Package 'RVFamSq'

March 6th, 2019

Type Package

Title RVFamSq: Rare Variant-Family-Based Score Test for Quantitative Traits

Version 1.0

Data 2019-3-6

Author Zhihui Zhang and Suzanne Leal

Contact Zhihui Zhang zhihuiz@bcm.edu

Maintainer Zhihui Zhang zhihuiz@bcm.edu

Depends R (>=3.0.0), bbmle (>=1.0.20), mvtnorm (>=1.0.10), rlist (>=0.4.6.1)

Imports bbmle, mvtnorm, rlist, kinship2

Description A package for family-based association analysis of rare variants and quantitative traits

License GPL-3

URL

LazyLoad yes

R topics documented:

RVFamSq-package *RVFamSq: Rare Variant-Family-Based Score Test for Quantitative Traits*

Description

The RVFamSq package provides an efficient approach to examine the association between the region-based rare variants and the quantitative traits in family-based data. The RVFamsSq can be broadly applied to diverse pedigree structures and families with members missing sequence data. In addition, qualitative and quantitative covariates, e.g., age, sex, and body mass index, can be flexibly included in the package. The RVFamSq extends the method proposed by Chen et al. [], and the speed of the package is significantly optimized to analyze large-scale data sets.

Details

Package: RVFamSq
Type: Package
Version: 1.0
Data: 2019-3-6
License: GPLv3

Author(s)

Zhihui Zhang zhihuiz@bcm.edu
Suzanne Leal

References

Chen, W.-M., and Abecasis, G.R. (2007). Family-Based Association Tests for Genomewide Association Scans. Am. J. Hum. Genet. *81*, 913–926.

Example.data *A small family data set for association analysis*

Description

test.ped A pedigree file containing 100 nuclear families and 100 SNPs for each individual. The first six columns of the file are family ID, individual ID, father ID, mother ID, covariant, trait, respectively. The rest columns of the file are genotypes of 400 individuals. The genotypes are coded as 0, 1, and 2.

test.sfs A file with descriptive information on 100 genetic variants of genotypes in test.ped. The file contains four columns representing the name of the interested gene, chromosome, position, and MAF of the variant, respectively.

test.fam A family file representing the relationship of all individuals in the pedigree. Five columns in the file indicate family ID, individual ID, father ID, mother ID, sex, respectively. Each family should have information of two founders, and all individuals in test.ped should be included in this file.

Usage

Download the files to you working directory. The path to these files is required to use the package.

RVFamSq *Rare Variant-Family-Based Score Test for Quantitative Traits*

Description

A regional association analysis of rare variants and quantitative train in family data

Usage

RV_FamSq (pedfile, maffile, parafile,out,kin, start_par=NULL)

Arguments

pedfile referring to the path of the .ped file. The columns in the file represent family ID, individual ID, father ID, mother ID, covariant, traits, and genotypes of each individual. The file does not have head.

maffile the path to the .sfs file describing the information of genotype variants in .ped file.

parafile the path to the file saving parameters estimated in null model. If the file is not existed, the package will estimate the parameters using the available data and save the values of the parameters in parafile.

out the directory that save the results. RVFamSq produces output file with the name of the gene, score, and p-value of the gene.

kin a kinship matrix calculated from the pedigree using the "kinship2" package.

start-par an optional argument that defines the starting values of the parameters fitting. User can define specific values to start the parameters fitting, otherwise, the starting values will generate randomly.

Details

RVFamSq estimates the parameter under the null model by Association can be tested by maximizing the multivariate normal likelihood:

$$L = \prod_i (2\pi)^{-n_i/2} |\Omega_i|^{-1/2} e^{[y_i - E(y_i)]' \Omega_i^{-1} [y_i - E(y_i)]}$$

where n_i is the number of individuals in family *i* and $|\Omega_i|$ is the determinant of matrix Ω_i. $E(Y_{ij})$ in the above function is a vector of expected phenotype of all individuals in each family and is calculated by:

$$E(Y_{ij}) = \mu + \beta_x x_{ij}$$

where x_{ij} is the covariant and the parameters μ and β_x are estimated by maximizing the above likelihood. The matrix Ω_i is calculated within each family i and defined as:

$$\Omega_{ijk} \begin{cases} \sigma_g^2 + \sigma_e^2 & if\ j = k \\ 2\varphi_{ijk}\sigma_g^2 & if\ j \neq k \end{cases}$$

Here, the parameter $\sigma_g^2, and\ \sigma_e^2$ are variance components that are account for background polygenic effects and environmental effects, respectively. Additionally, φ_{ijk} denote the kinship coefficient between individuals j and k. The values of parameters $\sigma_g^2, and\ \sigma_e^2$ are also estimated by maximizing the above likelihood. With the values of these four parameters (β_x, μ, $\sigma_g^2, and\ \sigma_e^2$), we can obtain the variance-covariance matrix $\Omega_i^{(base)}$ based on and vector $E(y_i)^{(base)}$ based on the above equations for each family. Using these two quantities, the statistic score is defined by:

$$T^{SCORE} = \frac{\{\sum_i [G_i - E(G_i)]' [\Omega_i^{(base)}]^{-1} [y_i - E(y_i)^{(base)}]\}^2}{\sum_i [G_i - E(G_i)]' [\Omega_i^{(base)}]^{-1} [G_i - E(G_i)^{(base)}]}$$

In the above equation, $E(G_i)$ is defined as:

$$E(G_i) = \sum_{m=1}^{M} 2p_m$$

where p_m is the minor allele frequency (MAF) which is obtain from the .sfs file.

Value

results a data frame results containing the name of the gene, the value of the score and p-value write into the file named by the name of the gene under the directory defined by the argument "out".

paras a list of parameters estimated by maximizing the likelihood. The values of the paras save to the file defined by the argument "parafile". If the "parafile" is existed, the values of the paras load from the file "parafile".

Author(s)

Zhihui Zhang zhihuiz@bcm.edu
Suzanne Leal

References

Chen, W.-M., and Abecasis, G.R. (2007). Family-Based Association Tests for Genomewide Association Scans. Am. J. Hum. Genet. *81*, 913–926.

Examples

```
library (kinship2)
library (RVFamSq)

##  Define variables pointing to the path to the example data
maffile<-"/directory-to-example-data/test.sfs"
pedfile<-"/ directory-to-example-data/test.ped"

##  Define output directory that save the results
out<-"/directory-to-save-the-results"

## Define the files that save the parameters estimated under the null model.
## If the file is not existed, the package will estimate the parameters based on the available data.
parafile<-"/Users/zhihuizhang/Desktop/postdoc/manu/speed_test/paras3.rds"
```

```
## Load phenotype data and calculate the kinship matrix of the pedigree.
ped_pheno<-read.table(phenofile)
## Detached “kinship” package if it is pre-loaded.
if ("kinship" %in% (.packages())){
    print("Detaching kinship library")
   detach("package:kinship", unload=TRUE)
}
kin_pre<-data.frame(id=ped_pheno[,2], mom=ped_pheno[,4], dad=ped_pheno[,3], sex=ped_pheno[,5])
tped <- with(kin_pre, pedigree(id, dad, mom, sex, famid=ped_pheno[,1]))
kin <- kinship(tped)

## Run RVFamsSq package and calculate the statistical score of the interested gene.
RVFamSq(pedfile, maffile, parafile,out,kin)
```