

Package ‘pedgene’

May 23, 2019

Type Package

Title Gene-Level Statistics for Pedigree Data

Version 3.2

Date 2019-5-22

Description Gene-level association tests with disease status for pedigree data: kernel and burden association statistics (2013, <doi: 10.1002/gepi.21727>).

License GPL (>= 2)

LazyLoad yes

Depends R (>= 3.2.0), stats, Matrix, CompQuadForm, survey, kinship2 (>= 1.6.0)

NeedsCompilation no

Maintainer Jason Sinnwell <sinnwell.jason@mayo.edu>

URL <http://bioinformaticstools.mayo.edu/research/pedgene>

Author Schaid Daniel [aut],
Visconti Alessia [ctb],
Jason Sinnwell [aut, cre]

Repository CRAN

Date/Publication 2019-05-23 05:00:23 UTC

R topics documented:

example.geno	2
pedgene	3
pedigreeChecks	6
quadfactor	7

Index	9
--------------	----------

 example.geno

Example datasets for pedgene

Description

example.geno: a data frame with minor allele count for subjects (rows) at variant positions (columns);
 example.ped: pedigree and trait data for subjects in example.geno; example.map: gene and chromosome for variant positions in example.geno; example.relation: special (twin) relationships for individuals in example.ped

Usage

```
data(example.geno)
data(example.ped)
data(example.map)
data(example.relation)
```

Format

example.geno -data frame with minor allele count for 20 variant positions:

- famid pedigree family ID, character or numeric
- person person ID, used with ped to match subjects to their row in example.ped
- AA.1-AA.10, AX.1-AX.10 genotype columns at 10 positions for each of 2 simulated genes

example.ped -data frame with pedigree structure and trait values in the following columns:

- famidpedigree family ID, character or numeric
- person person ID, a unique ID within each pedigree
- father father ID, 0 if no father
- mother mother ID, 0 if no mother
- sex coded as 1 for male, 2 for female
- trait phenotype, either case-control status coded as 1 for affected and 0 for unaffected, or a continuous value. Subjects with missing (NA) will be removed from the analysis
- trait.adjusted an optional variable for covariate-adjusted trait. If trait.adjusted is present in the data.frame, then gene-level tests are adjusted for covariates using residuals = (trait - trait.adjusted). Otherwise, gene-level tests are not adjusted for covariates, in which case residuals = trait - mean(trait)

example.map - data frame with columns for gene name and chromosome:

- chrom chromosome code (1-23,X allowed) where the gene is located
- gene gene identifier

example.relation - data.frame specifying special relationships between pairs of individuals in the following columns:

- famidpedigree family ID, character or numeric
- id1 person1 ID, used with ped to match subjects to their row in example.ped
- id2 person2 ID, used with ped to match subjects to their row in example.ped
- code any of the following: 1=Monozygotic twin, 2=Dizygotic twin, 3=Twin of unknown zygosity, and 4=Spouse

Source

Simulated data for testing purposes

Examples

```
data(example.geno)
data(example.ped)
data(example.map)
data(example.relation)
```

pedgene	<i>Compute Kernel and Burden Statistics for Pedigree Data (possibly with unrelated subjects)</i>
---------	--

Description

Compute linear kernel and burden statistics for gene-level analysis of data that includes pedigree-related subjects, and possibly unrelated subjects.

Usage

```
pedgene(ped, geno, map=NULL, male.dose=2,
        checkpeds=TRUE, verbose.return=FALSE,
        weights=NULL, weights.beta=c(1,25),
        weights.mb=FALSE, relation=NULL,
        method="kounen", acc.davies=1e-5)
```

Arguments

ped A data.frame with variables that define the pedigree structure (typical format used by LINKAGE and PLINK), trait (phenotype), and optionally a covariate-adjusted trait (for covariate-adjusted gene level statistics). The columns in the data.frame must be named as follows:

- ped: pedigree ID, character or numeric allowed
- person: person ID, a unique ID within each pedigree, numeric or character allowed
- father: father ID, 0 if no father
- mother: mother ID, 0 if no mother
- sex: coded as 1 for male, 2 for female

- trait: phenotype, either case-control status coded as 1 for affected and 0 for unaffected, or a continuous value. Subjects with missing (NA) will be removed from the analysis
- trait.adjusted: an optional variable for covariate-adjusted trait. If trait.adjusted is present in the data.frame, then gene-level tests are adjusted for covariates using residuals = (trait - trait.adjusted). Otherwise, gene-level tests are not adjusted for covariates, in which case residuals = trait - mean(trait) where the mean is taken on all subjects passed into pedgene before removing subjects who do not have genotype data.

geno	Data.frame or matrix with genotypes for subjects (rows) at each variant position (columns). The first two columns are required to be named ped and person, which are used to match subjects to their data in the ped data.frame. The genotypes are coded as 0, 1, 2 for autosomal markers (typically a count of the number of the less-frequent allele). For X-chromosome markers, females are coded 0, 1, 2, and males coded 0, 1. Missing genotypes (NA) are allowed.
map	Optional data.frame with columns "chrom" and "gene", one row per variant column in geno. The gene name can be any identifier for the gene. The chromosome can be either numeric or character, where the calculations will differ between autosomes vs X chromosome (allow "X"/"x"/23, converted to "X" in results)
male.dose	When analyzing the X-chromosome, male.dose defines how male genotypes should be analyzed. male.dose can be between 0 and 2, but is typically either 1 or 2. Ozbek and Clayton show that male.dose = 2 is powerful in the presence of X-chromosome dosage compensation in females.
checkped	logical, if FALSE, the method will skip the pedigree checking step, which can be intensive for large studies
verbose.return	logical, if TRUE, return the pedigree, geno, and map objects used in the tests after initial cleanup, e.g., the removal of monomorphic variants. They are returned in the pedgene object in a list called "save"
weights.beta	Weights based on a function of the minor allele frequency (maf) and the Beta distribution
weights	optional user-specified weights, a vector of weights for each variant column of geno. If none given, the Beta weights are applied
weights.mb	logical, if TRUE and no user-given weights, apply the Madsen-Browning weights per variant: $1/\sqrt{\text{maf}*(1-\text{maf})}$. The hierarchy of weights used is 1) user-specified weights, 2) Madsen-Browning if weights.mb=TRUE and weights=NULL, 3) Beta density weights (default if no other weights are set)
relation	Optional data.frame/matrix with 4 columns (id, person1, person2, code) specifying special relationships between pairs of individuals and used by the kinship function. Codes are : 1=Monozygotic twin, 2=Dizygotic twin, 3=Twin of unknown zygosity, and 4=Spouse. The last is necessary in order to place a marriage with no children into the plot. See kinship2::pedigree for details.
method	method for calculating the kernel test p-value. Kounen's saddlepoint approximation (default) is based on the survey package, and has been found to have less faults (e.g., returned missing value) than Davies' method (see Chen et al.,

2012). The Davies method, which computes an exact p-value for a mixture of chi-square distributions, is also provided. The accuracy of the Davies method depends on the numerical accuracy parameter (`acc.davies`), which can be difficult to specify ahead of time.

`acc.davies` Numerical accuracy parameter used in the Davies' method for calculating the kernel test p-value. In some instances, a p-value from the kernel test is out of range, in which case the p-value is set to 0 or 1, depending on which direction the p-value was out of range.

Details

The `pedgene` function is a wrapper function to call `pedgene.stats` on one gene at a time. The `pedgene.stats` function calculates gene-level tests for associations with a trait among subjects, accounting for relationships among subjects based on known pedigree relationships (see Schaid et al). This is achieved by the `kinship` function in the `kinship2` package. The kernel association statistic uses a weighted linear kernel, with default weights based on the beta distribution and the sample minor allele frequency. The burden statistic is based on a weighted sum of variants. If a gene only has one variant, the kernel test reduces to the burden statistic. Variant positions that have zero variance are removed from the analysis because they do not contribute information.

Note that if `ped` contains extra people that are not necessary to define relationships of people with genotype data, their trait value will still be used in `mean(trait)` in calculating `trait.adjusted` if `trait.adjusted` is not given as a column in `ped`.

Value

An object of the `pedgene` S3 class, with the following elements:

`call`: function call

`pgdf`: data.frame with gene name, chromosome, n-variants per gene(after removing unnecessary variants), n-variants removed per gene, kernel and burden test statistics and p-values. Kernel p-values are based on either Kuonen (1999) or Davies (1980) method. The burden statistic has a standard normal distribution, so the sign of the burden statistic gives information on the direction of association (positive value implies large burden score are positively associated with larger trait values). When a gene has only 1 variant, the kernel test reduces to the burden test. In this instance, the kernel statistic (chi-square) is the square of the burden statistic (standard normal), with both having the same p-value. When a gene has no markers after removing zero-variance markers, the gene test statistics and p-values are all NA.

`save`: If `verbose.return` was set to TRUE, a list containing the cleaned form of the data that was input to `pedgene` and is used in the tests: `ped`, `geno`, and `map`

Author(s)

Daniel J. Schaid, Jason P. Sinnwell, Mayo Clinic (contact: schaid@mayo.edu). Alessia Visconti, King's College London (alessia.visconti@kcl.ac.uk).

References

- Schaid DJ, McDonnell SK, Sinnwell JP, Thibodeau SN (2013) Multiple Genetic Variant Association Testing by Collapsing and Kernel Methods With Pedigree or Population Structured Data. *Genetic Epidemiology*, 37(5):409-18.
- Ozbek U (2012) Statistics for X-chromosome association. 62nd Annual Meeting of The American Society of Human Genetics; Program #22. San Francisco, California.
- Clayton D (2008) Testing for association on the X chromosome. *Biostatistics* 9:593-600
- Chen H, Meigs J, Dupuis J (2012) Sequence kernel association test for quantitative traits in family samples. *Genetic Epidem* 37:196-204
- Kounen D (1999) Saddle point approximatinos for distributions of quadratic forms in normal variables. *Biometrika* 86:929 -935
- Davies RB (1980) Algorithm AS 155: The Distribution of a Linear Combination of chi-2 Random Variables, *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 29(3):323-33

See Also

pedigreeChecks, example.ped

Examples

```
# example data with the same 10 variants for an autosome and X chromosome
# pedigree data on 39 subjects including 3 families and unrelateds
data(example.ped)
data(example.geno)
data(example.map)

# gene tests (chroms 1 and X) with male.dose=2
pg.m2 <- pedgene(example.ped, example.geno, example.map, male.dose=2)
# same genes, with male.dose=1
pg.m1 <- pedgene(example.ped, example.geno, example.map, male.dose=1)

## print and summary methods
print(pg.m2, digits=3)
summary(pg.m1, digits=3)
```

pedigreeChecks

Validity checks on pedigree data

Description

Checks for valid IDs, sex codes for data in a single pedigree

Usage

```
pedigreeChecks(pedigree, male.code = 1, female.code = 2)
```

Arguments

pedigree	data frame with variables named person, father, mother, sex.
male.code	sex code for males
female.code	sex code for females

Details

A series of basic pedigree checks

Value

valid = TRUE or FALSE for validity of pedigree data

Author(s)

Daniel J. Schaid (schaid@mayo.edu).

quadfactor	<i>Internal utility to calculate a constant quadratic factor for gene-level statistics variances, over all pedigrees</i>
------------	--

Description

Internal utility to calculate a constant quadratic factor for gene-level statistics variances over all pedigrees, for either autosomes or X chromosome

Usage

```
quadfactor(kinmat, chrom, resid, sex, male.dose)
```

Arguments

kinmat	A kinship matrix for one or more pedigrees
chrom	character string for chromosome number, if "X", the method accounts for sex code
resid	the residual based on the trait minus the group mean or adjusted trait value for each subject
sex	See pedgene
male.dose	See pedgene

Value

Constant quadratic factor for gene-level statistics variances, for autosomes and X chromosome

Author(s)

Daniel J. Schaid, Mayo Clinic (schaid@mayo.edu).

See Also

pedgene, kinship

Index

*Topic **datasets**

example.geno, [2](#)

*Topic **kinship**

pedgene, [3](#)

example.geno, [2](#)

example.map (example.geno), [2](#)

example.ped (example.geno), [2](#)

example.relation (example.geno), [2](#)

pedgene, [3](#)

pedigreeChecks, [6](#)

print.pedgene (pedgene), [3](#)

quadfactor, [7](#)

summary.pedgene (pedgene), [3](#)