Package 'Xvarhet'

December 22, 2020

Type Package

Title X-chromosome Variance Heterogeneity Analysis via a Generalized Levene's Test

Version 1.0

Author Wei Q. Deng

Maintainer Wei Q. Deng <deng@utstat.toronto.edu>

Description A collection of functions to produce robust quantitative trait variance heterogeneity p-values for X-chromosome SNPs.

License GPL (>= 2)

Encoding UTF-8

Imports methods,
ggplot2,
quantreg,
stats

Depends R (>= 3.4.0)

LazyData true

RoxygenNote 7.0.2

Suggests testthat (>= 2.1.0)

R topics documented:

`ConceptFigure` *Generating Conceptual Sex-Stratified Distributions as Shown in Xvarhet Manuscript*

Description

This function generates the conceptual distribution of a quantitative trait stratified by sex and genotypes under the null hypothesis and alternative hypothesis.

Usage

```
ConceptFigure(
  RAF,
  prp_m,
  length.OUT = 500,
  Yrange = 3,
  y1_vect = c(0, 1),
  y2_vect = c(-0.5, 0.5, 1, 1),
  y3_vect = c(0, 0, 0.8, 1.2),
  y4_vect = c(-0.5, 0.5, 0.8, 1.2),
  geno_sd = 0.2,
  cols = T
)
```

Arguments

RAF	a numeric between 0 and 1, the reference allele frequency of the SNP, assumed to be equal in females and males.
prp_m	a numeric between 0 and 1, giving the porportion of males in the sample.
length.OUT	an integer for the number of quantitative trait values at which to evaluate; the default is set to 500.
Yrange	a numeric for the maximum value of the quantitative trait to be evaluated assuming the distribution is symmetric about zero; the default is set to 3.
y1_vect	a vector for the sex-stratified mean and standard deviation in females and males under no sexual dimorphism; the default is c(0,1), with zero mean and standard deviation one in both sexes.
y2_vect	a vector for the sex-stratified mean and standard deviation in females and males under sexual dimorphism in mean only; the default is c(-0.5,0.5,1,1), with quantitative trait in female following a normal distribution $N(\mu = -0.5, \sigma = 1)$.
y3_vect	a vector for the sex-stratified mean and standard deviation in females and males under sexual dimorphism in variance only; the default is c(0,0,0.8,1.2), with quantitative trait in female following a normal distribution $N(\mu = 0, \sigma = 0.8)$.
y4_vect	a vector for the sex-stratified mean and standard deviation in females and males under sexual dimorphism in both mean and variance; the default is c(-0.5,0.5,0.8,1.2), with quantitative trait in female following a normal distribution $N(\mu = -0.5, \sigma = 0.8)$.
geno_sd	a numeric between 0 and the minimum of the sex-stratified standard deviation, giving the contribution to standard deviation per reference allele under no sexual dimorphism assuming an additive genetic variance effect. The default is 0.2.
cols	a logic indicating whether the outputted graphical object should be in colour or gray style.

Value

a ggplot2 graph object

Author(s)

Wei Q. Deng <deng@utstat.toronto.edu>

Examples

```
#ConceptFigure(RAF=0.2, prp_m=0.1); # not run
#ConceptFigure(RAF=0.2, prp_m=0.1, cols = FALSE); # not run
```

---

leveneReg *Levene's regression tests for variance homogeneity by SNP genotype*

Description

The function takes as input the genotype of a SNP (GENO), the SEX (SEX), and a quantitative trait (Y) in a sample population, and possibly additional covariates. It should be noted that these variables must be of the same length. The function then returns the variance heterogeneity *p*-values using the generalized Levene's test. The residual function could alternatively be replaced with the quantile regression quantreg::rq following the "fn" algorithm, for more details see ?quantreg::rq.

Usage

```
leveneReg(GENO, SEX, PLINK = FALSE, Y, COV = NULL, test_type = "REC")
```

Arguments

GENO	the genotype of a SNP, must be a vector of 0, 1, 2's indicating the number of reference alleles. The length of GENO should match that of SEX, Y, and any covariates COV.
SEX	the genetic sex of individuals in the sample population, must be a vector of 1 and 2 or 0 and 1, depending on whether the PLINK sex code is used. Note that the default sex code is 1 for males and 2 for females in PLINK.
PLINK	a logical indicating whether the SEX is coded following PLINK, i.e. female coded as 2 and male colded as 1; if not, SEX is assumed to be coded 0 for females and 1 for males.
Y	a vector of quantitative traits, such as human height.
COV	a vector or matrix of covariates that are used to reduce bias due to confounding, such as age.
test_type	a character of either "ALL", printing all strategies, or "M1", "M2", or "M3" for the 8 stratigies given the mean stage models, or "REC" printing only the recommended tests.

Value

a vector of Levene's test regression p-values according to the models specified.

Note

We recommend to quantile-normally transform Y to avoid 'scale-effect' where the variance values tend to be proportional to mean values when stratified by GENO.

Author(s)

Wei Q. Deng <deng@utstat.toronto.edu>

Examples

```
N <- 5000
GENO <- rbinom(N, 2, 0.3)
sex <- rbinom(N, 1, 0.5)
Y <- rnorm(N)
cov <- matrix(rnorm(N*10), ncol=10)
leveneReg(GENO=GENO, SEX=sex, Y=Y, COV=cov, test_type="ALL")
leveneReg(GENO=GENO, SEX=sex, Y=Y, COV=cov, test_type="REC")
```

`leveneTests` *Levene's test for variance homogeneity by SNP genotypes*

Description

The function takes as input the genotypes of a SNP (`GENO`), the sex (`SEX`), and a quantitative trait (`Y`) in a sample population, and possibly additional covariates. It should be noted that these variables must be of the same length. The function then returns the variance heterogeneity *p*-values for the model $Y \sim G$, $Y \sim G \times S$, the sex-specific results based on model $Y \sim G$, as well as that using Fisher's method to combine sex-specific results.

Usage

```
leveneTests(GENO, SEX, PLINK = FALSE, Y, centre = "median", COV = NULL)
```

Arguments

`GENO`	the genotype of a SNP, must be a vector of 0, 1, 2's indicating the number of reference alleles. The length of `GENO` should match that of `SEX`, `Y`, and any covariates.
`SEX`	the genetic sex of individuals in the sample population, must be a vector of 1 and 2 or 0 and 1, depending on whether the PLINK sex code is used. Note that the default sex code is 1 for male and 2 for female in PLINK.
`PLINK`	a logical indicating whether the SEX is coded following PLINK or not.
`Y`	a vector of quantitative traits, such as human height.
`centre`	a character indicating whether the absolute deviation should be calculated with respect to "median" or "mean".
`COV`	a vector or matrix of covariates that are used to reduce bias due to confounding, such as age.

Value

a vector of Levene's test p-values according to levels specified by `GENO`, the interaction between `GENO` and `SEX`, sex-specific Levene's test stratified by `GENO`, and the Fisher's method to combine the sex-specific Levene's test *p*-values.

Note

We recommend to quantile-normally transform `Y` to avoid 'scale-effect' where the variance values tend to be proportional to mean values when stratified by `G`.

Author(s)

Wei Q. Deng <deng@utstat.toronto.edu>

Examples

```
N <- 5000
geno <- rbinom(N, 2, 0.3)
sex <- rbinom(N, 1, 0.5)
y <- rnorm(N)
cov <- matrix(rnorm(N*10), ncol=10)
leveneTests(GENO=geno, SEX=sex, Y=y, COV=cov)
```

T1EsimDesign1 *Type I Error (T1E) Rates under Simulation Design I*

Description

This function returns the type I error rates of testing strategies under the simulation conditions specified by a linear regression model. The quantitative trait Y is simulated from the following model without any genetic effect: $Y = \beta_E E + \beta_S S + \beta_{SE} SE + \epsilon$, where E and ϵ are assumed to follow a standard normal distribution.

Usage

```
T1EsimDesign1(
  Nf,
  Nm,
  MAF_F,
  MAF_M,
  betaE,
  betaS,
  betaSE,
  nbSim = 100,
  alpha = 0.05,
  nbDigits = 3,
  MisspecX = FALSE,
  test_type = "ALL"
)
```

Arguments

Nf	an integer for the number of females in the sample population.
Nm	an integer for the number of males in the sample population.
MAF_F	a numerical between 0 and 0.5 for minor allele frequency of the simulated SNP in females.
MAF_M	a numerical between 0 and 0.5 for minor allele frequency of the simulated SNP in males.
betaE	a numerical value for the marginal environmental covariate effect on the simulated trait.

betaS	a numerical value for the marginal sex effect on the simulated trait.
betaSE	a numerical value for the interaction effect between sex and an environmental covariate.
nbSim	an integer for the number of simulations.
alpha	a numerical between 0 and 1 for the significance threshold, the default value is 0.05.
nbDigits	an integer indicating the number of printed digits for empirical T1E rates.
MisspecX	a logic indicating whether the T1E should be calculated under a mis-specified X-inactivation status, i.e. model generated under X-inactivation but tested under absence of X-inactivation.
test_type	a character of either "ALL", printing all strategies, or "M1", "M2", or "M3" for the 8 stratigies given the mean stage models, or "REC" printing only the recommended test.

Value

a vector of type I error rates of the specified testing strategies.

Note

We recommend to run the simulation in the background as it might take long for more than 500 simulations on a 1.6 GHz Core.

Author(s)

Wei Q. Deng <deng@utstat.toronto.edu>

Examples

```
T1EsimDesign1(Nf=1000, Nm=1000, MAF_F=0.2, MAF_M=0.2, betaE=0.1,
betaS=0.2, betaSE=0, nbSim=100, alpha=0.05, nbDigits = 3, MisspecX =FALSE,
test_type = "ALL")

T1EsimDesign1(Nf=1000, Nm=1000, MAF_F=0.2, MAF_M=0.2, betaE=0.1,
betaS=0.2, betaSE=0, nbSim=100, alpha=0.05, nbDigits = 3, MisspecX =FALSE,
test_type = "REC")

T1EsimDesign1(Nf=1000, Nm=2000, MAF_F=0.2, MAF_M=0.2, betaE=0.1,
betaS=0.2, betaSE=0, nbSim=100, alpha=0.05, nbDigits = 3, MisspecX=TRUE,
test_type = "ALL")
```

Index