

Package ‘tdthap’

November 29, 2023

Version 1.3

Date 2023-11-29

Title TDT Tests for Extended Haplotypes

Author David Clayton

Maintainer Jing Hua Zhao <jinghuazhao@hotmail.com>

URL <https://github.com/jinghuazhao/R>

BugReports <https://github.com/jinghuazhao/R/issues>

Depends R (>= 1.5.0)

Description Functions and examples are provided for Transmission/disequilibrium tests for extended marker haplotypes, as in Clayton, D. and Jones, H. (1999) “Transmission/disequilibrium tests for extended marker haplotypes”. Amer. J. Hum. Genet., 65:1161-1169, <doi:10.1086/302566>.

License Artistic-2.0

NeedsCompilation yes

Repository CRAN

Date/Publication 2023-11-29 22:40:02 UTC

R topics documented:

get.similarity	2
hap.transmit	2
set.similarity	3
tdt.quad	4
tdt.rr	5
tdt.select	7

Index	8
--------------	----------

get.similarity	<i>Interogate similarity function for Geary-Moran TDT tests</i>
----------------	---

Description

N/A

Usage

```
get.similarity(nloci=1)
```

Arguments

nloci	The number of loci.
-------	---------------------

See Also

[tdt.quad](#), [set.similarity](#)

hap.transmit	<i>Build parental haplotypes in nuclear families</i>
--------------	--

Description

This version only computes parental haplotypes in so far as they can be derived with complete certainty. Any locus with is uncertain in the final haplotype is coded as zero.

Usage

```
hap.transmit(pedfile, markers=1:((ncol(pedfile) - 6)/2),
            multiple.cases=0, use.affected=TRUE)
```

Arguments

pedfile	The input dataframe. The first six columns contain the pedigree id, the member id, the two parental id's, the sex, and the affectation status. Subsequent fields are in pairs and represent alleles at marker loci. All variables must take integer values, with zero being taken as "missing".
markers	Integer array indicating markers to be used and their order.
multiple.cases	The action to be taken if multiple affected offspring in any pedigree. Options are (0) include all, (1) include all, but whole family is duplicated and only one offspring is treated as affected in each repeated family, and (2) use only first affected offspring.
use.affected	If TRUE, data from affected offspring is used when imputing any missing parental data. Otherwise it is ignored.

Value

A dataframe with one row for each affected offspring. The first four columns identify the offspring by pedigree id, member id, and parental id's. The next block of columns hold the transmitted paternal haplotype. Following blocks contain the untransmitted paternal haplotype and maternal transmitted and untransmitted haplotypes.

References

Clayton, D. and Jones, H. (1999) Transmission/disequilibrium tests for extended marker haplotypes. *Am. J. Hum. Gen.*, 65:1161-1169.

See Also

[tdt.select](#), [tdt.rr](#), [tdt.quad](#)

Examples

```
## Not run:
# Read a pedfile (which includes the variable names in the top line)
# and build haplotypes using the markers which appear third, second, and
# first in the pedfile.

filespec <- system.file("tests/test.ped", package="tdthap")
ped <- read.table(filespec)

haps <- hap.transmit(ped, markers=c(3,2,1))

## End(Not run)
```

set.similarity

Set similarity function for Geary-Moran type TDT tests

Description

Haplotypes are only similar if they are IBS at the focal locus. The extent of the similar region to each side is determined by stepping outwards until the haplotypes are no longer IBS, the region being assumed to end midway between the last IBS locus and the first non-IBS locus. If the haplotypes are IBS at the last locus, half the "off-end" distance is scored. The similarity is defined as the total length of this shared region raised to some power.

Usage

```
set.similarity(nloci=1, spacing=rep(1, nloci + 1), focus=1, power=1)
```

Arguments

nloci	The number of loci.
spacing	A numeric array of length (nloci+1) giving marker spacings and "off-end" distances.
focus	An integer in the range 1:nloci indicating the "focus" for the similarity function.
power	The power to which the shared haplotype length is raised.

Value

A list of the values loaded.

Side Effects

Sets constants accessed by tdt.quad() when calculating Geary-Moran type statistics.

References

Clayton, D. and Jones, H. (1999) Transmission/disequilibrium tests for extended marker haplotypes. *Am.J.Hum.Gen.*, 65:1161-1169.

See Also

[tdt.quad](#), [get.similarity](#)

Examples

```
## Not run:
# To do a Geary_Moran test on a 10 marker haplotype
gaps <- c(0, 50, 60, 80, 20, 30, 50, 40, 50, 100, 0)
set.similarity(nloci=10, spacing=gaps, power=0.5)
test <- tdt.quad(hap.use, funct=T)

## End(Not run)
```

tdt.quad

TDT tests for extended haplotypes

Description

The function calculates the test statistic and then simulates its distribution under the null hypothesis by randomly transmitting parental haplotypes with probability 0.5. The test statistic is recalculated for each simulated dataset. For Geary-Moran tests in particular this can be quite slow.

Usage

```
tdt.quad(hap, nsim=5000, funct=FALSE, keep=TRUE, seeds=c(0, 0, 0))
```

Arguments

hap	A list containing the transmitted and untransmitted haplotypes. This would normally be computed using <code>tdt.select</code> .
nsim	The number of Monte Carlo simulations from the null hypothesis.
funct	If T, a similarity function is used and the test is a Geary-Moran test. Otherwise, the Pearsonian test, $\text{Sum } (O - E)^2 / E$, is used.
keep	If TRUE, all simulated values of the test statistic are kept. Otherwise only the realised value of the test statistic and the p-value are returned.
seeds	Three numbers to seed the random number generator. The default is to use a different three random numbers each time.

Value

A list containing, the number of distinct haplotypes (*n.hap*), the number of informative transmissions (*n.trans*), the test statistic (*test*), the p-value (*p.value*) and, optionally, all the simulated values of the test statistic under the null hypothesis (*sim*).

References

Clayton, D. and Jones, H. (1999) Transmission/disequilibrium tests for extended marker haplotypes. *Am.J.Hum.Gen.*, 65:1161-1169.

See Also

[hap.transmit](#), [tdt.select](#), [tdt.rr](#), [set.similarity](#), [get.similarity](#)

Examples

```
## Not run:
# Do a Pearsonian test using 10000 simulations and summarise the distribution
# of the statistic under the null hypothesis

test <- tdt.quad(hap.use, nsim=10000, keep=T)
test
summary(test$sim)

## End(Not run)
```

Description

The p-value is the conventional "exact" test based on the binomial distribution of transmissions. The estimated relative risks use a Bayesian method, recommended because of the multiplicity problem. the prior is a beta distribution of the second kind, defined by two "degrees of freedom" parameters. Note that the prior mean is $\text{prior.df}[1]/\text{prior.df}[2]$ and that Bayes estimates based on small numbers of transmissions are pulled in towards this. A "realistic" choice of these parameters is recommended, and to aid this, the function returns credible intervals using the prior alone as well as the a posteriori interval for each haplotype.

Usage

```
tdt.rr(hap, prior.df=c(0.5, 0.5), prob=c(0.05, 0.95))
```

Arguments

hap	A list containing the transmitted and untransmitted haplotypes. This would normally be computed using tdt.select .
prior.df	a vector of length two containing the degree of freedom parameters for the prior distribution of the haplotype relative risk - a beta distribution of the second kind.
prob	The probability levels for Bayesian credibility intervals for the haplotype relative risks.

Value

A matrix containing the numbers of transmitted and untransmitted haplotypes, the (binomial) p-values, the Bayes estimates of the haplotype relative risks, and the lower and upper bounds of the credible interval. The prior estimate and credible interval is also shown.

References

Spielman R., McGinnis R., and Ewens, W. (1993) Transmission tests for linkage disequilibrium. *American Journal of Human Genetics*, 52, 506-16.

See Also

[hap.transmit](#), [tdt.select](#), [tdt.quad](#)

Examples

```
## Not run:
# Select the sub-haplotype made up from the first two markers and
# print tables of TDT tests and haplotype relative risks

hap.use <- tdt.select(haps, markers=1:2)
rr <- tdt.rr(hap.use)
rr

## End(Not run)
```

tdt.select	<i>Select informative transmissions of sub-haplotypes for the TDT test</i>
------------	--

Description

This function is just a data handling intermediary between [hap.transmit](#), which computes haplotypes, and [tdt.quad](#) and [tdt.rr](#) which do TDT tests.

Usage

```
tdt.select(hap.data, markers=1:((ncol(hap.data) - 4)/4), complete=TRUE)
```

Arguments

hap.data	The input dataframe. This will usually have been created by hap.transmit .
markers	An integer array indicating which loci make up the relevant part of the haplotype.
complete	If TRUE, only "complete" haplotypes are used (ie no zero's will be included).

Value

A list of two arrays of class "factor". The first (trans) contains transmitted haplotypes and the second (untrans) contains untransmitted haplotypes. Rownames identify the transmission in terms of pedigree id, offspring id, father's id, mother's id, and whether it is a paternal transmission ("f") or a maternal transmission ("m").

References

Clayton, D. and Jones, H. (1999) Transmission/disequilibrium tests for extended marker haplotypes. *Am.J.Hum.Gen.*, 65:1161-1169.

See Also

[hap.transmit](#), [tdt.rr](#), [tdt.quad](#)

Examples

```
## Not run:  
# Select the sub-haplotype made up from the first two markers and print  
# tables of frequencies of transmitted and untransmitted haplotypes  
  
hap.use <- tdt.select(haps, markers=1:2)  
table(hap.use$trans)  
table(hap.use$untrans)  
  
## End(Not run)
```

Index

* **htest**

get.similarity, 2

hap.transmit, 2

set.similarity, 3

tdt.quad, 4

tdt.rr, 5

tdt.select, 7

get.similarity, 2, 4, 5

hap.transmit, 2, 5–7

set.similarity, 2, 3, 5

tdt.quad, 2–4, 4, 6, 7

tdt.rr, 3, 5, 5, 7

tdt.select, 3, 5, 6, 7