Package ‘GLIDE’

October 12, 2022

Type Package

Title Global and Individual Tests for Direct Effects

Version 1.0.5

Date 2022-05-20

Author James Dai [aut, cre],
Xiaoyu Wang [aut]

Maintainer James Dai `<jdai@fredhutch.org>`

Description Global and individual tests for pleiotropy and direct effects in Mendelian randomization studies. Refer to J. Y. Dai, U. Peters, X. Wang, J. Kocarnik et al. AJE (2018) <doi:10.1093/aje/kwy177>.

Depends R (>= 2.10)

Imports MASS,foreach,parallel,doParallel

LazyLoad no

License GPL (>= 2)

NeedsCompilation yes

Repository CRAN

Suggests rmarkdown, knitr

VignetteBuilder knitr

Date/Publication 2022-05-25 23:00:02 UTC

R topics documented:

`checkdata` *A function used to check the core inputs of function glide*

Description

It is used to check the core inputs of function glide

Usage

```
checkdata(formula,exposure_coeff,genotype_columns,data)
```

Arguments

`formula`	A an object of class "formula": outcome ~ adjusting covariates.
`exposure_coeff`	A named mumeric vector or a one-column dataframe for the external coefficients for genotype-exposure associations.
`genotype_columns`	An integer vector, column indices of data in which genotype data are stored.
`data`	A data frame composed of formula-related components and snp genotype columns.

Value

No return value, would generate error messages if input data are not right.

Author(s)

James Y. Dai and X. Wang

Examples

```
#load an example data; it includes a dataframe "simdat" and a one-column dataframe  "coeff"
data(simdata)
simdat=simdata$simdat
coeff=simdata$coeff
formula=as.formula("outcome~age+sex+pc1+pc2+pc3")
genotype_columns=which(grepl("^SNP",colnames(simdat)))
checkdata(formula=formula,exposure_coeff=coeff,genotype_columns,data=simdat)
```

`glide`

Evaluate individual and global pleiotropy among genotypes being used as instrumental variables in Mendelian randomization studies.

Description

Glide is a function to perform global and individual tests for assessing direct effects in Mendelian randomization studies.

Usage

```
glide(formula,exposure_coeff=NULL,genotype_columns=NULL,data,
np=100000,qcutoff=0.2,parallel=TRUE,corenumber=1,verbose=TRUE)
```

Arguments

`formula`	An object of class "formula": outcome ~ adjusting covariates, specifying the logistic regression of a binary study outcome with respect to covariates being adjusted for in addition to genotype data.
`exposure_coeff`	A named numeric vector or a one-column dataframe for the external regression coefficients for genotype-exposure associations. These coefficients are typically derived from previous genome-wide association studies. All the names or rownames must be found in "genotype_columns" (specified below) columns of the input data.
`genotype_columns`	An integer vector that contains column indices of data columns in which genotype data are stored.
`data`	A data frame composed of formula-related variables and SNP genotype columns.
`np`	An integer for the number of observations in the parametric simulation being generated to compute expected p-values and q-values.
`qcutoff`	The q-value cutoff to declare significance of the direct effects of individual variants. The default value is set to be 0.2.
`parallel`	If TRUE, use parallel foreach to speed up computation.
`corenumber`	An integer to set the number of computer cores used in parallel computation. If the requested number of cores is greater than number of cores available, GLIDE will user the latter number.
`verbose`	If TRUE, print on screen the computing time spent in each step of GLIDE.

Details

This function fits a logistic regression for each SNP being evaluated for instrumental variables, regressing outcome on one genotype at a time and adjusting covariates as specified in the formula (outcome ~ adjusting covariates). The p-values for the genotypes derived from these regression models were compared to the distribution of simulation-based null p-values to declare significance of individual variant and global pleiotropy. See Dai et al. (2017) for details of the algorithm.

Value

A data frame containing observed p-value, expected null p-value, family wise error rate (FWER), q-value, genetic association with exposure, and variance of the estimated association of each SNP.

Author(s)

James Y. Dai and X. Wang

References

J. Y. Dai, U. Peters, X. Wang, J. Kocarnik, J. et al. Diagnostics of pleiotropy in Mendelian randomization studies: Global and individual tests for direct effects. Am J Epidemiol 187(12):2672-80, 2018.

Examples

```
##load an example data.
data(simdata)
#The example dataset is a list composed of two dataframes.
#simat stores 20,000 observations of 81 variables, including outcome, 5 ajusting covariates,
#and 75 SNPs.
simdat=simdata$simdat
#coeff stores the 75 external regression cofficients.
coeff=simdata$coeff
#We define the regression formula for outcome and adjusting covariates
formula=as.formula("outcome~age+sex+pc1+pc2+pc3")
#We next define the columns in simdat that contain genotype data
genotype_columns=which(grepl("^SNP",colnames(simdat)))


##run glide
out=glide(formula=formula,exposure_coeff=coeff,genotype_columns,data=simdat,np=100000,
qcutoff=0.2,parallel=TRUE,corenumber=1,verbose=TRUE)

##use more cores
out=glide(formula=formula,exposure_coeff=coeff,genotype_columns,data=simdat,np=100000,
qcutoff=0.2,parallel=TRUE,corenumber=2,verbose=TRUE)
```

`plot.egger` *A function used to draw Egger plot*

Description

It is used to draw MR-Egger regression plot.

Usage

```
## S3 method for class 'egger'
plot(x,exposure_coeff,qcutoff=0.2,xlab="Genetic association with the exposure",
                       ylab="Genetic association with the outcome",...)
```

Arguments

`x`	x is a dataframe, it is the outcome of the glide function.
`exposure_coeff`	A named numeric vector or a one-column dataframe for the external regression coefficients for genotype-exposure associations. These coefficients are typically derived from previous genome-wide association studies. All the names or rownames must be found in "genotype_columns" (specified below) columns of the input data.
`qcutoff`	qcutoff is the q-value cutoff.
`xlab`	xlab is the default x label.
`ylab`	ylab is the default y label.
`...`	the rest of plot arguments.

Value

No return value, the Egger plot will be generated.

Author(s)

James Y. Dai and X. Wang

Examples

```
##first generate output using the glide function
#load an example data
data(simdata)
simdat=simdata$simdat
coeff=simdata$coeff

formula=as.formula("outcome~age+sex+pc1+pc2+pc3")
genotype_columns=which(grepl("^SNP",colnames(simdat)))

#run glide
out=glide(formula=formula,exposure_coeff=coeff,genotype_columns,data=simdat,np=100000,
qcutoff=0.2,parallel=TRUE,corenumber=1,verbose=TRUE)

##draw the Egger plot
plot.egger(out,exposure_coeff=coeff)
```

`plot.glide` *A function used to draw the q-q plot*

Description

It is used to draw the q-q plot

Usage

```
## S3 method for class 'glide'
plot(x,qcutoff=0.2,xlab="Expected null p-values (log base 10)",
                        ylab="Observed p-values (log base 10)",...)
```

Arguments

`x`	x is a dataframe, it is the outcome of the glide function.
`qcutoff`	qcutoff is the q-value cutoff.
`xlab`	xlab is the default x label.
`ylab`	ylab is the default y label.
`...`	the rest plot arguments.

Value

No return value, q-q plot will be generated.

Author(s)

James Y. Dai and X. Wang

Examples

```
##first generate output using the glide function
#load an example data
data(simdata)
simdat=simdata$simdat
coeff=simdata$coeff

formula=as.formula("outcome~age+sex+pc1+pc2+pc3")
genotype_columns=which(grepl("^SNP",colnames(simdat)))

#run glide
out=glide(formula=formula,exposure_coeff=coeff,genotype_columns,data=simdat,np=100000,
qcutoff=0.2,parallel=TRUE,corenumber=1,verbose=TRUE)

##draw the plot
plot.glide(out)
```

`remove_missingdata` *A function used to remove missing data*

Description

It is used to remove samples which have NA/missing data in covariates.

Usage

```
remove_missingdata(data)
```

Arguments

`data`	data is a dataframe.

Value

row index of the input data which don't have missing values.

`simdata` *An example dataset to demonstrate the usage of GLIDE*

Description

An example simulation dataset used to demonstrate the usage of GLIDE. It is a list composed of a dataframe "simdat" and a dataframe "coeff"

Usage

```
data("simdata")
```

Format

coeff is a one-column dataframe. It consists of exposure coefficients of genotypes of 75 SNPs. simdat is a data frame consisting of 20000 observations. It has 81 columns, with 75 columns for genotypes of 75 SNPs and 6 columns of following covariates:

`age`	an integer vector to indicate age
`sex`	a factor vector to indicate gender
`pc1`	a numeric vector of the first principal component of genotypes
`pc2`	a numeric vector of the second principal component of genotypes
`pc3`	a numeric vector of the third principal component of genotypes
`outcome`	a factor vector to indicate the outcome, 1/0

Examples

```
data("simdata")
ls()
```

Index