Package: MendelianRandomization (via r-universe)

October 28, 2026

Type Package

Title Mendelian Randomization Package

Version 0.10.0

Date 2024-04-12

Maintainer Stephen Burgess <sb452@medschl.cam.ac.uk>

Description Encodes several methods for performing Mendelian randomization analyses with summarized data. Summarized data on genetic associations with the exposure and with the outcome can be obtained from large consortia. These data can be used for obtaining causal estimates using instrumental variable methods.

License GPL-2 | GPL-3

Depends R (>= 3.0.1), methods

Imports knitr, rmarkdown, plotly (>= 3.6.0), ggplot2 (>= 1.0.1), robustbase (>= 0.92-6), Matrix (>= 1.2), iterpc (>= 0.3), quantreg (>= 5.01), rjson, glmnet, numDeriv, Rcpp

LinkingTo Rcpp, RcppArmadillo

VignetteBuilder knitr

RoxygenNote 7.3.1

NeedsCompilation yes

Author Stephen Burgess [aut, cre] (<https://orcid.org/0000-0001-5365-8760>), Olena Yavorska [aut], James Staley [ctb] (<https://orcid.org/0009-0001-9520-5011>), Fernando Hartwig [ctb] (<https://orcid.org/0000-0003-3729-0710>), Jim Broadbent [ctb] (<https://orcid.org/0000-0002-5685-6241>), Christopher Foley [ctb] (<https://orcid.org/0000-0002-0970-2610>), Andrew Grant [ctb], Amy Mason [ctb] (<https://orcid.org/0000-0002-8019-0777>), Ting Ye [ctb] (<https://orcid.org/0000-0001-6009-641X>), Haoran Xue [ctb] (<https://orcid.org/0000-0001-7923-6128>), Zhaotong Lin [ctb] (<https://orcid.org/0000-0001-8723-4392>), Siqi Xu [ctb],

Ashish Patel [ctb] (<https://orcid.org/0000-0002-0350-4365>),
Hyunseung Kang [ctb], Sheng Wang [ctb], Ville Karhunen [ctb]
(<https://orcid.org/0000-0001-6064-1588>)

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

Date/Publication 2024-04-13 02:27:20 UTC

RemoteUrl https://github.com/cran/MendelianRandomization

RemoteRef HEAD

RemoteSha 2e6f8aa6db24adc8a990448da898aba0f7819ca6

Contents

calcium . . . 3
CLR-class . . . 4
cML_estimate . . . 5
cML_estimate_random . . . 6
cML_SdTheta . . . 7
DIVW-class . . . 8
Egger-class . . . 8
invcov_mvmr . . . 10
IVW-class . . . 10
ldlc . . . 11
MaxLik-class . . . 12
mr_allmethods . . . 13
mr_clr . . . 14
mr_cML . . . 16
mr_conmix . . . 18
mr_divw . . . 21
mr_egger . . . 22
mr_forest . . . 25
mr_funnel . . . 27
mr_hetpen . . . 27
mr_input . . . 30
mr_ivw . . . 31
mr_lasso . . . 35
mr_loo . . . 37
mr_maxlik . . . 38
mr_mbe . . . 40
mr_median . . . 43
mr_mvcML . . . 45
mr_mvegger . . . 48
mr_mvgmm . . . 50
mr_mvinput . . . 52
mr_mvivw . . . 54
mr_mvivwme . . . 57
mr_mvlasso . . . 59
mr_mvmedian . . . 61

mr_mvpcgmm 63
mr_pcgmm 66
mr_pivw 68
mr_plot 71
MRAll-class 73
MRcML-class 73
MRConMix-class 74
MRHetPen-class 74
MRInput-class 75
MRLasso-class 76
MRMBE-class 77
MRMVInput-class 77
MVcML_SdTheta 78
MVEgger-class 79
MVGMM-class 80
MVIVW-class 80
MVIVWME-class 81
MVLasso-class 82
MVMedian-class 83
MVmr_cML 84
MVmr_cML_DP 85
MVMRcML-class 86
MVPCGMM-class 87
PCGMM-class 88
pheno_input 88
phenoscanner 91
PIVW-class 92
WeightedMedian-class 93

Index **94**

`calcium` *Data on effect of calcium on fasting glucose (correlated variants)*

Description

Two sets of example data are included in the package: one illustrating uncorrelated variants, and the other correlated variants. These are the data on correlated variants.

Usage

```
calcium

calciumse

fastgluc

fastglucse
```

calc.rho

Format

An object of class numeric of length 6.

An object of class numeric of length 6.

An object of class numeric of length 6.

An object of class numeric of length 6.

An object of class matrix (inherits from array) with 6 rows and 6 columns.

Details

The variables calcium, and fastgluc are the genetic associations with calcium and fasting glucose for 6 genetic variants reported by Burgess et al (2015). The respective standard errors of the associations are given as calciumse and fastglucse. The matrix of correlations between the genetic variants is given as calc.rho.

These data can be used to test out the various functions in the package.

References

Stephen Burgess, Robert A Scott, Nic J Timpson, George Davey Smith, Simon G Thompson. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. Eur J Epidemiol 2015; 30(7):543-552. doi: 10.1007/s10654-015-0011-z.

CLR-class *CLR Class*

Description

An object containing the confidence intervals produced using the conditional likelihood ratio method (CLR) and other identification-robust methods.

Slots

Exposure The name of the exposure variable.

Outcome The name of the outcome variable.

Correlation The matrix of correlations between genetic variants.

ARlower The lower bounds of the causal estimate based on inverting Anderson and Rubin's test.

ARupper The upper bounds of the causal estimate based on inverting Anderson and Rubin's test.

Klower The lower bounds of the causal estimate based on inverting Kleibergen's test.

Kupper The upper bounds of the causal estimate based on inverting Kleibergen's test.

CLRlower The lower bounds of the causal estimate based on inverting Moreira's conditional likelihood ratio test.

`CLRupper` The upper bounds of the causal estimate based on inverting Moreira's conditional likelihood ratio test.

`CIMin` The smallest value used in the search to find the confidence interval.

`CIMax` The largest value used in the search to find the confidence interval.

`CIStep` The step size used in the search to find the confidence interval.

`Alpha` The significance level used in constructing the confidence interval (default is 0.05).

`cML_estimate` *Estimate with Regular Likelihood*

Description

Internal function of mr_cML. Estimate theta, b vector, r vector with constrained maximum likelihood.

Usage

```
cML_estimate(
  b_exp,
  b_out,
  se_exp,
  se_out,
  K,
  initial_theta = 0,
  initial_mu = rep(0, length(b_exp)),
  maxit = 100
)
```

Arguments

`b_exp`	Vector of estimated effects for exposure.
`b_out`	Vector or estimated effects for outcome.
`se_exp`	Vector of standard errors for exposure.
`se_out`	Vector of standard errors for outcome.
`K`	Constraint parameter, number of invalid IVs.
`initial_theta`	Starting point for theta.
`initial_mu`	Starting point for mu.
`maxit`	Maximum number of iteration.

Value

A list contains: theta is the estimate causal effect, b_vec is the estimated vector of b, r_vec is the estimated vector of r.

Examples

```
cML_estimate(b_exp = ldlc,b_out = chdlodds,se_exp = ldlcse,
se_out = chdloddsse, K = 5)
```

`cML_estimate_random` *Estimate with Regular Likelihood Using Multiple Random Start Points*

Description

Internal function of mr_cML. Get estimated theta, se of estimated theta and negative log-likelihood, using multiple random starting points.

Usage

```
cML_estimate_random(
  b_exp,
  b_out,
  se_exp,
  se_out,
  K,
  random_start = 0,
  maxit = 100
)
```

Arguments

`b_exp`	Vector of estimated effects for exposure.
`b_out`	Vector or estimated effects for outcome.
`se_exp`	Vector of standard errors for exposure.
`se_out`	Vector of standard errors for outcome.
`K`	Constraint parameter, number of invalid IVs.
`random_start`	Number of random starting points, default is 0.
`maxit`	Maximum number of iteration.

Value

A list contains: theta is the estimate causal effect, se is standard error of estimated theta, l is negative log-likelihood, r_est is estimated r vector.

Examples

```
cML_estimate_random(b_exp = ldlc,b_out = chdlodds,se_exp = ldlcse,
se_out = chdloddsse, K = 5)
```

cML_SdTheta *Standard Error of Estimated Theta*

Description

Internal function of mr_cML. Get the standard error of estimated theta from constrained maximum likelihood.

Usage

```
cML_SdTheta(b_exp, b_out, se_exp, se_out, theta, b_vec, r_vec)
```

Arguments

b_exp	Vector of estimated effects for exposure.
b_out	Vector or estimated effects for outcome.
se_exp	Vector of standard errors for exposure.
se_out	Vector of standard errors for outcome.
theta	Estimated theta from cML.
b_vec	Estimated vector of b from cML.
r_vec	Estimated vector of r from cML.

Value

Standard error of theta.

Examples

```
# First get estimates:
MLE_result = cML_estimate(b_exp = ldlc,b_out = chdlodds,se_exp = ldlcse,
se_out = chdloddsse, K = 5)

# Calculate standard error:
cML_SdTheta(b_exp = ldlc,b_out = chdlodds,se_exp = ldlcse,
se_out = chdloddsse, theta = MLE_result$theta, b_vec = MLE_result$b_vec, r_vec = MLE_result$r_vec)
```

DIVW-class *DIVW Class*

Description

An object containing the estimate produced using the debiased inverse-variance weighted (dIVW) method as well as various statistics.

Slots

`Over.dispersion` Should the method consider overdispersion (balanced horizontal pleiotropy)? Default is TRUE.

`Exposure` The name of the exposure variable.

`Outcome` The name of the outcome variable.

`Estimate` The causal point estimate from the median-based method.

`StdError` The standard error associated with `Estimate` (obtained from bootstrapping).

`CILower` The lower bound of the confidence interval for `Estimate` based on `StdError`.

`CIUpper` The upper bound of the confidence interval for `Estimate` based on `StdError`.

`Alpha` The significance level used in constructing the confidence interval (default is 0.05).

`Pvalue` P-value associated with the causal estimate from the Wald method.

`SNPs` The number of SNPs that used in the calculation.

`Condition` A measure defined as (average F-statistic -1)*sqrt(# snps) that needs to be large for reliable asymptotic approximation based on the dIVW estimator. It is recommended to be greater than 20.

Egger-class *Egger Class*

Description

An object containing the estimate produced using the MR-Egger method as well as various statistics.

The MR-Egger model uses a random-effects model; a fixed-effect model does not make sense as pleiotropy leads to heterogeneity between the causal estimates targeted by the genetic variants. The (multiplicative) random-effects model allows over-dispersion in the regression model. Under-dispersion is not permitted (in case of under-dispersion, the residual standard error is set to 1).

Slots

`Model` Model always takes the value `random`, as only random-effects analyses are permitted.

`Exposure` The name of the exposure variable.

`Outcome` The name of the outcome variable.

`Correlation` The matrix of correlations between genetic variants.

`Robust` Whether robust regression was used in the regression model relating the genetic associations with the outcome and those with the exposure.

`Penalized` Whether weights in the regression model were penalized for variants with heterogeneous causal estimates.

`Estimate` The causal point estimate from the MR-Egger method.

`StdError.Est` The standard error associated with `Estimate`.

`Pvalue.Est` P-value associated with the causal estimate from the Wald method.

`CILower.Est` The lower bound of the confidence interval for `Estimate` based on `StdError.Est`.

`CIUpper.Est` The upper bound of the confidence interval for `Estimate` based on `StdError.Est`.

`Intercept` The intercept estimate from the MR-Egger method. Under the InSIDE assumption, the intercept represents the average pleiotropic effect (average direct effect on the outcome) of a genetic variant. If the intercept differs from zero, this is evidence that the genetic variants are not all valid instruments; specifically, there is directional pleiotropy.

`StdError.Int` The standard error associated with `Intercept`.

`Pvalue.Int` P-value associated with the intercept from the Wald method.

`CILower.Int` The lower bound of the confidence interval for `Intercept` based on `StdError.Int`.

`CIUpper.Int` The upper bound of the confidence interval for `Estimate` based on `StdError.Int`.

`Alpha` The significance level used in constructing the confidence interval (default is 0.05).

`SNPs` The number of SNPs that were used in the calculation.

`Causal.pval` P-value associated with the causal estimate.

`Pleio.pval` P-value associated with the intercept (p-value for the MR-Egger intercept test of directional pleiotropy).

`RSE` The estimated residual standard error from the regression model.

`Heter.Stat` Heterogeneity statistic (Cochran's Q statistic) and associated p-value: the null hypothesis is that the MR-Egger regression model describes the associations with the outcome with no excess heterogeneity.

`I.sq` A measure of heterogeneity between the genetic associations with the exposure (see Bowden IJE 2016: "Assessing the suitability of summary data for Mendelian randomization analyses using MR-Egger regression: The role of the I2 statistic."). Low values of `I.sq` relate both to large differences in precision between MR-Egger and IVW estimates, and to more weak instrument bias (in a two-sample setting, this is attenuation of MR-Egger estimate towards the null).

invcov_mvmr *Generate the list of inverse of covariance matrices used in MVMR-cML-DP*

Description

Generate the list of inverse of covariance matrices used in MVMR-cML-DP

Usage

```
invcov_mvmr(se_bx, se_by, rho_mat)
```

Arguments

se_bx	A m*L matrix of standard errors of SNP-exposure association
se_by	A vector of standard errors of SNP-outcome association
rho_mat	The correlation matrix among the L exposures and the outcome

Value

A list of inverse of covariance matrices with respect to each genetic variant, retaining the ordering in se_bx

IVW-class *IVW Class*

Description

An object containing the estimate produced using the inverse-variance weighted (IVW) method as well as various statistics.

Slots

Model The model used for estimation: random-effects ("random") or fixed-effect ("fixed"). The default option ("default") is to use a fixed-effect model when there are three or fewer genetic variants, and a random-effects model when there are four or more. The (multiplicative) random-effects model allows for heterogeneity between the causal estimates targeted by the genetic variants by allowing over-dispersion in the regression model. Under-dispersion is not permitted (in case of under-dispersion, the residual standard error is set to 1, as in a fixed-effect analysis).

Exposure The name of the exposure variable.

Outcome The name of the outcome variable.

Correlation The matrix of correlations between genetic variants.

`Robust` Whether robust regression was used in the regression model relating the genetic associations with the outcome and those with the exposure.

`Penalized` Whether weights in the regression model were penalized for variants with heterogeneous causal estimates.

`Estimate` The causal point estimate from the inverse-variance weighted method.

`StdError` The standard error associated with `Estimate`.

`CILower` The lower bound of the confidence interval for `Estimate` based on `StdError`.

`CIUpper` The upper bound of the confidence interval for `Estimate` based on `StdError`.

`Alpha` The significance level used in constructing the confidence interval (default is 0.05).

`Pvalue` P-value associated with the causal estimate.

`SNPs` The number of SNPs that were used in the calculation.

`RSE` The estimated residual standard error from the regression model.

`Heter.Stat` Heterogeneity statistic (Cochran's Q statistic) and associated p-value: the null hypothesis is that all genetic variants estimate the same causal parameter; rejection of the null is an indication that one or more variants may be pleiotropic.

`Fstat` An approximation of the first-stage F statistic for all variants based on the summarized data.

`ldlc` *Data on lipid effects on coronary artery disease (uncorrelated variants)*

Description

Two sets of example data are included in the package: one illustrating uncorrelated variants, and the other correlated variants. These are the data on uncorrelated variants.

The variables `ldlc`, `hdlc`, `trig`, and `chdlodds` are the genetic associations with (respectively) LDL-cholesterol, HDL-cholesterol, triglycerides, and coronary heart disease (CHD) risk for 28 genetic variants reported by Waterworth et al (2010). The respective standard errors of the associations are given as `ldlcse`, `hdlcse`, `trigse`, and `chdloddsse`.

These data can be used to test out the various functions in the package.

Usage

```
ldlc

hdlc

hdlcse

ldlcse

trig
```

`trigse`

`chdlodds`

`chdloddsse`

`lipid_effect`

`lipid_other`

`lipid_eaf`

Format

An object of class `numeric` of length 28.

An object of class `numeric` of length 28.

An object of class `numeric` of length 28.

An object of class `numeric` of length 28.

An object of class `numeric` of length 28.

An object of class `numeric` of length 28.

An object of class `numeric` of length 28.

An object of class `numeric` of length 28.

An object of class `character` of length 28.

An object of class `character` of length 28.

An object of class `numeric` of length 28.

References

Dawn Waterworth, Sally Ricketts, ..., Manj Sandhu: Genetic variants influencing circulating lipid levels and risk of coronary artery disease. Arterioscler Thromb Vasc Biol 2010; 30:2264-227. doi: 10.1161/atvbaha.109.201020.

`MaxLik-class` *MaxLik Class*

Description

An object containing the estimate produced using the maximum-likelihood method as well as various statistics.

Slots

`Model` The model used for estimation: fixed-effect (`"fixed"`) or random-effects (`"random"`).

`Exposure` The name of the exposure variable.

`Outcome` The name of the outcome variable.

`Correlation` The matrix of correlations between genetic variants.

`Psi` The correlations between genetic associations with the exposure and with the outcome.

`Estimate` The causal point estimate from the inverse-variance weighted method.

`StdError` The standard error associated with `Estimate`.

`CILower` The lower bound of the confidence interval for `Estimate` based on `StdError`.

`CIUpper` The upper bound of the confidence interval for `Estimate` based on `StdError`.

`Alpha` The significance level used in constructing the confidence interval (default is 0.05).

`Pvalue` P-value associated with the causal estimate.

`SNPs` The number of SNPs that were used in the calculation.

`RSE` The estimated residual standard error from the regression model.

`Heter.Stat` Heterogeneity statistic (likelihood ratio statistic) and associated p-value: the null hypothesis is that all genetic variants estimate the same causal parameter; rejection of the null is an indication that one or more variants may be pleiotropic.

`mr_allmethods` *Mendelian randomization estimation using all methods*

Description

The function `mr_allmethods` implements Mendelian randomization analyses using summarized data to calculate estimates (as well as standard errors and confidence interval limits) for all the methods included in the package (or alternatively for the group of methods chosen).

Usage

```
mr_allmethods(object, method = "all", ...)

## S4 method for signature 'MRInput'
mr_allmethods(object, method = "all", ...)
```

Arguments

`object`	An `MRInput` object.
`method`	Which estimation method should be included in the calculation. By default, all estimates are computed (`"all"`), but one can choose to show only the results of median-based, inverse-variance weighted, or MR-Egger methods separately through specifying `"median"`, `"ivw"`, `"egger"`, or `"main"` (gives main results only, that is simple and weighted median, IVW, and MR-Egger).
`...`	Additional arguments to be passed to other methods.

Details

See mr_median, mr_egger, and mr_ivw for details of how each of the methods is implemented.

Value

An object of type MRAll with the following slots :

Data	The MRInput object used to calculate the various values.
Values	A data.frame containing the various estimates.
Method	The choice of methods estimated (default is "all").

References

See mr_median, mr_egger, and mr_ivw.

Examples

```
mr_allmethods(mr_input(bx = ldlc, bxse = ldlcse,
  by = chdlodds, byse = chdloddsse), method="main", iterations = 100)
  # iterations is set to 100 to reduce runtime for the mr_median method,
  # at least 10000 iterations are recommended in practice
```

---

mr_clr *Conditional likelihood ratio (CLR) method*

Description

The mr_clr function calculates confidence intervals based on inverting the conditional likelihood ratio and other identification-robust tests.

Usage

```
mr_clr(object, nx, ny, alpha = 0.05, CIMin = -10, CIMax = 10, CIStep = 0.01)

## S4 method for signature 'MRInput'
mr_clr(object, nx, ny, alpha = 0.05, CIMin = -10, CIMax = 10, CIStep = 0.01)
```

Arguments

object	An MRInput object.
nx	The sample size used to compute genetic associations with the exposure.
ny	The sample size used to compute genetic associations with the outcome.
alpha	The significance level used to calculate the confidence interval. The default value is 0.05.

CIMin	The smallest value to use in the search to find the confidence interval (default is -10).
CIMax	The largest value to use in the search to find the confidence interval (default is +10).
CIStep	The step size to use in the search to find the confidence interval (default is 0.01). Using a lower value (such as 0.001) will give more precise confidence intervals, but increase run time.

Details

In weak instrument settings, usual inference based on point estimates and standard errors may not be accurate. This method calculates confidence intervals based on inverting identification-robust tests proposed in Wang and Kang (2021, Biometrics) that provide valid inferences regardless of instrument strength.

This includes conditional likelihood ratio (CLR), Kleibergen (K), and Anderson and Rubin (AR) tests.

Evidence from the econometrics literature suggests that CLR inference is the best option in terms of power under a wide range of settings.

Please note that these methods do not provide point estimates, only confidence intervals. While most examples provide a confidence interval that is a single range of values, in some cases the confidence interval may comprise multiple ranges of values. In other cases, a valid confidence interval may not exist.

Value

The output from the function is an CLR object containing:

Exposure	A character string with the name given to the exposure.
Outcome	A character string with the names given to the outcome.
Correlation	The matrix of genetic correlations.
ARlower	The lower bounds of the causal estimate based on inverting Anderson and Rubin's test.
ARupper	The upper bounds of the causal estimate based on inverting Anderson and Rubin's test.
Klower	The lower bounds of the causal estimate based on inverting Kleibergen's test.
Kupper	The upper bounds of the causal estimate based on inverting Kleibergen's test.
CLRlower	The lower bounds of the causal estimate based on inverting Moreira's conditional likelihood ratio test.
CLRupper	The upper bounds of the causal estimate based on inverting Moreira's conditional likelihood ratio test.
CIMin	The smallest value used in the search to find the confidence interval.
CIMax	The largest value used in the search to find the confidence interval.
CIStep	The step size used in the search to find the confidence interval.
Alpha	The significance level used when calculating the confidence intervals.

References

Description of the CLR method: "Weak-instrument robust tests in two-sample summary-data Mendelian randomization", S. Wang and H. Kang, Biometrics, 2021.

Examples

```
mr_clr(mr_input(bx = calcium, bxse = calciumse,
   by = fastgluc, byse = fastglucse, correl = calc.rho), nx=6351, ny=133010)
```

---

mr_cML *Constrained maximum likelihood (cML) method*

Description

Constrained maximum likelihood (cML) based Mendelian Randomization method robust to both correlated and uncorrelated pleiotropy.

Usage

```
mr_cML(
  object,
  MA = TRUE,
  DP = TRUE,
  K_vec = 0:(length(object@betaX) - 2),
  random_start = 0,
  num_pert = 200,
  random_start_pert = 0,
  maxit = 100,
  random_seed = 314,
  n,
  Alpha = 0.05
)

## S4 method for signature 'MRInput'
mr_cML(
  object,
  MA = TRUE,
  DP = TRUE,
  K_vec = 0:(length(object@betaX) - 2),
  random_start = 0,
  num_pert = 200,
  random_start_pert = 0,
  maxit = 100,
  random_seed = 314,
  n,
  Alpha = 0.05
)
```

Arguments

object	An MRInput object.
MA	Whether model average is applied or not. Default is TRUE.
DP	Whether data perturbation is applied or not. Default is TRUE.
K_vec	Set of candidate K's, the constraint parameter representing number of invalid IVs. Default is from 0 to (#IV - 2).
random_start	Number of random starting points for cML, default is 0.
num_pert	Number of perturbation when DP is TRUE, default is 200.
random_start_pert	Number of random start points for cML with data perturbation, default is 0.
maxit	Maximum number of iterations for each optimization. Default is 100.
random_seed	Random seed, default is 314. When random_seed=NULL, no random seed will be used and the results may not be reproducible.
n	Sample size. When sample sizes of GWAS for exposure and outcome are different, and/or when sample sizes of different SNPs are different, the smallest sample size is recommended to get conservative result and avoid type-I error. See reference for more discussions.
Alpha	Significance level for the confidence interval for estimate, default is 0.05.

Details

The MRcML method selects invalid IVs with correlated and/or uncorrelated peliotropic effects using constrained maximum likelihood. cML-BIC gives results of the selected model with original data, while cML-MA-BIC averages over all candidate models. cML-BIC-DP and cML-MA-BIC-DP are the versions with data-perturbation to account for selection uncertainty when many invalid IVs have weak pleiotropic effects.

When DP is performed, two goodness-of-fit (GOF) tests are developed to check whether the model-based and DP- based variance estimates converge to the same estimate. Small p-values of GOF tests indicate selection uncertainty is not ignorable, and results from DP is more reliable. See reference for more details.

As the constrained maximum likelihood function is non-convex, multiple random starting points could be used to find a global minimum. For some starting points the algorithm may not converge and a warning message will be prompted, typically this will not affect the results.

Value

The output from the function is an MRcML object containing:

Exposure	A character string giving the name given to the exposure.
Outcome	A character string giving the name given to the outcome.
Estimate	Estimate of theta.
StdError	Standard error of estimate.
Pvalue	p-value of estimate.

BIC_invalid	Set of selected invalid IVs if cML-BIC is performed, i.e. without MA or DP.
GOF1_p	p-value of the first goodness-of-fit test.
GOF2_p	p-value of the second goodness-of-fit test.
SNPs	The number of SNPs that were used in the calculation.
Alpha	Significance level for the confidence interval for estimate, default is 0.05.
CILower	Lower bound of the confidence interval for estimate.
CIUpper	Upper bound of the confidence interval for estimate.
MA	Indicator of whether model average is applied.
DP	Indicator of whether data perturbation is applied.

References

Xue, H., Shen, X., & Pan, W. (2021). Constrained maximum likelihood-based Mendelian randomization robust to both correlated and uncorrelated pleiotropic effects. The American Journal of Human Genetics, 108(7), 1251-1269.

Examples

```
# Perform cML-MA-BIC-DP:
mr_cML(mr_input(bx = ldlc, bxse = ldlcse, by = chdlodds,
byse = chdloddsse), num_pert=5, MA = TRUE, DP = TRUE, n = 17723)
# num_pert is set to 5 to reduce computational time
# the default value of 200 is recommended in practice

# Perform cML-BIC-DP:
mr_cML(mr_input(bx = ldlc, bxse = ldlcse, by = chdlodds,
byse = chdloddsse), MA = TRUE, DP = FALSE,, n = 17723)
```

---

mr_conmix *Contamination mixture method*

Description

Contamination mixture method for robust and efficient estimation under the 'plurality valid' assumption.

Usage

```
mr_conmix(object, psi = 0, CIMin = NA, CIMax = NA, CIStep = 0.01, alpha = 0.05)

## S4 method for signature 'MRInput'
mr_conmix(object, psi = 0, CIMin = NA, CIMax = NA, CIStep = 0.01, alpha = 0.05)
```

Arguments

object	An MRInput object.
psi	The value of the standard deviation of the distribution of invalid estimands (default value is 0, corresponding to 1.5 times the standard deviation of the ratio estimates).
CIMin	The smallest value to use in the search to find the confidence interval. The default value is NA, which means that the method uses the smallest value of the lower bound of the 95% confidence interval for the variant-specific ratio estimates as the smallest value.
CIMax	The largest value to use in the search to find the confidence interval. The default value is NA, which means that the method uses the greatest value of the upper bound of the 95% confidence interval for the variant-specific ratio estimates as the largest value.
CIStep	The step size to use in the search to find the confidence interval (default is 0.01). The confidence interval is determined by a grid search algorithm. Using the default settings, we calculate the likelihood at all values from -1 to +1 increasing in units of 0.01. If this range is too large or the step size is too small, then the method will take a long time to run.
alpha	The significance level used to calculate the confidence interval. The default value is 0.05.

Details

The contamination mixture method is implemented by constructing a likelihood function based on the variant-specific causal estimates. If a genetic variant is a valid instrument, then its causal estimate will be normally distributed about the true value of the causal effect. If a genetic variant is not a valid instrument, then its causal estimate will be normally distributed about some other value. We assume that the values estimated by invalid instruments are normally distributed about zero with a large standard deviation. This enables a likelihood function to be specified that is a product of two-component mixture distributions, with one mixture distribution for each variant. The computational time for maximizing this likelihood directly is exponential in the number of genetic variants. We use a profile likelihood approach to reduce the computational complexity to be linear in the number of variants.

We consider different values of the causal effect in turn. For each value, we calculate the contribution to the likelihood for each genetic variant as a valid instrument and as an invalid instrument. If the contribution to the likelihood as a valid instrument is greater, then we take the variant's contribution as a valid instrument; if less, then its contribution is taken as an invalid instrument. This gives us the configuration of valid and invalid instruments that maximizes the likelihood for the given value of the causal effect. This is a profile likelihood, a one-dimensional function of the causal effect. The point estimate is then taken as the value of the causal effect that maximizes the profile likelihood.

Confidence intervals are evaluated by calculating the log-likelihood function, and finding all points within a given vertical distance of the maximum of the log-likelihood function (which is the causal estimate). As such, if the log-likelihood function is multimodal, then the confidence interval may include multiple disjoint ranges. This may indicate the presence of multiple causal mechanisms by which the exposure may influence the outcome with different magnitudes of causal effect. As the

confidence interval is determined by a grid search, care must be taken when chosing the minimum (CIMin) and maximum (CIMax) values in the search, as well as the step size (CIStep). The default values will not be suitable for all applications.

Value

The output from the function is an MRConMix object containing:

Exposure	A character string giving the name given to the exposure.
Outcome	A character string giving the name given to the outcome.
Psi	The value of the standard deviation parameter.
Estimate	The value of the causal estimate.
CIRange	The range of values in the confidence interval based on a grid search between the minimum and maximum values for the causal effect provided.
CILower	The lower limit of the confidence interval. If the confidence interval contains multiple ranges, then lower limits of all ranges will be reported.
CIUpper	The upper limit of the confidence interval. If the confidence interval contains multiple ranges, then upper limits of all ranges will be reported.
CIMin	The smallest value used in the search to find the confidence interval.
CIMax	The largest value used in the search to find the confidence interval.
CIStep	The step size used in the search to find the confidence interval.
Pvalue	The p-value associated with the estimate calculated using the likelihood function and a chi-squared distribution.
Valid	The numbers of genetic variants that were considered valid instruments at the causal estimate.
ValidSNPs	The names of genetic variants that were considered valid instruments at the causal estimate.
Alpha	The significance level used when calculating the confidence intervals.
SNPs	The number of genetic variants (SNPs) included in the analysis.

References

Stephen Burgess, Christopher N Foley, Elias Allara, Joanna Howson. A robust and efficient method for Mendelian randomization with hundreds of genetic variants: unravelling mechanisms linking HDL-cholesterol and coronary heart disease. Nat Comms 2020. doi: 10.1038/s41467-019-14156-4.

Examples

```
mr_conmix(mr_input(bx = ldlc, bxse = ldlcse, by = chdlodds,
   byse = chdloddsse), psi = 3, CIMin = -1, CIMax = 5, CIStep = 0.01)
```

mr_divw *Debiased inverse-variance weighted method*

Description

The mr_divw function implements the debiased inverse-variance weighted method.

Usage

```
mr_divw(object, over.dispersion = TRUE, alpha = 0.05, diagnostics = FALSE)

## S4 method for signature 'MRInput'
mr_divw(object, over.dispersion = TRUE, alpha = 0.05, diagnostics = FALSE)
```

Arguments

object	An MRInput object.
over.dispersion	Should the method consider overdispersion (balanced horizontal pleiotropy)? Default is TRUE.
alpha	The significance level used to calculate the confidence intervals. The default value is 0.05.
diagnostics	Should the function returns the q-q plot for assumption diagnosis. Default is FALSE.

Details

The debiased inverse-variance weighted method (dIVW) removes the weak instrument bias of the IVW method and is more robust under many weak instruments.

Value

The output from the function is a DIVW object containing:

Over.dispersion	TRUE if the method has considered balanced horizontal pleiotropy, FALSE otherwise.
Exposure	A character string giving the name given to the exposure.
Outcome	A character string giving the name given to the outcome.
Estimate	The value of the causal estimate.
StdError	Standard error of the causal estimate calculated using bootstrapping.
CILower	The lower bound for the causal estimate based on the estimated standard error and the significance level provided.
CIUpper	The upper bound for the causal estimate based on the estimated standard error and the significance level provided.

Alpha	The significance level used when calculating the confidence intervals.
Pvalue	The p-value associated with the estimate (calculated using Estimate/StdError as per a Wald test) using a normal distribution.
SNPs	The number of genetic variants (SNPs) included in the analysis.
Condition	A measure (average F-statistic -1)*sqrt(# snps) that needs to be large for reliable asymptotic approximation based on the dIVW estimator. It is recommended to be greater than 20.

References

Ting Ye, Jun Shao, Hyunseung Kang (2021). Debiased Inverse-Variance Weighted Estimator in Two-Sample Summary-Data Mendelian Randomization. The Annals of Statistics, 49(4), 2079-2100. Also available at https://arxiv.org/abs/1911.09802.

Examples

```
mr_divw(mr_input(bx = ldlc, bxse = ldlcse, by = chdlodds, byse = chdloddsse))
```

---

mr_egger *MR-Egger method*

Description

The mr_egger function implements the MR-Egger method introduced by Bowden et al (2015).

This method provides: 1) a test of the for directional pleiotropy (the MR-Egger intercept test), 2) a test for a causal effect, and 3) an estimate of the causal effect. If the intercept term differs from zero, then the genetic variants are not all valid instrumental variables and the standard (inverse-variance weighted) estimate is biased. If the InSIDE (Instrument Strength Independent of Direct Effect) assumption holds, then the MR-Egger slope parameter provides a test for a causal effect, and a consistent estimate of the causal effect even if the intercept differs from zero.

Usage

```
mr_egger(
  object,
  robust = FALSE,
  penalized = FALSE,
  correl = FALSE,
  distribution = "normal",
  alpha = 0.05,
  ...
)

## S4 method for signature 'MRInput'
mr_egger(
```

```
  object,
  robust = FALSE,
  penalized = FALSE,
  correl = FALSE,
  distribution = "normal",
  alpha = 0.05,
  ...
)
```

Arguments

object	An MRInput object.
robust	Indicates whether robust regression using the lmrob() function from the package robustbase should be used in the method.
penalized	Indicates whether a penalty should be applied to the weights to downweight the contribution of genetic variants with outlying ratio estimates to the analysis.
correl	If the genetic variants are correlated, then this correlation can be accounted for. The matrix of correlations between must be provided: the elements of this matrix are the correlations between the individual variants (diagonal elements are 1). If a correlation is specified, then the values of "robust" and "penalized" are taken as FALSE.
distribution	The type of distribution used to calculate the confidence intervals, can be "normal" (the default option) or "t-dist". If the distribution is "t-dist", then a t-distribution is used in case of over-dispersion. In case of under-dispersion, the confidence interval is the wider of that using the estimated residual standard error and a t-distribution, or that using a residual standard error of 1 and a normal distribution. This ensures that under-dispersion is not "doubly penalized" by setting the residual standard error to 1 and using a t-distribution, and also that the random-effects analysis is no more precise than a fixed-effect analysis would be.
alpha	The significance level used to calculate the confidence interval. The default value is 0.05.
...	Additional arguments to be passed to the regression method.

Details

The causal estimate is obtained by regression of the associations with the outcome on the associations with the risk factor, with weights being the inverse-variances of the associations with the outcome. The intercept is estimated (in contrast with the inverse-variance weighted method, where the intercept is set to zero).

As part of the analysis, the genetic variants are orientated so that all of the associations with the risk factor are positive (and signs of associations with the outcome are changed to keep the orientation consistent if required). Re-orientation of the genetic variants is performed automatically as part of the function.

The MR-Egger model uses a random-effects model ("random"); a fixed-effect model does not make sense as pleiotropy leads to heterogeneity between the causal estimates targeted by the genetic

variants. The (multiplicative) random-effects model allows over-dispersion in the regression model. Under-dispersion is not permitted (in case of under-dispersion, the residual standard error is set to 1).

Value

The output of the function is an `Egger` object containing:

`Model`	A character string giving the type of model used (`"random"`).
`Exposure`	A character string giving the name given to the exposure.
`Outcome`	A character string giving the name given to the outcome.
`Correlation`	The matrix of genetic correlations.
`Robust`	`TRUE` if robust estimate has been calculated, `FALSE` otherwise.
`Penalized`	`TRUE` if weights have been penalized, `FALSE` otherwise.
`Estimate`	The value of the causal estimate (slope coefficient).
`StdError.Est`	Standard error of the causal estimate.
`Pvalue.Est`	The p-value associated with the estimate (calculated as Estimate/StdError as per Wald test) using a normal or t-distribution (as specified in `distribution`).
`CILower.Est`	The lower bound of the causal estimate based on the estimated standard error and the significance level provided.
`CIUpper.Est`	The upper bound of the causal estimate based on the estimated standard error and the significance level provided.
`Intercept`	The value of the intercept estimate.
`StdError.Int`	Standard error of the intercept estimate.
`Pvalue.Int`	The p-value associated with the intercept.
`CILower.Int`	The lower bound of the intercept based on the estimated standard error and the significance level provided.
`CIUpper.Int`	The upper bound of the intercept based on the estimated standard error and the significance level provided.
`Alpha`	The significance level used when calculating the confidence intervals (same as `alpha` above).
`SNPs`	The number of genetic variants (SNPs) included in the analysis.
`Causal.pval`	The p-value for the MR-Egger causal estimate.
`Pleio.pval`	The p-value for the MR-Egger intercept test (a low p-value suggests either directional pleiotropy or failure of the InSIDE assumption, and indicates that the IVW estimate is biased).
`RSE`	The estimated residual standard error from the regression model.
`Heter.Stat`	Heterogeneity statistic (Cochran's Q statistic) and associated p-value: the null hypothesis is that the regression model (including an intercept) fits the regression model with no additional variability. Rejection of the null hypothesis is expected if genetic variants are pleiotropic, and doesn't mean that the MR-Egger analysis or the InSIDE assumption is invalid.

`I.sq` A measure of heterogeneity between the genetic associations with the exposure (see Bowden IJE 2016). Low values of `I.sq` relate both to large differences in precision between MR-Egger and IVW estimates, and to more weak instrument bias (in a two-sample setting, this is attenuation of MR-Egger estimate towards the null).

References

Jack Bowden, George Davey Smith, Stephen Burgess. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. International Journal of Epidemiology 2015; 44:512–525. doi: 10.1093/ije/dyv080.

Confidence intervals, and robust and penalized weights: Stephen Burgess, Jack Bowden, Frank Dudbridge, Simon G Thompson. Robust instrumental variable methods using multiple candidate instruments with application to Mendelian randomization. arXiv 2016; 1606.03729.

I-squared statistic: Jack Bowden and others. Assessing the suitability of summary data for Mendelian randomization analyses using MR-Egger regression: The role of the I2 statistic. Int J Epidemiol 2016 (to appear).

Examples

```
mr_egger(mr_input(bx = ldlc, bxse = ldlcse, by = chdlodds, byse = chdloddsse))
mr_egger(mr_input(bx = ldlc, bxse = ldlcse, by = chdlodds, byse = chdloddsse),
  robust = TRUE)
mr_egger(mr_input(bx = ldlc, bxse = ldlcse, by = chdlodds, byse = chdloddsse),
  penalized = TRUE)
mr_egger(mr_input(calcium, calciumse, fastgluc, fastglucse, corr=calc.rho))
  ## correlated variants
```

---

mr_forest *Draw a forest plot of causal estimates*

Description

The `mr_forest` function draws a forest plot of causal estimates. The default option plots the variant-specific causal estimates (`by/bx`) and the estimate from the `mr_ivw` function using default settings (assuming variants are uncorrelated, random-effects for 4+ variants). Options allow users to plot estimates from a variety of different methods.

Usage

```
mr_forest(
  object,
  alpha = 0.05,
  snp_estimates = TRUE,
  methods = "ivw",
  ordered = FALSE
```

```
)

## S4 method for signature 'MRInput'
mr_forest(
  object,
  alpha = 0.05,
  snp_estimates = TRUE,
  methods = "ivw",
  ordered = FALSE
)
```

Arguments

object	An MRInput object.
alpha	The significance level used to calculate the confidence intervals. The default value is 0.05, corresponding to 95% confidence intervals.
snp_estimates	Whether to plot the variant-specific estimates. Defaults to TRUE.
methods	Takes a string of computation methods used to calculate estimates. Defaults to "ivw". Options are: "median" (simple median estimate), "wmedian" (weighted median estimate), "egger" (MR-Egger estimate), "mbe" (mode-based estimate), "conmix" (contamination mixture estimate), and "maxlik" (maximum likelihood estimate).
ordered	Determines by whether to arrange the variant-specific estimates in ascending order. Defaults to FALSE.

Details

As the function produces a ggplot object, graphical parameters can be changed by adding commands from the ggplot2 package.

Examples

```
mr_forest(mr_input(bx = ldlc, bxse = ldlcse, by = chdlodds, byse = chdloddsse),
  alpha = 0.01, ordered = TRUE)
mr_forest(mr_input(bx = ldlc, bxse = ldlcse, by = chdlodds, byse = chdloddsse),
  methods = c("ivw", "wmedian", "egger"), snp_estimates = FALSE)
forest = mr_forest(mr_input(bx = ldlc, bxse = ldlcse, by = chdlodds, byse = chdloddsse))
# how to change x-axis limits
# library(ggplot2)
# forest2 = forest + coord_cartesian(xlim=c(-5,5))
# forest2
```

mr_funnel *Draw a funnel plot of variant-specific estimates*

Description

The mr_funnel function draws a funnel plot of variant-specific causal estimates. Estimates (by/bx) are plotted against the precision of the estimates (abs(bx)/byse). Precision is the reciprocal of standard error. A vertical dashed line is plotted at the estimate from the mr_ivw function.

Usage

```
mr_funnel(object, CI = TRUE)

## S4 method for signature 'MRInput'
mr_funnel(object, CI = TRUE)
```

Arguments

object	An MRInput object.
CI	A logical variable dicating as to whether to plot the confidence interval associated with each point. Default value is TRUE.

Details

As the function produces a ggplot object, graphical parameters can be changed by adding commands from the ggplot2 package.

Examples

```
mr_funnel(mr_input(bx = ldlc, bxse = ldlcse, by = chdlodds, byse = chdloddsse))
```

mr_hetpen *Heterogeneity-penalized method*

Description

Heterogeneity-penalized model-averaging method for efficient modal-based estimation.

Usage

```
mr_hetpen(
  object,
  prior = 0.5,
  CIMin = -1,
  CIMax = 1,
  CIStep = 0.001,
  alpha = 0.05
)

## S4 method for signature 'MRInput'
mr_hetpen(
  object,
  prior = 0.5,
  CIMin = -1,
  CIMax = 1,
  CIStep = 0.001,
  alpha = 0.05
)
```

Arguments

object	An MRInput object.
prior	The prior probability of a genetic variant being a valid instrument (default is 0.5).
CIMin	The smallest value to use in the search to find the confidence interval (default is -1).
CIMax	The largest value to use in the search to find the confidence interval (default is +1).
CIStep	The step size to use in the search to find the confidence interval (default is 0.001). The confidence interval is determined by a grid search algorithm. Using the default settings, we calculate the likelihood at all values from -1 to +1 increasing in units of 0.001. If this range is too large or the step size is too small, then the grid search algorithm will take a long time to converge.
alpha	The significance level used to calculate the confidence interval. The default value is 0.05.

Details

This method was developed as a more efficient version of the mode-based estimation method of Hartwig et al. It proceeds by evaluating weights for all subsets of genetic variants (excluding the null set and singletons). Subsets receive greater weight if they include more variants, but are severely downweighted if the variants in the subset have heterogeneous causal estimates. As such, the method will identify the subset with the largest number (by weight) of variants having similar causal estimates.

Confidence intervals are evaluated by calculating a log-likelihood function, and finding all points within a given vertical distance of the maximum of the log-likelihood function (which is the causal

estimate). As such, if the log-likelihood function is multimodal, then the confidence interval may include multiple disjoint ranges. This may indicate the presence of multiple causal mechanisms by which the exposure may influence the outcome with different magnitudes of causal effect. As the confidence interval is determined by a grid search, care must be taken when chosing the minimum (`CIMin`) and maximum (`CIMax`) values in the search, as well as the step size (`CIStep`). The default values will not be suitable for all applications.

The method should give consistent estimates as the sample size increases if a weighted plurality of the genetic variants are valid instruments. This means that the largest group of variants with the same causal estimate in the asymptotic limit are the valid instruments.

The current implementation of the method evaluates a weight and an estimate for each of the subsets of genetic variants. This means that the method complexity doubles for each additional genetic variant included in the analysis. Currently, the method provides a warning message when used with 25+ variants, and fails to run with 30+.

Value

The output from the function is an `MRHetPen` object containing:

`Exposure`	A character string giving the name given to the exposure.
`Outcome`	A character string giving the name given to the outcome.
`Prior`	The value of the bandwidth factor.
`Estimate`	The value of the causal estimate.
`CIRange`	The range of values in the confidence interval based on a grid search between the minimum and maximum values for the causal effect provided.
`CILower`	The lower limit of the confidence interval. If the confidence interval contains multiple ranges, then lower limits of all ranges will be reported.
`CIUpper`	The upper limit of the confidence interval. If the confidence interval contains multiple ranges, then upper limits of all ranges will be reported.
`CIMin`	The smallest value used in the search to find the confidence interval.
`CIMax`	The largest value used in the search to find the confidence interval.
`CIStep`	The step size used in the search to find the confidence interval.
`Alpha`	The significance level used when calculating the confidence intervals.
`SNPs`	The number of genetic variants (SNPs) included in the analysis.

References

Stephen Burgess, Verena Zuber, Apostolos Gkatzionis, Christopher N Foley. Improving on a modal-based estimation method: model averaging for consistent and efficient estimation in Mendelian randomization when a plurality of candidate instruments are valid. bioRxiv 2017. doi: 10.1101/175372.

Examples

```
mr_hetpen(mr_input(bx = ldlc[1:10], bxse = ldlcse[1:10], by = chdlodds[1:10],
   byse = chdloddsse[1:10]), CIMin = -1, CIMax = 5, CIStep = 0.01)
```

mr_input *Inputting and formatting data for use in causal estimation*

Description

The mr_input function is required for inputting and formatting data for use in any of the estimation functions provided in this package. The MRInput class outputted by the function can also be viewed graphically using the mr_plot function.

Usage

```
mr_input(
  bx = 0,
  bxse = 0,
  by = 0,
  byse = 0,
  correlation = matrix(),
  exposure = "exposure",
  outcome = "outcome",
  snps = "snp",
  effect_allele = NA,
  other_allele = NA,
  eaf = NA
)
```

Arguments

bx	A numeric vector of beta-coefficient values for genetic associations with the first variable (often referred to as the exposure, risk factor, or modifiable phenotype).
bxse	The standard errors associated with the beta-coefficients bx.
by	A numeric vector of beta-coefficient values for genetic associations with the second variable (often referred to as the outcome). For a disease outcome, the beta coefficients are log odds estimates from logistic regression analyses.
byse	The standard errors associated with the beta-coefficients in by.
correlation	The matrix of correlations between genetic variants. If this variable is not provided, then we assume that genetic variants are uncorrelated.
exposure	The name of the exposure variable.
outcome	The name of the outcome variable.
snps	The names of the genetic variants (SNPs) included in the analysis. The inputs exposure, outcome, and snps are not required, but may be useful for keeping track of various MRInput objects. They are also used by the mr_plot function.
effect_allele	The name of the effect allele for each SNP. The beta-coefficients are the associations with the exposure and outcome per additional copy of the effect allele.
other_allele	The name of the non-effect allele.

eaf	The expected allele frequencies (numeric). The slots effect_allele, other_allele, and eaf are neither required, nor currently used in the MendelianRandomization package. They are included for future compatibility with the MR-Base suite of functions.

Details

The beta-coefficients are assumed to be estimated for uncorrelated (independent) genetic variants, although a correlation matrix can be specified if the variants are correlated in their distributions. We also assume that the beta-coefficients for associations with the exposure and with the outcome are uncorrelated (corresponding to a two-sample Mendelian randomization analysis), although correlation between associations with the exposure and with the outcome generally have little impact on causal estimates or standard errors.

If the four variables are not all the same length, then an error message will be reported. The analyses will still try to run, but the output may be misleading. However, in some analyses (for example, the standard IVW and MR-Egger methods), the values of bxse are not used in the analysis, and can therefore safely be omitted (provided that the other variables are correctly labelled).

Value

An MRInput object containing:

betaX	The genetic associations with the exposure.
betaXse	The corresponding standard errors.
betaY	The genetic associations with the outcome.
betaYse	The corresponding standard errors.
correlation	The matrix of genetic correlations.
exposure	A character string giving the name given to the exposure.
outcome	A character string giving the name given to the outcome.
snps	A vector of character strings with the names of the genetic variants.
effect_allele	A vector of character strings with the names of the effect alleles.
other_allele	A vector of character strings with the names of the non-effect alleles.
eaf	A numeric vector with the effect allele frequencies.

---

mr_ivw *Inverse-variance weighted method*

Description

The mr_ivw function implements the inverse-variance method, informally known as the "Toby Johnson" method. With a single genetic variant, this is simply the ratio method.

Usage

```
mr_ivw(
  object,
  model = "default",
  robust = FALSE,
  penalized = FALSE,
  weights = "simple",
  psi = 0,
  correl = FALSE,
  distribution = "normal",
  alpha = 0.05,
  ...
)

## S4 method for signature 'MRInput'
mr_ivw(
  object,
  model = "default",
  robust = FALSE,
  penalized = FALSE,
  weights = "simple",
  psi = 0,
  correl = FALSE,
  distribution = "normal",
  alpha = 0.05,
  ...
)
```

Arguments

object	An MRInput object.
model	What type of model should be used: "default", "random" or "fixed". The random-effects model ("random") is a multiplicative random-effects model, allowing overdispersion in the weighted linear regression (the residual standard error is not fixed to be 1, but is not allowed to take values below 1). The fixed-effect model ("fixed") sets the residual standard error to be 1. The "default" setting is to use a fixed-effect model with 3 genetic variants or fewer, and otherwise to use a random-effects model.
robust	Indicates whether robust regression using the lmrob() function from the package robustbase should be used in the method rather than standard linear regression (lm).
penalized	Indicates whether a penalty should be applied to the weights to downweight the contribution of genetic variants with outlying ratio estimates to the analysis.
weights	Which weights to use in the weighted regression. If "simple" (the default option), then the IVW estimate is equivalent to meta-analysing the ratio estimates from each variant using inverse-variance weights based on the simplest expression of the variance for the ratio estimate (first-order term from the delta

expansion - standard error of the association with the outcome divided by the association with the exposure). If "delta", then the variance expression is the second-order term from the delta expansion. The second-order term incorporates uncertainty in the genetic association with the exposure – this uncertainty is ignored using the simple weighting.

psi — The correlation between the genetic associations with the exposure and the association with the outcome for each variant resulting from sample overlap. The default value is 0, corresponding to a strict two-sample Mendelian randomization analysis (no overlap). If there is complete overlap between the samples, then the correlation should be set to the observational correlation between the exposure and the outcome. This correlation is only used in the calculation of standard errors if the option weights is set to "delta".

correl — If the genetic variants are correlated, then this correlation can be accounted for. The matrix of correlations between must be provided in the MRInput object: the elements of this matrix are the correlations between the individual variants (diagonal elements are 1). If a correlation matrix is specified in the MRInput object, then correl is set to TRUE. If correl is set to TRUE, then the values of robust and penalized are taken as FALSE, and weights is set to "simple".

distribution — The type of distribution used to calculate the confidence intervals. Options are "normal" (default) or "t-dist".

alpha — The significance level used to calculate the confidence interval. The default value is 0.05.

... — Additional arguments to be passed to the regression method.

Details

With multiple uncorrelated genetic variants, this estimate can be thought of as: 1) the inverse-variance weighted combination of the ratio estimates from a meta-analysis; 2) the ratio estimate from combining the genetic variants into a weighted score and then using this score as an instrumental variable (the same estimate is obtained from the two-stage least squares method using individual-level data); 3) the coefficient from weighted linear regression of the associations with the outcome on the associations with the risk factor fixing the intercept to zero and using the inverse-variance weights.

Here, we implement the method using weighted linear regression. If the variants are correlated, the method is implemented using generalized weighted linear regression; this is hard coded using matrix algebra.

The causal estimate is obtained by regression of the associations with the outcome on the associations with the risk factor, with the intercept set to zero and weights being the inverse-variances of the associations with the outcome.

With a single genetic variant, the estimate is the ratio of coefficients betaY/betaX and the standard error is the first term of the delta method approximation betaYse/betaX.

Value

The output from the function is an IVW object containing:

Model — A character string giving the type of model used ("fixed", "random", or "default").

Exposure	A character string giving the name given to the exposure.
Outcome	A character string giving the name given to the outcome.
Correlation	The matrix of genetic correlations.
Robust	TRUE if robust regression has been used to calculate the estimate, FALSE otherwise.
Penalized	TRUE if weights have been penalized, FALSE otherwise.
Estimate	The value of the causal estimate.
StdError	Standard error of the causal estimate.
CILower	The lower bound of the causal estimate based on the estimated standard error and the significance level provided.
CIUpper	The upper bound of the causal estimate based on the estimated standard error and the significance level provided.
Alpha	The significance level used when calculating the confidence intervals.
Pvalue	The p-value associated with the estimate (calculated as Estimate/StdError as per Wald test) using a normal or t-distribution (as specified in distribution).
SNPs	The number of genetic variants (SNPs) included in the analysis.
RSE	The estimated residual standard error from the regression model.
Heter.Stat	Heterogeneity statistic (Cochran's Q statistic) and associated p-value: the null hypothesis is that all genetic variants estimate the same causal parameter; rejection of the null is an indication that one or more variants may be pleiotropic.
Fstat	An approximation of the first-stage F statistic for all variants based on the summarized data.

References

Original implementation: The International Consortium for Blood Pressure Genome-Wide Association Studies. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature 2011; 478:103-109. doi: 10.1038/nature10405.

Detailed description of method: Stephen Burgess, Adam S Butterworth, Simon G Thompson. Mendelian randomization analysis with multiple genetic variants using summarized data. Genetic Epidemiology 2013; 37:658-665. doi: 10.1002/gepi.21758.

Robust and penalized weights: Stephen Burgess, Jack Bowden, Frank Dudbridge, Simon G Thompson. Robust instrumental variable methods using multiple candidate instruments with application to Mendelian randomization. arXiv 2016; 1606.03729.

Heterogeneity test: Fabiola del Greco, Cosetta Minelli, Nuala A Sheehan, John R Thompson. Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. Stat Med 2015; 34(21):2926-2940. doi: 10.1002/sim.6522.

Simple versus delta weights (first-order versus second-order): Stephen Burgess, Jack Bowden. Integrating summarized data from multiple genetic variants in Mendelian randomization: bias and coverage properties of inverse-variance weighted methods. arXiv:1512.04486.

Examples

```
mr_ivw(mr_input(bx = ldlc, bxse = ldlcse, by = chdlodds, byse = chdloddsse))
mr_ivw(mr_input(bx = ldlc, bxse = ldlcse, by = chdlodds, byse = chdloddsse),
  robust = TRUE)
mr_ivw(mr_input(bx = ldlc, bxse = ldlcse, by = chdlodds, byse = chdloddsse),
  penalized = TRUE)
mr_ivw(mr_input(calcium, calciumse, fastgluc, fastglucse, corr=calc.rho))
  ## correlated variants
```

---

mr_lasso *MR-Lasso method*

Description

The `mr_lasso` function performs the MR-Lasso method, which applies lasso-type penalization to the direct effects of genetic variants on the outcome. The causal estimate is described as a post-lasso estimate, and is obtained by performing the IVW method using only those genetic variants that are identified as valid by the lasso procedure.

Usage

```
mr_lasso(object, distribution = "normal", alpha = 0.05, lambda = numeric(0))

## S4 method for signature 'MRInput'
mr_lasso(object, distribution = "normal", alpha = 0.05, lambda = numeric(0))
```

Arguments

`object`	An `MRInput` object.
`distribution`	The type of distribution used to calculate the confidence intervals. Options are `"normal"` (default) or `"t-dist"`.
`alpha`	The significance level used to calculate the confidence intervals. The default value is 0.05.
`lambda`	The value of the tuning parameter used by the lasso procedure which controls the level of sparsity. If not specified, the tuning parameter will be calculated by the heterogeneity stopping rule.

Details

MR-Lasso extends the IVW model to include an intercept term for each genetic variant. These intercept terms represent associations between the genetic variants and the outcome which bypass the risk factor. The causal effect estimates are estimated by weighted linear regression where the intercept terms are subject to lasso-type penalization. The lasso penalization will tend to shrink the intercept terms corresponding to the valid instruments to zero.

The lasso penalty relies on a tuning parameter which controls the level of sparsity. The default is to use a heterogeneity stopping rule, but a fixed value may be specified.

As part of the analysis, the genetic variants are orientated so that all of the associations with the risk factor are positive (and signs of associations with the outcome are changed to keep the orientation consistent if required). Re-orientation of the genetic variants is performed automatically as part of the function.

The MR-Lasso method is performed in two steps. First, a regularized regression model is fitted, and some genetic variants are identified as valid instruments. Second, the causal effect is estimated using standard IVW with only the valid genetic variants. The post-lasso method will be performed as long as at least two genetic variants are identified as valid instruments. The default heterogeneity stopping rule will always return at least two genetic variants as valid instruments. The main estimate given by the method is the post-lasso estimate. However, parameter estimates from the regularized regression model used to identify invalid variants are also provided for completeness.

If a substantial proportion of genetic variants are removed from the analysis, the MR-Lasso method may give a false impression of confidence in the causal estimate due to homogeneity of the variant-specific causal estimates amongst the remaining variants. However, it is not reasonable to claim that there is strong evidence for a causal effect after a large number of variants with heterogeneous estimates have been removed from the analysis.

Value

The output from the function is an `MRLasso` object containing:

`Exposure`	A character vector with the names given to the exposure.
`Outcome`	A character string with the names given to the outcome.
`Estimate`	The causal estimate from the MR-Lasso method. This is the post-lasso estimate.
`StdError`	The standard error of the causal estimate from the MR-Lasso method.
`CILower`	The lower bound of the confidence interval for the causal estimate based on the estimated standard error and the significance level provided.
`CIUpper`	The upper bound of the confidence interval for the causal estimate based on the estimated standard error and the significance level provided.
`Alpha`	The significance level used when calculating the confidence intervals.
`Pvalue`	The p-value associated with the causal estimate using a normal or t-distribution (as specified in `distribution`).
`SNPs`	The number of genetic variants (SNPs) included in the analysis.
`RegEstimate`	The estimate from the regularized regression model used in the MR-Lasso method.
`RegIntercept`	The intercept estimates from the regularized regression model used in the MR-Lasso method.
`Valid`	The number of genetic variants that have been identified as valid instruments.
`ValidSNPs`	The names of genetic variants that have been identified as valid instruments.
`Lambda`	The value of the tuning parameter used to compute `RegEstimate`

References

Jessica MB Rees, Angela M Wood, Frank Dudbridge, Stephen Burgess. Robust methods in Mendelian randomization via penalization of heterogeneous causal estimates. PLoS ONE 2019; 14(9):e0222362

Examples

```
mr_lasso(mr_input(bx = ldlc, bxse = ldlcse, by = chdlodds, byse = chdloddsse))
```

`mr_loo` *Leave-one-out estimates*

Description

The `mr_loo` function draws a forest plot of causal estimates from the `mr_ivw` function using default settings (assuming variants are uncorrelated, random-effects for 4+ variants) omitting each variant in turn. So the estimate labelled `snp_1` includes all variants except the labelled variant, and so on. The `mr_ivw` estimate including all variants ("IVW estimate") is also provided for reference.

Usage

```
mr_loo(object, alpha = 0.05)

## S4 method for signature 'MRInput'
mr_loo(object, alpha = 0.05)
```

Arguments

`object`	An `MRInput` object.
`alpha`	The significance level used to calculate the confidence intervals. The default value is 0.05, corresponding to 95% confidence intervals.

Details

As the function produces a `ggplot` object, graphical parameters can be changed by adding commands from the `ggplot2` package.

Examples

```
mr_loo(mr_input(bx = ldlc, bxse = ldlcse, by = chdlodds, byse = chdloddsse),
  alpha = 0.01)
```

mr_maxlik *Maximum-likelihood method*

Description

The mr_maxlik function implements the maximum-likelihood method introduced by Burgess et al (2013).

Usage

```
mr_maxlik(
  object,
  model = "default",
  correl = FALSE,
  psi = 0,
  distribution = "normal",
  alpha = 0.05,
  ...
)

## S4 method for signature 'MRInput'
mr_maxlik(
  object,
  model = "default",
  correl = FALSE,
  psi = 0,
  distribution = "normal",
  alpha = 0.05,
  ...
)
```

Arguments

object	An MRInput object.
model	What type of model should be used: "default", "random" or "fixed". The method naturally estimates a fixed-effect model, assuming that the same causal effect is estimated by each of the genetic variants. However, if there is heterogeneity in the causal estimates of the different variants, then confidence intervals under a fixed-effect model will be overly narrow. The random-effects model adds additional uncertainty by multiplying the standard error by the square-root of the likelihood ratio heterogeneity statistic divided by the number of genetic variants less one (unless this quantity is less than 1, in which case no modification to the standard error is made). This parallels the residual standard error in a regression model (the Cochran Q heterogeneity test statistic is equal to the square of the RSE multiplied by the number of genetic variants less one). The default setting ("default") is to use a fixed-effect model with 3 genetic variants or fewer, and otherwise to use a random-effects model.

correl	If the genetic variants are correlated, then this correlation can be accounted for. The matrix of correlations between must be provided in the MRInput object: the elements of this matrix are the correlations between the individual variants (diagonal elements are 1).
psi	The correlation between the association with the exposure and the association with the outcome for each variant resulting from sample overlap.
distribution	The type of distribution used to calculate the confidence intervals, can be "normal" (the default option) or "t-dist".
alpha	The significance level used to calculate the confidence interval. The default value is 0.05.
...	Additional arguments to be passed to the optimization method.

Details

A likelihood function is defined by assuming that the summarized data for each genetic variant are normally distributed. A bivariate normal distribution is assumed for the associations of each genetic variant with the exposure and with the outcome. The mean of the association with the outcome is taken as the mean association with the exposure multiplied by the causal effect parameter.

Thus, if there are K genetic variants, then K+1 parameters are estimated by the method: one for each gene–exposure association, plus the causal parameter. If the number of genetic variants is large, then maximization of this function may be an issue. If the maximum likelihood estimate substantially differs from the inverse-variance weighted estimate, this may indicate that convergence has not occurred in the optimization algorithm.

The variance-covariance matrices for the bivariate normal distributions are obtained from the standard error estimates provided. The correlation psi between genetic associations with the exposure and with the outcome due to sample overlap can be specified; its default value is zero.

Two features why this method may be preferred over the inverse-variance weighted method are the incorporation in the model of uncertainty in the genetic associations with the exposure, and of correlation between the genetic association estimates with exposure and outcome for each variant. The method is implemented both for uncorrelated and correlated genetic variants. It can also be used for a single genetic variant.

The original version of the maximum-likelihood method assumed that all genetic variants identify the same causal estimate; a fixed-effect model. The causal estimate may be overly precise if the fixed-effect model is incorrect and there is substantial heterogeneity in the causal estimates from the different variants. The random-effects analysis implemented here is an ad hoc solution to the problem of heterogeneity, but one that should result in reasonable confidence intervals that incorporate this heterogeneity.

Value

The output from the function is an MaxLik object containing:

Model	A character string giving the type of model used ("fixed", "random", or "default").
Exposure	A character string giving the name given to the exposure.
Outcome	A character string giving the name given to the outcome.
Correlation	The matrix of genetic correlations.

Psi	The correlation between genetic associations with the exposure and with the outcome.
Estimate	The value of the causal estimate.
StdError	Standard error of the causal estimate.
CILower	The lower bound of the causal estimate based on the estimated standard error and the significance level provided.
CIUpper	The upper bound of the causal estimate based on the estimated standard error and the significance level provided.
Alpha	The significance level used when calculating the confidence intervals.
Pvalue	The p-value associated with the estimate (calculated as Estimate/StdError as per Wald test) using a normal or t-distribution (as specified in `distribution`).
SNPs	The number of genetic variants (SNPs) included in the analysis.
RSE	The estimated residual standard error from the regression model (always equal to 1, as a fixed-effect model is required.
Heter.Stat	Heterogeneity statistic (likelihood ratio statistic) and associated p-value: the null hypothesis is that all genetic variants estimate the same causal parameter; rejection of the null is an indication that one or more variants may be pleiotropic.

References

Stephen Burgess, Adam S Butterworth, Simon G Thompson. Mendelian randomization analysis with multiple genetic variants using summarized data. Genetic Epidemiology 2013; 37:658-665. doi: 10.1002/gepi.21758.

Examples

```
mr_maxlik(mr_input(bx = ldlc[1:10], bxse = ldlcse[1:10],
  by = chdlodds[1:10], byse = chdloddsse[1:10]))
mr_maxlik(mr_input(bx = ldlc[1:10], bxse = ldlcse[1:10],
  by = chdlodds[1:10], byse = chdloddsse[1:10]), psi=0.2)
mr_maxlik(mr_input(calcium, calciumse, fastgluc, fastglucse, corr=calc.rho))
  ## correlated variants
```

---

mr_mbe *Mode-based method of Hartwig*

Description

The `mr_mbe` function implements the mode-based method introduced by Hartwig, Bowden and Davey Smith (2017).

Usage

```
mr_mbe(
  object,
  weighting = "weighted",
  stderror = "simple",
  phi = 1,
  seed = 314159265,
  iterations = 10000,
  distribution = "normal",
  alpha = 0.05
)

## S4 method for signature 'MRInput'
mr_mbe(
  object,
  weighting = "weighted",
  stderror = "delta",
  phi = 1,
  seed = 314159265,
  iterations = 10000,
  distribution = "normal",
  alpha = 0.05
)
```

Arguments

object	An MRInput object.
weighting	Whether the analysis should be "weighted" (the default option) or "unweighted".
stderror	Whether standard error estimates should be i) "simple" - calculated as the first-order term from the delta expansion - standard error of the association with the outcome divided by the association with the exposure), or ii) "delta" - calculated as the second-order term from the delta expansion (the default option). The second-order term incorporates uncertainty in the genetic association with the exposure – this uncertainty is ignored using the simple weighting. The "simple" option is referred to by Hartwig et al as "assuming NOME", and the "delta" option as "not assuming NOME".
phi	The choice of bandwidth in the kernel-smoothly density method. A value of 1 (the default value) represents the bandwidth value selected by the modified Silverman's bandwidth rule, as recommended by Hartwig et al. A value of 0.5 represents half that value, and so on.
seed	The random seed to use when generating the bootstrap samples used to calculate the confidence intervals (for reproducibility). The default value is 314159265. If set to NA, the random seed will not be set (for example, if the function is used as part of a larger simulation).
iterations	Number of iterations to use in the bootstrap procedure.
distribution	The type of distribution used to calculate the confidence intervals, can be "normal" (the default option) or "t-dist".

alpha — The significance level used to calculate the confidence interval. The default value is 0.05.

Details

The mode-based estimation (MBE) method takes the variant-specific ratio estimates from each genetic variant in turn, and calculates the modal estimate. This is implemented by constructing a kernel-smoothed density out of the ratio estimates, and taking the maximum value as the modal estimate. The standard error is calculated by a bootstrap procedure, and confidence intervals based on the estimate having a normal distribution.

The method should give consistent estimates as the sample size increases if a plurality (or weighted plurality) of the genetic variants are valid instruments. This means that the largest group of variants with the same causal estimate in the asymptotic limit are the valid instruments.

Value

The output from the function is an MRMBE object containing:

Exposure	A character string giving the name given to the exposure.
Outcome	A character string giving the name given to the outcome.
Weighting	A character string "weighted" or "unweighted".
StdErr	A character string "simple" or "delta".
Phi	The value of the bandwidth factor.
Estimate	The value of the causal estimate.
StdError	Standard error of the causal estimate.
CILower	The lower bound of the causal estimate based on the estimated standard error and the significance level provided.
CIUpper	The upper bound of the causal estimate based on the estimated standard error and the significance level provided.
Alpha	The significance level used when calculating the confidence intervals.
Pvalue	The p-value associated with the estimate (calculated as Estimate/StdError as per Wald test) using a normal or t-distribution (as specified in distribution).
SNPs	The number of genetic variants (SNPs) included in the analysis.

References

Fernando Pires Hartwig, George Davey Smith, Jack Bowden. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. International Journal of Epidemiology 2017; 46(6): 1985-1998. doi: 10.1093/ije/dyx102.

Examples

```
mr_mbe(mr_input(bx = ldlc, bxse = ldlcse, by = chdlodds, byse = chdloddsse), iterations=100)
mr_mbe(mr_input(bx = ldlc, bxse = ldlcse, by = chdlodds, byse = chdloddsse),
   phi=0.5, iterations=100)
mr_mbe(mr_input(bx = ldlc, bxse = ldlcse, by = chdlodds, byse = chdloddsse),
```

```
   weighting="weighted", stderror="delta", iterations=100)
# iterations set to 100 to reduce computational time,
#  more iterations are recommended in practice
```

`mr_median` *Median-based method*

Description

The `mr_median` function implements the weighted median (default) or simple median method introduced by Bowden et al (2016) to calculate the median of the ratio instrumental variable estimates evaluated using each genetic variant individually.

Usage

```
mr_median(
  object,
  weighting = "weighted",
  distribution = "normal",
  alpha = 0.05,
  iterations = 10000,
  seed = 314159265
)

## S4 method for signature 'MRInput'
mr_median(
  object,
  weighting = "weighted",
  distribution = "normal",
  alpha = 0.05,
  iterations = 10000,
  seed = 314159265
)
```

Arguments

`object`	An `MRInput` object.
`weighting`	The type of weighting applied. The default option is to calculate the weighted median (`"weighted"`); other options are `"simple"` and `"penalized"`.
`distribution`	The type of distribution to use to calculate the 95% confidence intervals, can be `"normal"` or `"t-dist"`.
`alpha`	The significance level used to calculate the confidence intervals. The default value is 0.05.
`iterations`	The number of bootstrap samples to generate when calculating the estimated standard error. The default value is 10000.

seed — The random seed to use when generating the bootstrap samples (for reproducibility). The default value is 314159265. If set to NA, the random seed will not be set (for example, if the function is used as part of a larger simulation).

Details

The median-based methods have greater robustness to individual genetic variants with strongly outlying causal estimates compared with the inverse-variance weighted and MR-Egger methods. Formally, the simple median method gives a consistent estimate of the causal effect when at least 50% of the genetic variants are valid instrumental variables (for the weighted median method, when 50% of the weight comes from valid instrumental variables).

When the weighting is "simple", the estimate is obtained by calculating the ratio causal estimates from each genetic variants theta = betaY/betaX, and finding the median estimate.

When the weighting is "weighted", the estimate is obtained by:

1. Calculating the ratio causal estimates and ordering the genetic variants according to the magnitude of their estimates, i.e.
$$\theta_1 < \theta_2 < ... < \theta_J$$
2. Calculate normalized inverse-variance weights for each genetic variant $w_1, w_2, ..., w_J$, as:
$$w_j = \frac{\beta_{Xj}^2}{se(\beta_{Yj})^2} / \sum_{i=1}^{J} \frac{\beta_{Xi}^2}{se(\beta_{Yi})^2}$$
3. Find k such that
$$s_k = \sum_{i=1}^{k} w_i < 0.5$$
and
$$s_{k+1} = \sum_{i=1}^{k+1} w_i > 0.5$$
4. Calculate the weighted median estimate by extrapolation as:
$$\theta_{WM} = \theta_k + (\theta_{k+1} - \theta_k) \times \frac{0.5 - s_k}{s_{k+1} - s_k}$$

The simple median estimate is the same as the weighted median estimate when all the weights are equal. Standard errors for both the simple and weighted median methods are calculated through bootstrapping.

When the weighting is "penalized", the weighted method is used, but the contribution of genetic variants with outlying (heterogeneous) ratio estimates to the analysis is downweighted.

Value

The output from the function is a WeightedMedian object containing:

Type — The type of weights used: "weighted", "simple", or "penalized".

Exposure — A character string giving the name given to the exposure.

Outcome	A character string giving the name given to the outcome.
Estimate	The value of the causal estimate.
StdError	Standard error of the causal estimate calculated using bootstrapping.
CILower	The lower bound for the causal estimate based on the estimated bootstrapped standard error and the significance level provided.
CIUpper	The upper bound for the causal estimate based on the estimated bootstrapped standard error and the significance level provided.
Alpha	The significance level used when calculating the confidence intervals.
Pvalue	The p-value associated with the estimate (calculated using Estimate/StdError as per a Wald test) using a normal or t-distribution (as specified in distribution).
SNPs	The number of genetic variants (SNPs) included in the analysis.

References

Jack Bowden, George Davey Smith, Philip C Haycock, Stephen Burgess. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. Genetic Epidemiology 2016; 40(4):304-314. doi: 10.1002/gepi.21965.

Examples

```
mr_median(mr_input(bx = ldlc, bxse = ldlcse, by = chdlodds, byse = chdloddsse),
  weighting = "weighted", iterations = 100)
  # iterations is set to 100 to reduce runtime for the mr_median method,
  # 10000 iterations are recommended in practice
mr_median(mr_input(bx = ldlc, bxse = ldlcse, by = chdlodds, byse = chdloddsse),
  weighting = "simple", iterations = 100)
mr_median(mr_input(bx = ldlc, bxse = ldlcse, by = chdlodds, byse = chdloddsse),
  weighting = "penalized", iterations = 100)
```

---

mr_mvcML *Multivariable constrained maximum likelihood method*

Description

The mr_mvcML function performs multivariable Mendelian randomization via the constrained maximum likelihood method, which is robust to both correlated and uncorrelated pleiotropy.

Usage

```
mr_mvcML(
  object,
  n,
  DP = TRUE,
  rho_mat = diag(ncol(object@betaX) + 1),
  K_vec = 0:(ceiling(nrow(object@betaX)/2)),
```

```
  random_start = 0,
  num_pert = 100,
  min_theta_range = -0.5,
  max_theta_range = 0.5,
  maxit = 100,
  alpha = 0.05,
  seed = 314159265
)

## S4 method for signature 'MRMVInput'
mr_mvcML(
  object,
  n,
  DP = TRUE,
  rho_mat = diag(ncol(object@betaX) + 1),
  K_vec = 0:(ceiling(nrow(object@betaX)/2)),
  random_start = 0,
  num_pert = 200,
  min_theta_range = -0.5,
  max_theta_range = 0.5,
  maxit = 100,
  alpha = 0.05,
  seed = 314159265
)
```

Arguments

object	An MRMVInput object.
n	Sample size. The smallest sample size among all (both exposures and outcome) GWAS used in the analysis is recommended.
DP	Whether data perturbation is applied or not. Default is TRUE.
rho_mat	The correlation matrix among the exposures and outcome GWAS estimates, which can be estimated by the intercept term from bivariate LDSC. See reference for more discussions. Default is the identify matrix, for example, in the absence of overlapping samples among GWAS datasets.
K_vec	Set of candidate K's, the constraint parameter representing number of invalid IVs. It can range from 0 up to #IV - (#exposure + 1). Default is from 0 to (#IV/2).
random_start	Number of random starting points for MVMRcML, default is 0.
num_pert	Number of perturbation when DP is TRUE, default is 200.
min_theta_range	The lower bound of the uniform distribution for each initial value for theta generated from, default is -0.5.
max_theta_range	The uppder bound of the uniform distribution for each initial value for theta generated from, default is 0.5.

maxit	Maximum number of iterations for each optimization. Default is 100.
alpha	Significance level for the confidence interval for estimate, default is 0.05.
seed	The random seed to use when generating the perturbed samples (for reproducibility). The default value is 314159265. If set to NA, the random seed will not be set (for example, if the function is used as part of a larger simulation).

Details

Multivariable MRcML (MVMRcML) is an extension of MRcML to deal with multiple exposures of interest. It is robust to both correlated and uncorrelated pleiotropy as its univariable version.

In practice, the data perturbation (DP) version is preferred in practice for a more robust inference as it can account for the uncertainty in model selection. However, it may take a longer time especially when the number of IVs is large (so the range of K_vec can be large too). One strategy is to try a small range of K (the number of invalid IVs) first (with a small num_pert), then adjust it if the number of selected invalid IVs fall close to the boundary. You can also use other methods, e.g. mr_mvlasso, to get a rough sense of the number of invalid IVs.

Similar to mr_cML, multiple random starting points could be used to find a global minimum.

Value

The output from the function is an MVMRcML object containing:

Exposure	A character vector with the names given to the exposure.
Outcome	A character string with the names given to the outcome.
Estimate	A vector of causal estimates.
StdError	A vector of standard errors of the causal estimates.
CILower	The lower bounds of the causal estimates based on the estimated standard errors and the significance level provided.
CIUpper	The upper bounds of the causal estimates based on the estimated standard errors and the significance level provided.
Alpha	The significance level used when calculating the confidence intervals.
Pvalue	The p-values associated with the estimates (calculated as Estimate/StdError as per Wald test) using a normal distribution.
BIC_invalid	Set of selected invalid IVs by MVMRcML-BIC.
K_hat	The number of selected invalid IVs by MVMRcML-BIC, or a vector for each data perturbation in MVMRcML-DP.
eff_DP_B	The number of data perturbations with successful convergence in MVMRcML-DP.
SNPs	The number of genetic variants (SNPs) included in the analysis.

References

Lin, Z., Xue, H., & Pan, W. (2023). Robust multivariable Mendelian randomization based on constrained maximum likelihood. The American Journal of Human Genetics, 110(4), 592-605.

Examples

```
# Perform MVMRcML-DP:
mr_mvcML(mr_mvinput(bx = cbind(ldlc, hdlc, trig), bxse = cbind(ldlcse, hdlcse, trigse),
   by = chdlodds, byse = chdloddsse), n = 17723, num_pert = 5, random_start = 5)
  # num_pert is set to 5 to reduce runtime for the mr_mvcML method,
  # At least 100 perturbations should be used and more is preferred for a stable result.

rho_mat = matrix(c(1,-0.1,0.2,0,-0.1,1,-0.3,0,
       0.2,-0.3,1,0,0,0,0,1),ncol=4) ## Toy example of rho_mat
mr_mvcML(mr_mvinput(bx = cbind(ldlc, hdlc, trig), bxse = cbind(ldlcse, hdlcse, trigse),
   by = chdlodds, byse = chdloddsse), n = 17723, num_pert = 5, rho_mat = rho_mat)

# Perform MVMRcML-BIC:
mr_mvcML(mr_mvinput(bx = cbind(ldlc, hdlc, trig), bxse = cbind(ldlcse, hdlcse, trigse),
   by = chdlodds, byse = chdloddsse), n = 17723, DP = FALSE)
```

`mr_mvegger` *Multivariable MR-Egger method*

Description

The `mr_mvegger` function performs multivariable Mendelian randomization via the MR-Egger method. This is implemented by multivariable weighted linear regression.

Usage

```
mr_mvegger(
  object,
  orientate = 1,
  correl = FALSE,
  distribution = "normal",
  alpha = 0.05
)

## S4 method for signature 'MRMVInput'
mr_mvegger(
  object,
  orientate = 1,
  correl = FALSE,
  distribution = "normal",
  alpha = 0.05
)
```

Arguments

`object`	An `MRMVInput` object.

orientate	The risk factor that genetic associations are orientated to. The univariable and multivariable versions of MR-Egger are both sensitive to the choice of parameterization of the genetic associations - which allele the associations are orientated with respect to (in other words, which allele is the effect allele). For univariable MR-Egger, this is resolved by setting the genetic associations with the exposure all to be positive. In multivariable MR-Egger, we have to choose which of the exposures to orientate the genetic associations to. The default option is 1, meaning that genetic associations with the first exposure are set to be positive.
correl	If the genetic variants are correlated, then this correlation can be accounted for. The matrix of correlations between must be provided in the MRInput object: the elements of this matrix are the correlations between the individual variants (diagonal elements are 1). If a correlation matrix is specified in the MRInput object, then correl is set to TRUE.
distribution	The type of distribution used to calculate the confidence intervals. Options are "normal" (default) or "t-dist".
alpha	The significance level used to calculate the confidence interval. The default value is 0.05.

Details

Multivariable MR-Egger is an extension of the MR-Egger method to deal with genetic variants that are associated with multiple risk factors.

We implement the method using multivariable weighted linear regression. If the variants are correlated, the method is implemented using generalized weighted linear regression; this is hard coded using matrix algebra.

The causal estimate is obtained by regression of the associations with the outcome on the associations with the risk factors, with the intercept estimated and weights being the inverse-variances of the associations with the outcome.

Value

The output from the function is an MVEgger object containing:

Model	A character string giving the type of model used ("random").
Orientate	The number corresponding to the risk factor that the genetic associations are orientated to.
Exposure	A character vector with the names given to the exposure.
Outcome	A character string with the names given to the outcome.
Correlation	The matrix of genetic correlations.
Estimate	A vector of the causal estimates (slope coefficient).
StdError.Est	Standard errors of the causal estimates.
Pvalue.Est	The p-values associated with the estimates using a normal or t-distribution (as specified in distribution).
CILower.Est	The lower bound of the causal estimates based on the estimated standard error and the significance level provided.

CIUpper.Est	The upper bound of the causal estimates based on the estimated standard error and the significance level provided.
Intercept	The value of the intercept estimate.
StdError.Int	Standard error of the intercept estimate.
Pvalue.Int	The p-value associated with the intercept.
CILower.Int	The lower bound of the intercept based on the estimated standard error and the significance level provided.
CIUpper.Int	The upper bound of the intercept based on the estimated standard error and the significance level provided.
Alpha	The significance level used when calculating the confidence intervals.
Pvalue	The p-values associated with the estimates (calculated as Estimate/StdError as per Wald test) using a normal or t-distribution (as specified in distribution).
SNPs	The number of genetic variants (SNPs) included in the analysis.
RSE	The estimated residual standard error from the regression model.
Heter.Stat	Heterogeneity statistic (Cochran's Q statistic) and associated p-value: the null hypothesis is that all genetic variants estimate the same causal parameter; rejection of the null is an indication that one or more variants may be pleiotropic.

References

Jessica Rees, Angela Wood, Stephen Burgess. Extending the MR-Egger method for multivariable Mendelian randomization to correct for both measured and unmeasured pleiotropy. Statistics in Medicine 2017; 36(29): 4705-4718. doi: 10.1002/sim.7492.

Examples

```
mr_mvegger(mr_mvinput(bx = cbind(ldlc, hdlc, trig), bxse = cbind(ldlcse, hdlcse, trigse),
   by = chdlodds, byse = chdloddsse), orientate = 1)
```

---

mr_mvgmm *Multivariable generalized method of moments (GMM) method*

Description

The mr_mvgmm function performs multivariable Mendelian randomization via the generalized method of moments method.

Usage

```
mr_mvgmm(object, nx, ny, cor.x = NULL, robust = TRUE, alpha = 0.05, ...)

## S4 method for signature 'MRMVInput'
mr_mvgmm(object, nx, ny, cor.x = NULL, robust = TRUE, alpha = 0.05, ...)
```

Arguments

object	An MRMVInput object.
nx	Vector of sample sizes used to compute genetic associations with the exposure (one for each exposure).
ny	The sample size used to compute genetic associations with the outcome.
cor.x	Correlation matrix for exposures. Default is to assume the exposures are uncorrelated.
robust	Indicates whether overdispersion heterogeneity is accounted for in the model. Default is TRUE.
alpha	The significance level used to calculate the confidence interval. The default value is 0.05.
...	Additional arguments to be passed to the optimization routines to calculate GMM estimates and overdispersion parameter.

Details

Robust inference in two-sample multivariable Mendelian randomization using the generalized method of moments. The method accounts for overdispersion heterogeneity in genetic variant-outcome associations.

Value

The output from the function is an MVGMM object containing:

Robust	TRUE if overdispersion heterogeneity was included in the model, FALSE otherwise.
Exposure	A character vector with the names given to the exposure.
Outcome	A character string with the names given to the outcome.
Correlation	The matrix of genetic correlations if supplied. If not supplied, then an identity matrix will be returned.
ExpCorrelation	TRUE if an exposure correlation matrix was specified, FALSE otherwise.
CondFstat	A vector of conditional F-statistics (one for each exposure).
Estimate	A vector of causal estimates.
StdError	A vector of standard errors of the causal estimates.
CILower	The lower bounds of the causal estimates based on the estimated standard errors and the significance level provided.
CIUpper	The upper bounds of the causal estimates based on the estimated standard errors and the significance level provided.
Overdispersion	The estimate of the overdispersion parameter. If this is negative, then a value of zero is used in the method.
Pvalue	The p-values associated with the estimates (calculated as Estimate/StdError as per Wald test) using a normal distribution.
Alpha	The significance level used when calculating the confidence intervals.

Heter.Stat — Heterogeneity statistic (Cochran's Q statistic) and associated p-value (for non-robust model only): the null hypothesis is that all genetic variants estimate the same causal parameter; rejection of the null is an indication that one or more genetic variants may be pleiotropic. If the robust option is set to TRUE, then this represents excess heterogeneity beyond the overdispersion model.

References

Description of the generalized method of moments: Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. Econometrica, pp.1029-1054.

Examples

```
mr_mvgmm(mr_mvinput(bx = cbind(ldlc, hdlc, trig), bxse = cbind(ldlcse, hdlcse, trigse),
   by = chdlodds, byse = chdloddsse), nx=rep(17723,3), ny=17723)
```

---

mr_mvinput *Inputting and formatting data for use in causal estimation*

Description

The mr_mvinput function is required for inputting and formatting data for use in the multivariable Mendelian randomization functions provided in this package.

Usage

```
mr_mvinput(
  bx = matrix(),
  bxse = matrix(),
  by = 0,
  byse = 0,
  correlation = matrix(),
  exposure = "exposure",
  outcome = "outcome",
  snps = "snp",
  effect_allele = NA,
  other_allele = NA,
  eaf = NA
)
```

Arguments

bx — A matrix of beta-coefficient values for genetic associations with the risk factor variables. These should be arranged so that column 1 are the beta-coefficients for risk factor 1, and row 1 are the beta-coefficients for genetic variant 1.

bxse — The matrix of standard errors associated with the beta-coefficients bx.

by	A numeric vector of beta-coefficient values for genetic associations with the second variable (often referred to as the outcome). For a disease outcome, the beta coefficients are log odds estimates from logistic regression analyses.
byse	The vector standard errors associated with the beta-coefficients in by.
correlation	The matrix of correlations between genetic variants. If this variable is not provided, then we assume that genetic variants are uncorrelated.
exposure	The names of the exposure variables.
outcome	The name of the outcome variable.
snps	The names of the genetic variants (SNPs) included in the analysis. The inputs exposure, outcome, and snps are not required, but may be useful for keeping track of various MRInput objects. They are also used by the mr_plot function.
effect_allele	The name of the effect allele for each SNP. The beta-coefficients are the associations with the exposure and outcome per additional copy of the effect allele.
other_allele	The name of the non-effect allele.
eaf	The expected allele frequencies (numeric). The slots effect_allele, other_allele, and eaf are neither required, nor currently used in the MendelianRandomization package. They are included for future compatibility with the MR-Base suite of functions.

Details

The beta-coefficients are assumed to be estimated for uncorrelated (independent) genetic variants, although a correlation matrix can be specified if the variants are correlated in their distributions. We also assume that the beta-coefficients for associations with the exposure and with the outcome are uncorrelated (corresponding to a two-sample Mendelian randomization analysis), although correlation between associations with the exposure and with the outcome generally have little impact on causal estimates or standard errors.

If the variables are not all the same length, then an error message will be reported. The analyses will still try to run, but the output may be misleading. However, in some analyses (for example, the standard IVW and MR-Egger methods), the values of bxse are not used in the analysis, and can therefore safely be omitted (provided that the other variables are correctly labelled).

Value

An MRMVInput object containing:

betaX	The genetic associations with the exposures.
betaXse	The corresponding standard errors.
betaY	The genetic associations with the outcome.
betaYse	The corresponding standard errors.
correlation	The matrix of genetic correlations.
exposure	Character strings with the names given to the exposures.
outcome	A character string giving the name given to the outcome.
snps	A vector of character strings with the names of the genetic variants.

effect_allele	A vector of character strings with the names of the effect alleles.
other_allele	A vector of character strings with the names of the non-effect alleles.
eaf	A numeric vector with the effect allele frequencies.

---

mr_mvivw *Multivariable inverse-variance weighted method*

Description

The mr_mvivw function performs multivariable Mendelian randomization via the inverse-variance method. This is implemented by multivariable weighted linear regression.

Usage

```
mr_mvivw(
  object,
  model = "default",
  robust = FALSE,
  correl = FALSE,
  correl.x = NULL,
  nx = NA,
  distribution = "normal",
  alpha = 0.05,
  ...
)

## S4 method for signature 'MRMVInput'
mr_mvivw(
  object,
  model = "default",
  robust = FALSE,
  correl = FALSE,
  correl.x = NULL,
  nx = NA,
  distribution = "normal",
  alpha = 0.05,
  ...
)
```

Arguments

object	An MRMVInput object.
model	What type of model should be used: "default", "random" or "fixed". The random-effects model ("random") is a multiplicative random-effects model, allowing overdispersion in the weighted linear regression (the residual standard

	error is not fixed to be 1, but is not allowed to take values below 1). The fixed-effect model ("fixed") sets the residual standard error to be 1. The "default" setting is to use a fixed-effect model with 3 genetic variants or fewer, and otherwise to use a random-effects model.
robust	Indicates whether robust regression using the lmrob() function from the package robustbase should be used in the method rather than standard linear regression (lm).
correl	If the genetic variants are correlated, then this correlation can be accounted for. The matrix of correlations between must be provided in the MRMVInput object: the elements of this matrix are the correlations between the individual variants (diagonal elements are 1). If a correlation matrix is specified in the MRMVInput object, then correl is set to TRUE.
correl.x	Correlation matrix for exposures (Optional). Default is to assume the exposures are uncorrelated. This is only used in the computation of conditional F-statistics.
nx	Either a single value, or a vector of sample sizes for the genetic associations with the exposures (one for each exposure, and assumed equal for all variants). If a single value is provided, it is assumed this is the sample size for all exposures. This is optional, and it is only used in the calculation of conditional F statistics. If not supplied, then conditional F statistics are not reported.
distribution	The type of distribution used to calculate the confidence intervals. Options are "normal" (default) or "t-dist".
alpha	The significance level used to calculate the confidence interval. The default value is 0.05.
...	Additional arguments to be passed to the regression method.

Details

Multivariable Mendelian randomization is an extension of Mendelian randomization to deal with genetic variants that are associated with multiple risk factors. Two scenarios are envisioned for its use: 1) risk factors that are biologically related, such as lipid fractions; and 2) risk factors where there is potentially a network of causal effects (mediation) from one risk factor to another. In both cases, under the extended assumptions of multivariable Mendelian randomization, coefficients represent the direct causal effects of each risk factor in turn with the other risk factors being fixed.

We implement the method using multivariable weighted linear regression. If the variants are correlated, the method is implemented using generalized weighted linear regression; this is hard coded using matrix algebra.

The causal estimate is obtained by regression of the associations with the outcome on the associations with the risk factors, with the intercept set to zero and weights being the inverse-variances of the associations with the outcome.

Value

The output from the function is an MVIVW object containing:

Model	A character string giving the type of model used ("fixed", "random", or "default").
Exposure	A character vector with the names given to the exposure.

Outcome	A character string with the names given to the outcome.
Robust	TRUE if robust regression has been used to calculate the estimate, FALSE otherwise.
Correlation	The matrix of genetic correlations.
Estimate	A vector of causal estimates.
StdError	A vector of standard errors of the causal estimates.
CILower	The lower bounds of the causal estimates based on the estimated standard errors and the significance level provided.
CIUpper	The upper bounds of the causal estimates based on the estimated standard errors and the significance level provided.
Alpha	The significance level used when calculating the confidence intervals.
Pvalue	The p-values associated with the estimates (calculated as Estimate/StdError as per Wald test) using a normal or t-distribution (as specified in distribution).
SNPs	The number of genetic variants (SNPs) included in the analysis.
RSE	The estimated residual standard error from the regression model.
Heter.Stat	Heterogeneity statistic (Cochran's Q statistic) and associated p-value: the null hypothesis is that all genetic variants estimate the same causal parameter; rejection of the null is an indication that one or more variants may be pleiotropic.
CondFstat	Conditional F statistics: An approximation of the first-stage conditional F statistics for all variants based on the summarized data. This represents the instrument strength for each exposure conditional on other exposures in the model. This is only reported if the sample sizes for the genetic associations with the exposures are provided.

References

Description of approach: Stephen Burgess, Simon G Thompson. Multivariable Mendelian Randomization: the use of pleiotropic genetic variants to estimate causal effects. American Journal of Epidemiology 2015; 181(4):251-260. doi: 10.1093/aje/kwu283.

Description of inverse-variance weighted method: Stephen Burgess, Frank Dudbridge, Simon G Thompson. Re: "Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects." American Journal of Epidemiology 2015; 181(4):290-291. doi: 10.1093/aje/kwv017.

Use for mediation analysis: Stephen Burgess, Deborah J Thompson, Jessica MB Rees, Felix R Day, John R Perry, Ken K Ong. Dissecting causal pathways using Mendelian randomization with summarized genetic data: Application to age at menarche and risk of breast cancer. Genetics 2017; 207(2):481-487. doi: 10.1534/genetics.117.300191.

Calculation of conditional F statistics: Ashish Patel, Dipender Gill, Dmitry Shungin, Christos Mantzoros, Lotte Bjerre Knudsen, Jack Bowden, Stephen Burgess. Robust use of phenotypic heterogeneity at drug target genes for mechanistic insights: application of multivariable Mendelian randomization to *GLP1R* gene region. Pre-print.

Examples

```
mr_mvivw(mr_mvinput(bx = cbind(ldlc, hdlc, trig), bxse = cbind(ldlcse, hdlcse, trigse),
   by = chdlodds, byse = chdloddsse))
```

`mr_mvivwme` *Multivariable inverse-variance weighted method with measurement error*

Description

The `mr_mvivwme` function performs multivariable Mendelian randomization via the inverse-variance method with measurement error.

Usage

```
mr_mvivwme(
  object,
  model = "default",
  correl = FALSE,
  correl.x = NULL,
  distribution = "normal",
  alpha = 0.05,
  max_iter = 100,
  no_ini = 1,
  seed = 20201201,
  ...
)

## S4 method for signature 'MRMVInput'
mr_mvivwme(
  object,
  model = "default",
  correl = FALSE,
  correl.x = NULL,
  distribution = "normal",
  alpha = 0.05,
  max_iter = 100,
  no_ini = 1,
  seed = 20201201,
  ...
)
```

Arguments

`object`	An `MRMVInput` object.

model	What type of model should be used: "default", "random" or "fixed". The random-effects model ("random") is a multiplicative random-effects model, allowing overdispersion in the weighted linear regression (the residual standard error is not fixed to be 1, but is not allowed to take values below 1). The fixed-effect model ("fixed") sets the residual standard error to be 1. The "default" setting is to use a fixed-effect model with 3 genetic variants or fewer, and otherwise to use a random-effects model.
correl	If the genetic variants are correlated, then this correlation can be accounted for.
correl.x	Correlation matrix for exposures. Default is to assume the exposures are uncorrelated.
distribution	The type of distribution used to calculate the confidence intervals. Options are "normal" (default) or "t-dist".
alpha	The significance level used to calculate the confidence interval. The default value is 0.05.
max_iter	The maximum number of iterations in the optimisation procedure.
no_ini	The number of initial values for the optimisation procedure.
seed	The random seed to use for the optimisation procedure. The default value is 20201201. If set to NA, the random seed will not be set (for example, if the function is used as part of a larger simulation).
...	Additional arguments to be passed to the regression method.

Details

The extension of multivariable Mendelian randomization to account for measurement error in the genetic associations with the exposure traits.

Value

The output from the function is an MVIVWME object containing:

Model	A character string giving the type of model used ("fixed", "random", or "default").
Exposure	A character vector with the names given to the exposure.
Outcome	A character string with the names given to the outcome.
Estimate	A vector of causal estimates.
StdError	A vector of standard errors of the causal estimates.
CILower	The lower bounds of the causal estimates based on the estimated standard errors and the significance level provided.
CIUpper	The upper bounds of the causal estimates based on the estimated standard errors and the significance level provided.
Alpha	The significance level used when calculating the confidence intervals.
Pvalue	The p-values associated with the estimates (calculated as Estimate/StdError as per Wald test) using a normal or t-distribution (as specified in distribution).
Correlation	The matrix of genetic correlations.
SNPs	The number of genetic variants (SNPs) included in the analysis.

RSE	The estimated residual standard error from the regression model.
Heter.Stat	Heterogeneity statistic (Cochran's Q statistic) and associated p-value: the null hypothesis is that all genetic variants estimate the same causal parameter; rejection of the null is an indication that one or more variants may be pleiotropic.

References

Zhu, Jiazheng, Stephen Burgess, and Andrew J. Grant. Bias in Multivariable Mendelian Randomization Studies Due to Measurement Error on Exposures, 2022. https://doi.org/10.48550/arXiv.2203.08668.

Examples

```
mr_mvivwme(mr_mvinput(bx = cbind(ldlc, hdlc, trig), bxse = cbind(ldlcse, hdlcse, trigse),
   by = chdlodds, byse = chdloddsse))
```

`mr_mvlasso` *Multivariable MR-Lasso method*

Description

The `mr_mvlasso` function performs the multivariable MR-Lasso method, which applies lasso-type penalization to the direct effects of genetic variants on the outcome. The causal estimates are described as post-lasso estimates, and are obtained by performing the multivariable IVW method using only those genetic variants that are identified as valid by the lasso procedure.

Usage

```
mr_mvlasso(
  object,
  orientate = 1,
  distribution = "normal",
  alpha = 0.05,
  lambda = numeric(0)
)

## S4 method for signature 'MRMVInput'
mr_mvlasso(
  object,
  orientate = 1,
  distribution = "normal",
  alpha = 0.05,
  lambda = numeric(0)
)
```

Arguments

object	An MRMVInput object.
orientate	The risk factor that genetic associations are orientated to. The default option is 1, meaning that genetic associations with the first risk factor are set to be positive.
distribution	The type of distribution used to calculate the confidence intervals. Options are "normal" (default) or "t-dist".
alpha	The significance level used to calculate the confidence intervals. The default value is 0.05.
lambda	The value of the tuning parameter used by the lasso procedure which controls the level of sparsity. If not specified, the tuning parameter will be calculated by the heterogeneity stopping rule.

Details

Multivariable MR-Lasso extends the multivariable IVW model to include an intercept term for each genetic variant. These intercept terms represent associations between the genetic variants and the outcome which bypass the risk factors. The regularized regression model is estimated by multivariable weighted linear regression where the intercept terms are subject to lasso-type penalization. The lasso penalization will tend to shrink the intercept terms corresponding to the valid instruments to zero.

The lasso penalty relies on a tuning parameter which controls the level of sparsity. The default is to use a heterogeneity stopping rule, but a fixed value may be specified.

As part of the analysis, the genetic variants are orientated so that all of the associations with one of the risk factors are positive (the first risk factor is used by default). Re-orientation of the genetic variants is performed automatically as part of the function.

The MR-Lasso method is performed in two steps. First, a regularized regression model is fitted, and some genetic variants are identified as valid instruments. Second, causal effects are estimated using standard multivariable IVW with only the valid genetic variants. The post-lasso method will be performed as long as the number of genetic variants identified as valid instruments is greater than the number of risk factors. The default heterogeneity stopping rule will always return more genetic variants as valid instruments than risk factors for identification. The main estimates given by the method are the post-lasso estimates. However, parameter estimates from the regularized regression model used to identify invalid variants are also provided for completeness.

If a substantial proportion of genetic variants are removed from the analysis, the multivariable MR-Lasso method may give a false impression of confidence in the causal estimate due to homogeneity of the variant-specific causal estimates amongst the remaining variants. However, it is not reasonable to claim that there is strong evidence for a causal effect after a large number of variants with heterogeneous estimates have been removed from the analysis.

Value

The output from the function is an MVLasso object containing:

Exposure	A character vector with the names given to the exposure.
Outcome	A character string with the names given to the outcome.

Estimate	A vector of causal estimates from the multivariable MR-Lasso method. These are the post-lasso estimates.
StdError	A vector of standard errors of the causal estimates from the multivariable MR-Lasso method.
CILower	The lower bounds of the confidence intervals for the causal estimates based on the estimated standard errors and the significance level provided.
CIUpper	The upper bounds of the confidence intervals for the causal estimates based on the estimated standard errors and the significance level provided.
Alpha	The significance level used when calculating the confidence intervals.
Pvalue	The p-values associated with the (post-lasso) causal estimates using a normal or t-distribution (as specified in distribution).
SNPs	The number of genetic variants (SNPs) included in the analysis.
RegEstimate	The estimates from the regularized regression model used in the multivariable MR-Lasso method.
RegIntercept	The intercept estimates estimates from the regularized regression model used in the multivariable MR-Lasso method.
Valid	The number of genetic variants that have been identified as valid instruments.
ValidSNPs	The names of genetic variants that have been identified as valid instruments.
Lambda	The value of the tuning parameter used to compute RegEstimate.

References

Andrew J Grant, Stephen Burgess. Pleiotropy robust methods for multivariable Mendelian randomization. arXiv 2020; 2008.11997

Examples

```
mr_mvlasso(mr_mvinput(bx = cbind(ldlc, hdlc, trig), bxse = cbind(ldlcse, hdlcse, trigse),
   by = chdlodds, byse = chdloddsse))
```

---

mr_mvmedian *Multivariable median-based method*

Description

The mr_mvmedian function performs multivariable Mendelian randomization via the median method. This is implemented by multivariable weighted quantile regression, with the quantile set to 0.5 (the median).

Usage

```
mr_mvmedian(
  object,
  distribution = "normal",
  alpha = 0.05,
  iterations = 10000,
  seed = 314159265
)

## S4 method for signature 'MRMVInput'
mr_mvmedian(
  object,
  distribution = "normal",
  alpha = 0.05,
  iterations = 10000,
  seed = 314159265
)
```

Arguments

object	An MRMVInput object.
distribution	The type of distribution used to calculate the confidence intervals. Options are "normal" (default) or "t-dist".
alpha	The significance level used to calculate the confidence intervals. The default value is 0.05.
iterations	The number of bootstrap samples to generate when calculating the estimated standard error. The default value is 10000.
seed	The random seed to use when generating the bootstrap samples (for reproducibility). The default value is 314159265. If set to NA, the random seed will not be set (for example, if the function is used as part of a larger simulation).

Details

The multivariable median method is similar to the univariable weighted median method, except that it is implemented using quantile regression. The regression model is multivariable and weighted by the inverse of the variances of the variant-specific estimates. Confidence intervals are calculated by parametric bootstrap to estimate the standard error of the estimates, and then using quantiles of a normal or t-distribution (depending on the value of distribution).

Value

The output from the function is an MVMedian object containing:

Exposure	A character vector with the names given to the exposure.
Outcome	A character string with the names given to the outcome.
Estimate	A vector of causal estimates.
StdError	A vector of standard errors of the causal estimates.

CILower	The lower bounds of the causal estimates based on the estimated standard errors and the significance level provided.
CIUpper	The upper bounds of the causal estimates based on the estimated standard errors and the significance level provided.
Alpha	The significance level used when calculating the confidence intervals.
Pvalue	The p-values associated with the estimates (calculated as Estimate/StdError as per Wald test) using a normal or t-distribution (as specified in distribution).
SNPs	The number of genetic variants (SNPs) included in the analysis.

Examples

```
mr_mvmedian(mr_mvinput(bx = cbind(ldlc, hdlc, trig), bxse = cbind(ldlcse, hdlcse, trigse),
   by = chdlodds, byse = chdloddsse), iterations = 100)
  # iterations is set to 100 to reduce runtime for the mr_mvmedian method,
  # 10000 iterations are recommended in practice
```

---

mr_mvpcgmm *Multivariable principal components generalized method of moments (PC-GMM) method*

Description

The mr_mvpcgmm function performs multivariable Mendelian randomization via the principal components generalized method of moments method.

Usage

```
mr_mvpcgmm(
  object,
  nx,
  ny,
  cor.x = NULL,
  r = NULL,
  thres = 0.99,
  robust = TRUE,
  alpha = 0.05,
  ...
)

## S4 method for signature 'MRMVInput'
mr_mvpcgmm(
  object,
  nx,
  ny,
  cor.x = NULL,
```

```
  r = NULL,
  thres = 0.99,
  robust = TRUE,
  alpha = 0.05,
  ...
)
```

Arguments

object	An MRMVInput object.
nx	Vector of sample sizes used to compute genetic associations with the exposure (one for each exposure).
ny	The sample size used to compute genetic associations with the outcome.
cor.x	Correlation matrix for exposures. Default is to assume the exposures are uncorrelated.
r	The number of genetic principal components to be used to instrument the exposures. Default chooses r to explain 99.9% of variation in a sample weighted genetic correlation matrix (this can be varied by setting the thres parameter).
thres	The threshold value of variation in the sample weighted genetic correlation matrix explained by the genetic principal components. The default value is 0.99, indicating that 99% of variation is explained by the principal components. Note that if r and thres are both specified, then r will take precedence and thres will be ignored.
robust	Indicates whether overdispersion heterogeneity is accounted for in the model. Default is TRUE.
alpha	The significance level used to calculate the confidence interval. The default value is 0.05.
...	Additional arguments to be passed to the optimization routines to calculate GMM estimates and overdispersion parameter.

Details

When a Mendelian randomization analysis is performed using correlated genetic variants from a single gene region, there is a tradeoff between using too few variants (and compromising on power) and using too many variants (in which case, estimates can be highly sensitive to small variation in the correlation matrix). This method performs dimension reduction on a weighted version of the genetic correlation matrix to form principal components based on the genetic variants, which are then used as instruments. It is recommended not to include very highly correlated variants in this method (say, r^2 > 0.95), but the method should cope well with variants correlated below this level.

This function runs a multivariable version of the PC-GMM method, which can be used when there are distinct exposures associated with variants at a single gene region. Phenotypic heterogeneity (that is, the genetic associations with the exposures are not collinear) at genomic loci encoding drug targets can be exploited by multivariable Mendelian randomization to provide insight on the pathways by which pharmacological interventions may affect disease risk.

This method provides two-sample multivariable Mendelian randomization estimates and associated confidence intervals that account for overdispersion heterogeneity in dimension-reduced genetic associations (when robust = TRUE).

Value

The output from the function is an MVPCGMM object containing:

Robust	TRUE if overdispersion heterogeneity was included in the model, FALSE otherwise.
Exposure	A character vector with the names given to the exposure.
Outcome	A character string with the names given to the outcome.
Correlation	The matrix of genetic correlations.
ExpCorrelation	TRUE if an exposure correlation matrix was specified, FALSE otherwise.
CondFstat	A vector of conditional F-statistics (one for each exposure).
Estimate	A vector of causal estimates.
StdError	A vector of standard errors of the causal estimates.
CILower	The lower bounds of the causal estimates based on the estimated standard errors and the significance level provided.
CIUpper	The upper bounds of the causal estimates based on the estimated standard errors and the significance level provided.
Overdispersion	The estimate of the overdispersion parameter.
PCs	The number of genetic principal components used to instrument the exposures.
Pvalue	The p-values associated with the estimates (calculated as Estimate/StdError as per Wald test) using a normal distribution.
Alpha	The significance level used when calculating the confidence intervals.
Heter.Stat	Heterogeneity statistic (Cochran's Q statistic) and associated p-value (for non-robust model only): the null hypothesis is that all genetic principal components estimate the same causal parameter; rejection of the null is an indication that one or more principal components may be pleiotropic.

References

Description of multivariable Mendelian randomization: Stephen Burgess, Simon G Thompson. Multivariable Mendelian Randomization: the use of pleiotropic genetic variants to estimate causal effects. American Journal of Epidemiology 2015; 181(4):251-260. doi: 10.1093/aje/kwu283.

Description of the PC-GMM method: "Robust use of phenotypic heterogeneity at drug target genes for mechanistic insights: application of cis-multivariable Mendelian randomization to GLP1R gene region" (Preprint).

Examples

```
mr_mvpcgmm(mr_mvinput(bx = cbind(ldlc, hdlc, trig), bxse = cbind(ldlcse, hdlcse, trigse),
  by = chdlodds, byse = chdloddsse, correlation = diag(length(ldlc))), nx=rep(17723,3), ny=17723)
# Note this example does not use variants from a single gene region, and is provided
#  to demonstrate that the code works, rather than to illustrate a recommended use case.
```

mr_pcgmm *Univariable principal components generalized method of moments (PC-GMM) method*

Description

The mr_pcgmm function performs multivariable Mendelian randomization via the principal components generalized method of moments method.

Usage

```
mr_pcgmm(
  object,
  nx,
  ny,
  r = NULL,
  thres = 0.99,
  robust = TRUE,
  alpha = 0.05,
  ...
)

## S4 method for signature 'MRInput'
mr_pcgmm(
  object,
  nx,
  ny,
  r = NULL,
  thres = 0.99,
  robust = TRUE,
  alpha = 0.05,
  ...
)
```

Arguments

object	An MRInput object.
nx	The sample size used to compute genetic associations with the exposure.
ny	The sample size used to compute genetic associations with the outcome.
r	The number of genetic principal components to be used to instrument the exposure. Default chooses r to explain 99.9% of variation in a sample weighted genetic correlation matrix (this can be varied by setting the thres parameter).
thres	The threshold value of variation in the sample weighted genetic correlation matrix explained by the genetic principal components. The default value is 0.99, indicating that 99% of variation is explained by the principal components. Note that if r and thres are both specified, then r will take precedence and thres will be ignored.

robust	Indicates whether overdispersion heterogeneity is accounted for in the model. Default is TRUE.
alpha	The significance level used to calculate the confidence interval. The default value is 0.05.
...	Additional arguments to be passed to the optimization routines to calculate the GMM estimate and overdispersion parameter.

Details

When a Mendelian randomization analysis is performed using correlated genetic variants from a single gene region, there is a tradeoff between using too few variants (and compromising on power) and using too many variants (in which case, estimates can be highly sensitive to small variation in the correlation matrix). This method performs dimension reduction on a weighted version of the genetic correlation matrix to form principal components based on the genetic variants, which are then used as instruments. It is recommended not to include very highly correlated variants in this method (say, `r^2 > 0.95`), but the method should cope well with variants correlated below this level.

This function runs a univariable version of the PC-GMM method, which can be used when there are a single exposure associated with variants at a given gene region.

This method provides two-sample univariable Mendelian randomization estimates and associated confidence intervals that account for overdispersion heterogeneity in dimension-reduced genetic associations (when `robust = TRUE`).

Value

The output from the function is an `PCGMM` object containing:

Robust	TRUE if overdispersion heterogeneity was included in the model, FALSE otherwise.
Exposure	A character string with the name given to the exposure.
Outcome	A character string with the names given to the outcome.
Correlation	The matrix of genetic correlations.
Estimate	The causal estimate.
StdError	The standard error of the causal estimate.
CILower	The lower bound of the causal estimate based on the estimated standard error and the significance level provided.
CIUpper	The upper bound of the causal estimate based on the estimated standard error and the significance level provided.
Fstat	The first-stage F statistic for all genetic principal components used as instruments.
Overdispersion	The estimate of the overdispersion parameter.
PCs	The number of genetic principal components used to instrument the exposure.
Pvalue	The p-values associated with the estimates (calculated as Estimate/StdError as per Wald test) using a normal distribution.
Alpha	The significance level used when calculating the confidence intervals.

`Heter.Stat` Heterogeneity statistic (Cochran's Q statistic) and associated p-value (for non-robust model only): the null hypothesis is that all genetic principal components estimate the same causal parameter; rejection of the null is an indication that one or more principal components may be pleiotropic.

References

Description of the PC-GMM method: "Robust use of phenotypic heterogeneity at drug target genes for mechanistic insights: application of cis-multivariable Mendelian randomization to GLP1R gene region" (Preprint).

Examples

```
mr_pcgmm(mr_input(bx = calcium, bxse = calciumse,
   by = fastgluc, byse = fastglucse, correlation = calc.rho), nx=6351, ny=133010)
```

`mr_pivw` *Penalized inverse-variance weighted method*

Description

The `mr_pivw` function implements the penalized inverse-variance weighted (pIVW) method.

Usage

```
mr_pivw(
  object,
  lambda = 1,
  over.dispersion = TRUE,
  delta = 0,
  sel.pval = NULL,
  Boot.Fieller = TRUE,
  alpha = 0.05
)

## S4 method for signature 'MRInput'
mr_pivw(
  object,
  lambda = 1,
  over.dispersion = TRUE,
  delta = 0,
  sel.pval = NULL,
  Boot.Fieller = TRUE,
  alpha = 0.05
)
```

Arguments

object	An MRInput object.
lambda	The penalty parameter in the pIVW estimator. It plays a role in the bias-variance trade-off of the estimator. It is recommended to choose lambda=1 to achieve the smallest bias and valid inference. By default, lambda=1.
over.dispersion	Should the method consider overdispersion (balanced horizontal pleiotropy)? Default is TRUE.
delta	The z-score threshold for IV selection. delta should be greater than or equal to zero. By default, delta=0 (i.e., no IV selection will be conducted). See 'Details'.
sel.pval	A numeric vector containing the P-values of the SNP effects on the exposure, which will be used for the IV selection. sel.pval should be provided when delta is not zero. See 'Details'.
Boot.Fieller	If Boot.Fieller=TRUE, then the P-value and the confidence interval of the causal effect will be calculated based on the bootstrapping Fieller method. Otherwise, the P-value and the confidence interval of the causal effect will be calculated from the normal distribution. It is recommended to use the bootstrapping Fieller method when Condition is smaller than 10 (see 'Details'). By default, Boot.Fieller=TRUE.
alpha	The significance level used to calculate the confidence intervals. The default value is 0.05.

Details

The penalized inverse-variance weighted (pIVW) estimator accounts for weak instruments and balanced horizontal pleiotropy simultaneously in two-sample MR with summary statistics, i.e., an exposure sample (with IV-exposure effect Bx and standard error Bxse) and an outcome sample (with IV-outcome effect By and standard error Byse).

The pIVW estimator also allows for IV selection in three-sample MR, where weak IVs are screened out using an extra sample (with IV-exposure effect Bx* and standard error Bxse*) independent of the exposure sample and outcome sample. Generally, the P-value for Bx* can be computed by sel.pval=2*pnorm(abs(Bx*/Bxse*), lower.tail = FALSE), Given sel.pval and a z-score threshold delta, the variants kept in the analysis will be those with sel.pval<2*pnorm(delta,lower.tail = FALSE).

The mr_pivw function outputs a measure Condition that needs to be large for reliable asymptotic properties of the pIVW estimator. We also refer to Condition as effective sample size, which is a function of a measure of IV strength and the number of IVs. When delta is zero (i.e., no IV selection), Condition = (average F-statistic -1)*sqrt(# snps). When delta is not zero (i.e., IV selection is conducted), Condition = [(average F-statistic -1)*sqrt(# snps)]/c, where the numerator is computed using the selected variants, and the denominator c involves the selection probabilities of all variants (see more details in the paper). We suggest that Condition should be greater than 5 for the pIVW estimator to achieve reliable asymptotic properties.

Value

The output from the function is a PIVW object containing:

Over.dispersion	TRUE if the method has considered balanced horizontal pleiotropy, FALSE otherwise.
Boot.Fieller	TRUE if the bootstrapping Fieller method is used to calculate the P-value and the confidence interval of the causal effect, FALSE otherwise.
Lambda	The penalty parameter in the pIVW estimator.
Delta	The z-score threshold for IV selection.
Exposure	A character string giving the name given to the exposure.
Outcome	A character string giving the name given to the outcome.
Estimate	The causal point estimate from the pIVW estimator.
StdError	The standard error associated with Estimate.
CILower	The lower bound of the confidence interval for Estimate, which is derived from the bootstrapping Fieller method or normal distribution. For the bootstrapping Fieller's interval, if it contains multiple ranges, then lower limits of all ranges will be reported.
CIUpper	The upper bound of the confidence interval for Estimate, which is derived from the bootstrapping Fieller method or normal distribution. For the bootstrapping Fieller's interval, if it contains multiple ranges, then upper limits of all ranges will be reported.
Alpha	The significance level used in constructing the confidence interval.
Pvalue	P-value associated with the causal estimate from the pIVW estimator, which is derived from the bootstrapping Fieller method or normal distribution.
Tau2	The variance of the balanced horizontal pleiotropy. Tau2 is calculated by using all IVs in the data before conducting the IV selection.
SNPs	The number of SNPs after IV selection.
Condition	The estimated effective sample size. It is recommended to be greater than 5 for the pIVW estimator to achieve reliable asymptotic properties. See 'Details'.

References

Xu S., Wang P., Fung W.K. and Liu Z. (2022). A Novel Penalized Inverse-Variance Weighted Estimator for Mendelian Randomization with Applications to COVID-19 Outcomes. Biometrics. doi: 10.1111/biom.13732.

Examples

```
mr_pivw(mr_input(bx = ldlc, bxse = ldlcse, by = chdlodds, byse = chdloddsse))
```

mr_plot *Draw a scatter plot of the genetic associations and/or causal estimates*

Description

The function mr_plot has three functionalities. It can generate a visual representation of MRInput, MRMVInput and MRAll objects.

Usage

```
mr_plot(
  object,
  error = TRUE,
  line = "ivw",
  orientate = FALSE,
  interactive = TRUE,
  labels = FALSE
)

## S4 method for signature 'MRInput'
mr_plot(
  object,
  error = TRUE,
  line = "ivw",
  orientate = FALSE,
  interactive = TRUE,
  labels = FALSE
)

## S4 method for signature 'MRAll'
mr_plot(object)

## S4 method for signature 'MRMVInput'
mr_plot(
  object,
  error = TRUE,
  line = TRUE,
  orientate = FALSE,
  interactive = TRUE,
  labels = FALSE
)
```

Arguments

object	An MRInput object or an MRMVInput object or an MRAll object.

error	When viewing an MRInput or MRMVInput object, one can choose whether to include error bars (default is to include). For an MRMVInput object, the horizontal error bars only take into account uncertainty in the causal estimates.
line	When viewing an MRInput object, one can choose whether to include the IVW estimate (line = "ivw") or the MR-Egger estimate (line = "egger"). When viewing an MRMVInput, one can choose whether to include a line through the origin with gradient 1 (line = TRUE) or not.
orientate	When viewing an MRInput or MRMVInput object, one can choose whether to orientate all genetic variants so that the associations with the risk factor are all positive. This is recommended particularly when plotting the MR-Egger estimate, although the default setting is FALSE.
interactive	When viewing an MRInput or MRMVInput object, one can choose whether to produce an interactive graph using the plotly package, or a static graph using the regular plot command.
labels	When viewing an MRInput or MRMVInput object with interactive set to FALSE, setting labels to TRUE means that the name of each genetic variants appears above the corresponding datapoint.

Details

The result is dependent on the type of object passed to mr_plot. When the object is an MRInput object, the function uses either the plot command (if interactive is set to FALSE) or plotly syntax (if interactive is set to TRUE) to plot the association estimates against each other. When the object is an MRMVInput object, functionality is similar except that we plot the estimated associations with the outcome on the y-axis, and fitted values of the associations with the outcome from the inverse-variance weighted method on the x-axis. If interactive is set to FALSE, then a static graph is produced. By setting labels to TRUE, the names of the genetic variants appear above the points. This produces a less visually appealing graph, but one where it is easier to identify the individual genetic variants. If interactive is set to TRUE, then the plot is interactive and the user can hover over the various points to see the name of the associated genetic variant and its association estimates. When the object is an MRAll object, the function generates a ggplot to compare the causal estimates proposed by different methods.

Examples

```
mr_plot(mr_input(bx = ldlc, bxse = ldlcse, by = chdlodds, byse = chdloddsse),
  line="egger", orientate = TRUE)
mr_plot(mr_input(bx = ldlc, bxse = ldlcse, by = chdlodds, byse = chdloddsse),
  line="ivw", interactive=FALSE) # produces a static graph
mr_plot(mr_allmethods(mr_input(bx = ldlc, bxse = ldlcse,
   by = chdlodds, byse = chdloddsse), method="all", iterations = 50))
  # iterations is set to 50 to reduce runtime for the mr_median method,
  # 10000 iterations are recommended in practice
```

MRAll-class *MRAll Class*

Description

An object containing the estimates produced using the `mr_allmethods` function.

Slots

`Data` The `mr_input` object that was used as an input to the `mr_allmethods` function. This includes the original data, so that a call to `mr_plot` can plot the original data and the various causal estimates.

`Values` A data.frame object comprising estimates from the various methods called by the `mr_allmethods` function. The first column gives the names of the methods, then the causal estimates, standard errors, 95% confidence intervals, and p-values.

`Method` A string indicating whether all methods are implemented (`"all"`, the default option), or just main methods (`"main"`), or only a subset of methods (`"ivw"`, `"egger"`, or `"median"`).

MRcML-class *MRcML Class*

Description

An object containing the results of MRcML.

Slots

`Exposure` The names of the exposure variables.

`Outcome` The name of the outcome variable.

`Estimate` Estimate of theta.

`StdError` Standard error of estimate.

`Pvalue` p-value of estimate.

`BIC_invalid` Set of selected invalid IVs if cML-BIC is performed, i.e. without MA or DP.

`GOF1_p` p-value of the first goodness-of-fit test.

`GOF2_p` p-value of the second goodness-of-fit test.

`SNPs` The number of SNPs that were used in the calculation.

`Alpha` Significance level for the confidence interval for estimate, default is 0.05.

`CILower` Lower bound of the confidence interval for estimate.

`CIUpper` Upper bound of the confidence interval for estimate.

`MA` Indicator of whether model average is applied.

`DP` Indicator of whether data perturbation is applied.

MRConMix-class *MRConMix Class*

Description

An object containing the estimate produced using the contamination mixture method as well as various statistics.

Slots

`Exposure` The names of the exposure variables.

`Outcome` The name of the outcome variable.

`Psi` The value of the standard deviation of the distribution of invalid estimands (default is 1.5 times the standard deviation of the ratio estimates).

`Estimate` The causal estimate from the contamination mixture method.

`CIRange` The confidence interval for `Estimate` based on a grid search.

`CILower` The lower limit of the confidence interval. If the confidence interval contains multiple ranges, then lower limits of all ranges will be reported.

`CIUpper` The upper limit of the confidence interval. If the confidence interval contains multiple ranges, then upper limits of all ranges will be reported.

`CIMin` The smallest value used in the search to find the confidence interval.

`CIMax` The largest value used in the search to find the confidence interval.

`CIStep` The step size used in the search to find the confidence interval.

`Pvalue` The p-value associated with the estimate calculated using the likelihood function and a chi-squared distribution.

`Valid` The numbers of genetic variants that were considered valid instruments at the causal estimate.

`ValidSNPs` The names of genetic variants that were considered valid instruments at the causal estimate.

`Alpha` The significance level used in constructing the confidence interval (default is 0.05).

`SNPs` The number of SNPs that were used in the calculation.

MRHetPen-class *MRHetPen Class*

Description

An object containing the estimate produced using the heterogeneity-penalized model-averaging mode-based estimation method as well as various statistics.

Slots

`Exposure` The names of the exposure variables.

`Outcome` The name of the outcome variable.

`Prior` The value of the prior probability of a genetic variant being a valid instrument (default is 0.5).

`Estimate` The causal estimate from the heterogeneity-penalized method.

`CIRange` The confidence interval for `Estimate` based on a grid search.

`CILower` The lower limit of the confidence interval. If the confidence interval contains multiple ranges, then lower limits of all ranges will be reported.

`CIUpper` The upper limit of the confidence interval. If the confidence interval contains multiple ranges, then upper limits of all ranges will be reported.

`CIMin` The smallest value used in the search to find the confidence interval.

`CIMax` The largest value used in the search to find the confidence interval.

`CIStep` The step size used in the search to find the confidence interval.

`Alpha` The significance level used in constructing the confidence interval (default is 0.05).

`SNPs` The number of SNPs that were used in the calculation.

`MRInput-class` *MRInput Class*

Description

An object containing the four vectors of summary statistics required to calculate Mendelian randomization estimates.

Details

The beta-coefficients are assumed to be estimated for uncorrelated (independent) genetic variants, although a correlation matrix can be specified if the variants are correlated in their distributions. We also assume that the beta-coefficients for associations with the exposure and with the outcome are uncorrelated (corresponding to a two-sample Mendelian randomization analysis), although correlation between associations with the exposure and with the outcome generally have little impact on causal estimates or standard errors. Estimates can either be specified by the user, or extracted from the PhenoScanner tool.

Slots

`betaX` A numeric vector of beta-coefficient values for genetic associations with the first variable (often referred to as the exposure, risk factor, or modifiable phenotype).

`betaY` A numeric vector of beta-coefficient values for genetic associations with the second variable (often referred to as the outcome). For a disease outcome, the beta coefficients are log odds estimates from logistic regression analyses.

`betaXse` The standard errors associated with the beta-coefficients in `betaX`.

betaYse The standard errors associated with the beta-coefficients in betaY.

correlation The matrix of correlations between genetic variants. If this variable is not provided, then we assume that genetic variants are uncorrelated.

exposure The name of the exposure variable.

outcome The name of the outcome variable.

snps The names of the genetic variants (SNPs) included in the analysis. The slots exposure, outcome, and snps are not required, but may be useful for keeping track of various MRInput objects. They are also used by the mr_plot function.

effect_allele The name of the effect allele for each SNP. The beta-coefficients are the associations with the exposure and outcome per additional copy of the effect allele.

other_allele The name of the non-effect allele.

eaf The expected allele frequencies (numeric). The slots effect_allele, other_allele, and eaf are neither required, nor currently used in the MendelianRandomization package. They are included for future compatibility with the MR-Base suite of functions.

MRLasso-class *MRLasso class*

Description

An object containing the estimates produced using the MR-Lasso method as well as various statistics.

Slots

Exposure The names of the exposure variables.

Outcome The name of the outcome variable.

Estimate The causal estimate from the MR-Lasso method.

StdError The standard error associated with Estimate.

CILower The lower bounds of the confidence intervals for Estimate based on StdError.

CIUpper The upper bounds of the confidence intervals for Estimate based on StdError.

Alpha The significance level used in constructing the confidence interval (default is 0.05).

Pvalue P-value associated with the causal estimate from the MR-Lasso method.

SNPs The number of SNPs used in the calculation.

RegEstimate The estimate from the regularized regression model used in the MR-Lasso method.

RegIntercept The intercept estimates from the regularized regression model used in the MR-Lasso method. An intercept estimate of zero identifies the corresponding genetic variant as a valid instrument. Genetic variants with non-zero intercept estimates will be excluded from the post-lasso estimator.

Valid The number of genetic variants that have been identified as valid instruments.

ValidSNPs The names of genetic variants that have been identified as valid instruments.

Lambda The value of the tuning parameter used to compute RegEstimate (default is to calulate Lambda using the heterogeneity stopping rule).

MRMBE-class *MRMBE Class*

Description

An object containing the estimate produced using the mode-based estimation method of Hartwig et al as well as various statistics.

Slots

`Exposure` The names of the exposure variables.

`Outcome` The name of the outcome variable.

`Weighting` Whether the analysis was `weighted` or `unweighted`.

`StdErr` Whether the `simple` or `delta` version of the standard errors were used.

`Phi` The value of the bandwidth factor.

`Estimate` The causal estimate from the mode-based estimation method.

`StdError` The standard errors associated with `Estimate`.

`CILower` The lower bounds of the confidence interval for `Estimate` based on `StdError`.

`CIUpper` The upper bounds of the confidence interval for `Estimate` based on `StdError`.

`Alpha` The significance level used in constructing the confidence interval (default is 0.05).

`Pvalue` P-value associated with the causal estimate.

`SNPs` The number of SNPs that were used in the calculation.

MRMVInput-class *MRMVInput Class*

Description

An object containing the summary statistics required to calculate multivariable Mendelian randomization estimates.

Details

The beta-coefficients are assumed to be estimated for uncorrelated (independent) genetic variants, although a correlation matrix can be specified if the variants are correlated in their distributions. We also assume that the beta-coefficients for associations with the exposure and with the outcome are uncorrelated (corresponding to a two-sample Mendelian randomization analysis), although correlation between associations with the exposure and with the outcome generally have little impact on causal estimates or standard errors.

Slots

betaX A matrix of beta-coefficient values for genetic associations with the risk factor variables. These should be arranged so that column 1 are the beta-coefficients for risk factor 1, and row 1 are the beta-coefficients for genetic variant 1.

betaY A numeric vector of beta-coefficient values for genetic associations with the second variable (often referred to as the outcome). For a disease outcome, the beta coefficients are log odds estimates from logistic regression analyses.

betaXse The matrix of standard errors associated with the beta-coefficients in betaX.

betaYse The vector of standard errors associated with the beta-coefficients in betaY.

correlation The matrix of correlations between genetic variants. If this variable is not provided, then we assume that genetic variants are uncorrelated.

exposure The names of the exposure variables.

outcome The name of the outcome variable.

snps The names of the genetic variants (SNPs) included in the analysis. The slots exposure, outcome, and snps are not required, but may be useful for keeping track of various MRInput objects. They are also used by the mr_plot function.

effect_allele The name of the effect allele for each SNP. The beta-coefficients are the associations with the exposure and outcome per additional copy of the effect allele.

other_allele The name of the non-effect allele.

eaf The expected allele frequencies (numeric). The slots effect_allele, other_allele, and eaf are neither required, nor currently used in the MendelianRandomization package. They are included for future compatibility with the MR-Base suite of functions.

MVcML_SdTheta *Standard error estimate for MVMR-cML-BIC*

Description

This is based on the profile likelihood of the set of valid IVs, which is not robust to uncertainty in model selection.

Usage

```
MVcML_SdTheta(b_exp, b_out, Sig_inv_l, theta, zero_ind, r_vec = NULL)
```

Arguments

b_exp	A matrix of SNP effects on the exposure variable.
b_out	A vector of SNP effects on the outcome variable.
Sig_inv_l	A list of inverse of covariance matrix.
theta	A vector of final estimates of causal effect of each exposure by MVMR-cML-BIC obtained from MVmr_cML_DP.
zero_ind	A vector of the index of valid IVs.
r_vec	A vector of estimated horizontal pleiotropic effects.

Value

A vector

`MVEgger-class` *MVEgger Class*

Description

An object containing the estimates produced using the multivariable MR-Egger method as well as various statistics.

Slots

`Model` Model always takes the value `random`, as only random-effects analyses are permitted.

`Orientate` The number of the risk factor that genetic associations are orientated to. The default value is `1`, meaning that genetic associations with the first risk factor are set to be positive.

`Exposure` The names of the exposure variables.

`Outcome` The name of the outcome variable.

`Correlation` The matrix of correlations between genetic variants.

`Estimate` The causal estimates from the inverse-variance weighted method.

`StdError.Est` The standard errors associated with `Estimate`.

`CILower.Est` The lower bounds of the confidence interval for `Estimate` based on `StdError`.

`CIUpper.Est` The upper bounds of the confidence interval for `Estimate` based on `StdError`.

`Pvalue.Est` P-value associated with the causal estimate.

`Intercept` The intercept estimate from the MR-Egger method. Under the InSIDE assumption, the intercept represents the average pleiotropic effect (average direct effect on the outcome) of a genetic variant. If the intercept differs from zero, this is evidence that the genetic variants are not all valid instruments; specifically, there is directional pleiotropy.

`StdError.Int` The standard error associated with `Intercept`.

`CILower.Int` The lower bound of the confidence interval for `Intercept` based on `StdError.Int`.

`CIUpper.Int` The upper bound of the confidence interval for `Estimate` based on `StdError.Int`.

`Pvalue.Int` P-value associated with the intercept.

`Alpha` The significance level used in constructing the confidence interval (default is 0.05).

`SNPs` The number of SNPs that were used in the calculation.

`RSE` The estimated residual standard error from the regression model.

`Heter.Stat` Heterogeneity statistic (Cochran's Q statistic) and associated p-value: the null hypothesis is that all genetic variants estimate the same causal parameter; rejection of the null is an indication that one or more variants may be pleiotropic.

MVGMM-class *MVGMM Class*

Description

An object containing the estimates produced using the multivariable generalized method of methods (GMM) method.

Slots

`robust` Whether the robust model with overdispersion heterogeneity is estimated.

`Exposure` The names of the exposure variables.

`Outcome` The name of the outcome variable.

`Correlation` The matrix of correlations between genetic variants if specified. If not specified, an identity matrix will be returned.

`ExpCorrelation` Whether an exposure correlation matrix was specified.

`CondFstat` The conditional F-statistic for each exposure.

`Estimate` The causal estimates from the GMM method.

`StdError` The standard errors associated with `Estimate`.

`CILower` The lower bounds of the confidence interval for `Estimate` based on `StdError`.

`CIUpper` The upper bounds of the confidence interval for `Estimate` based on `StdError`.

`Overdispersion` The estimate of the overdispersion parameter for the robust model. If this is negative, then a value of zero is used in the method.

`Pvalue` P-value associated with the causal estimates.

`Alpha` The significance level used in constructing confidence intervals (default is 0.05).

`Heter.Stat` Heterogeneity statistic (Cochran's Q statistic) and associated p-value (for non-robust model): the null hypothesis is that all genetic variants estimate the same causal parameter; rejection of the null is an indication that one or more genetic variants may be pleiotropic.

MVIVW-class *MVIVW Class*

Description

An object containing the estimates produced using the multivariable inverse-variance weighted (IVW) method as well as various statistics.

Slots

`Model` The model used for estimation: random-effects (`"random"`) or fixed-effect (`"fixed"`). The default option (`"default"`) is to use a fixed-effect model when there are three or fewer genetic variants, and a random-effects model when there are four or more. The (multiplicative) random-effects model allows for heterogeneity between the causal estimates targeted by the genetic variants by allowing over-dispersion in the regression model. Under-dispersion is not permitted (in case of under-dispersion, the residual standard error is set to 1, as in a fixed-effect analysis).

`Exposure` The names of the exposure variables.

`Outcome` The name of the outcome variable.

`Robust` Whether robust regression was used in the regression model relating the genetic associations with the outcome and those with the exposure.

`Correlation` The matrix of correlations between genetic variants.

`Estimate` The causal estimates from the inverse-variance weighted method.

`StdError` The standard errors associated with `Estimate`.

`CILower` The lower bounds of the confidence interval for `Estimate` based on `StdError`.

`CIUpper` The upper bounds of the confidence interval for `Estimate` based on `StdError`.

`Alpha` The significance level used in constructing the confidence interval (default is 0.05).

`Pvalue` P-value associated with the causal estimate.

`SNPs` The number of SNPs that were used in the calculation.

`RSE` The estimated residual standard error from the regression model.

`Heter.Stat` Heterogeneity statistic (Cochran's Q statistic) and associated p-value: the null hypothesis is that all genetic variants estimate the same causal parameter; rejection of the null is an indication that one or more variants may be pleiotropic.

`CondFstat` Conditional F statistics: An approximation of the first-stage conditional F statistics for all variants based on the summarized data. This represents the instrument strength for each exposure conditional on other exposures in the model. This is only reported if the sample sizes for the genetic associations with the exposures are provided.

`MVIVWME-class` *MVIVWME Class*

Description

An object containing the estimates produced using the multivariable inverse-variance weighted (IVW) method with measurement error, as well as various statistics.

Slots

`Model` The model used for estimation: random-effects (`"random"`) or fixed-effect (`"fixed"`). The default option (`"default"`) is to use a fixed-effect model when there are three or fewer genetic variants, and a random-effects model when there are four or more. The (multiplicative) random-effects model allows for heterogeneity between the causal estimates targeted by the genetic variants by allowing over-dispersion in the regression model. Under-dispersion is not permitted (in case of under-dispersion, the residual standard error is set to 1, as in a fixed-effect analysis).

`Exposure` The names of the exposure variables.

`Outcome` The name of the outcome variable.

`Correlation` The matrix of correlations between genetic variants.

`Estimate` The causal estimates from the inverse-variance weighted method.

`StdError` The standard errors associated with `Estimate`.

`CILower` The lower bounds of the confidence interval for `Estimate` based on `StdError`.

`CIUpper` The upper bounds of the confidence interval for `Estimate` based on `StdError`.

`Alpha` The significance level used in constructing the confidence interval (default is 0.05).

`Pvalue` P-value associated with the causal estimate.

`SNPs` The number of SNPs that were used in the calculation.

`RSE` The estimated residual standard error from the regression model.

`Heter.Stat` Heterogeneity statistic (Cochran's Q statistic) and associated p-value: the null hypothesis is that all genetic variants estimate the same causal parameter; rejection of the null is an indication that one or more variants may be pleiotropic.

`MVLasso-class` *MRMVLasso class*

Description

An object containing the estimates produced using the multivariable MR-Lasso method as well as various statistics.

Slots

`Exposure` The names of the exposure variables.

`Outcome` The name of the outcome variable.

`Orientate` The number of the risk factor that genetic associations are orientated to. The default value is `1`, meaning that genetic associations with the first risk factor are set to be positive.

`Estimate` The causal estimates from the multivariable MR-Lasso method.

`StdError` The standard errors associated with `Estimate`.

`CILower` The lower bounds of the confidence intervals for `Estimate` based on `StdError`.

`CIUpper` The upper bounds of the confidence intervals for `Estimate` based on `StdError`.

Alpha The significance level used in constructing the confidence interval (default is 0.05).

Pvalue P-values associated with the causal estimates from the multivariable MR-Lasso method.

SNPs The number of SNPs used in the calculation.

RegEstimate The estimates from the regularized regression model used in the multivariable MR-Lasso method.

RegIntercept The intercept estimates from the regularized regression model used in the multivariable MR-Lasso method. An intercept estimate of zero identifies the corresponding genetic variant as a valid instrument. Genetic variants with non-zero intercept estimates will be excluded from the MR-Lasso method, which is obtained as a post-lasso estimator.

Valid The number of genetic variants that have been identified as valid instruments.

ValidSNPs The names of genetic variants that have been identified as valid instruments.

Lambda The value of the tuning parameter used to compute RegEstimate (default is to calulate Lambda using the heterogeneity stopping rule).

MVMedian-class *MRMVMedian class*

Description

An object containing the estimates produced using the multivariable median method as well as various statistics.

Slots

Exposure The names of the exposure variables.

Outcome The name of the outcome variable.

Estimate The causal estimates from the multivariable median method.

StdError The standard errors associated with Estimate (obtained from bootstrapping).

CILower The lower bounds of the confidence intervals for Estimate based on StdError.

CIUpper The upper bounds of the confidence intervals for Estimate based on StdError.

Alpha The significance level used in constructing the confidence interval (default is 0.05).

Pvalue P-values associated with the causal estimates from the Wald method.

SNPs The number of SNPs that used in the calculation.

MVmr_cML *MVMRcML method with Data Perturbation*

Description

This is the internal MVMRcML-BIC function of mr_mvcML.

Usage

```
MVmr_cML(
  b_exp,
  b_out,
  se_bx,
  Sig_inv_l,
  n,
  K_vec = as.numeric(c()),
  random_start = 1L,
  min_theta_range = -0.5,
  max_theta_range = 0.5,
  maxit = 100L,
  thres = 1e-04
)
```

Arguments

b_exp	A m*L matrix of SNP effects on the exposure variable.
b_out	A m*1 matrix of SNP effects on the outcome variable.
se_bx	A m*L matrix of standard errors of b_exp.
Sig_inv_l	A list of the inverse of m covariance matrices.
n	The smallest sample size of the L+1 GWAS dataset.
K_vec	Sets of candidate K's, the constraint parameter representing number of invalid IVs.
random_start	Number of random start points, default is 1.
min_theta_range	The lower bound of the uniform distribution for each initial value for theta generated from.
max_theta_range	The upper bound of the uniform distribution for each initial value for theta generated from.
maxit	Maximum number of iterations for each optimization, default is 100.
thres	Threshold for convergence criterion.

Value

A list

BIC_theta Estimated causal effect from MVMR-cML-BIC

BIC_invalid Invalid IVs selected by MVMR-cML-BIC

l_vec A vector of negative log-likelihood corresponding to each K.

K_vec A vector of candidate K's

theta_vec A matrix of causal parameter estimates, each column corresponds to a candidate K.

Conv_vec A vector of successful convergence indicators corresponding to each K.

Converge Indicator of successful convergence, 0 means success, 1 means failure.

BIC_vec Data perturbation with successful convergence

Khat The length of BIC_invalid.

MVmr_cML_DP *MVMRcML method with Data Perturbation*

Description

This is the internal MVMRcML-DP function of mr_mvcML.

Usage

```
MVmr_cML_DP(
  b_exp,
  b_out,
  se_bx,
  Sig_inv_l,
  n,
  K_vec = as.numeric(c()),
  random_start = 1L,
  num_pert = 100L,
  min_theta_range = -0.5,
  max_theta_range = 0.5,
  maxit = 100L,
  thres = 1e-04
)
```

Arguments

b_exp	A m*L matrix of SNP effects on the exposure variable.
b_out	A m*1 matrix of SNP effects on the outcome variable.
se_bx	A m*L matrix of standard errors of b_exp.
Sig_inv_l	A list of the inverse of m covariance matrices.

n	The smallest sample size of the L+1 GWAS dataset.
K_vec	Sets of candidate K's, the constraint parameter representing number of invalid IVs.
random_start	Number of random start points, default is 1.
num_pert	Number of perturbation, default is 100.
min_theta_range	The lower bound of the uniform distribution for each initial value for theta generated from.
max_theta_range	The upper bound of the uniform distribution for each initial value for theta generated from.
maxit	Maximum number of iterations for each optimization, default is 100.
thres	Threshold for convergence criterion.

Value

A list

BIC_theta Estimated causal effect from MVMR-cML-BIC

BIC_invalid Invalid IVs selected by MVMR-cML-BIC

BIC_DP_theta Estimated causal effect from MVMR-cML-DP

BIC_DP_se Estimate standard error for `BIC_DP_theta`

eff_DP_B Data perturbation with successful convergence

DP_ninvalid A vector of the number of selected invalid IVs by MVMRcML-BIC in each data perturbation.

`MVMRcML-class` *MVMRcML Class*

Description

An object containing the results of MVMRcML.

Slots

`Exposure` The names of the exposure variables.

`Outcome` The name of the outcome variable.

`Estimate` The causal estimates from the multivariable MRcML method.

`StdError` The standard errors associated with `Estimate`.

`CILower` The lower bounds of the confidence intervals for `Estimate` based on `StdError`.

`CIUpper` The upper bounds of the confidence intervals for `Estimate` based on `StdError`.

`Alpha` The significance level used in constructing the confidence interval (default is 0.05).

`Pvalue` P-values associated with the causal estimates from the multivariable MRcML method.

`BIC_invalid` The index set of selected invalid IVs by MVMRcML-BIC.

`K_hat` The number of selected invalid IVs by MVMRcML-BIC, or a vector for each data perturbation in MVMRcML-DP.

`eff_DP_B` The number of data perturbations with successful convergence in MVMRcML-DP.

`SNPs` The number of SNPs that were used in the calculation.

`DP` Indicator of whether data perturbation is applied.

`MVPCGMM-class` *MVPCGMM Class*

Description

An object containing the estimates produced using the multivariable principal components generalized method of methods (PC-GMM) method as well as various statistics.

Slots

`robust` Whether the robust model with overdispersion heterogeneity is estimated.

`Exposure` The names of the exposure variables.

`Outcome` The name of the outcome variable.

`Correlation` The matrix of correlations between genetic variants.

`ExpCorrelation` Whether an exposure correlation matrix was specified.

`CondFstat` The conditional F-statistic for each exposure.

`Estimate` The causal estimates from the PC-GMM method.

`StdError` The standard errors associated with `Estimate`.

`CILower` The lower bounds of the confidence interval for `Estimate` based on `StdError`.

`CIUpper` The upper bounds of the confidence interval for `Estimate` based on `StdError`.

`Overdispersion` The estimate of the overdispersion parameter for the robust model.

`PCs` The number of genetic principal components used to instrument the exposures.

`Pvalue` P-value associated with the causal estimates.

`Alpha` The significance level used in constructing confidence intervals (default is 0.05).

`Heter.Stat` Heterogeneity statistic (Cochran's Q statistic) and associated p-value (for non-robust model): the null hypothesis is that all principal components estimate the same causal parameter; rejection of the null is an indication that one or more principal components may be pleiotropic.

PCGMM-class *PCGMM Class*

Description

An object containing the estimates produced using the univariable principal components generalized method of methods (PC-GMM) method as well as various statistics.

Slots

`robust` Whether the robust model with overdispersion heterogeneity is estimated.

`Exposure` The name of the exposure variable.

`Outcome` The name of the outcome variable.

`Correlation` The matrix of correlations between genetic variants.

`Estimate` The causal estimate from the PC-GMM method.

`StdError` The standard error associated with `Estimate`.

`CILower` The lower bound of the confidence interval for `Estimate` based on `StdError`.

`CIUpper` The upper bound of the confidence interval for `Estimate` based on `StdError`.

`Fstat` The first-stage F statistic for all genetic principal components used as instruments.

`Overdispersion` The estimate of the overdispersion parameter for the robust model.

`PCs` The number of genetic principal components used to instrument the exposure.

`Pvalue` P-value associated with the causal estimate.

`Alpha` The significance level used in constructing the confidence interval (default is 0.05).

`Heter.Stat` Heterogeneity statistic (Cochran's Q statistic) and associated p-value (for non-robust model): the null hypothesis is that all principal components estimate the same causal parameter; rejection of the null is an indication that one or more principal components may be pleiotropic.

pheno_input *Extract summarized data from PhenoScanner*

Description

The function `pheno_input` extracts summarized data on associations with named exposure and outcome variables from PhenoScanner.

Usage

```
pheno_input(
  snps,
  exposure,
  pmidE,
  ancestryE,
  outcome,
  pmidO,
  ancestryO,
  correl = NULL
)
```

Arguments

snps	The names (rsid) of the genetic variants to be included in the analysis.
exposure	The name of the exposure variable.
pmidE	The PubMed ID (PMID) of the publication in which the genetic association estimates with the exposure were originally reported. Some variables are reported in multiple consortia (for example, associations with coronary artery disease by CARDIoGRAM in 2011 [PMID:21378990], by CARDIoGRAMplusC4D in 2013, and again by CARDIoGRAMplusC4D in 2015 [PMID:26343387]). Equally, some publications reported associations on multiple variables (for example, CARDIoGRAMplusC4D in 2015 [PMID:26343387] reported associations with coronary artery disease and with myocardial infarction). By providing the variable name and the PubMed ID, the set of associations is (almost) uniquely identified.
ancestryE	The ancestry of individuals in which estimates were obtained. A small number of studies reported genetic association estimates for a single variable in a single publication for multiple ethnicities (for example, associations with log(eGFR creatinine) from CKD-Gen in 2016 [PMID:26831199] were reported for both Europeans and Africans). The combination of exposure name, PubMed ID, and ancestry uniquely defines the set of associations. Providing the ancestry also reminds analysts of the additional complication of conducting Mendelian randomization when associations with the exposure and with the outcome are in individuals of different ancestry. Most association estimates are obtained in `"European"` or `"Mixed"` populations, although some are obtained in `"African"`, `"Asian"`, or `"Hispanic"` populations.
outcome	The name of the outcome variable.
pmidO	The PubMed ID of the publication in which the genetic association estimates with the outcome were originally reported.
ancestryO	The ancestry of individuals in which genetic association estimates with the outcome were obtained.
correl	The correlations between the genetic variants. If this is not specified, then the genetic variants are assumed to be uncorrelated. Note that for the correlations to reference the correct variants, the list of genetic variants needs to be in alphabetical order.

Details

The PhenoScanner bioinformatic tool is a curated database of publicly available results from large-scale genetic association studies. Queries can be made for individual genetic variants (SNPs and small indels), or for multiple variants in a single batch query. These association estimates and their standard errors can be used in Mendelian randomization analyses.

The `phenoscanner` command is included in the `MendelianRandomization` package with permission of James Staley. The function is also available in a standalone package from github: https://github.com/phenoscanner/phenoscanner.

Value

The output of the `pheno_input` function is an `MRInput` object that can be used directly in any of the estimation functions (such as `mr_ivw`) or in the plotting function `mr_plot`. The output contains:

`bx`	The genetic associations with the exposure.
`bxse`	The corresponding standard errors.
`by`	The genetic associations with the outcome.
`byse`	The corresponding standard errors.
`correlation`	The matrix of genetic correlations as specified by the user.
`exposure`	A character string giving the name of the exposure as provided in the PhenoScanner database.
`outcome`	A character string giving the name of the outcome as provided in the PhenoScanner database.
`snps`	A vector of character strings with the names of the genetic variants.

References

James R Staley, James Blackshow, Mihir A Kamat, Steve Ellis, Prvaeen Surendran, Benjamin B Sun, Dirk S Paul, Daniel Freitag, Stephen Burgess, John Danesh, Robin Young, and Adam S Butterworth. PhenoScanner: a database of human genotype–phenotype associations. Bioinformatics 2016. doi: 10.1093/bioinformatics/btw373.

Examples

```
# pheno_input(snps=c("rs12916", "rs2479409", "rs217434", "rs1367117"),
# exposure = "Low density lipoprotein", pmidE = "24097068", ancestryE = "European",
# outcome = "Coronary artery disease", pmidO = "26343387", ancestryO = "Mixed")
```

phenoscanner *PhenoScanner*

Description

The phenoscanner function queries the PhenoScanner database of genotype-phenotype associations from inside R.

Usage

```
phenoscanner(
  snpquery = NULL,
  genequery = NULL,
  regionquery = NULL,
  catalogue = "GWAS",
  pvalue = 1e-05,
  proxies = "None",
  r2 = 0.8,
  build = 37
)
```

Arguments

snpquery	a vector of SNPs.
genequery	a vector of gene names.
regionquery	a vector of genomic regions.
catalogue	the catalogue to be searched (options: None, GWAS, eQTL, pQTl, mQTL, methQTL).
pvalue	the p-value threshold.
proxies	the proxies database to be searched (options: None, AFR, AMR, EAS, EUR, SAS).
r2	the r2 threshold.
build	the genome build (options: 37, 38).

Value

a list containing a data.frame of association results and a data.frame of SNP/Region/Gene information from PhenoScanner.

Author(s)

PhenoScanner <phenoscanner@gmail.com>

Examples

```
# SNP
# res <- phenoscanner(snpquery="rs10840293")
# head(res$results)
# res$snps

# Gene
# res <- phenoscanner(genequery="SWAP70")
# head(res$results)
# res$snps

# Region
# res <- phenoscanner(regionquery="chr11:9685624-9774538")
# head(res$results)
# res$regions
```

`PIVW-class` *PIVW Class*

Description

An object containing the estimate produced using the penalized inverse-variance weighted (pIVW) method as well as various statistics.

Slots

`Over.dispersion` Should the method consider overdispersion (balanced horizontal pleiotropy)? Default is TRUE.

`Boot.Fieller` If `Boot.Fieller=TRUE`, then the P-value and the confidence interval of the causal effect will be calculated based on the bootstrapping Fieller method. Otherwise, the P-value and the confidence interval of the causal effect will be calculated from the normal distribution. It is recommended to use the bootstrapping Fieller method when `Condition` (the estimated effective sample size) is smaller than 10. By default, `Boot.Fieller=TRUE`.

`Lambda` The penalty parameter in the pIVW estimator. The penalty parameter plays a role in the bias-variance trade-off of the estimator. It is recommended to choose `lambda=1` to achieve the smallest bias and valid inference. By default, `lambda=1`.

`Delta` The z-score threshold for IV selection. By default, `delta=0` (i.e., no IV selection will be conducted).

`Exposure` The name of the exposure variable.

`Outcome` The name of the outcome variable.

`Estimate` The causal point estimate from the pIVW estimator.

`StdError` The standard error associated with `Estimate`.

`CILower` The lower bound of the confidence interval for `Estimate`, which is derived from the bootstrapping Fieller method or normal distribution. For the bootstrapping Fieller's interval, if it contains multiple ranges, then lower limits of all ranges will be reported.

`CIUpper` The upper bound of the confidence interval for `Estimate`, which is derived from the bootstrapping Fieller method or normal distribution. For the bootstrapping Fieller's interval, if it contains multiple ranges, then upper limits of all ranges will be reported.

`Alpha` The significance level used in constructing the confidence interval (default is 0.05).

`Pvalue` P-value associated with the causal estimate from the pIVW estimator.

`Tau2` The variance of the balanced horizontal pleiotropy. `Tau2` is calculated by using all IVs in the data before conducting the IV selection.

`SNPs` The number of SNPs after IV selection.

`Condition` The estimated effective sample size. It is recommended to be greater than 5 for the pIVW estimator to achieve reliable asymptotic properties.

`WeightedMedian-class` *WeightedMedian Class*

Description

An object containing the estimate produced using the median-based method as well as various statistics.

Slots

`Type` The type of median that has been calculated, `"simple"`, `"weighted"`, or `"penalized"`.

`Exposure` The name of the exposure variable.

`Outcome` The name of the outcome variable.

`Estimate` The causal point estimate from the median-based method.

`StdError` The standard error associated with `Estimate` (obtained from bootstrapping).

`CILower` The lower bound of the confidence interval for `Estimate` based on `StdError`.

`CIUpper` The upper bound of the confidence interval for `Estimate` based on `StdError`.

`Alpha` The significance level used in constructing the confidence interval (default is 0.05).

`Pvalue` P-value associated with the causal estimate from the Wald method.

`SNPs` The number of SNPs that used in the calculation.

Index

∗ **datasets**
    calcium, 3
    ldlc, 11

calc.rho *(*calcium*)*, 3
calcium, 3
calciumse *(*calcium*)*, 3
chdlodds *(*ldlc*)*, 11
chdloddsse *(*ldlc*)*, 11
CLR-class, 4
cML_estimate, 5
cML_estimate_random, 6
cML_SdTheta, 7

DIVW-class, 8

Egger-class, 8

fastgluc *(*calcium*)*, 3
fastglucse *(*calcium*)*, 3

hdlc *(*ldlc*)*, 11
hdlcse *(*ldlc*)*, 11

invcov_mvmr, 10
IVW-class, 10

ldlc, 11
ldlcse *(*ldlc*)*, 11
lipid_eaf *(*ldlc*)*, 11
lipid_effect *(*ldlc*)*, 11
lipid_other *(*ldlc*)*, 11

MaxLik-class, 12
mr_allmethods, 13
mr_allmethods,MRInput-method *(*mr_allmethods*)*, 13
mr_clr, 14
mr_clr,MRInput-method *(*mr_clr*)*, 14
mr_cML, 16
mr_cML,MRInput-method *(*mr_cML*)*, 16
mr_conmix, 18
mr_conmix,MRInput-method *(*mr_conmix*)*, 18
mr_divw, 21
mr_divw,MRInput-method *(*mr_divw*)*, 21
mr_egger, 22
mr_egger,MRInput-method *(*mr_egger*)*, 22
mr_forest, 25
mr_forest,MRInput-method *(*mr_forest*)*, 25
mr_funnel, 27
mr_funnel,MRInput-method *(*mr_funnel*)*, 27
mr_hetpen, 27
mr_hetpen,MRInput-method *(*mr_hetpen*)*, 27
mr_input, 30
mr_ivw, 31
mr_ivw,MRInput-method *(*mr_ivw*)*, 31
mr_lasso, 35
mr_lasso,MRInput-method *(*mr_lasso*)*, 35
mr_loo, 37
mr_loo,MRInput-method *(*mr_loo*)*, 37
mr_maxlik, 38
mr_maxlik,MRInput-method *(*mr_maxlik*)*, 38
mr_mbe, 40
mr_mbe,MRInput-method *(*mr_mbe*)*, 40
mr_median, 43
mr_median,MRInput-method *(*mr_median*)*, 43
mr_mvcML, 45
mr_mvcML,MRMVInput-method *(*mr_mvcML*)*, 45
mr_mvegger, 48
mr_mvegger,MRMVInput-method *(*mr_mvegger*)*, 48
mr_mvgmm, 50
mr_mvgmm,MRMVInput-method *(*mr_mvgmm*)*, 50
mr_mvinput, 52
mr_mvivw, 54
mr_mvivw,MRMVInput-method *(*mr_mvivw*)*, 54
mr_mvivwme, 57
mr_mvivwme,MRMVInput-method *(*mr_mvivwme*)*, 57
mr_mvlasso, 59

mr_mvlasso,MRMVInput-method
(mr_mvlasso), 59
mr_mvmedian, 61
mr_mvmedian,MRMVInput-method
(mr_mvmedian), 61
mr_mvpcgmm, 63
mr_mvpcgmm,MRMVInput-method
(mr_mvpcgmm), 63
mr_pcgmm, 66
mr_pcgmm,MRInput-method (mr_pcgmm), 66
mr_pivw, 68
mr_pivw,MRInput-method (mr_pivw), 68
mr_plot, 71
mr_plot,MRAll-method (mr_plot), 71
mr_plot,MRInput-method (mr_plot), 71
mr_plot,MRMVInput-method (mr_plot), 71
MRAll-class, 73
MRcML-class, 73
MRConMix-class, 74
MRHetPen-class, 74
MRInput-class, 75
MRLasso-class, 76
MRMBE-class, 77
MRMVInput-class, 77
MVcML_SdTheta, 78
MVEgger-class, 79
MVGMM-class, 80
MVIVW-class, 80
MVIVWME-class, 81
MVLasso-class, 82
MVMedian-class, 83
MVmr_cML, 84
MVmr_cML_DP, 85
MVMRcML-class, 86
MVPCGMM-class, 87

PCGMM-class, 88
pheno_input, 88
phenoscanner, 91
PIVW-class, 92

trig (ldlc), 11
trigse (ldlc), 11

WeightedMedian-class, 93