Package 'GeneticsDesign'

April 11, 2018

Title Functions for designing genetics studies

Version 1.46.0

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:

`Depreciated` *Depreciated functions*

Description

These functions are depreciated.

Usage

```
power.casectrl(...)
```

Arguments

...	All arguments are ignored

Details

The power.casectl function contained serious errors and has been replaced by GPC, GeneticPower.Quantitative.Fac or GeneticPower.Quantitative.Numeric as appropriate.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

```
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)
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')
freq	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 beforee 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.
mat.aFreq	A 2 by 2 matrix of expected allele frequencies Pr(B|D), Pr(b|D), Pr(B|non D), Pr(b|non D).
mat.gFreq	A 3 by 2 matrix of expected genotype frequencies Pr(BB|D), Pr(Bb|D), Pr(bb|D), Pr(BB|non D)
mat.stat	Power estimates for a sequence of Type I errors.

Author(s)

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

References

Armitage, P. (1955) Tests for linear trends in proportions and frequencies. *Biometrics*, 11, 375-386.

Cochran, W.G. (1954) Some methods for strengthening the common chi-squared tests. *Biometrics*, 10, 417-451.

Gordon D, Finch SJ, Nothnagel M, Ott J (2002) Power and sample size calculations for case-control genetic association tests when errors are present: application to single nucleotide polymorphisms. *Hum. Hered.*, 54:22-33.

Gordon D, Haynes C, Blumenfeld J, Finch SJ (2005) PAWE-3D: visualizing Power for Association With Error in case/control genetic studies of complex traits. *Bioinformatics*, 21:3935-3937.

Purcell S, Cherny SS, Sham PC. (2003). Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics*, 19(1):149-150.

Sham P. (1998). *Statistics in Human Genetics.* Arnold Applications of Statistics.

Examples

```
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)
```

---

gregorius — *Probability of Observing All Alleles with a Given Frequency in a Sample of a Specified Size.*

Description

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

Usage

```
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 freq and N are provided, but missprob is omitted, this function computes the probability of failing to observe all alleles with true underlying frequency 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 tol to the result.

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

Value

A list containing the following values:

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

Gregorius, H.R. 1980. The probability of losing an allele when diploid genotypes are sampled. Biometrics 36, 643-652.

Examples

```
# 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 *power for genetic studies using baseline measure*

Description

Estimate power for genetice studies using baseline measurements via simulation.

Usage

```
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 simulatin 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 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

Frison and Pocock (1992) "Repeated measures in clinical trials: analysis using mean summary statistics and its implications for design" Statistics in Medicine 11:1685-1704

Vickers (2001) "The use of percentage change from baseline as an outcome in a controlled trial is statistically inefficient: a simulation study" BMC Med Res Methodol. 2001; 1 (1): 6

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