

Package ‘choplump’

October 12, 2022

Type Package

Title Permutation Test for Some Positive and Many Zero Responses

Version 1.1.1

Date 2022-05-09

Author Michael P. Fay

Maintainer Michael P. Fay <mfay@niaid.nih.gov>

Depends R (>= 2.5.0), stats

Suggests coin, survival

Description Calculates permutation tests that can be powerful for comparing two groups with some positive but many zero responses (see Follmann, Fay, and Proschan <[DOI:10.1111/j.1541-0420.2008.01131.x](https://doi.org/10.1111/j.1541-0420.2008.01131.x)>).

License GPL-3

NeedsCompilation no

Repository CRAN

Date/Publication 2022-05-09 15:50:02 UTC

R topics documented:

| | |
|----------------------------------|---|
| choplump-package | 2 |
| chooseMatrix | 3 |
| choplump | 3 |
| choplumpGeneral | 6 |
| methodRule1 | 7 |
| wilcox.manyzeros.exact | 8 |

| | |
|--------------|----------|
| Index | 9 |
|--------------|----------|

Description

This package has basically one important function, `choplump` for performing the choplump test, which is for comparing two groups with some positive response and many zero responses. These tests can often be more powerful than simpler permutation tests. Exact and approximation methods are available for calculating p-values.

Details

Package: choplump
Type: Package
Version: 1.1.1
Date: 2022-05-09
License: GPL

See example below. There is also two vignettes. The vignette computation (see `vignette("choplumpComputation")`) gives computational details, and the vignette validation (see `vignette("choplumpValidation")`) details the way we have validated the function.

Author(s)

Michael P. Fay

Maintainer: Michael P. Fay <mfay@niaid.nih.gov>

References

Follmann, DA, Fay, MP, and Proschan, MA. (2009) "Chop-lump tests for Vaccine trials" *Biometrics* 65: 885-893. (see `/doc/choplump.pdf`)

Examples

```
set.seed(13921)
Ntotal<-200
Mtotal<-54
Z<-rep(0,Ntotal)
Z[sample(1:Ntotal,Ntotal/2,replace=FALSE)]<-1
test<-data.frame(W=c(rep(0,Ntotal-Mtotal),abs(rnorm(Mtotal))),Z=Z)
## defaults to asymptotic approximation if the number
## of calculations of the test statistic
## is >methodRuleParms=10^4
choplump(W~Z,data=test,use.ranks=TRUE,exact=FALSE)
```

| | |
|--------------|---|
| chooseMatrix | <i>Create an (n choose m) by n matrix with unique rows.</i> |
|--------------|---|

Description

Create a choose(n,m) by n matrix. The matrix has unique rows with m ones in each row and the rest zeros.

Usage

```
chooseMatrix(n, m)
```

Arguments

| | |
|---|---------------|
| n | an integer |
| m | an integer<=n |

Value

A matrix with choose(n,m) rows n columns. The matrix has unique rows with m ones in each row and the rest zeros.

Note

Used for exact test method for [choplump](#)

Author(s)

M.P.Fay

Examples

```
chooseMatrix(5,2)
```

| | |
|----------|----------------------|
| choplump | <i>Choplump Test</i> |
|----------|----------------------|

Description

The choplump test is a two-sample permutation test, that is used when there are many responses that are zero with some positive.

Usage

```

choplump(x, ...)

## Default S3 method:
choplump(x, y, alternative = c("two.sided", "less", "greater"),
         use.ranks=TRUE, exact = NULL, method=NULL,
         methodRule=methodRule1, methodRuleParms=c(10^4),
         nMC=10^4-1, seed=1234321, printNumCalcs=TRUE, ...)

## S3 method for class 'formula'
choplump(formula, data, subset, na.action, ...)

```

Arguments

| | |
|-----------------|--|
| x | a numeric vector of responses in first group, or a formula. Should have some zeros and the rest positive. |
| y | numeric vector of responses in second group |
| alternative | a character string specifying the alternative hypothesis, must be one of "two.sided" (default), "greater" or "less". |
| use.ranks | a logical indicating whether to use ranks for the responses |
| exact | a logical indicating whether an exact p-value should be computed (see details) |
| method | a character value, one of 'approx', 'exact', 'exactMC'. If NULL method chosen by methodRule |
| methodRule | a function used to choose the method (see details). Ignored if method is not NULL |
| methodRuleParms | a vector of parameters passed to methodRule. Ignored if method is not NULL |
| nMC | number of Monte Carlo replications, used if method='exactMC', ignored otherwise |
| seed | value used in set.seed if method='exactMC', ignored otherwise |
| printNumCalcs | logical, print number of calculations of test statistic for exact tests |
| formula | a formula of the form lhs~rhs where lhs is a numeric variable giving the data values and rhs a factor with two levels giving the corresponding groups. |
| data | an optional matrix or data frame containing the variables in the formula |
| subset | an optional vector specifying a subset of observations to be used. |
| na.action | a function which indicates what should happen when the data contain NAs. Defaults to getOption("na.action"). |
| ... | further arguments to be passed to or from methods. |

Details

Consider a randomized trial where one wants to compare the responses in two groups, but there are many zeros in both groups. For example, in an HIV vaccine trial the response could be level of virus in the blood and very many in both groups will have zero values for the response. In order to gain

power, the choplump test removes the same proportion of zeros from both groups, and compares the standardized means between the values left. The test can use ranks to obtain a Wilcoxon-like test. The choplump is a formal permutation test (in other words for each permutation, the chopping is redone) so the type I error is less than the nominal significance level either exactly (for exact methods) or approximately (for the approximate method).

There are a choice of 3 different methods to calculate the p-values: `approx`, an approximation method, see `vignette("choplumpComputation")`; `exact`, an exact method, see `vignette("choplumpComputation")`; `exactMC`, exact method using Monte Carlo resampling with `nMC` resamples.

The associated functions for the above methods (`choplumpApprox`, `choplumpExact`, `choplumpExactMC`), are internal and not to be called directly.

A `methodRule` function has 4 input values: `W` (a vector of all responses), `Z` (a vector of 0 or 1 denoting group membership), `exact` (a logical value, same as `exact` in the `choplump` call), and `parms` (the vector of parameters, same as `methodRuleParms` in the `choplump` call). The `methodRule` function returns a character vector with one of the allowed methods. The default method rule is `methodRule1`. It gives a result of `'approx'` if either `exact=FALSE` or `exact=NULL` and there are more than `parms` calculations of the test statistic. It gives a result of `'exact'` if there are less than `methodRuleParms` calculations of the test statistic, and it gives a result of `'exactMC'` if `exact=TRUE` and there are more than `methodRuleParms` calculations of the test statistic.

Value

A `htest` object, a list with elements

| | |
|--------------------------|--|
| <code>p.value</code> | p value associated with alternative |
| <code>alternative</code> | description of alternative hypothesis |
| <code>p.values</code> | a vector giving lower, upper, and two-sided p-values |
| <code>METHOD</code> | a character vector describing the test |
| <code>data.name</code> | a character vector describing the two groups |

Author(s)

M.P. Fay

References

Follmann, DA, Fay, MP, and Proschan, MA. (2009) "Chop-lump tests for Vaccine trials" *Biometrics* 65: 885-893. (see `/doc/choplump.pdf`)

Examples

```
set.seed(1)
Ntotal<-200
Mtotal<-12
Z<-rep(0,Ntotal)
Z[sample(1:Ntotal,Ntotal/2,replace=FALSE)]<-1
test<-data.frame(W=c(rep(0,Ntotal-Mtotal),abs(rnorm(Mtotal))),Z=Z)
## defaults to asymptotic approximation if
## the number of calculations of the test
```

```
## statistic is greater than parms
## see help for methodRule1
choplump(W~Z,data=test,use.ranks=TRUE)
## alternate form
cout<-choplump(test$W[test$Z==0],test$W[test$Z==1],use.ranks=TRUE,exact=TRUE)
cout
cout$p.values
```

choplumpGeneral *General choplump test*

Description

This function does a general choplump test. For simple difference in standardized means (on the responses or on the ranks), use the much faster [choplump](#) function.

Usage

```
choplumpGeneral(W, Z, testfunc=testfunc.wilcox.ties.general)
```

Arguments

| | |
|----------|---|
| W | numeric vector of responses, some should be zero |
| Z | numeric vector of group membership, values either 0 or 1 |
| testfunc | test function, inputs a data frame with two columns labeled W and Z, outputs test statistic |

Value

Returns a p-value vector of length 3, with 3 named values: p.lower, p.upper, p.2sided.

Examples

```
### compare speed and results using two different functions
W<-c(0,0,0,0,0,0,0,0,2,4,6)
Z<-c(0,0,0,0,1,1,1,1,0,1,1)
Testfunc<-function(d){
  W<-d$W
  Z<-d$Z
  N<-length(Z)
  sqrt(N-1)*(sum(W*(1-Z)) - N*mean(W)*mean(1-Z) )/
  sqrt(var(W)*var(1-Z))
}
time0<-proc.time()
choplumpGeneral(W,Z,Testfunc)
time1<-proc.time()
choplump(W~Z,use.ranks=FALSE)$p.values
time2<-proc.time()
time1-time0
time2-time1
```

| | |
|-------------|--|
| methodRule1 | <i>Rule for determining method for choplump function</i> |
|-------------|--|

Description

This is the default function which determines which method to use in [choplump](#).

Usage

```
methodRule1(W,Z, exact, parms)
```

Arguments

| | |
|-------|---|
| W | numeric vector of response scores, usually many zeros and the rest positive |
| Z | group membership vector, values all 0 (control) or 1 (treated) |
| exact | logical, TRUE=exact method, FALSE=approximate method, NULL=see below |
| parms | numeric value of maximum number of calculations of test statistic, if number of calculations greater than parms then use Monte Carlo for exact method |

Details

This function determines which of several methods will be used in [choplump](#); see that help for description of methods.

When exact=FALSE then returns 'approx'. When exact=TRUE then returns either 'exact' if the number of calculations of the test statistic is less than or equal to parms or 'exactMC' otherwise. When exact=NULL then returns either 'exact' if the number of calculations of the test statistic is less than or equal to parms or 'approx' otherwise.

Value

a character vector with one of the following values: "approx","exact","exactMC"

See Also

[choplump](#)

`wilcox.manyzeros.exact`*Wilcoxon Rank Sum Test*

Description

This function gives exact p-values for the Wilcoxon rank sum. This algorithm is designed for the case when the responses are either positive or zero, and there are many zero responses. Its purpose is mostly for the validation of the [choplump](#) function (see `vignette("choplumpValidation")`).

Usage

```
wilcox.manyzeros.exact(W, Z)
```

Arguments

| | |
|---|---|
| W | a vector of responses, should have some zeros and all rest positive |
| Z | a vector of group membership, should be either 0 or 1 |

Value

A vector of three types of p-values: `p.lower`, `p.upper`, and `p.2sided`.

Author(s)

M.P. Fay

See Also

[choplump](#)

Index

- * **array**
 - chooseMatrix, 3
- * **htest**
 - choplump, 3
 - choplumpGeneral, 6
 - wilcox.manyzeros.exact, 8
- * **misc**
 - methodRule1, 7
- * **package**
 - choplump-package, 2

chooseMatrix, 3
choplump, 2, 3, 3, 6–8
choplump-package, 2
choplumpGeneral, 6

methodRule1, 5, 7

wilcox.manyzeros.exact, 8