Package 'QMVtest'

March 26, 2022

Type Package

Title Robust association tests for quantitative traits on the X chromosome

Version 1.0

Date 2022-03-26

Author Yu-Xin Yuan, Zi-Ying Yang and Ji-Yuan Zhou

Maintainer Yu-Xin Yuan <smuyuanyuxin@gmail.com>, Zi-Ying Yang <pdlma123@163.com> and Ji-Yuan Zhou <zhoujiyuan5460@hotmail.com>

Description This code contains the location tests QXcat and QZ-max, the scale test wM3VNA3.3 and the location-scale tests QMVXcat and QMVZ-max, which test for the association between a SNP on the X chromosome and a quantitative trait, with or without covariates using unrelated individuals.

LazyData true

License GPL-3

Encoding UTF-8

Imports methods, stats, quantreg, mvtnorm, expm, bigsnpr, xgboost

NeedsCompilation no

RoxygenNote 7.1.2

Depends R (>= 3.6.0)

R topics documented:

Genotype — *a dataset of simulated genotype matrix*

Description

A dataset of simulated genotype matrix with each row as a different individual and each column as a separate SNP. Each genotype is coded as 0, 1 or 2 for females and coded as 0 or 1 for males, indicating the number of reference allele.

Usage

```
Genotype
```

Format

a dataset for 4000 unrelated individuals and 31 SNPs.

location_test — *A function to obtain the p-values and the test statistics of QXcat and QZmax for testing the mean differences of the trait value across genotypes*

Description

A function to obtain the *p*-values and the test statistics of QXcat and QZmax in Yang et al. (2021) for testing the mean differences of the trait value across genotypes. This function takes the genotype of SNPs (`Genotype`), the sex (`Sex`), the quantitative trait (`Y`) in the sample population, and possibly additional covariates, such as age and BMI, as input.

Usage

```
location_test(Genotype,Y,Sex,
              Covariate=NULL,
              missing_cutoff=0.15,
              MAF_Cutoff=NULL,
              MGC_Cutoff=30)
```

Arguments

`Genotype`	A numeric genotype matrix with each row as a different individual and each column as a separate SNP. Each genotype is coded as 0, 1 or 2 for females and coded as 0 or 1 for males, indicating the number of reference allele. The length/dimension of `Genotype` should be the same to those of `Y`, `Sex` and `Covariate`.
`Y`	A numeric vector of a quantitative trait, such as human height.
`Sex`	A vector of the genetic sex following PLINK default coding, where males are coded as 1 and females are coded as 2.
`Covariate`	Optional: a vector or a matrix of covariates, such as age and BMI.
`missing_cutoff`	Cutoff of the missing rates of SNPs (default=0.15). Any SNPs with missing rates higher than the cutoff will be excluded from the analysis.

`MAF_Cutoff` MAF cutoff for common vs. rare variants (default=NULL). It should be a numeric value between 0 and 0.5, or NULL. When it is NULL, 1/ sqrt(2 SampleSize) will be used (Ionita-Laza et al. 2013). Only common variants are included in the analysis.

`MGC_Cutoff` Cutoff for the minimum genotype count in either females or males (default=30), SNPs whose minimum genotype counts are less than this cutoff will not be included in the analysis. This is based on the quality control that SNPs with a minimum count below 30 should be removed to avoid inflated type I errors (Deng et al., 2019; Soave et al., 2015).

Details

QXcat tests for the mean differences of the quantitative trait across different genotypes on the X chromosome by considering various X chromosome inactivation (XCI) patterns. QZmax tests for the mean differences of the quantitative trait across different genotypes on the X chromosome by considering different dosage compensation (DC) patterns.

Value

the *p*-values and the test statistics of the location tests QXcat and QZmax for each SNP.

Author(s)

Yu-Xin Yuan, Zi-Ying Yang and Ji-Yuan Zhou

References

Yang ZY, Liu W, Yuan YX, et al. Robust association tests for quantitative traits on the X chromosome. 2022

Deng WQ, Mao S, Kalnapenkis A, et al. Analytical strategies to include the X-chromosome in variance heterogeneity analyses: evidence for trait-specific polygenic variance structure. *Genetic Epidemiology*, 2019, **43**: 815-830.

Ionita-Laza I, Lee S, Makarov V, et al. Sequence kernel association tests for the combined effect of rare and common variants. *The American Journal of Human Genetics*, 2013, **92**: 841-853.

Soave D, Corvol H, Panjwani N, et al. A joint location-scale test improves power to detect associated SNPs, gene sets, and pathways. *The American Journal of Human Genetics*, 2015, **97**: 125–138.

Examples

```
#Phedata: a dataset for 4000 unrelated individuals and four variables,
          including one phenotype and three covariates (i.e., Sex,age and BMI).
data(Phedata)
#Genotype: a dataset for 4000 unrelated individuals and 31 SNPs.
data(Genotype)

#the location tests (i.e., QXcat and QZmax)
#no covariate
#set "Covariate=NULL"
location_test(Genotype,Phedata$Y,
              Phedata$Sex,
              Covariate=NULL,
              missing_cutoff=0.15,
```

```
              MAF_Cutoff=NULL,
              MGC_Cutoff=30)

#age and BMI as covariates
location_test(Genotype,Phedata$Y,
              Phedata$Sex,
              Covariate=Phedata[,c(-1,-2)],
              missing_cutoff=0.15,
              MAF_Cutoff=NULL,
              MGC_Cutoff=30)
```

`Phedata` *a dataset of simulated phenotype and covariates*

Description

A dataset of simulated phenotype and covariates including sex, age and BMI.

Usage

```
Phedata
```

Format

a dataset for 4000 unrelated individuals and four variables, including one phenotype and three covariates.

Y the values of the quantitative trait of individuals in the simulated dataset.

Sex the genetic sex of individuals, coded as 1 for males and coded as 2 for females, following the PLINK default coding.

age age of individuals in the simulated dataset.

BMI body mass index of individuals in the simulated dataset.

`QMV_test` *A function to obtain the p-values and the test statistics of the location tests, the scale test, the location-scale tests or all*

Description

A function to obtain the *p*-values and the test statistics of the location tests (i.e., QXcat and QZmax), the scale test (i.e.,wM3VNA3.3), the location-scale tests (i.e., QMVXcat and QMVZmax) or all.This function takes the genotype of SNPs (`Genotype`), the sex (`Sex`), the quantitative trait (`Y`) in the sample population, and possibly additional covariates, such as age and BMI, as input.

Usage

```
QMV_test(Genotype,Y,Sex,
         Covariate=NULL,
         missing_cutoff=0.15,
         MAF_Cutoff=NULL,
         MGC_Cutoff=30,
         method='joint')
```

Arguments

`Genotype`	A numeric genotype matrix with each row as a different individual and each column as a separate SNP. Each genotype is coded as 0, 1 or 2 for females and coded as 0 or 1 for males, indicating the number of reference allele. The length/dimension of `Genotype` should be the same to those of `Y`, `Sex` and `Covariate`.
`Y`	A numeric vector of a quantitative trait, such as human height.
`Sex`	A vector of the genetic sex following PLINK default coding, where males are coded as 1 and females are coded as 2.
`Covariate`	Optional: a vector or a matrix of covariates, such as age and BMI.
`missing_cutoff`	Cutoff of the missing rates of SNPs (default=0.15). Any SNPs with missing rates higher than the cutoff will be excluded from the analysis.
`MAF_Cutoff`	MAF cutoff for common vs. rare variants (default=NULL). It should be a numeric value between 0 and 0.5, or NULL. When it is NULL, 1/ sqrt(2 SampleSize) will be used (Ionita-Laza et al. 2013). Only common variants are included in the analysis.
`MGC_Cutoff`	Cutoff for the minimum genotype count in either females or males (default=30), SNPs whose minimum genotype counts are less than this cutoff will not be included in the analysis. This is based on the quality control that SNPs with a minimum count below 30 should be removed to avoid inflated type I errors (Deng et al., 2019; Soave et al., 2015).
`method`	A character string indicating which kind of association tests is to be conducted. There are four options: "location", "scale", "joint" (default) and "all". method="location": QXcat and QZmax; method="scale": wM3VNA3.3; method="joint": QMVXcat and QMVZmax; method="all": all of the above association tests.

Details

QMVXcat and QMVZmax are designed to test for both the mean differences and the variance heterogeneity of the trait value across genotypes. QXcat and QZmax are used for testing the mean differences of the trait value only. wM3VNA3.3 is for testing the variance heterogeneity only.

Value

the *p*-values and the test statistics of association tests selected by the method option for each SNP.

Author(s)

Yu-Xin Yuan, Zi-Ying Yang and Ji-Yuan Zhou

References

Yang ZY, Liu W, Yuan YX, et al. Robust association tests for quantitative traits on the X chromosome. 2022

Deng WQ, Mao S, Kalnapenkis A, et al. Analytical strategies to include the X-chromosome in variance heterogeneity analyses: evidence for trait-specific polygenic variance structure. *Genetic Epidemiology*, 2019, **43**: 815-830.

Ionita-Laza I, Lee S, Makarov V, et al. Sequence kernel association tests for the combined effect of rare and common variants. *The American Journal of Human Genetics*, 2013, **92**: 841-853.

Soave D, Corvol H, Panjwani N, et al. A joint location-scale test improves power to detect associated SNPs, gene sets, and pathways. *The American Journal of Human Genetics*, 2015, **97**: 125–138.

Examples

```
#Phedata: a dataset for 4000 unrelated individuals and four variables,
          including one phenotype and three covariates (i.e., Sex,age and BMI).
data(Phedata)
#Genotype: a dataset for 4000 unrelated individuals and 31 SNPs.
data(Genotype)

#the location tests (i.e., QXcat and QZmax)
QMV_test(Genotype,Phedata$Y,
         Phedata$Sex,
         Covariate=Phedata[,c(-1,-2)],
         missing_cutoff=0.15,
         MAF_Cutoff=NULL,
         MGC_Cutoff=30,
         method='location')

#the scale test (i.e., wM3VNA3.3)
QMV_test(Genotype,Phedata$Y,
         Phedata$Sex,
         Covariate=Phedata[,c(-1,-2)],
         missing_cutoff=0.15,
         MAF_Cutoff=NULL,
         MGC_Cutoff=30,
         method='scale')

#the joint tests (i.e., QMVXcat and QMVZmax)
QMV_test(Genotype,Phedata$Y,
         Phedata$Sex,
         Covariate=Phedata[,c(-1,-2)],
         missing_cutoff=0.15,
         MAF_Cutoff=NULL,
         MGC_Cutoff=30,
         method='joint')

#all of the above association tests
QMV_test(Genotype,Phedata$Y,
         Phedata$Sex,
         Covariate=Phedata[,c(-1,-2)],
         missing_cutoff=0.15,
         MAF_Cutoff=NULL,
         MGC_Cutoff=30,
         method='all')
```

`QMV_test_bed` *Read PLINK bed files into R, and compute the p-values and the test statistics of the location tests, the scale test, the location-scale tests or all*

Description

A function to obtain the *p*-values and the test statistics of the location tests (i.e., QXcat and QZmax), the scale test (i.e.,wM3VNA3.3), the location-scale tests (i.e., QMVXcat and QMVZmax) or all. This function takes the path to the file of genotype with extension ".bed" (`bedfile`), and possibly additional path to the file of covariates (`Covariate_path`) in the sample population, as input.

Usage

```
QMV_test_bed(bedfile,fastImpute=FALSE,
             Covariate_path=NULL,
             Covariate_missing = NA,
             Covariate.header=FALSE,
             missing_cutoff=0.15,
             MAF_Cutoff=NULL,
             MGC_Cutoff=30,
             method='joint')
```

Arguments

bedfile	Path to the file of genotype with extension ".bed" to read. The corresponding files with extensions ".bim" and ".fam" should be in the same directory.
fastImpute	Logical scalar defaulting to False (or F) indicating no imputation for missing genotypes. If fastImpute=TURE, imputation is conducted by a fast imputation algorithm based on local XGBoost models (bigsnpr::snp_fastImpute).
Covariate_path	Path to the file of possibly additional covariates to read. This file should be a text file, one line per sample, with C+6 columns, where C is the number of covariates. The former six columns should be Family ID (FID), Individual ID (IID), Paternal within-family ID (PID), Maternal within-family ID (MID), Sex and Phenotype.
Covariate_missing	The input variable "covariate_missing" is the missing value for the covariates in the data file, and the default value is NA.
Covariate.header	Logical scalar defaulting to False (or F) indicating whether the covariate file contains variable names or not.
missing_cutoff	Cutoff of the missing rates of SNPs (default=0.15). Any SNPs with missing rates higher than the cutoff will be excluded from the analysis.
MAF_Cutoff	MAF cutoff for common vs. rare variants (default=NULL). It should be a numeric value between 0 and 0.5, or NULL. When it is NULL, 1/ sqrt(2 SampleSize) will be used (Ionita-Laza et al. 2013). Only common variants are included in the analysis.
MGC_Cutoff	Cutoff for the minimum genotype count in either females or males (default=30), SNPs whose minimum genotype counts are less than this cutoff will not be included in the analysis. This is based on the quality control that SNPs with a minimum count below 30 should be removed to avoid inflated type I errors (Deng et al., 2019; Soave et al., 2015).
method	A character string indicating which kind of association tests is to be conducted. There are four options: "location", "scale", "joint" (default) and "all". method="location": QXcat and QZmax; method="scale": wM3VNA3.3; method="joint": QMVXcat and QMVZmax; method="all": all of the above association tests.

Details

QMVXcat and QMVZmax are designed to test for both the mean differences and the variance heterogeneity of the trait value across genotypes. QXcat and QZmax are used for testing the mean differences of the trait value only. wM3VNA3.3 is for testing the variance heterogeneity only.

Value

the *p*-values and the test statistics of association tests selected by the method option for each SNP.

Author(s)

Yu-Xin Yuan, Zi-Ying Yang and Ji-Yuan Zhou

References

Yang ZY, Liu W, Yuan YX, et al. Robust association tests for quantitative traits on the X chromosome. 2022

Deng WQ, Mao S, Kalnapenkis A, et al. Analytical strategies to include the X-chromosome in variance heterogeneity analyses: evidence for trait-specific polygenic variance structure. *Genetic Epidemiology*, 2019, **43**: 815-830.

Ionita-Laza I, Lee S, Makarov V, et al. Sequence kernel association tests for the combined effect of rare and common variants. *The American Journal of Human Genetics*, 2013, **92**: 841-853.

Soave D, Corvol H, Panjwani N, et al. A joint location-scale test improves power to detect associated SNPs, gene sets, and pathways. *The American Journal of Human Genetics*, 2015, **97**: 125–138.

Examples

```
path <- system.file("extdata", package = "QMVtest")

QMV_test_bed(paste(path,"example-missing.bed",sep = '/'),
             fastImpute=FALSE,
             Covariate_path=paste(path,"Covariate.txt",sep = '/'),
             Covariate_missing = NA,
             Covariate.header=TRUE,
             missing_cutoff=0.15,
             MAF_Cutoff=NULL,
             MGC_Cutoff=30,
             method='joint')
```

`QMV_test_ped` *Read PLINK raw files into R, and compute the p-values and the test statistics of the location tests, the scale test, the location-scale tests or all*

Description

A function to obtain the *p*-values and the test statistics of the location tests (i.e., QXcat and QZmax), the scale test (i.e.,wM3VNA3.3), the location-scale tests (i.e., QMVXcat and QMVZmax) or all. This function takes the path to the file of genotype with extension ".raw" (`ped_raw_path`), the path to the file of SNP information with extension ".map" (`map_path`), and possibly additional path to the file of covariates (`Covariate_path`) in the sample population, as input.

Usage

```
QMV_test_ped(ped_raw_path,trait_missing=NA,
             Genotype_missing=NA,
             ped_raw.header=FALSE,
             Covariate_path=NULL,
             Covariate_missing = NA,
             Covariate.header=FALSE,
             map_path=NULL,map_header=FALSE,
             missing_cutoff=0.15,
             MAF_Cutoff=NULL,
             MGC_Cutoff=30,
             method='joint')
```

Arguments

ped_raw_path	Path to the file of genotype with extension ".raw" to read, which is produced by "–recode A" in PLINK for use with R. This file should be a text file, one line per sample, with V+6 columns, where V is the number of variants. The first six columns are: Family ID (FID), Individual ID (IID), Paternal within-family ID (PID), Maternal within-family ID (MID), Sex and Phenotype, followed by one column per variant.
trait_missing	The input variable "trait_missing" is the missing value for the trait in the raw file, and the default value is NA. It may be -9 in some raw files; or other numeric value.
Genotype_missing	The input variable "Genotype_missing" represents that the genotype at the locus in the raw file is missing, and the default value is NA. It may be -9 in some raw files; or other numeric value.
ped_raw.header	Logical scalar defaulting to False (or F) indicating whether the raw file contains variable names or not.
Covariate_path	Path to the file of possibly additional covariates to read. This file should be a text file, one line per sample, with C+6 columns, where C is the number of covariates. The former six columns should be Family ID (FID), Individual ID (IID), Paternal within-family ID (PID), Maternal within-family ID (MID), Sex and Phenotype.
Covariate_missing	The input variable "covariate_missing" is the missing value for the covariates in the data file, and the default value is NA.
Covariate.header	Logical scalar defaulting to False (or F) indicating whether the covariate file contains variable names or not.
map_path	Path to the file of SNP information with extension ".map" to read. Each line of the map file describes a SNP and must contain four columns: chromosome (1-22, X, Y or 0 if unplaced), rs# or SNP identifier, genetic distance (morgans) and base-pair position (bp units).
map_header	Logical scalar defaulting to False (or F) indicating whether the map file contains variable names or not.
missing_cutoff	Cutoff of the missing rates of SNPs (default=0.15). Any SNPs with missing rates higher than the cutoff will be excluded from the analysis.

MAF_Cutoff	MAF cutoff for common vs. rare variants (default=NULL). It should be a numeric value between 0 and 0.5, or NULL. When it is NULL, 1/ sqrt(2 SampleSize) will be used (Ionita-Laza et al. 2013). Only common variants are included in the analysis.
MGC_Cutoff	Cutoff for the minimum genotype count in either females or males (default=30), SNPs whose minimum genotype counts are less than this cutoff will not be included in the analysis. This is based on the quality control that SNPs with a minimum count below 30 should be removed to avoid inflated type I errors (Deng et al., 2019; Soave et al., 2015).
method	A character string indicating which kind of association tests is to be conducted. There are four options: "location", "scale", "joint" (default) and "all". method="location": QXcat and QZmax; method="scale": wM3VNA3.3; method="joint": QMVXcat and QMVZmax; method="all": all of the above association tests.

Details

QMVXcat and QMVZmax are designed to test for both the mean differences and the variance heterogeneity of the trait value across genotypes. QXcat and QZmax are used for testing the mean differences of the trait value only. wM3VNA3.3 is for testing the variance heterogeneity only.

Value

the *p*-values and the test statistics of association tests selected by the method option for each SNP.

Author(s)

Yu-Xin Yuan, Zi-Ying Yang and Ji-Yuan Zhou

References

Yang ZY, Liu W, Yuan YX, et al. Robust association tests for quantitative traits on the X chromosome. 2022

Deng WQ, Mao S, Kalnapenkis A, et al. Analytical strategies to include the X-chromosome in variance heterogeneity analyses: evidence for trait-specific polygenic variance structure. *Genetic Epidemiology*, 2019, **43**: 815-830.

Ionita-Laza I, Lee S, Makarov V, et al. Sequence kernel association tests for the combined effect of rare and common variants. *The American Journal of Human Genetics*, 2013, **92**: 841-853.

Soave D, Corvol H, Panjwani N, et al. A joint location-scale test improves power to detect associated SNPs, gene sets, and pathways. *The American Journal of Human Genetics*, 2015, **97**: 125–138.

Examples

```
path <- system.file("extdata", package = "QMVtest")

QMV_test_ped(paste(path,"example.raw",sep = '/'),trait_missing=NA,
             Genotype_missing=NA,
             ped_raw.header=FALSE,
             Covariate_path=paste(path,"Covariate.txt",sep = '/'),
             Covariate_missing = NA,
             Covariate.header=TRUE,
             map_path=paste(path,"example.map",sep = '/'),
             map_header=FALSE,
```

```
              missing_cutoff=0.15,
              MAF_Cutoff=NULL,
              MGC_Cutoff=30,
              method='joint')
```

`scale_test` *A function to obtain the p-value and the test statistic of wM3VNA3.3 testing for the variance heterogeneity of the trait value across genotypes*

Description

A function to obtain the *p*–value and the test statistic of wM3VNA3.3 in Deng et al. (2019). This function takes the genotype of SNPs (`Genotype`), the sex (`Sex`), the quantitative trait (`Y`) in the sample population, and possibly additional covariates, such as age and BMI, as input.

Usage

```
scale_test(Genotype,Y,Sex,
           Covariate=NULL,
           missing_cutoff=0.15,
           MAF_Cutoff=NULL,
           MGC_Cutoff=30)
```

Arguments

`Genotype`	A numeric genotype matrix with each row as a different individual and each column as a separate SNP. Each genotype is coded as 0, 1 or 2 for females and coded as 0 or 1 for males, indicating the number of reference allele. The length/dimension of `Genotype` should be the same to those of `Y`, `Sex` and `Covariate`.
`Y`	A numeric vector of a quantitative trait, such as human height.
`Sex`	A vector of the genetic sex following PLINK default coding, where males are coded as 1 and females are coded as 2.
`Covariate`	Optional: a vector or a matrix of covariates, such as age and BMI.
`missing_cutoff`	Cutoff of the missing rates of SNPs (default=0.15). Any SNPs with missing rates higher than the cutoff will be excluded from the analysis.
`MAF_Cutoff`	MAF cutoff for common vs. rare variants (default=NULL). It should be a numeric value between 0 and 0.5, or NULL. When it is NULL, 1/ sqrt(2 SampleSize) will be used (Ionita-Laza et al. 2013). Only common variants are included in the analysis.
`MGC_Cutoff`	Cutoff for the minimum genotype count in either females or males (default=30), SNPs whose minimum genotype counts are less than this cutoff will not be included in the analysis. This is based on the quality control that SNPs with a minimum count below 30 should be removed to avoid inflated type I errors (Deng et al., 2019; Soave et al., 2015).

Value

the *p*-value and the test statistic of the scale test wM3VNA3.3 in Deng et al. (2019).

Author(s)

Wei Q. Deng and Lei Sun

References

Yang ZY, Liu W, Yuan YX, et al. Robust association tests for quantitative traits on the X chromosome. 2022

Deng WQ, Mao S, Kalnapenkis A, et al. Analytical strategies to include the X-chromosome in variance heterogeneity analyses: evidence for trait-specific polygenic variance structure. *Genetic Epidemiology*, 2019, **43**: 815-830.

Ionita-Laza I, Lee S, Makarov V, et al. Sequence kernel association tests for the combined effect of rare and common variants. *The American Journal of Human Genetics*, 2013, **92**: 841-853.

Soave D, Corvol H, Panjwani N, et al. A joint location-scale test improves power to detect associated SNPs, gene sets, and pathways. *The American Journal of Human Genetics*, 2015, **97**: 125–138.

Examples

```
#Phedata: a dataset for 4000 unrelated individuals and four variables,
          including one phenotype and three covariates (i.e., Sex,age and BMI).
data(Phedata)
#Genotype: a dataset for 4000 unrelated individuals and 31 SNPs.
data(Genotype)

#the scale test (i.e., wM3VNA3.3)
#no covariate
#set "Covariate=NULL"
scale_test(Genotype,Phedata$Y,
           Phedata$Sex,
           Covariate=NULL,
           missing_cutoff=0.15,
           MAF_Cutoff=NULL,
           MGC_Cutoff=30)

#age and BMI as covariates
scale_test(Genotype,Phedata$Y,
           Phedata$Sex,
           Covariate=Phedata[,c(-1,-2)],
           missing_cutoff=0.15,
           MAF_Cutoff=NULL,
           MGC_Cutoff=30)
```

Index