>	Optional. Font size for the output table (default is 13).
<code>source.note.font.size</code>	Optional. Font size for the source notes in the output table (default is 11).

Details

The function employs statistical techniques appropriate for stratified data analysis.

The odds ratio for each stratum and the combined strata (marginal table) are calculated. Confidence intervals are derived based on the standard error of the log odds ratio.

Chi-squared tests are conducted for each table to assess the association between the variables at each level of the stratifying variable.

The Cochran-Mantel-Haenszel (CMH) test assesses the overall association while accounting for stratification.

The Mantel-Haenszel (MH) test and the Breslow-Day-Tarone (BDT) test evaluate the homogeneity of odds ratios across strata. The latter test uses the code implemented by Michael Hoehle (see References).

The output includes a detailed breakdown of the odds ratios, confidence intervals, and test statistics for each stratum and the marginal table. It also presents combined results with annotations explaining the significance and implications of the tests.

Interpretational Scenarios:

1. Significant CMH test with homogeneity (non-significant MH and BDT test): Indicates conditional dependence and consistent association across strata. The common odds ratio is a reliable summary of the association.
2. Significant CMH test with heterogeneity (significant MH and BDT test): Suggests conditional dependence and varying strength or direction of association across strata (interaction), cautioning against a simple summary of the association.
3. Non-significant CMH test with homogeneity (non-significant MH and BDT test): Implies conditional independence and consistent association across strata. The common odds ratio is a reliable summary of the association.
4. Non-significant CMH test with heterogeneity (significant MH and BDT test): Since not all the conditional odds ratios are in the same direction, the result of the CMH test might not be reliable. Evaluating the stratum-specific chi-squared tests should be considered.

Note that:

(1) the interpretation guidelines provided by the function are suggested based on the statistical tests' outcomes and should be further evaluated within the context of your study;

(2) in case a table features a zero along any of the diagonals, the Haldane-Anscombe correction is applied for the calculation of the odds ratios and to internally compute the weights used in the Mantel-Haenszel test of homogeneity. The correction consists in adding 0.5 to every cell of the table. The original data, on which all the other tests are based, remain unchanged.

On the correction, see for example: Fleiss et al 2003; Pagano-Gauvreau 2018. See the latter also for the application of the correction in the calculation of the weights used in the Mantel-Haenszel test.

Value

A list containing the following elements:

- `odds.ratios`: A data frame of odds ratios and their 95 percent confidence intervals for each stratum and the marginal table.
- `partial.chisq.result`: A character vector storing the result of the chi-squared test for each partial table.

- `marginal.chisq.result`: A character vector storing the result of the chi-squared test for the marginal table.
- `cmh.test.result`: Result of the Cochran-Mantel-Haenszel test, providing an overall measure of association while accounting for stratification.
- `mh.test.result`: Result of the Mantel-Haenszel test, used to assess the homogeneity of the odds ratios across strata.
- `bdt.test.result`: Result of the Breslow-Day-Tarone test, used to assess the homogeneity of the odds ratios across strata.

References

Azen, R., & Walker, C. M. (2021). *Categorical data analysis for the behavioral and social sciences* (2nd ed.). New York: Routledge.

Breslow, N. E., & Day, N. E. (1980). *Statistical methods in cancer research. Volume I - The analysis of case-control studies*. IARC Scientific Publications.

Fleiss, J. L., Levin, B., & Paik, M. C. 2003. *Statistical Methods for Rates and Proportions* (3rd ed.). Wiley.

Hoehle, M. (2000). Breslow-Day-Tarone Test. Retrieved from <https://online.stat.psu.edu/onlinecourses/sites/stat504/files/les>

Lachin, J. M. (2000). *Biostatistical methods: The assessment of relative risks*. Wiley.

Pagano, M., & Gauvreau, K. (2018). *Principles of Biostatistics* (2nd ed.). Chapman and Hall/CRC.

Sheskin, D. J. (2011). *Handbook of parametric and nonparametric statistical procedures* (5th ed.). Chapman & Hall/CRC

See Also

Refer to [chisq.test](#) for chi-squared tests, and to [mantelhaen.test](#) for the Cochran-Mantel-Haenszel test.

Examples

```
# EXAMPLE 1
# Survival on the Titanic

# create three individual partial tables

table1 <- matrix(c(118,5,61,139), byrow = TRUE, ncol=2)
table2 <- matrix(c(146,12,25,94), byrow = TRUE, ncol=2)
table3 <- matrix(c(418,110,75,106), byrow = TRUE, ncol=2)

# make a list

tables <- list(table1, table2, table3)

# specify the variable names

varnames <- c("Survival", "Gender")
stratvar <- "Class"
```

```
# carry out the analysis
results <- stratastats(input.data = tables, variable.names = varnames,
stratifying.variable.name = stratvar)

# EXAMPLE 2
# Smoking status and breathing test results (after Azen-Walker 2021)

table1 <- matrix(c(577, 34, 682, 57), byrow = TRUE, ncol=2)
table2 <- matrix(c(164,4,245,74), byrow = TRUE, ncol=2)

tables <- list(table1, table2)

varnames <- c("Smoking Status", "Breathing Test Result")
stratvar <- "Age"

results <-stratastats(input.data = tables, variable.names = varnames,
stratifying.variable.name = stratvar)

# EXAMPLE 3
# Admission to graduate school at Berkeley in 1973 (3-dimensional array).
# Since the array contains variables name and the name of the stratifying variable,
# all we need to feed into the function is the dataset name 'UCBAdmissions'. However,
# the name of the variables can be customised using either the 'variable.names'
# or 'stratifying.variable.name' # parameter, or both.

results <- stratastats(input.data = UCBAdmissions)
```

Index

* **stratastats**

stratastats, 2

chisq.test, 4

gt, 2

mantelhaen.test, 4

stratastats, 2