

# Package ‘Rassoc’

February 14, 2012

**Type** Package

**Title** Robust tests for case-control genetic association studies

**Version** 1.03

**Date** 2010-05-10

**Author**

Yong ZANG <zangyong@hku.hk>, Wingkam FUNG <wingfung@hku.hk>,Gang ZHENG <zhenng@nhlbi.nih.gov>

**Maintainer** Yong ZANG <zangyong@hku.hk>

**Depends** R (>= 2.6.0), mvtnorm

**Description** This package supplies several robust tests for case-control genetic association studies. The tests contained in this package are: allelic based test, Cochran-Armitage trend test, maximin efficiency robust test, MAX3 test and genetic model selection test. For each test, the corresponding R code reports the test statistics and the associated p-value.

**License** GPL (>= 2)

**LazyLoad** yes

**Repository** CRAN

**Date/Publication** 2010-05-10 05:36:00

## R topics documented:

Rassoc-package . . . . .	2
ABT . . . . .	2
caco . . . . .	4
CATT . . . . .	5
GMS . . . . .	6
MAX3 . . . . .	7
MERT . . . . .	9

<b>Index</b>	<b>11</b>
--------------	-----------

---

Rassoc-package

*Robust tests for case-control genetic association studies*

---

### Description

This package supplies several robust tests for case-control genetic association studies. The tests contained in this package are: allelic based test, Cochran-Armitage trend test, maximin efficiency robust test, MAX3 test and genetic model selection test. For each test, the corresponding R code reports the test statistics and the associated p-value.

### Details

Package: Rassoc  
Type: Package  
Version: 1.01  
Date: 2009-11-06  
License: GPL (>=2)  
LazyLoad: yes

### Author(s)

Yong ZANG <zanngyong@hku.hk>, Wingkam FUNG <wingfung@hku.hk>, Gang ZHENG <zhengg@nhlbi.nih.gov>  
Maintainer: Yong ZANG <zanngyong@hku.hk>

### References

Zang Y, Fung WK and Zheng G(2010). Simple algorithms to calculate the asymptotic null distributions of robust tests in case-control genetic association studies in R. Journal of Statistical software 33(8).

### See Also

[ABT,CATT,GMS,MAX3,MERT](#)

---

ABT

*The allelic based test*

---

### Description

This function conducts the allelic based test to a 2 by 3 case-control contingency table and reports the test statistics and associated p-value.

**Usage**

```
ABT(data)
```

**Arguments**

`data` data is a 2 by 3 case-control contingency table. The first and second rows represent the case group and control group respectively. The first, second and third columns represent the genotypes of a susceptibility diallelic marker containing 0, 1 and 2 risk allele respectively. Thus, the numbers in this table represent the genotype counts belonging to the corresponding genotypes and case-control status.

**Value**

`statistic` the statistic of the allelic based test  
`p.value` the associated p-value of the allelic based test

**References**

Sasieni, PD(1997). From genotypes to genes: doubling the sample size. *Biometrics* 53, 1253-1261.

Zang Y, Fung WK and Zheng G(2010). Simple algorithms to calculate the asymptotic null distributions of robust tests in case-control genetic association studies in R. *Journal of Statistical software* 33(8).

**See Also**

[caco](#)

**Examples**

```
library(Rassoc)
data(caco)
ex=matrix(caco[1,],nrow=2,byrow=TRUE)
## ex is an example of a 2 by 3 case-control contingency table.
ABT(ex)
## Conduct the allelic based test to dataset ex.
## The allelic based test
## data:  ex
## statistic = -5.4903, p-value = 4.013e-08
## The statistic of the test is -5.4903.
## The associated p-value of the test is 4.013e-08.
```

---

caco

*The case-control dataset of 17 SNPs*

---

## Description

caco is a matrix containing the case-control samples of 17 SNPs reported from four genome-wide association studies for age-related macular degeneration (AMD) (Klein et al., 2005), cancer studies (Hunter et al., 2007 and Yeager et al., 2007) and a hypertension study (WTCCC, 2007).

## Usage

```
data(caco)
```

## Details

caco is a 17 by 6 matrix. Each row represents a different SNP. The 1st, 2nd and 3rd columns represent the genotypes of case group containing 2, 1 and 0 risk alleles; the 4th, 5th and 6th columns represent the genotypes of control group containing 2, 1 and 0 risk alleles. Thus, each number in the caco matrix represents the genotype counts belonging to the corresponding SNP, genotype and case-control status.

## Source

Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager N, Hankinson SE, Wacholder S, Wang Z, Welch R, Hutchinson A, et al (2007). A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nature Genetics* 39, 870-874.

Klein RJ, Zeiss C, Chew EY, Tsai J-Y, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL et al (2005). Complement factor H polymorphism in aged-related macular degeneration. *Science* 308, 385-389.

The Wellcome Trust Case Control Consortium (WTCCC) (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447, 661-683.

Yeager M, Orr N, Hayes RB, Jacobs KB, Kraft P, Wacholder S, Minichiello MJ, Fearnhead P, Yu K, Chatterjee N et al (2007). Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nature Genetics* 39, 645-649.

## Examples

```
data(caco)
ex=matrix(caco[1,],nrow=2,byrow=TRUE)
## ex is a 2 by 3 case-control contingency table of SNP rs380390.
ABT(ex)
## Conduct the allelic based test to dataset ex.
## The allelic based test
## data:  ex
## statistic = -5.4903, p-value = 4.013e-08
## The statistic of the test is -5.4903.
## The associated p-value of the test is 4.013e-08.
```

---

CATT

*The Cochran-Armitage trend test*

---

### Description

This function conducts the Cochran-Armitage trend test to a 2 by 3 case-control contingency table and reports the test statistics, associated p-value and conclusion for the hypothesis test respectively.

### Usage

```
CATT(data, x)
```

### Arguments

data	data is a 2 by 3 case-control contingency table. The first and second rows represent the case group and control group respectively. The first, second and third columns represent the genotypes of a susceptibility diallelic marker containing 0, 1 and 2 risk allele respectively. Thus, the numbers in this table represent the genotype counts belonging to the corresponding genotypes and case-control status.
x	x is the score of the Cochran-Armitage trend test. It can be any real number between 0 and 1. Specifically, $x=0$ , 0.5 and 1 are optimal for recessive, additive/multiplicative and dominant genetic models respectively.

### Value

statistic	the statistic of the Cochran-Armitage trend test
p.value	the associated p-value of the Cochran-Armitage trend test

### References

Sasieni, PD(1997). From genotypes to genes: doubling the sample size. *Biometrics* 53, 1253-1261.

Zang Y, Fung WK and Zheng G(2010). Simple algorithms to calculate the asymptotic null distributions of robust tests in case-control genetic association studies in R. *Journal of Statistical software* 33(8).

Zheng, G, Freidlin, B, Li, Z and Gastwirth, JL (2003). Choice of scores in trend tests for case-control studies of candidate-gene associations. *Biometrical Journal* 45, 335-348.

### See Also

[caco](#)

## Examples

```

library(Rassoc)
data(caco)
ex=matrix(caco[1,],nrow=2,byrow=TRUE)
## ex is an example of a 2 by 3 case-control contingency table.
CATT(ex,0.5)
## Conduct the Cochran-Armitage trend test to dataset ex. x is chosen as 0.5, ##which is optimal under the additive/m
## The Cochran-Armitage trend test
## data:  ex
## statistic = -5.1171, p-value = 3.102e-07
## The statistic of the test is -5.1171.
## The associated p-value of the test is 3.102e-07.

```

---

GMS

*The genetic model selection test*

---

## Description

This function conducts the genetic model selection (GMS) test to a 2 by 3 case-control contingency table using either empirical or asymptotic methods. GMS is a two-phase test. In phase 1, the Hardy-weinberg disequilibrium trend test (HWDTT) is used to detect the underlying genetic model. In phase 2, an optimal Cochran-Armitage trend test corresponding to the selected genetic model is used for testing association.

## Usage

```
GMS(data, method, m)
```

## Arguments

data	data is a 2 by 3 case-control contingency table. The first and second rows represent the case group and control group respectively. The first, second and third columns represent the genotypes of a susceptibility diallelic marker containing 0, 1 and 2 risk allele respectively. Thus, the numbers in this table represent the genotype counts belonging to the corresponding genotypes and case-control status.
method	method is the approach used for calculating the p-value of GMS statistics. boot and bvn are empirical methods and asy is asymptotic method.
m	m is the replication times used for calculating the empirical p-values (boot and bvn). It can be any positive integer (1, for example) while choosing asymptotic method (asy).

## Details

boot method re-samples case-control data under null hypothesis and applies GMS to each replicate. bvn directly generates statistics of GMS rather than case-control data under null hypothesis. asy calculates the p-value from the asymptotic distribution of GMS under null hypothesis.

**Value**

method	the method chosen to calculate the p-value of GMS
statistic	the statistic of the GMS test
p.value	the associated p-value of the GMS test

**References**

Joo, J, Kwak, M and Zheng, G (2009). Improving power for testing genetic association in case-control studies by reducing alternative space. *Biometrics* (in press).

Zang Y, Fung WK and Zheng G(2010). Simple algorithms to calculate the asymptotic null distributions of robust tests in case-control genetic association studies in R. *Journal of Statistical software* 33(8).

Zheng, G and Ng HKT (2008). Genetic model selection in two-phase analysis for case-control association studies. *Biostatistics* 9, 391-399.

**See Also**

[CATT](#)

**Examples**

```
library(Rassoc)
ca=c(139,249,112)
co=c(136,244,120)
a=rbind(ca,co)

GMS(a,"boot",100000)
## The GMS test using the boot method
## data: a
## statistic = 0.4894, p-value = 0.6658

GMS(a,"bvn",100000)
## The GMS test using the bvn method
## data: a
## statistic = 0.4894, p-value = 0.6598

GMS(a,"asy",100000)
## The GMS test using the asy method
## data: a
## statistic = 0.4894, p-value = 0.6621
```

## Description

This function conducts the MAX3 test to a 2 by 3 case-control contingency table using either empirical or asymptotic methods. This test takes the maximum of the absolute values of the Cochran-Armitage trend tests optimal for recessive, additive/multiplicative and dominant models respectively.

## Usage

```
MAX3(data, method, m)
```

## Arguments

data	data is a 2 by 3 case-control contingency table. The first and second rows represent the case group and control group respectively. The first, second and third columns represent the genotypes of a susceptibility diallelic marker containing 0, 1 and 2 risk allele respectively. Thus, the numbers in this table represent the genotype counts belonging to the corresponding genotypes and case-control status.
method	method is the approach used for calculating the p-value of MAX3 statistics. boot and bvn are empirical methods and asy is asymptotic method.
m	m is the replication times used for calculating the empirical p-values (boot and bvn). It can be any positive integer (1, for example) while choosing asymptotic method (asy).

## Details

boot method re-samples case-control data under null hypothesis and applies MAX3 to each replicate. bvn directly generates statistics of MAX3 rather than case-control data under null hypothesis. asy calculates the p-value from the asymptotic distribution of MAX3 under null hypothesis.

## Value

method	the method chosen to calculate the p-value of MAX3
statistic	the statistic of the MAX3 test
p.value	the associated p-value of the MAX3 test

## References

Freidlin, B, Zheng, G, Li, Z and Gastwirth, JL (2002). Trend tests for case-control studies of genetic markers: power, sample size and robustness. *Human Heredity* 53, 146-152.

Zang Y, Fung WK and Zheng G(2010). Simple algorithms to calculate the asymptotic null distributions of robust tests in case-control genetic association studies in R. *Journal of Statistical software* 33(8).

## See Also

[CATT](#)

### Examples

```
library(Rassoc)
ca=c(139,249,112)
co=c(136,244,120)
a=rbind(ca,co)

MAX3(a,"boot",100000)
## The MAX3 test using the boot method
## data: a
## statistic = 0.5993, p-value = 0.7936

MAX3(a,"bvn",100000)
## The MAX3 test using the bvn method
## data: a
## statistic = 0.5993, p-value = 0.792

MAX3(a,"asy",1)
## The MAX3 test using the asy method
## data: a
## statistic = 0.5993, p-value = 0.7933
```

---

MERT

*The maximin efficiency robust test*

---

### Description

This function conducts the maximin efficiency robust test to a 2 by 3 case-control contingency table and reports the test statistics, associated p-value and conclusion for the hypothesis test respectively.

### Usage

```
MERT(data)
```

### Arguments

**data** data is a 2 by 3 case-control contingency table. The first and second rows represent the case group and control group respectively. The first, second and third columns represent the genotypes of a susceptibility diallelic marker containing 0, 1 and 2 risk allele respectively. Thus, the numbers in this table represent the genotype counts belonging to the corresponding genotypes and case-control status.

### Value

**statistic** the statistic of the maximin efficiency robust test  
**p.value** the associated p-value of the maximin efficiency robust test

## References

Freidlin, B, Zheng, G, Li, Z and Gastwirth, JL (2002). Trend tests for case-control studies of genetic markers: power, sample size and robustness. *Human Heredity* 53, 146-152.

Gastwirth JL (1966). On robust procedures. *Journal of American Statistical Association* 61, 929-948.

Gastwirth JL (1985). The use of maximin efficiency robust tests in combining contingency tables and survival analysis. *Journal of American Statistical Association* 80, 380-384.

Zang Y, Fung WK and Zheng G(2010). Simple algorithms to calculate the asymptotic null distributions of robust tests in case-control genetic association studies in R. *Journal of Statistical software* 33(8).

## See Also

[caco](#)

## Examples

```
library(Rassoc)
data(caco)
ex=matrix(caco[1,],nrow=2,byrow=TRUE)
## ex is an example of a 2 by 3 case-control contingency table.
MERT(ex)
## Conduct the maximin efficiency robust test to dataset ex.
## The maximin efficiency robust test
## data:  ex
## statistic = -5.0735, p-value = 3.906e-07
## The statistic of the test is -5.0735.
## The associated p-value of the test is 3.906e-07.
```

# Index

ABT, [2](#), [2](#)

caco, [3](#), [4](#), [5](#), [10](#)

CATT, [2](#), [5](#), [7](#), [8](#)

GMS, [2](#), [6](#)

MAX3, [2](#), [7](#)

MERT, [2](#), [9](#)

Rassoc (Rassoc-package), [2](#)

Rassoc-package, [2](#)