Package 'CCpop'

February 19, 2015

Type Package

Title One and two locus GWAS of binary phenotype with case-control-population design

Version 1.0

Date 2014-03-18

Author Shachar Kaufman

Maintainer Shachar Kaufman <shachark@post.tau.ac.il>

Description Tests of association between SNPs or pairs of SNPs and binary phenotypes, in case-control / case-population / case-control-population studies.

Depends nloptr

License GPL-2

NeedsCompilation no

Repository CRAN

Date/Publication 2014-03-24 12:03:06

R topics documented:

`CCpop-package` *One and two locus GWAS of binary phenotype with case-control-population design*

Description

Tests for marginal and pairwise SNP associations with binary phenotypes in case-control, case-population, and case-control-population studies.

Details

Package:	CCpop
Type:	Package
Version:	1.0
Date:	2014-03-18
License:	GPL-2

Inputs to test functions are vectors (for marginal tests) and matrices (for pairwise tests) of genotypic counts and are given separately for cases, controls, and population cohorts (where relevant). The constrained maximum likelihood estimation based tests (Kaufman and Rosset, 2014) also require a value for the known phenotypic prevalence in the population (a crude estimate is usually sufficient). When applicable, tests assuming Hardy-Weinberg equilibrium (HWE) and linkage equilibrium (LE) will be considerably more powerful.

See marginal.assoc.tests for single SNP tests, and pairwise.assoc.tests for pairwise joint, pure-interaction, and conditional tests.

Author(s)

Shachar Kaufman <shachark@post.tau.ac.il>

References

Kaufman, S., & Rosset, S. (2014). Exploiting Population Samples To Enhance Genome-Wide Association Studies of Disease. Genetics, genetics-114.

Examples

```
## An example marginal/pairwise association

# Controls
t0 = matrix(c(375,  240,   46,
              640,  405,   62,
              300,  169,   19), nrow = 3, byrow = TRUE)

# Cases
t1 = matrix(c(317,  162,   15,
              459,  209,   22,
              120,   76,   13), nrow = 3, byrow = TRUE)

# Independent population sample, marginalized for SNP1 and SNP2
tp1 = c(2410, 4253, 1945)
tp2 = c(4972, 3140,  496)

## The prevalence of the studied disease in the population
prevalence = 0.001

marginal.assoc.test.pop.hwe.kpy(t0 = rowSums(t0), t1 = rowSums(t1), tp = tp1, prevalence)
marginal.assoc.test.pop.hwe.kpy(t0 = colSums(t0), t1 = colSums(t1), tp = tp2, prevalence)
pairwise.assoc.test.pop.hwe.le.kpy(t0, t1, tp1, tp2, prevalence)
conditional.assoc.test.pure.pop.hwe.le.kpy(t0, t1, tp1, tp2, prevalence)
```

`marginal.assoc.test.unconstrained.chisq` *Marginal association tests*

Description

Tests for association of single SNP locus and a binary phenotype in case-control, case-population, and case-control-population designs, under various assumptions.

Usage

```
marginal.assoc.test.unconstrained.chisq(t0, t1)
marginal.assoc.test.unconstrained.gsq(t0, t1)
marginal.assoc.test.hwe.in.controls(t0, t1)
marginal.assoc.test.pop.hwe.kpy(t0, t1, tp, prevalence,
  pen.initial = NULL, f.initial = NULL)
marginal.assoc.test.kpx.kpy(t0, t1, prevalence, px, pen.initial = NULL)
```

Arguments

`t0`	A 3-component vector of genotype counts in the controls cohort
`t1`	A 3-component vector of genotype counts in the cases cohort
`tp`	A 3-component vector of genotype counts in the population cohort
`prevalence`	The population prevalence of the phenotype `Pr(y=1) < 0.5`
`px`	A 3-component vector of known genotypic distribution in the population `Pr(x)`
`pen.initial`	A 3-component vector of initial guess for phenotypic penetrance `Pr(y=1|x)`
`f.initial`	Initial guess for the SNP's minor allele frequency (MAF)

Details

marginal.assoc.test.unconstrained.chisq – Pearson's chi-squared test for the 2x3 contingency table of a standard case-control design.

marginal.assoc.test.unconstrained.gsq – Likelihood ratio ("G") test for the 2x3 contingency table of a standard case-control design.

marginal.assoc.test.hwe.in.controls – A test for association in a standard case-control design, assuming HWE among the controls (Chen and Chatterjee, 2007).

marginal.assoc.test.pop.hwe.kpy – GLRT based on constrained maximum likelihood estimation assuming HWE in the general population, and that disease prevalence is known.

marginal.assoc.test.kpx.kpy – GLRT based on constrained maximum likelihood estimation assuming the marginal genotypic and phenotypic distributions are known.

Value

p.value	P-value indicating significance of the association
statistic	Value of the relevant test statistic
pen	The estimated case-control peneterance Pr(y=1|x,ascertained)

Note

The implementation assumes that contingency tables are provided with homozygote major allele as the first element.

In order to perform case-population testing, the user may set t0 to a vector of zeros.

An inferior alternative to case-control-population analysis which may be desirable in the interest of analysis speed and can be acceptable for low prevalence phenotypes is to extend the controls cohort with any population samples. This can be achieved by providing t0.new = t0.old + tp.old and using one of the case-control tests.

Tests based on the 2x9 contingency table treat the two SNPs as one 9-category variable.

Author(s)

Shachar Kaufman

References

Chen J., & Chatterjee, N. (2007). Exploiting hardy-weinberg equilibrium for effcient screening of single SNP associations from case-control studies. Human heredity, 63(3-4), 196-204.

Kaufman, S., & Rosset, S. (2014). Exploiting Population Samples To Enhance Genome-Wide Association Studies of Disease. Genetics, 114 (early electronic).

See Also

pairwise.assoc.tests,

Examples

```
## An example marginal/pairwise association

# Controls
t0 = matrix(c(375,  240,   46,
              640,  405,   62,
              300,  169,   19), nrow = 3, byrow = TRUE)

# Cases
t1 = matrix(c(317,  162,   15,
              459,  209,   22,
              120,   76,   13), nrow = 3, byrow = TRUE)

# Independent population sample, marginalized for SNP1 and SNP2
tp1 = c(2410, 4253, 1945)
tp2 = c(4972, 3140,  496)
```

```
## The prevalence of the studied disease in the population
prevalence = 0.001

marginal.assoc.test.pop.hwe.kpy(t0 = rowSums(t0), t1 = rowSums(t1), tp = tp1, prevalence)
marginal.assoc.test.pop.hwe.kpy(t0 = colSums(t0), t1 = colSums(t1), tp = tp2, prevalence)
pairwise.assoc.test.pop.hwe.le.kpy(t0, t1, tp1, tp2, prevalence)
conditional.assoc.test.pure.pop.hwe.le.kpy(t0, t1, tp1, tp2, prevalence)
```

`pairwise.assoc.test.unconstrained.chisq`
Pairwise association tests

Description

Tests for joint association of a pair of SNP loci and a binary phenotype in case-only, case-control, case-population, and case-control-population designs, under various assumptions.

Usage

```
pairwise.assoc.test.unconstrained.chisq(t0, t1)
pairwise.assoc.test.unconstrained.gsq(t0, t1)
pairwise.assoc.test.case.only(t1)
pairwise.assoc.test.ind.3d(t0, t1)
pairwise.assoc.test.pure.unconstrained(t0, t1)
pairwise.assoc.test.kpy(t0, t1, prevalence, pen.initial = NULL, pxx.initial = NULL)
pairwise.assoc.test.hwe.le.kpy(t0, t1, prevalence,
  pen.initial = NULL, f1.initial = NULL, f2.initial = NULL)
pairwise.assoc.test.kpx.kpy(t0, t1, prevalence, pxx, pen.initial = NULL)
pairwise.assoc.test.pop.kpy(t0, t1, tp, prevalence,
  pen.initial = NULL, pxx.initial = NULL)
pairwise.assoc.test.pop.hwe.le.kpy(t0, t1, tp1, tp2, prevalence,
  pen.initial = NULL, f1.initial = NULL, f2.initial = NULL)
pairwise.assoc.test.pure.pop.kpy(t0, t1, tp, prevalence,
  pen.initial = NULL, pxx.initial = NULL)
pairwise.assoc.test.pure.pop.hwe.le.kpy(t0, t1, tp1, tp2, prevalence,
  pen.initial = NULL, f1.initial = NULL, f2.initial = NULL)
conditional.assoc.test.pure.pop.hwe.le.kpy(t0, t1, tp1, tp2, prevalence,
  pen.initial = NULL, f1.initial = NULL, f2.initial = NULL)
```

Arguments

`t0`	A 3x3 matrix of genotype counts in the controls cohort
`t1`	A 3x3 matrix of genotype counts in the cases cohort
`tp`	A 3x3 matrix of genotype counts in the population cohort
`tp1`	A 3-component vector of genotype counts for the first (`t0`/`t1` rows) locus in the population cohort

tp2	A 3-component vector of genotype counts for the second (t0/t1 columns) locus in the population cohort
prevalence	The population prevalence of the phenotype Pr(y=1) < 0.5
pxx	A 3x3 matrix of known genotypic distribution in the population Pr(x1,x2)
pen.initial	A 3x3 matrix of initial guess for the phenotypic penetrance Pr(y=1|x1,x2)
pxx.initial	A 3x3 matrix of initial guess for the genotypic distribution in the population Pr(x1,x2)
f1.initial	Initial guess for the MAF at the first (rows) locus
f2.initial	Initial guess for the MAF at the second (columns) locus

Details

pairwise.assoc.test.unconstrained.chisq – Pearson's chi-squared test for the 2x9 contingency table of a standard case-control design.

pairwise.assoc.test.unconstrained.gsq – Likelihood ratio ("G") test for the 2x9 contingency table of a standard case-control design.

pairwise.assoc.test.case.only – The likelihood ratio ("G") test for the 3x3 contingency table of the two SNPs in cases only.

pairwise.assoc.test.ind.3d – The likelihood ratio ("G") test for the 2x3x3 contingency table of the two SNPs and phenotype value (i.e., a case-control design).

pairwise.assoc.test.pure.unconstrained – The logistic regression based GLRT for full model vs. main effects model (within a standard case-control design).

pairwise.assoc.test.kpy – The GLRT based on constrained maximum likelihood estimation assuming that phenotype prevalence in the population is known (within a standard case-control design).

pairwise.assoc.test.hwe.le.kpy – The GLRT based on constrained maximum likelihood estimation assuming Hardy-Weinberg equilibrium (HWE) and linkage equilibrium (LE) in the general population, and that phenotype prevalence in the population is known (within a standard case-control design).

pairwise.assoc.test.kpx.kpy – The GLRT based on constrained maximum likelihood estimation assuming that marginal genotypic and phenotypic distributions in the population are both known (within a standard case-control design).

pairwise.assoc.test.pop.kpy – The GLRT based on constrained maximum likelihood estimation assuming that phenotype prevalence in the population is known (within a case-control-population design).

pairwise.assoc.test.pop.hwe.le.kpy – The GLRT based on constrained maximum likelihood estimation assuming Hardy-Weinberg equilibrium (HWE) and linkage equilibrium (LE) in the general population, and that phenotype prevalence in the population is known (within a case-control-population design).

pairwise.assoc.test.pure.pop.kpy – A GLRT based on constrained maximum likelihood estimation assuming that phenotype prevalence is known (within a case-control-population design). The test contrasts a general pairwise association versus a logit-additive "main-effects" model.

pairwise.assoc.test.pure.pop.hwe.le.kpy – Like pairwise.assoc.test.pure.pop.kpy, but done under the assumption of HWE and LE.

conditional.assoc.test.pure.pop.hwe.le.kpy – A GLRT based on constrained maximum likelihood estimation assuming that phenotype prevalence is known and assuming HWE and LE (within a case-control-population design). This tests for a pairwise association above and beyond a marginal association of x1 alone, or of x2 alone (and the *maximum* p-value of these two options is returned, i.e., the test is conservative)

Value

`p.value`	P-value indicating significance of the association
`statistic`	Value of the relevant test statistic
`pen`	The estimated case-control peneterance `Pr(y=1|x,ascertained)`

Note

The implementation assumes that contingency tables are provided with homozygote major allele as the first row and column.

In order to perform case-population testing, the user may set `t0` to a matrix of zeros.

An inferior alternative to case-control-population analysis which may be desirable in the interest of analysis speed and can be acceptable for low prevalence phenotypes is to extend the controls cohort with any population samples. This can be achieved by providing `t0.new = t0.old + tp.old` and using one of the case-control tests.

`pen` is not returned by the case-only test.

Author(s)

Shachar Kaufman

References

Kaufman, S., & Rosset, S. (2014). Exploiting Population Samples To Enhance Genome-Wide Association Studies of Disease. Genetics, 114 (early electronic).

See Also

`marginal.assoc.tests`,

Examples

```
## An example marginal/pairwise association

# Controls
t0 = matrix(c(375,  240,   46,
              640,  405,   62,
              300,  169,   19), nrow = 3, byrow = TRUE)

# Cases
t1 = matrix(c(317,  162,   15,
              459,  209,   22,
              120,   76,   13), nrow = 3, byrow = TRUE)
```

```
# Independent population sample, marginalized for SNP1 and SNP2
tp1 = c(2410, 4253, 1945)
tp2 = c(4972, 3140,  496)

## The prevalence of the studied disease in the population
prevalence = 0.001

marginal.assoc.test.pop.hwe.kpy(t0 = rowSums(t0), t1 = rowSums(t1), tp = tp1, prevalence)
marginal.assoc.test.pop.hwe.kpy(t0 = colSums(t0), t1 = colSums(t1), tp = tp2, prevalence)
pairwise.assoc.test.pop.hwe.le.kpy(t0, t1, tp1, tp2, prevalence)
conditional.assoc.test.pure.pop.hwe.le.kpy(t0, t1, tp1, tp2, prevalence)
```

Index