Description

These functions provide code for genetic epidemiology methods developed at MSKCC. They currently include estimating haplotype disease risk and two stage designs for GWAS.

Details
Package: genepi
Type: Package
Version: 1.0
Date: 2010-09-17
License: GPL version 2 or later
LazyLoad: yes

~~ An overview of how to use the package, including the most important ~~ ~~ functions ~~

Author(s)

Venkatraman E. Seshan
Maintainer: Venkatraman E. Seshan <seshanv@mskcc.org>

References


===

**haplotypeOddsRatio**

*Calculate haplotype disease risk.*

Description

Haplotype disease risk is calculated resolving haplotype ambiguity and adjusting for covariates and population stratification.

Usage

```
haplotypeOddsRatio(formula, gtypevar, data, stratvar=NULL, nsim=100, tol=1e-8)
```

# S3 method for class 'haploOR'
print(x, ...)

Arguments

- `formula` The formula for logistic regression without the haplotype variable.
- `gtypevar` The variable names in the data frame corresponding to the loci of interest. Each variable counts the number of mutant genotypes a subject has at that locus. Legal values are 0, 1, 2 & NA.
- `data` The name of the dataframe being analyzed. It should have all the variables in the formula as well as those in genotype and stratvar.
- `stratvar` Name of the stratification variable. This is used to account for population stratification. The haplotype frequencies are estimated within each stratum.
haplotypeOddsRatio

nsim

Variance should be inflated to account for inferred ambiguous haplotypes. The estimates are recalculated by simulating the disease haplotype copy number and variance added to average.

tol

Tolerance limit for the EM algorithm convergence.

x

Object of class haploOR.

Details

This implements the method in the reference below.

Value

It is a list of class haploOR

call

The function call that produced this output.

coef

Table with estimated coefficients, standard error, Z-statistic and p-value.

var

Covariance matrix of the estimated log odds-ratios.

deviance

Average of the simulated deviances. Its theoretical properties are unknown.

aic

Average of the simulated aic.

null.deviance

Deviance of null model.

df.null

Degrees of freedom of null model.

df.residual

Degrees of freedom of full model.

The "print" method formats the results into a user-friendly table.

Author(s)

Venkatraman E. Seshan

References


Examples

```R
# simulated data with 2 loci haplotypes 1=00, 2=01, 3=10, 4=11
# control haplotype probabilities p[i] i=1,2,3,4
# haplotype pairs (i<=j) i=j: probs = p[i]^2 ; i<j: p[i]*p[j]
# p <- c(0.25, 0.2, 0.2, 0.35)
p0 <- rep(0, 10)
l <- 0
for(i in 1:4) {for(j in i:4) {l <- l+1; p0[i] <- 2*p[i]*p[j]/(1+1*(i==j))}}
controls <- as.numeric(cut(runif(1000), cumsum(c(0,p0)), labels=1:10))
# case probabilities disease haplotype is 11
or <- c(2, 5)
```
twostagepower

Calculate the power of a two stage design for GWAS

description

Calculate the power of a two stage design for GWAS under sample size or cost constraints. Implements methods in the references below.

usage

twostagepower(n=NULL, Cost=NULL, m=5000, mu=0.045, mu.loc=0.5, p=0.10, f=0.5, relcost=1, true.needed=1, rho=0, rho0=0, nsim=2000)

arguments

n The maximum sample size for the study.
Cost Maximum available resource. One of Cost or n must be specified.
m The number of markers in the study. Default is 5000. It will take a long time to compute power for very large numbers e.g. 100000
mu The mean vector for the markers that are associated with endpoint.
mu.loc The locations of the true markers. Since the chromosome is mapped to the unit interval (0,1) the numbers should be between 0 and 1.
p The proportion of markers taken to the second stage. The default is 0.1 which is found to be optimal.
f The fraction of Cost or sample size allocated to the first stage. If not specified it uses 0.75 for the Cost constraint scenario and 0.5 for the sample size constraint scenario.
relcost Specifies how expensive it is to genotype in the second stage compared to the first stage.
true.needed The number of markers selected in the end. Can be a maximum of length of mu.loc (or mu).
rho, rho0 correlation between markers
nsim Number of Monte Carlo replications to compute power.
twoStagePower

Details

This implements the method in the reference below.

Value

It returns the power as a single numeric value

Author(s)

Jaya M. Satagopan & Venkatraman E. Seshan

References


Examples

twoStagePower(n=1000)
twoStagePower(Cost=1000)
Index

*Topic design
twoStagePower, 4
*Topic odds-ratio
haplotypeOddsRatio, 2
*Topic package
genepi-package, 1
*Topic regression
haplotypeOddsRatio, 2

genepi (genepi-package), 1
genepi-package, 1

haplotypeOddsRatio, 2

print.haploOR (haplotypeOddsRatio), 2

twoStagePower, 4