Package ‘GeneticsDesign’

April 25, 2017

Title  Functions for designing genetics studies
Version 1.44.0
Date 2010-04-15
Author Gregory Warnes <warnes@bst.rochester.edu>
David Duffy <davidD@qumr.edu.au>,
Michael Man <michael.man@pfizer.com>
Weiliang Qiu <stwxq@channing.harvard.edu>
Ross Lazarus <ross.lazarus@channing.harvard.edu>
Maintainer The R Genetics Project <r-genetics-talk@lists.sourceforge.net>
Imports gmodels, graphics, gtools (>= 2.4.0), mvtnorm, stats
Description This package contains functions useful for designing genetics
studies, including power and sample-size calculations.
biocViews Genetics
License GPL-2
NeedsCompilation no

R topics documented:

<table>
<thead>
<tr>
<th>Deprecated</th>
<th>Depreciated functions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>GeneticPower.Quantitative.Numeric</td>
<td>2</td>
</tr>
<tr>
<td>GPC</td>
<td>4</td>
</tr>
<tr>
<td>gregorius</td>
<td>5</td>
</tr>
<tr>
<td>power.genotype.conti</td>
<td>7</td>
</tr>
</tbody>
</table>

Index 9

Depreciated

These functions are deprecated.

Usage

power.casectrl(...)
Arguments

... All arguments are ignored

Details

The `power.casectl` function contained serious errors and has been replaced by `GPC.GeneticPower.Quantitative.Factor` or `GeneticPower.Quantitative.Numeric` as appropriate.

In specific, the `power.casectl` function used an expected contingency table to create the test statistic that was erroneously based on the underlying null, rather than on the marginal totals of the observed table. In addition, the modeling of dominant and recessive modes of inheritance had assumed a "perfect" genotype with no disease, whereas in reality a dominant or recessive mode of inheritance simply means that two of the genotypes will have an identical odds ratio compared to the 3rd genotype (the other homozygote).

---

**GeneticPower.Quantitative.Numeric**

*Power of Genetics Study*

---

Description

Compute power of quantitative genetics studies, when the genotype is handled as a numeric value (0,1,2) `GeneticPower.Quantitative.Numeric` or as a factor `GeneticPower.Quantitative.Factor`.

Usage

```r
GeneticPower.Quantitative.Numeric(
  N=1000,
  delta=1,
  freq=0.15,
  minh=c("additive", "dominant", "recessive"),
  sigma=1,
  OtherParms=0,
  alpha=0.05,
  numtests=1,
  moi=NULL,
  rsquared=NULL)
```

```r
GeneticPower.Quantitative.Factor(
  N=1000,
  delta=1,
  freq=0.15,
  minh=c("additive","dominant","recessive"),
  sigma=1,
  OtherParms=0,
  alpha=0.05,
  numtests=1,
  moi=NULL,
  rsquared=NULL)
```
Arguments

N: total samples in the analysis

delta: Treatment effect for an individual homozygote for the disease allele ('b') relative to an individual homozygote for the reference allele ('A')

delta: allele frequency of disease allele 'b'

minh: mode of inheritance: "additive","dominant","recessive". Default is "additive". This parameter is OVER-RIDDEN by moi.

sigma: standard deviation of the response phenotype

OtherParms: number of additional parameters (really, DOF) in the model that will reduce your overall DOF

alpha: desired significance level

numtests: number of tests to be corrected by Bonferroni adjustment before achieving 'alpha'

moi: continuous value between 0 and 1 (inclusive) specifying the mode of inheritance: 0 for recessive, 0.5 for additive, 1.0 for dominant. This parameter OVER-RIDES minh.

rsquared: fraction of total sum-of-squares explained by fit. This parameter OVER-RIDES delta AND sigma.

Details

The value of moi overrides any value specified for minh. Specifying a minh="recessive" is equivalent to specifying moi=0, minh="additive" is equivalent to moi=0.5, and minh="dominant" is equivalent to moi=1.0.

Author(s)

Craig L. Hyde <Craig.L.Hyde@pfizer.com> and Feng Gao <feng.gao1@pfizer.com>

Examples

GeneticPower.Quantitative.Numeric(
    N=50,
    freq=0.1,
    minh="recessive",
    alpha=0.05
)

GeneticPower.Quantitative.Factor(
    N=50,
    freq=0.1,
    minh="recessive",
    alpha=0.05
)

##
GPC

Genetics power calculator for linear trend association studies

Description

Genetics power calculator for linear trend association studies.

Usage

GPC(pA, pD, RRAa, RRAA, r2, pB,  
nCase=500, ratio=1, alpha=0.05, quiet=FALSE)
GPC.default(pA, pD, RRAa, RRAA, Dprime, pB,  
nCase=500, ratio=1, alpha=0.05, quiet=FALSE)

Arguments

pA High risk allele frequency (A).
pD Disease prevalence.
RRAa Genotype relative risk (Aa) = RR(Aa|aa)=Pr(D|Aa)/Pr(D|aa).
RRAA Genotype relative risk (AA) = RR(AA|aa)=Pr(D|AA)/Pr(D|aa).
r2 LD measure. Assume that D > 0.
Dprime LD measure.
pB Marker allele frequency (B).
nCase Number of cases.
ratio Control:case ratio = nControl/nCase.
alpha User-defined type I error rate.
quiet Print some intermediate results if quiet=FALSE.

Details

The power is for the test that disease is associated with a marker, given high risk allele frequency (A), disease prevalence, genotype relative risk (Aa), genotype relative risk (AA), LD measure (D’ or \( r^2 \)), marker allele frequency (B), number of cases, control:case ratio, and probability of the Type I error. The linear trend test (Cochran 1954; Armitage 1955) is used.

Value

power The estimated power for the association test.
ncp Non-centrality parameter.
mat.para A matrix of case-control parameters, including number of cases, number of controls, high risk allele frequency, prevalence, genotypic relative risk (Aa), genotypic relative risk (AA), genotypic risk for aa (baseline).
mat.B A matrix of marker locus B parameters, including marker allele frequency, linkage disequilibrium (D’), penetrance at marker genotype bb, penetrance at marker genotype Bb, penetrance at marker genotype BB, genotypic odds ratio Bb, genotypic odds ratio BB.
mag.fFreq A 2 by 2 matrix of expected allele frequencies Pr(B|D), Pr(b|D), Pr(B|non D), Pr(b|non D).
mag.gFreq A 3 by 2 matrix of expected genotype frequencies Pr(BB|D), Pr(Bb|D), Pr(bb|D), Pr(BB|non D), Pr(Bb|non D), Pr(bb|non D).
mag.stat Power estimates for a sequence of Type I errors.
**gregorius**

**Author(s)**

Weiliang Qiu <stwxq@channing.harvard.edu>, Ross Lazarus <ross.lazarus@channing.harvard.edu>

**References**


**Examples**

```r
res1<-GPC(pA=0.05, pD=0.1, RRAa=1.414, RRAA=2, r2=0.9, pB=0.06, nCase=500, ratio=1, alpha=0.05, quiet=FALSE)
res2<-GPC.default(pA=0.05, pD=0.1, RRAa=1.414, RRAA=2, Dprime=0.9, pB=0.06, nCase=500, ratio=1, alpha=0.05, quiet=FALSE)
```

**Description**

Probability of observing all alleles with a given frequency in a sample of a specified size.

**Usage**

```r
gregorius(freq, N, missprob, tol = 1e-10, maxN = 10000, maxiter=100, showiter = FALSE)
```

**Arguments**

- `freq` (Minimum) Allele frequency (required)
- `N` Number of sampled genotypes
- `missprob` Desired maximum probability of failing to observe an allele.
- `tol` Omit computation for terms which contribute less than this value.
- `maxN` Largest value to consider when searching for N.
- `maxiter` Maximum number of iterations to use when searching for N.
- `showiter` Boolean flag indicating whether to show the iterations performed when searching for N.
Details

If \( \text{freq} \) and \( N \) are provided, but \( \text{missprob} \) is omitted, this function computes the probability of failing to observe all alleles with true underlying frequency \( \text{freq} \) when \( N \) diploid genotypes are sampled. This is accomplished using the sum provided in Corollary 2 of Gregorius (1980), omitting terms which contribute less than \( \text{tol} \) to the result.

When \( \text{freq} \) and \( \text{missprob} \) are provide, but \( N \) is omitted. A binary search on the range of \([1, \text{maxN}]\) is performed to locate the smallest sample size, \( N \), for which the probability of failing to observe all alleles with true underlying frequency \( \text{freq} \) is at most \( \text{missprob} \). In this case, \( \text{maxiter} \) specifies the largest number of iterations to use in the binary search, and \( \text{showiter} \) controls whether the iterations of the search are displayed.

Value

A list containing the following values:

- \text{call} Function call used to generate this object.
- \text{method} One of the strings, "Compute missprob given \( N \) and \( \text{freq} \)", or "Determine minimal \( N \) given \( \text{missprob} \) and \( \text{freq} \)", indicating which type of computation was performed.
- \text{retval.freq} Specified allele frequency.
- \text{retval.N} Specified or computed sample size.
- \text{retval.missprob} Computed probability of failing to observe all of the alleles with frequency \( \text{freq} \).

Note

This code produces sample sizes that are slightly larger than those given in table 1 of Gregorius (1980). This appears to be due to rounding of the computed \text{missprobs} by the authors of that paper.

Author(s)

Code submitted by David Duffy <davidD@qumr.edu.au>, substantially enhanced by Gregory R. Warnes <warnes@bst.rochester.edu>.

References


Examples

```r
# Compute the probability of missing an allele with frequency 0.15 when
# 20 genotypes are sampled:
gregorius(freq=0.15, N=20)

# Determine what sample size is required to observe all alleles with true
# frequency 0.15 with probability 0.95
# gregorius(freq=0.15, missprob=1-0.95)
```
**power.genotype.conti**

**Description**

Estimate power for genetic studies using baseline measurements via simulation.

**Usage**

```r
power.genotype.conti(N, Rep = 2000, alpha = 0.05, ...)  
simu.genotype.conti(N, p=0.15, pi=0, me1=50, me2=me1, delta=-5,  
sd1=10, sd2=10, verbose=FALSE,  
minh=c('additive', 'dominant', 'recessive'),  
genotype.delta=TRUE, Factor=FALSE)
```

**Arguments**

- **N**  
  total number of subjects
- **p**  
  frequency of A (affected) allele
- **Rep**  
  number of simulation runs used to estimate power
- **alpha**  
  significance level
- **pi**  
  correlation coefficient
- **me1, me2**  
  mean of control and treatment groups
- **delta**  
  treatment/genotype effect
- **sd1, sd2**  
  standard deviation of the control and treatment groups
- **minh**  
  mode of inheritance, one of 'additive', 'dominant', or 'recessive'
- **genotype.delta**  
  logical indicating whether the treatment effect occurs only for an individual genotype (genotype.delta=TRUE) or for all genotypes (genotype.delta=FALSE)
- **Factor**  
  Should the simulated treatment variable 'Trt' be treated as a factor variable (Factor=TRUE) or as a numeric variable (Factor=FALSE).
- **verbose**  
  Should information about each simulated data set and model fit be displayed.
- **...**  
  Arguments to be passed to simu.genotype.conti

**Value**

- Describe the value returned If it is a LIST, use
  - comp1 Description of 'comp1'
  - comp2 Description of 'comp2'
- ...

**Author(s)**

Michael Man, minor changes by Gregory R. Warnes <greg@random-technologies-llc.com>
References


See Also

power.casectrl

Examples

## Not run:
    # use defaults, 100 subjects
power.genotype.conti(N=100)

    # same calculation, specifying all values
power.genotype.conti(N=100, Rep=2000, p=0.15, pi=0, me1=50, me2=50, delta=-5,
    sd1=10, sd2=10, verbose=FALSE, minh="additive",
genotype.delta=TRUE, Factor=FALSE)

    # Show details for small simulation study
power.genotype.conti(N=10, verbose=TRUE)

## End(Not run)
Index

*Topic **design**
  GeneticPower.Quantitative.Numeric, 2
  power.genotype.conti, 7
*Topic **htest**
  GPC, 4
*Topic **misc**
  Deprecated, 1
  gregorius, 5

Deprecated, 1

GeneticPower.Quantitative.Factor, 2
GeneticPower.Quantitative.Factor
  (GeneticPower.Quantitative.Numeric), 2
GeneticPower.Quantitative.Numeric, 2, 2
GPC, 2, 4
gregorius, 5

power.casectrl, 8
power.casectrl (Deprecated), 1
power.genotype.conti, 7

simu.genotype.conti
  (power.genotype.conti), 7