Package 'HIMA'

August 24, 2024

Type Package

Title High-Dimensional Mediation Analysis

Version 2.2.2

Date 2024-08-24

Description Allows to estimate and test high-dimensional mediation effects based on advanced mediator screening and penalized regression techniques. Methods used in the package refer to Zhang H, Zheng Y, Zhang Z, Gao T, Joyce B, Yoon G, Zhang W, Schwartz J, Just A, Colicino E, Vokonas P, Zhao L, Lv J, Baccarelli A, Hou L, Liu L. Estimating and Testing High-dimensional Mediation Effects in Epigenetic Studies. Bioinformatics. (2016) <doi:10.1093/bioinformatics/btw351>. PMID: 27357171.

License GPL-3

Depends R (>= 3.4.0), ncvreg, glmnet

Imports utils, stats, MASS, survival, HDMT, hdi, conquer, quantreg, hommel, iterators, parallel, foreach, doParallel

Collate utils.R hima.R survHIMA.R microHIMA.R dblassoHIMA.R qHIMA.R eHIMA.R hima2.R himaDat.R onAttach.R HIMA-package.R

Encoding UTF-8

LazyData true

URL https://github.com/YinanZheng/HIMA/

BugReports https://github.com/YinanZheng/HIMA/issues/

RoxygenNote 7.3.2

NeedsCompilation no

Author Yinan Zheng [aut, cre] (<https://orcid.org/0000-0002-2006-7320>),
Haixiang Zhang [aut],
Lifang Hou [aut],
Lei Liu [aut, cph]

Maintainer Yinan Zheng <y-zheng@northwestern.edu>

Repository CRAN

Date/Publication 2024-08-24 18:00:02 UTC

Contents

`HIMA-package` *High-Dimensional Mediation Analysis for 'Omic' Data*

Description

HIMA is an R package for estimating and testing high-dimensional mediation effects in omic studies. HIMA can perform high-dimensional mediation analysis on a wide range of omic data types as potential mediators, including epigenetics, transcriptomics, proteomics, metabolomics, and microbiomics. HIMA can also handle survival data mediation analysis and perform quantile mediation analysis.

Package:	HIMA
Type:	Package
Version:	2.2.2
Date:	2024-08-24
License:	GPL-3

Details

If package "qvalue" is not found during installation, please first install "qvalue" package # through Bioconductor: https://www.bioconductor.org/packages/release/bioc/html/qvalue.html

Author(s)

Yinan Zheng <y-zheng@northwestern.edu>, Haixiang Zhang <haixiang.zhang@tju.edu.cn>, Lei liu (Contact) <lei.liu@wustl.edu>

Maintainer: Yinan Zheng <y-zheng@northwestern.edu>

References

1. Zhang H, Zheng Y, Zhang Z, Gao T, Joyce B, Yoon G, Zhang W, Schwartz J, Just A, Colicino E, Vokonas P, Zhao L, Lv J, Baccarelli A, Hou L, Liu L. Estimating and Testing High-dimensional Mediation Effects in Epigenetic Studies. Bioinformatics. 2016. DOI: 10.1093/bioinformatics/btw351. PMID: 27357171; PMCID: PMC5048064

2. Zhang H, Zheng Y, Hou L, Zheng C, Liu L. Mediation Analysis for Survival Data with High-Dimensional Mediators. Bioinformatics. 2021. DOI: 10.1093/bioinformatics/btab564. PMID: 34343267; PMCID: PMC8570823

3. Zhang H, Chen J, Feng Y, Wang C, Li H, Liu L. Mediation Effect Selection in High-dimensional and Compositional Microbiome data. Stat Med. 2021. DOI: 10.1002/sim.8808. PMID: 33205470; PMCID: PMC7855955

4. Zhang H, Chen J, Li Z, Liu L. Testing for Mediation Effect with Application to Human Microbiome Data. Stat Biosci. 2021. DOI: 10.1007/s12561-019-09253-3. PMID: 34093887; PMCID: PMC8177450

5. Perera C, Zhang H, Zheng Y, Hou L, Qu A, Zheng C, Xie K, Liu L. HIMA2: High-dimensional Mediation Analysis and Its Application in Epigenome-wide DNA Methylation Data. BMC Bioinformatics. 2022. DOI: 10.1186/s12859-022-04748-1. PMID: 35879655; PMCID: PMC9310002

6. Zhang H, Hong X, Zheng Y, Hou L, Zheng C, Wang X, Liu L. High-Dimensional Quantile Mediation Analysis with Application to a Birth Cohort Study of Mother–Newborn Pairs. Bioinformatics. 2024. DOI: 10.1093/bioinformatics/btae055. PMID: 38290773; PMCID: PMC10873903

7. Bai X, Zheng Y, Hou L, Zheng C, Liu L, Zhang H. An Efficient Testing Procedure for High-dimensional Mediators with FDR Control. Statistics in Biosciences. 2024. DOI: 10.1007/s12561-024-09447-4.

`dblassoHIMA` *This is the function for high-dimensional mediation analysis using de-biased lasso HIMA with de-biased lasso*

Description

`dblassoHIMA` is used to estimate and test high-dimensional mediation effects using de-biased lasso penalty.

Usage

```
dblassoHIMA(
  X,
  Y,
  M,
  COV = NULL,
  Y.family = c("gaussian", "binomial"),
  topN = NULL,
  scale = TRUE,
  FDRcut = 0.05,
  verbose = FALSE
)
```

Arguments

X	a vector of exposure.
Y	a vector of outcome. Can be either continuous or binary (0-1).
M	a data.frame or matrix of high-dimensional mediators. Rows represent samples, columns represent variables.
COV	a data.frame or matrix of covariates dataset for testing the association M ~ X and Y ~ M.
Y.family	either 'gaussian' (default) or 'binomial', depending on the data type of outcome (Y). This parameter is passed to function lasso.proj in R package hdi for debiased lasso penalization.
topN	an integer specifying the number of top markers from sure independent screening. Default = NULL. If NULL, topN will be either ceiling(n/log(n)) if Y.family = 'gaussian', or ceiling(n/(2*log(n))) if Y.family = 'binomial', where n is the sample size. If the sample size is greater than topN (pre-specified or calculated), all mediators will be included in the test (i.e. low-dimensional scenario).
scale	logical. Should the function scale the data? Default = TRUE.
FDRcut	HDMT pointwise FDR cutoff applied to select significant mediators. Default = 0.05.
verbose	logical. Should the function be verbose? Default = FALSE.

Value

A data.frame containing mediation testing results of significant mediators (FDR <FDRcut).

Index: mediation name of selected significant mediator.

alpha_hat: coefficient estimates of exposure (X) –> mediators (M) (adjusted for covariates).

alpha_se: standard error for alpha.

beta_hat: coefficient estimates of mediators (M) –> outcome (Y) (adjusted for covariates and exposure).

beta_se: standard error for beta.

IDE: mediation (indirect) effect, i.e., alpha*beta.

rimp: relative importance of the mediator.

pmax: joint raw p-value of selected significant mediator (based on HDMT pointwise FDR method).

References

Perera C, Zhang H, Zheng Y, Hou L, Qu A, Zheng C, Xie K, Liu L. HIMA2: high-dimensional mediation analysis and its application in epigenome-wide DNA methylation data. BMC Bioinformatics. 2022. DOI: 10.1186/s12859-022-04748-1. PMID: 35879655; PMCID: PMC9310002

Examples

```
## Not run:
# Note: In the following examples, M1, M2, and M3 are true mediators.
data(himaDat)

# When Y is continuous and normally distributed
# Example 1 (continuous outcome):
head(himaDat$Example1$PhenoData)

dblassohima.fit <- dblassoHIMA(X = himaDat$Example1$PhenoData$Treatment,
                               Y = himaDat$Example1$PhenoData$Outcome,
                               M = himaDat$Example1$Mediator,
                               COV = himaDat$Example1$PhenoData[, c("Sex", "Age")],
                               Y.family = 'gaussian',
                               scale = FALSE, # Disabled only for simulation data
                               verbose = TRUE)
dblassohima.fit

# When Y is binary (should specify Y.family)
# Example 2 (binary outcome):
head(himaDat$Example2$PhenoData)

dblassohima.logistic.fit <- dblassoHIMA(X = himaDat$Example2$PhenoData$Treatment,
                                        Y = himaDat$Example2$PhenoData$Disease,
                                        M = himaDat$Example2$Mediator,
                                      COV = himaDat$Example2$PhenoData[, c("Sex", "Age")],
                                        Y.family = 'binomial',
                                      scale = FALSE, # Disabled only for simulation data
                                        verbose = TRUE)
dblassohima.logistic.fit

## End(Not run)
```

---

eHIMA *Efficient high-dimensional mediation analysis*

Description

eHIMA is used to estimate and test high-dimensional mediation effects using an efficient algorithm. It provides higher statistical power than the standard hima. Note: efficient HIMA is only applicable to mediators and outcomes that are both continuous and normally distributed.

Usage

```
eHIMA(
  X,
  M,
  Y,
```

```
  COV = NULL,
  topN = NULL,
  scale = TRUE,
  FDRcut = 0.05,
  verbose = FALSE
)
```

Arguments

X	a vector of exposure.
M	a data.frame or matrix of high-dimensional mediators. Rows represent samples, columns represent mediator variables. M has to be continuous and normally distributed.
Y	a vector of continuous outcome. Do not use data.frame or matrix.
COV	a matrix of adjusting covariates. Rows represent samples, columns represent variables. Can be NULL.
topN	an integer specifying the number of top markers from sure independent screening. Default = NULL. If NULL, topN will be 2*ceiling(n/log(n)), where n is the sample size. If the sample size is greater than topN (pre-specified or calculated), all mediators will be included in the test (i.e. low-dimensional scenario).
scale	logical. Should the function scale the data? Default = TRUE.
FDRcut	Benjamini-Hochberg FDR cutoff applied to select significant mediators. Default = 0.05.
verbose	logical. Should the function be verbose? Default = FALSE.

Value

A data.frame containing mediation testing results of significant mediators (FDR <FDRcut).

Index: mediation name of selected significant mediator.

alpha_hat: coefficient estimates of exposure (X) –> mediators (M) (adjusted for covariates).

alpha_se: standard error for alpha.

beta_hat: coefficient estimates of mediators (M) –> outcome (Y) (adjusted for covariates and exposure).

beta_se: standard error for beta.

IDE: mediation (indirect) effect, i.e., alpha*beta.

rimp: relative importance of the mediator.

pmax: joint raw p-value of selected significant mediator (based on divide-aggregate composite-null test [DACT] method).

References

Bai X, Zheng Y, Hou L, Zheng C, Liu L, Zhang H. An Efficient Testing Procedure for High-dimensional Mediators with FDR Control. Statistics in Biosciences. 2024. DOI: 10.1007/s12561-024-09447-4.

Examples

```
## Not run:
# Note: In the following example, M1, M2, and M3 are true mediators.
data(himaDat)

# Y is continuous and normally distributed
# Example (continuous outcome):
head(himaDat$Example1$PhenoData)

eHIMA.fit <- eHIMA(X = himaDat$Example1$PhenoData$Treatment,
                   Y = himaDat$Example1$PhenoData$Outcome,
                   M = himaDat$Example1$Mediator,
                   COV = himaDat$Example1$PhenoData[, c("Sex", "Age")],
                   scale = FALSE, # Disabled only for simulation data
                   verbose = TRUE)
eHIMA.fit

## End(Not run)
```

`hima` *High-dimensional Mediation Analysis*

Description

`hima` is used to estimate and test high-dimensional mediation effects.

Usage

```
hima(
  X,
  Y,
  M,
  COV.XM = NULL,
  COV.MY = COV.XM,
  Y.family = c("gaussian", "binomial"),
  M.family = c("gaussian", "negbin"),
  penalty = c("MCP", "SCAD", "lasso"),
  topN = NULL,
  parallel = FALSE,
  ncore = 1,
  scale = TRUE,
  Bonfcut = 0.05,
  verbose = FALSE,
  ...
)
```

Arguments

X	a vector of exposure. Do not use data.frame or matrix.
Y	a vector of outcome. Can be either continuous or binary (0-1). Do not use data.frame or matrix.
M	a data.frame or matrix of high-dimensional mediators. Rows represent samples, columns represent variables.
COV.XM	a data.frame or matrix of covariates dataset for testing the association M ~ X. Covariates specified here will not participate penalization. Default = NULL. If the covariates contain mixed data types, please make sure all categorical variables are properly formatted as factor type.
COV.MY	a data.frame or matrix of covariates dataset for testing the association Y ~ M. Covariates specified here will not participate penalization. If not specified, the same set of covariates for M ~ X will be applied. Using different sets of covariates is allowed but this needs to be handled carefully.
Y.family	either 'gaussian' (default) or 'binomial', depending on the data type of outcome (Y). See ncvreg
M.family	either 'gaussian' (default) or 'negbin' (i.e., negative binomial), depending on the data type of mediator (M).
penalty	the penalty to be applied to the model. Either 'MCP' (the default), 'SCAD', or 'lasso'.
topN	an integer specifying the number of top markers from sure independent screening. Default = NULL. If NULL, topN will be either ceiling(n/log(n)) if Y.family = 'gaussian', or ceiling(n/(2*log(n))) if Y.family = 'binomial', where n is the sample size. If the sample size is greater than topN (pre-specified or calculated), all mediators will be included in the test (i.e. low-dimensional scenario).
parallel	logical. Enable parallel computing feature? Default = FALSE.
ncore	number of cores to run parallel computing Valid when parallel == TRUE. By default max number of cores available in the machine will be utilized.
scale	logical. Should the function scale the data? Default = TRUE.
Bonfcut	Bonferroni-corrected p value cutoff applied to select significant mediators. Default = 0.05.
verbose	logical. Should the function be verbose? Default = FALSE.
...	other arguments passed to ncvreg.

Value

A data.frame containing mediation testing results of selected mediators.

Index: mediation name of selected significant mediator.

alpha_hat: coefficient estimates of exposure (X) –> mediators (M) (adjusted for covariates).

beta_hat: coefficient estimates of mediators (M) –> outcome (Y) (adjusted for covariates and exposure).

IDE: mediation (indirect) effect, i.e., alpha*beta.

rimp: relative importance of the mediator.

pmax: joint raw p-value of selected significant mediator (based on Bonferroni method).

References

Zhang H, Zheng Y, Zhang Z, Gao T, Joyce B, Yoon G, Zhang W, Schwartz J, Just A, Colicino E, Vokonas P, Zhao L, Lv J, Baccarelli A, Hou L, Liu L. Estimating and Testing High-dimensional Mediation Effects in Epigenetic Studies. Bioinformatics. 2016. DOI: 10.1093/bioinformatics/btw351. PMID: 27357171; PMCID: PMC5048064

Examples

```
## Not run:
# Note: In the following examples, M1, M2, and M3 are true mediators.
data(himaDat)

# When Y is continuous and normally distributed
# Example 1 (continuous outcome):
head(himaDat$Example1$PhenoData)

hima.fit <- hima(X = himaDat$Example1$PhenoData$Treatment,
                 Y = himaDat$Example1$PhenoData$Outcome,
                 M = himaDat$Example1$Mediator,
                 COV.XM = himaDat$Example1$PhenoData[, c("Sex", "Age")],
                 Y.family = 'gaussian',
                 scale = FALSE, # Disabled only for simulation data
                 verbose = TRUE)
hima.fit

# When Y is binary (should specify Y.family)
# Example 2 (binary outcome):
head(himaDat$Example2$PhenoData)

hima.logistic.fit <- hima(X = himaDat$Example2$PhenoData$Treatment,
                          Y = himaDat$Example2$PhenoData$Disease,
                          M = himaDat$Example2$Mediator,
                          COV.XM = himaDat$Example2$PhenoData[, c("Sex", "Age")],
                          Y.family = 'binomial',
                          scale = FALSE, # Disabled only for simulation data
                          verbose = TRUE)
hima.logistic.fit

## End(Not run)
```

hima2 *High-dimensional Mediation Analysis*

Description

hima2 is a wrapper function designed to perform various HIMA methods for estimating and testing high-dimensional mediation effects. hima2 can automatically select the appropriate HIMA method based on the outcome and mediator data type specified by the user.

Usage

```
hima2(
  formula,
  data.pheno,
  data.M,
  outcome.family = c("gaussian", "binomial", "survival", "quantile"),
  mediator.family = c("gaussian", "negbin", "compositional"),
  penalty = c("DBlasso", "MCP", "SCAD", "lasso"),
  efficient = FALSE,
  scale = TRUE,
  Sigcut = 0.05,
  verbose = FALSE,
  ...
)
```

Arguments

formula	an object of class formula: a symbolic description of the overall effect model, i.e., outcome ~ exposure + covariates, to be fitted. Make sure the "exposure" is the variable of interest, which must be listed as the first variable in the right hand side of the formula.
data.pheno	a data frame containing exposure and covariates that are listed in the right hand side of the formula. The variable names must match those listed in formula. By default hima2 will scale data.pheno.
data.M	a data.frame or matrix of high-dimensional mediators. Rows represent samples, columns represent variables. By default hima2 will scale data.M.
outcome.family	either 'gaussian' (default, for normally distributed continuous outcome), 'binomial' (for binary outcome), 'survival' (for time-to-event outcome), or 'quantile' (for quantile mediation analysis)
mediator.family	either 'gaussian' (default, for continuous mediators), 'negbin' (i.e., negative binomial, for RNA-seq data as mediators), or 'compositional' (for microbiome data as mediators), depending on the data type of high-dimensional mediators (data.M).

penalty	the penalty to be applied to the model. Either 'DBlasso' (De-biased LASSO, default), 'MCP', 'SCAD', or 'lasso'. Please note, survival HIMA and microbiome HIMA can be only performed with 'DBlasso'; Quantile HIMA cannot be performed with 'DBlasso'. Not applicable for efficient HIMA (when efficient = TRUE), as it will always apply 'MCP'.
efficient	use efficient HIMA (eHIMA). Only applicable for linear HIMA with continuous outcome (i.e., outcome.family = "gaussian" and mediator.family = "gaussian"). Default = FALSE.
scale	logical. Should the function scale the data (exposure, mediators, and covariates)? Default = TRUE. Note: for simulation study, scale can be turned off to avoid estimate compression.
Sigcut	cutoff applied to select significant mediators. Default = 0.05.
verbose	logical. Should the function be verbose and show the progression? Default = FALSE.
...	reserved passing parameter.

Value

A data.frame containing mediation testing results of selected mediators.

ID: Mediator ID/name.

alpha: Coefficient estimates of exposure (X) –> mediators (M) (adjusted for covariates).

beta: Coefficient estimates of mediators (M) –> outcome (Y) (adjusted for covariates and exposure).

alpha*beta: Mediation (indirect) effect.

Relative Importance: Relative importance of the mediator. It is the proportion of the mediation effect for each mediator out of the sum of the mediation effect (absolute value) across all significant mediators selected.

p-value: Joint raw p-value of significant mediators selected based on corresponding approach.

tau: Quantile level of the outcome (applicable only to the quantile mediation model).

References

1. Zhang H, Zheng Y, Zhang Z, Gao T, Joyce B, Yoon G, Zhang W, Schwartz J, Just A, Colicino E, Vokonas P, Zhao L, Lv J, Baccarelli A, Hou L, Liu L. Estimating and Testing High-dimensional Mediation Effects in Epigenetic Studies. Bioinformatics. 2016. DOI: 10.1093/bioinformatics/btw351. PMID: 27357171; PMCID: PMC5048064

2. Zhang H, Zheng Y, Hou L, Zheng C, Liu L. Mediation Analysis for Survival Data with High-Dimensional Mediators. Bioinformatics. 2021. DOI: 10.1093/bioinformatics/btab564. PMID: 34343267; PMCID: PMC8570823

3. Zhang H, Chen J, Feng Y, Wang C, Li H, Liu L. Mediation Effect Selection in High-dimensional and Compositional Microbiome data. Stat Med. 2021. DOI: 10.1002/sim.8808. PMID: 33205470; PMCID: PMC7855955

4. Zhang H, Chen J, Li Z, Liu L. Testing for Mediation Effect with Application to Human Microbiome Data. Stat Biosci. 2021. DOI: 10.1007/s12561-019-09253-3. PMID: 34093887; PMCID: PMC8177450

5. Perera C, Zhang H, Zheng Y, Hou L, Qu A, Zheng C, Xie K, Liu L. HIMA2: High-dimensional Mediation Analysis and Its Application in Epigenome-wide DNA Methylation Data. BMC Bioinformatics. 2022. DOI: 10.1186/s12859-022-04748-1. PMID: 35879655; PMCID: PMC9310002

6. Zhang H, Hong X, Zheng Y, Hou L, Zheng C, Wang X, Liu L. High-Dimensional Quantile Mediation Analysis with Application to a Birth Cohort Study of Mother–Newborn Pairs. Bioinformatics. 2024. DOI: 10.1093/bioinformatics/btae055. PMID: 38290773; PMCID: PMC10873903

7. Bai X, Zheng Y, Hou L, Zheng C, Liu L, Zhang H. An Efficient Testing Procedure for High-dimensional Mediators with FDR Control. Statistics in Biosciences. 2024. DOI: 10.1007/s12561-024-09447-4.

Examples

```
## Not run:
# Note: In the following examples, M1, M2, and M3 are true mediators.
data(himaDat)

# Example 1 (continous outcome - linear hima):
head(himaDat$Example1$PhenoData)

e1 <- hima2(Outcome ~ Treatment + Sex + Age,
      data.pheno = himaDat$Example1$PhenoData,
      data.M = himaDat$Example1$Mediator,
      outcome.family = "gaussian",
      mediator.family = "gaussian",
      penalty = "MCP", # Can be "DBlasso" for dblassoHIMA
      scale = FALSE) # Disabled only for simulation data
e1
attributes(e1)$variable.labels

# Efficient HIMA (only applicable to mediators and outcomes that are
# both continuous and normally distributed.)
e1e <- hima2(Outcome ~ Treatment + Sex + Age,
             data.pheno = himaDat$Example1$PhenoData,
             data.M = himaDat$Example1$Mediator,
             outcome.family = "gaussian",
             mediator.family = "gaussian",
             efficient = TRUE,
             scale = FALSE) # Disabled only for simulation data
e1e
attributes(e1e)$variable.labels

# Example 2 (binary outcome - logistic hima):
head(himaDat$Example2$PhenoData)

e2 <- hima2(Disease ~ Treatment + Sex + Age,
      data.pheno = himaDat$Example2$PhenoData,
      data.M = himaDat$Example2$Mediator,
      outcome.family = "binomial",
      mediator.family = "gaussian",
      penalty = "MCP",
      scale = FALSE) # Disabled only for simulation data
```

```
e2
attributes(e2)$variable.labels

# Example 3 (time-to-event outcome - survival hima):
head(himaDat$Example3$PhenoData)

e3 <- hima2(Surv(Status, Time) ~ Treatment + Sex + Age,
      data.pheno = himaDat$Example3$PhenoData,
      data.M = himaDat$Example3$Mediator,
      outcome.family = "survival",
      mediator.family = "gaussian",
      penalty = "DBlasso",
      scale = FALSE) # Disabled only for simulation data
e3
attributes(e3)$variable.labels

# Example 4 (compositional data as mediator, e.g., microbiome):
head(himaDat$Example4$PhenoData)

e4 <- hima2(Outcome ~ Treatment + Sex + Age,
      data.pheno = himaDat$Example4$PhenoData,
      data.M = himaDat$Example4$Mediator,
      outcome.family = "gaussian",
      mediator.family = "compositional",
      penalty = "DBlasso") # Scaling is always enabled for microHIMA
e4
attributes(e4)$variable.labels

#' # Example 5 (quantile mediation anlaysis - quantile hima):
head(himaDat$Example5$PhenoData)

# Note that the function will prompt input for quantile level.
e5 <- hima2(Outcome ~ Treatment + Sex + Age,
      data.pheno = himaDat$Example5$PhenoData,
      data.M = himaDat$Example5$Mediator,
      outcome.family = "quantile",
      mediator.family = "gaussian",
      penalty = "MCP", # Quantile HIMA does not support DBlasso
      scale = FALSE, # Disabled only for simulation data
      tau = c(0.3, 0.5, 0.7)) # Specify multiple quantile level
e5
attributes(e5)$variable.labels

## End(Not run)
```

---

himaDat *Example datasets for HIMA demo*

Description

A `list` containing datasets for various scenarios of HIMA. Each dataset contains a phenotype data frame and a high-dimension mediator data matrix. The datasets are simulated using parameters generated from real datasets. The code used to generate the data can be found in "/inst/script" folder of the source package.

Usage

```
himaDat
```

Format

An object of class `list` of length 5.

Value

A list of example datasets for HIMA demo and testing.

–» **Example dataset 1 for HIMA: Continuous outcome** «–

Treatment: treated (value = 1) or not treated (value = 0)

Outcome: outcome of the treatment- a normally distributed continuous variable

Sex: female (value = 1) or male (value = 0)

Age: Age of the participant

–» **Example dataset 2 for HIMA: Binary outcome** «–

Treatment: treated (value = 1) or not treated (value = 0)

Disease: diseased (value = 1) or healthy (value = 0)

Sex: female (value = 1) or male (value = 0)

Age: Age of the participant

–» **Example dataset 3 for HIMA: Survival data outcome** «–

Treatment: treated (value = 1) or not treated (value = 0)

Status: Status indicator: dead (value = 1) or alive (value = 0)

Time: time to event

Sex: female (value = 1) or male (value = 0)

Age: Age of the participant

–» **Example dataset 4 for HIMA: Compositional mediator (e.g., microbiome)** «–

Treatment: treated (value = 1) or not treated (value = 0)

Outcome: outcome of the treatment- a normally distributed continuous variable

Sex: female (value = 1) or male (value = 0)

Age: Age of the participant

–» **Example dataset 5 for HIMA: High-dimensional quantile mediation analysis** «–

Treatment: treated (value = 1) or not treated (value = 0)

Outcome: outcome of the treatment- abnormally distributed continuous variable

Sex: female (value = 1) or male (value = 0)

Age: Age of the participant

`microHIMA` *High-dimensional mediation analysis for compositional microbiome data*

Description

`microHIMA` is used to estimate and test high-dimensional mediation effects for compositional microbiome data.

Usage

```
microHIMA(X, Y, OTU, COV = NULL, FDRcut = 0.05, verbose = FALSE)
```

Arguments

`X`	a vector of exposure.
`Y`	a vector of continuous outcome. Binary outcome is not allowed.
`OTU`	a `data.frame` or `matrix` of high-dimensional Operational Taxonomic Unit (OTU) data (mediators). Rows represent samples, columns represent variables.
`COV`	a `data.frame` or `matrix` of adjusting covariates. Rows represent samples, columns represent microbiome variables. Can be `NULL`.
`FDRcut`	Hommel FDR cutoff applied to select significant mediators. Default = `0.05`.
`verbose`	logical. Should the function be verbose? Default = `FALSE`.

Value

A data.frame containing mediation testing results of significant mediators (FDR <`FDRcut`).

Index: mediation name of selected significant mediator.

alpha_hat: coefficient estimates of exposure (X) –> mediators (M) (adjusted for covariates).

alpha_se: standard error for alpha.

beta_hat: coefficient estimates of mediators (M) –> outcome (Y) (adjusted for covariates and exposure).

beta_se: standard error for beta.

IDE: mediation (indirect) effect, i.e., alpha*beta.

rimp: relative importance of the mediator.

pmax: joint raw p-value of selected significant mediator (based on Hommel FDR method).

References

1. Zhang H, Chen J, Feng Y, Wang C, Li H, Liu L. Mediation effect selection in high-dimensional and compositional microbiome data. Stat Med. 2021. DOI: 10.1002/sim.8808. PMID: 33205470; PMCID: PMC7855955

2. Zhang H, Chen J, Li Z, Liu L. Testing for mediation effect with application to human microbiome data. Stat Biosci. 2021. DOI: 10.1007/s12561-019-09253-3. PMID: 34093887; PMCID: PMC8177450

Examples

```
## Not run:
# Note: In the following example, M1, M2, and M3 are true mediators.
data(himaDat)

head(himaDat$Example4$PhenoData)

microHIMA.fit <- microHIMA(X = himaDat$Example4$PhenoData$Treatment,
                           Y = himaDat$Example4$PhenoData$Outcome,
                           OTU = himaDat$Example4$Mediator,
                           COV = himaDat$Example4$PhenoData[, c("Sex", "Age")])
microHIMA.fit

## End(Not run)
```

---

qHIMA *High-dimensional quantile mediation analysis*

Description

qHIMA is used to estimate and test high-dimensional quantile mediation effects.

Usage

```
qHIMA(
  X,
  M,
  Y,
  COV = NULL,
  penalty = c("MCP", "SCAD", "lasso"),
  topN = NULL,
  tau = 0.5,
  scale = TRUE,
  Bonfcut = 0.05,
  verbose = FALSE,
  ...
)
```

Arguments

X	a vector of exposure.
M	a data.frame or matrix of high-dimensional mediators. Rows represent samples, columns represent mediator variables.
Y	a vector of continuous outcome. Do not use data.frame or matrix.
COV	a matrix of adjusting covariates. Rows represent samples, columns represent variables. Can be NULL.
penalty	the penalty to be applied to the model (a parameter passed to function conquer.cv.reg in package conquer. Either 'MCP' (the default), 'SCAD', or 'lasso'.
topN	an integer specifying the number of top markers from sure independent screening. Default = NULL. If NULL, topN will be 2*ceiling(n/log(n)), where n is the sample size. If the sample size is greater than topN (pre-specified or calculated), all mediators will be included in the test (i.e. low-dimensional scenario).
tau	quantile level of outcome. Default = 0.5. A vector of tau is accepted.
scale	logical. Should the function scale the data? Default = TRUE.
Bonfcut	Bonferroni-corrected p value cutoff applied to select significant mediators. Default = 0.05.
verbose	logical. Should the function be verbose? Default = FALSE.
...	reserved passing parameter.

Value

A data.frame containing mediation testing results of selected mediators (Bonferroni-adjusted p value <Bonfcut).

Index: mediation name of selected significant mediator.

alpha_hat: coefficient estimates of exposure (X) –> mediators (M) (adjusted for covariates).

alpha_se: standard error for alpha.

beta_hat: coefficient estimates of mediators (M) –> outcome (Y) (adjusted for covariates and exposure).

beta_se: standard error for beta.

IDE: mediation (indirect) effect, i.e., alpha*beta.

rimp: relative importance of the mediator.

pmax: joint raw p-value of selected significant mediator (based on Bonferroni method).

References

Zhang H, Hong X, Zheng Y, Hou L, Zheng C, Wang X, Liu L. High-Dimensional Quantile Mediation Analysis with Application to a Birth Cohort Study of Mother–Newborn Pairs. Bioinformatics. 2024. DOI: 10.1093/bioinformatics/btae055. PMID: 38290773; PMCID: PMC10873903

Examples

```
## Not run:
# Note: In the following example, M1, M2, and M3 are true mediators.
data(himaDat)

head(himaDat$Example5$PhenoData)

qHIMA.fit <- qHIMA(X = himaDat$Example5$PhenoData$Treatment,
                M = himaDat$Example5$Mediator,
                Y = himaDat$Example5$PhenoData$Outcome,
                COV = himaDat$Example5$PhenoData[, c("Sex", "Age")],
                Bonfcut = 0.05,
                tau = c(0.3, 0.5, 0.7),
                scale = FALSE, # Disabled only for simulation data
                verbose = TRUE)
qHIMA.fit

## End(Not run)
```

---

survHIMA *High-dimensional mediation analysis for survival data*

Description

survHIMA is used to estimate and test high-dimensional mediation effects for survival data.

Usage

```
survHIMA(
  X,
  M,
  COV = NULL,
  OT,
  status,
  FDRcut = 0.05,
  scale = TRUE,
  verbose = FALSE
)
```

Arguments

X	a vector of exposure.
M	a data.frame or matrix of high-dimensional mediators. Rows represent samples, columns represent mediator variables.
COV	a matrix of adjusting covariates. Rows represent samples, columns represent variables. Can be NULL.

OT	a vector of observed failure times.
status	a vector of censoring indicator (status = 1: uncensored; status = 0: censored)
FDRcut	HDMT pointwise FDR cutoff applied to select significant mediators. Default = 0.05.
scale	logical. Should the function scale the data? Default = TRUE.
verbose	logical. Should the function be verbose? Default = FALSE.

Value

A data.frame containing mediation testing results of significant mediators (FDR <FDRcut).

Index: mediation name of selected significant mediator.

alpha_hat: coefficient estimates of exposure (X) –> mediators (M) (adjusted for covariates).

alpha_se: standard error for alpha.

beta_hat: coefficient estimates of mediators (M) –> outcome (Y) (adjusted for covariates and exposure).

beta_se: standard error for beta.

IDE: mediation (indirect) effect, i.e., alpha*beta.

rimp: relative importance of the mediator.

pmax: joint raw p-value of selected significant mediator (based on HDMT pointwise FDR method).

References

Zhang H, Zheng Y, Hou L, Zheng C, Liu L. Mediation Analysis for Survival Data with High-Dimensional Mediators. Bioinformatics. 2021. DOI: 10.1093/bioinformatics/btab564. PMID: 34343267; PMCID: PMC8570823

Examples

```
## Not run:
# Note: In the following example, M1, M2, and M3 are true mediators.
data(himaDat)

head(himaDat$Example3$PhenoData)

survHIMA.fit <- survHIMA(X = himaDat$Example3$PhenoData$Treatment,
                M = himaDat$Example3$Mediator,
                COV = himaDat$Example3$PhenoData[, c("Sex", "Age")],
                OT = himaDat$Example3$PhenoData$Time,
                status = himaDat$Example3$PhenoData$Status,
                FDRcut = 0.05,
                scale = FALSE, # Disabled only for simulation data
                verbose = TRUE)
survHIMA.fit

## End(Not run)
```

Index