Package 'MendelianRandomization'

March 7, 2019

Type Package

Title Mendelian Randomization Package

Version 0.4.0

Date 2019-03-06

Author Olena Yavorska <olena.yavorska@ludwig.ox.ac.uk>
James Staley <james.staley@bristol.ac.uk>

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), rjson

VignetteBuilder knitr

LazyData TRUE

RoxygenNote 6.1.1

NeedsCompilation no

Repository CRAN

Date/Publication 2019-03-07 11:33:29 UTC

R topics documented:

`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.

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.

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).

`extract.pheno.csv` *Extract summarized data from a PhenoScanner .csv file (legacy)*

Description

The function `extract.pheno.csv` extracts summarized data on associations with named exposure and outcome variables from a .csv file provided by PhenoScanner.

Usage

```
extract.pheno.csv(exposure, pmidE, ancestryE, outcome, pmidO, ancestryO,
  file, rsq.proxy = 1, snps = "all")
```

Arguments

`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.

`file` The file path where the PhenoScanner .csv file can be found.

`rsq.proxy` A proxy variant is a genetic variant in close correlation (high linkage disequilibrium) with the named variant. If PhenoScanner is run with proxies included, then proxies can be included in the analysis. In the second example below, with log(eGFR creatinine) as the exposure and Tanner stage as the outcome, the association of variant rs12785878 with the outcome is not reported. Instead, rs4944958 is used as a proxy for rs12785878. The association of rs4944958 with the outcome is used in the resulting `MRInput` object. The correlation between the two variants is reported as `R^2 = 1.000`. A message will always appear when a proxy variant is included in an analysis in place of the primary variant. The value of `rsq.proxy` is used as a threshold in the analysis; a variant is only included in the analysis if the value of `R^2` equals or excedes this threshold. The default option is `rsq.proxy = 1`, meaning that only perfect proxies are used in the analysis.

`snps` The names (rsIDs) of the genetic variants to be included in the analysis. The default option is `"all"`, indicating that all the genetic variants in the .csv file with beta-coefficients and standard errors for their associations with the risk factor and with the outcome should be used in the analysis. Otherwise, only variants whose names are included in the vector of character strings provided as `snps` will be included in the analysis.

Details

Note that this function was written for a previous version of PhenoScanner. It has not been updated, as it has been overtaken by the `pheno_input` function that queries PhenoScanner directly from R.

The PhenoScanner bioinformatic tool (http://phenoscanner.medschl.cam.ac.uk) 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. One of the output files is a .csv file containing all associations of variables with each of the SNPs. For commonly genotyped variants, associations with up to 200 variables may be reported. These association estimates and their standard errors can be used in Mendelian randomization analyses.

The plan is to enable PhenoScanner to be queried directly from the MendelianRandomization package. However, this functionality is currently unavailable.

The `extract.pheno.csv` function takes the output from the web version of PhenoScanner, and converts this into an `MRInput` object. PhenoScanner is still under development. This function is designed for output from PhenoScanner version 1.1 (Little Miss Sunshine).

Value

The output of the `extract.pheno.csv` 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. Currently, this is set to the empty matrix (`matrix()`), meaning that only uncorrelated variants can be used in the `extract.pheno.csv` function.
`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

```
path.noproxy <- system.file("extdata", "vitD_snps_PhenoScanner.csv",
package = "MendelianRandomization")
path.proxies <- system.file("extdata", "vitD_snps_PhenoScanner_proxies.csv",
 package = "MendelianRandomization")
 # these two files from PhenoScanner are provided
 # as part of the MendelianRandomization package

extract.pheno.csv(
 exposure = "log(eGFR creatinine)", pmidE = 26831199, ancestryE = "European",
 outcome = "Tanner stage", pmidO = 24770850, ancestryO = "European",
 file = path.noproxy)

extract.pheno.csv(
 exposure = "log(eGFR creatinine)", pmidE = 26831199, ancestryE = "European",
 outcome = "Tanner stage", pmidO = 24770850, ancestryO = "European",
 rsq.proxy = 0.6, file = path.proxies)

extract.pheno.csv(
 exposure = "log(eGFR creatinine)", pmidE = 26831199, ancestryE = "European",
 outcome = "Asthma", pmidO = 20860503, ancestryO = "European",
 rsq.proxy = 0.6, file = path.proxies)
```

`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.

`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.

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/atbvaha.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.

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"`).

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.

`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.

See Also

extract.pheno.csv() for a description of how the above values can be extracted from PhenoScanner http://www.phenoscanner.medschl.cam.ac.uk/.

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.

---

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_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 = -1, CIMax = 1, CIStep = 0.001,
  alpha = 0.05)

## S4 method for signature 'MRInput'
mr_conmix(object, psi = 0, CIMin = -1, CIMax = 1,
  CIStep = 0.001, 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 (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

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.

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. bioRxiv 2019. doi: [to add].

Examples

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

---

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_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.

See Also

extract.pheno.csv() for a description of how an MRInput object can be extracted from PhenoScanner (http://www.phenoscanner.medschl.cam.ac.uk/).

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.

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_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, bxse = ldlcse, by = chdlodds, byse = chdloddsse))
mr_maxlik(mr_input(bx = ldlc, bxse = ldlcse, by = chdlodds, byse = chdloddsse), 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_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_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", correl = FALSE,
  distribution = "normal", alpha = 0.05, ...)

## S4 method for signature 'MRMVInput'
mr_mvivw(object, model = "default",
  correl = FALSE, 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.
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.
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.
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.

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.

Examples

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

---

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

Description

The function mr_plot has two functionalities. It can generate a visual representation of both MRInput 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)
```

Arguments

object	An MRInput object or an MRAll object.
error	When viewing an MRInput object, one can choose whether to include error bars (default is to include).
line	When viewing an MRInput object, one can choose whether to include the IVW estimate (line = "ivw") or the MR-Egger estimate (line = "egger").
orientate	When viewing an MRInput 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 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 object with interactive set to FALSE, setting labels to TRUE means that the name of each genetic variants appears above the corresponding datapoint.
...	Additional arguments to be passed to other methods.

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. 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
```

`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.

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.

`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.

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
```

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 (http://phenoscanner.medschl.cam.ac.uk) 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")
```

`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