Package: MRAPSS (via r-universe)

September 24, 2024

Title The MRAPSS package implement the MR-APPSS approach to test for the causal effect of an exposure on a outcome disease.

Version 0.0.0.9000

Description The MRAPSS package implement the MR-APPSS approach to test for the causal effects between an exposure and a outcome disease. The MR-APPSS is a unified approach to Mendelian Randomization accounting for polygenicity, pleiotropy and sample structure using genome-wide summary statistics. Specifically, MR-APPSS uses a background-foreground model to characterize both SNP-exposure effects and SNP-outcome effects estimates, where the background model accounts for confounding from genetic correlation and sample structure and the foreground model captures the valid signal for causal inference.

License What license it uses

Encoding UTF-8

LazyData true

Imports ggplot2, ggnewscale, dplyr, magrittr, ieugwasr (>= 0.1.4), readr, expm

Remotes mrcieu/ieugwasr

RoxygenNote 7.0.2

Depends R (>= 2.10)

Suggests knitr, rmarkdown

VignetteBuilder knitr

Repository https://mrcieu.r-universe.dev

RemoteUrl https://github.com/YangLabHKUST/MR-APSS

RemoteRef HEAD

RemoteSha ad5827e4f121f26a0ec63874c2f2b395ff7debf9

Contents

`clump` *Perform LD clumping*

Description

Peform LD clumping, to prune SNPs in LD within a window. Keep the most significant ones.

Usage

```
clump(
  dat,
  IV.Threshold = 5e-05,
  SNP_col = "SNP",
  pval_col = "pval.exp",
  clump_kb = 1000,
  clump_r2 = 0.001,
  clump_p = 0.999,
  pop = "EUR",
  bfile = NULL,
  plink_bin = NULL
)
```

Arguments

`dat`	a data frame must have columns with information about SNPs and p values
`SNP_col`	column with SNP rsid. The default is ‘"SNP"‘
`pval_col`	column with p value. The default is ‘"pval"‘
`clump_kb`	clumping window in kb. Default is 1000.
`clump_r2`	clumping r2 threshold. Default is 0.001.
`clump_p`	clumping significance level for index variants. Default = 5e-05
`bfile`	bfile as LD reference panel. If this is provided, then will use local PLINK. Default = NULL.
`plink_bin`	path to local plink binary. Default = NULL.

Value

data frame of clumped SNPs

`est_paras` *A function harmonising datasets and estimate background parameters by LD score regression.*

Description

A function harmonising datasets and estimate background parameters by LD score regression.

Usage

```
est_paras(
  dat1,
  dat2,
  trait1.name = "exposure",
  trait2.name = "outcome",
  LDSC = T,
  h2.fix.intercept = F,
  ldscore.dir = NULLL
)
```

Arguments

`dat1:`	formmated summary statistics for trait 1.
`dat2:`	formmated summary statistics for trait 2.
`trait1.name:`	specify the name of trait 1, default 'exposure'.
`trait2.name:`	specify the name of trait 2, default 'outcome'.
`LDSC:`	whether to run LD score regression, default 'TRUE'. If 'FALSE', the function will not give the parameter estimates but will do harmonising.
`h2.fix.intercept:`	whether to fix LD score regression intercept to 1, default 'FALSE'.
`ldscore.dir:`	specify the path to the LD score files.

Value

List with the following elements:

Mdat Homonised data set

C the estimated C matrix capturing the effects of sample structure

Omega the estimated variance-covariance matrix for polygenic effects

format_data *Format GWAS summary data.*

Description

Reads in GWAS summary data. Infer Zscores from p-values and signed satatistics. This function is adapted from the format_data() function in MRCIEU/TwoSampleMR.

Usage

```
format_data(
  dat,
  snps.merge = w_hm3.snplist,
  snps.remove = MHC.SNPs,
  snp_col = "SNP",
  b_col = "b",
  or_col = "or",
  se_col = "se",
  freq_col = "freq",
  A1_col = "A1",
  A2_col = "A2",
  p_col = "p",
  ncase_col = "ncase",
  ncontrol_col = "ncontrol",
  n_col = "n",
  n = NULL,
  z_col = "z",
  info_col = "INFO",
  log_pval = FALSE,
  chi2_max = NULL,
  min_freq = 0.05
)
```

Arguments

dat	Data frame. Must have header with at least SNP A1 A2 signed statistics pvalue and sample size.
snps.merge	Data frame with SNPs to extract. must have headers: SNP A1 and A2. For example, the hapmap3 SNPlist.
snps.remove	a set of SNPs needed to be removed. For example, the SNPs in MHC region.
snp_col	column with SNP rs IDs. The default is SNP.
b_col	Name of column with effect sizes. The default is b.
se_col	Name of column with standard errors. The default is se.
freq_col	Name of column with effect allele frequency. The default is frew.

A1_col	Name of column with effect allele. Must contain only the characters "A", "C", "T" or "G". The default is A1.
A2_col	Name of column with non effect allele. Must contain only the characters "A", "C", "T" or "G". The default is A2.
p_col	Name of column with p-value. The default is p.
ncase_col	Name of column with number of cases. The default is ncase.
ncontrol_col	Name of column with number of controls. The default is ncontrol.
n_col	Name of column with sample size. The default is n.
n	Sample size
z_col	Name of column with Zscore. The default is z.
info_col	Name of column with inputation Info. The default is info.
log_pval	The pval is -log10(p_col). The default is FALSE.
chi2_max	SNPs with tested chi^2 statistics large than chi2_max will be removed.The default is 80
min_freq	SNPs with allele frequecy less than min_freq will be removed.The default is 0.05
or_col:	Name of column with odds ratio. The default is or.
n_qc	Whether to remove SNPs according to the sample size of SNPs. The default is FALSE.

Value

data frame wih headers: SNP: rsid; A1: effect allele; A2: non effect allel; Z: Z score; N: sample size; chi2: chi square statistics; P: p-value.

MRAPSS *A function for implementing MR-APSS.*

Description

MR-APSS: a unified approach to Mendelian Randomization accounting for pleiotropy and sample structure using genome-wide summary statistics. MA-APSS uses a variantional EM algorithm for estimation of parameters. MR-APSS uses likelihood ratio test for inference.

Usage

```
MRAPSS(
  MRdat = NULL,
  exposure = "exposure",
  outcome = "outcome",
  pi0 = NULL,
  sigma.sq = NULL,
  tau.sq = NULL,
```

```
  C = matrix(c(1, 0, 0, 1), 2, 2),
  Omega = matrix(0, 2, 2),
  Cor.SelectionBias = T,
  tol = 1e-08,
  ELBO = F
)
```

Arguments

MRdat	data frame at least contain the following varaibles: b.exp b.out se.exp se.out L2 Threshold. L2:LD score, Threshold: modified IV selection threshold for correction of selection bias
exposure	exposure name
outcome	outcome name
pi0	initial value for pi0, default 'NULL' will use the default initialize procedure.
sigma.sq	initial value for sigma.sq , default 'NULL'will use the default initialize procedure.
tau.sq	initial value for tau.sq , default 'NULL' will use the default initialize procedure.
C	the estimated C matrix capturing the effects of sample structure. default 'diag(2)'.
Omega	the estimated variance-covariance matrix of polygenic effects. default 'matrix(0,2,2)'.
Cor.SelectionBias	Whether use the selection Threshold for correction of selection bias. If FALSE, the model won't correct for selection bias.
tol	tolerence, default '1e-08'
ELBO	Whether check the evidence lower bound or not, if 'FALSE', check the maximum likelihood instead. default 'FALSE'.

Value

a list with the following elements:

MRdat: Input data frame

exposure: exposure of interest

outcome: outcome of interest

beta: causal effect estimate

beta.se: standard error

pval: p-value

sigma.sq: variance of forground exposure effect

tau.sq: variance of forground outcome effect

pi0: The probability of a SNP with forground signal after selection

post: Posterior estimates of latent varaibles

method: "MR-APSS"

Examples

```
library(MRAPSS)
exposure = "BMI"
outcome = "T2D"
Threshold = 5e-05  # IV selection Threshold
data(C)
data(Omega)
data(MRdat)
MRres = MRAPSS(MRdat,
               exposure = "BMI",
               outcome = "T2D",
               C = C,
               Omega =  Omega ,
               Cor.SelectionBias = T)
MRplot(MRres, exposure = "BMI", outcome = "T2D")
```

`MRplot` *Visualize the MRAPSS results*

Description

Visualize the MRAPSS results

Usage

```
MRplot(MRres, exposure = "trait 1", outcome = "trait 2")
```

Arguments

`outcome`	: outcome name
`MRres:`	MRAPSS fit results
`exposure:`	exposure name

Value

Plot of SNP-exposure effect and SNP-outcome effect with the causal effect and 95% confidence interval.

Index