Package 'HIMA'

May 7, 2026

Type Package

Title High-Dimensional Mediation Analysis

Version 2.3.3

Date 2025-11-17

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, Hou L, Liu L, HIMA: An R Package for High-Dimensional Mediation Analysis. Journal of Data Science. (2025). <doi:10.6339/25-JDS1192>.

License GPL-3

Depends R (>= 3.5), ncvreg, glmnet

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

Collate utils.R hima_classic.R hima_dblasso.R hima_survival.R hima_survival_long.R hima_microbiome.R hima_quantile.R hima_efficient.R hima.R hima_data.R onAttach.R HIMA-package.R

VignetteBuilder knitr

Suggests knitr, rmarkdown, testthat

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] (ORCID: <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 2025-11-17 21:51:56 UTC

Contents

HIMA-package . . . 2
BinaryOutcome . . . 3
ContinuousOutcome . . . 4
hima . . . 5
hima_classic . . . 10
hima_dblasso . . . 12
hima_efficient . . . 14
hima_microbiome . . . 16
hima_quantile . . . 18
hima_survival . . . 20
hima_survival_long . . . 22
MicrobiomeData . . . 24
QuantileData . . . 25
SurvivalData . . . 25
SurvivalLongData . . . 26

Index **28**

`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:	2.3.3
Date:	2025-11-17
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, Hou L, Liu L, HIMA: An R Package for High-Dimensional Mediation Analysis. Journal of Data Science. 2025. DOI: 10.6339/25-JDS1192

2. 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

3. 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

4. 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

5. 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

6. 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

7. 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

8. 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.

9. Liu L, Zhang H, Zheng Y, Gao T, Zheng C, Zhang K, Hou L, Liu L. High-dimensional mediation analysis for longitudinal mediators and survival outcomes. Briefings in Bioinformatics. 2025. DOI: 10.1093/bib/bbaf206. PMID: 40350699 PMCID: PMC12066418

`BinaryOutcome` *Binary Outcome Dataset for HIMA Demo*

Description

A dataset containing phenotype data and high-dimensional mediators for binary outcome analysis. The dataset was simulated using parameters generated from real data.

Usage

```
BinaryOutcome
```

Format

A list with the following components:

PhenoData A data frame containing:

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

Disease binary outcome: diseased (value = 1) or healthy (value = 0).

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

Age age of the participant.

Mediator A matrix of high-dimensional mediators (rows: samples, columns: variables).

Examples

```
data(BinaryOutcome)
head(BinaryOutcome$PhenoData)
```

ContinuousOutcome *Continuous Outcome Dataset for HIMA Demo*

Description

A dataset containing phenotype data and high-dimensional mediators for continuous outcome analysis. The dataset was simulated using parameters generated from real data.

Usage

```
ContinuousOutcome
```

Format

A list with the following components:

PhenoData A data frame containing:

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

Outcome a normally distributed continuous outcome variable.

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

Age age of the participant.

Mediator A matrix of high-dimensional mediators (rows: samples, columns: variables).

Examples

```
data(ContinuousOutcome)
head(ContinuousOutcome$PhenoData)
```

hima *High-dimensional Mediation Analysis*

Description

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

Usage

```
hima(
  formula,
  data.pheno,
  data.M,
  mediator.type = c("gaussian", "negbin", "compositional"),
  penalty = c("DBlasso", "MCP", "SCAD", "lasso"),
  quantile = FALSE,
  efficient = FALSE,
  longitudinal = FALSE,
  id.var = NULL,
  scale = TRUE,
  sigcut = 0.05,
  contrast = NULL,
  subset = NULL,
  verbose = FALSE,
  parallel = FALSE,
  ncore = 1,
  ...
)
```

Arguments

formula	an object of class formula representing the overall effect model to be fitted, specified as outcome ~ exposure + covariates. The "exposure" variable (the variable of interest) must be listed first on the right-hand side of the formula. For survival outcomes specified using Surv(), the exposure should be the first variable after the ~.
data.pheno	a data frame containing the exposure, outcome, and covariates specified in the formula. Variable names in data.pheno must match those in the formula. When scale = TRUE, the exposure and covariates will be scaled (the outcome retains its original scale).
data.M	a data.frame or matrix of high-dimensional mediators, with rows representing samples and columns representing mediator variables. When scale = TRUE, data.M will be scaled.

mediator.type	a character string indicating the data type of the high-dimensional mediators (data.M). Options are: 'gaussian' (default): for continuous mediators. 'negbin': for count data mediators modeled using the negative binomial distribution (e.g., RNA-seq data). 'compositional': for compositional data mediators (e.g., microbiome data).
penalty	a character string specifying the penalty method to apply in the model. Options are: 'DBlasso': De-biased LASSO (default). 'MCP': Minimax Concave Penalty. 'SCAD': Smoothly Clipped Absolute Deviation. 'lasso': Least Absolute Shrinkage and Selection Operator. Note: Survival HIMA and microbiome HIMA can only be performed with 'DBlasso'. Quantile HIMA and efficient HIMA cannot use 'DBlasso'; they always apply 'MCP'.
quantile	logical. Indicates whether to use quantile HIMA (hima_quantile). Default is FALSE. Applicable only for classic HIMA with a continuous outcome and mediator.type = 'gaussian'. If TRUE, specify the desired quantile(s) using the tau parameter; otherwise, the default tau = 0.5 (i.e., median) is used.
efficient	logical. Indicates whether to use efficient HIMA (hima_efficient). Default is FALSE. Applicable only for classic HIMA with a continuous outcome and mediator.type = 'gaussian'.
longitudinal	logical. Indicates whether to run the longitudinal survival mediation model hima_survival_long (requires a Surv(tstart, tstop, status) outcome). Default = FALSE.
id.var	Character string specifying the column name in data.pheno that stores subject identifiers. Required when longitudinal = TRUE.
scale	logical. Determines whether the function scales the data (exposure, mediators, and covariates). Default is TRUE. Note: For simulation studies, set scale = FALSE to avoid estimate compression (i.e., shrinkage of estimates toward zero due to scaling).
sigcut	numeric. The significance cutoff for selecting mediators. Default is 0.05.
contrast	a named list of contrasts to be applied to factor variables in the covariates (cannot be the variable of interest).
subset	an optional vector specifying a subset of observations to use in the analysis.
verbose	logical. Determines whether the function displays progress messages. Default is FALSE.
parallel	logical. Enable parallel computing feature? Default = FALSE.
ncore	number of cores to run parallel computing Valid when parallel = TRUE.
...	reserved passing parameter (or for future use).

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: The estimated indirect (mediation) effect of exposure on outcome through each mediator.

rimp: Relative importance- the proportion of each mediator's mediation effect relative to the sum of the absolute mediation effects of all significant mediators.

p-value: The joint p-value assessing the significance of each mediator's indirect effect, calculated based on the corresponding statistical approach.

tau: The quantile level of the outcome (applicable only when using the quantile mediation model).

References

1. Zhang H, Zheng Y, Hou L, Liu L, HIMA: An R Package for High-Dimensional Mediation Analysis. Journal of Data Science. 2025. DOI: 10.6339/25-JDS1192

2. 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

3. 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

4. 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

5. 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

6. 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

7. 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

8. 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.

9. Liu L, Zhang H, Zheng Y, Gao T, Zheng C, Zhang K, Hou L, Liu L. High-dimensional mediation analysis for longitudinal mediators and survival outcomes. Briefings in Bioinformatics. 2025. DOI: 10.1093/bib/bbaf206. PMID: 40350699 PMCID: PMC12066418

Examples

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

# Example 1 (continuous outcome - linear HIMA):
data(ContinuousOutcome)
pheno_data <- ContinuousOutcome$PhenoData
```

```
mediator_data <- ContinuousOutcome$Mediator

e1 <- hima(Outcome ~ Treatment + Sex + Age,
  data.pheno = pheno_data,
  data.M = mediator_data,
  mediator.type = "gaussian",
  penalty = "MCP", # Can be "DBlasso" for hima_dblasso
  scale = FALSE, # Disabled only for simulation data
  verbose = TRUE
)
summary(e1)

# Efficient HIMA (only applicable to mediators and outcomes that are
# both continuous and normally distributed.)
e1e <- hima(Outcome ~ Treatment + Sex + Age,
  data.pheno = pheno_data,
  data.M = mediator_data,
  mediator.type = "gaussian",
  efficient = TRUE,
  penalty = "MCP", # Efficient HIMA does not support DBlasso
  scale = FALSE, # Disabled only for simulation data
  verbose = TRUE
)
summary(e1e)

# Example 2 (binary outcome - logistic HIMA):
data(BinaryOutcome)
pheno_data <- BinaryOutcome$PhenoData
mediator_data <- BinaryOutcome$Mediator

e2 <- hima(Disease ~ Treatment + Sex + Age,
  data.pheno = pheno_data,
  data.M = mediator_data,
  mediator.type = "gaussian",
  penalty = "MCP",
  scale = FALSE, # Disabled only for simulation data
  verbose = TRUE
)
summary(e2)

# Example 3 (time-to-event outcome - survival HIMA):
data(SurvivalData)
pheno_data <- SurvivalData$PhenoData
mediator_data <- SurvivalData$Mediator

e3 <- hima(Surv(Time, Status) ~ Treatment + Sex + Age,
  data.pheno = pheno_data,
  data.M = mediator_data,
  mediator.type = "gaussian",
  penalty = "DBlasso",
  scale = FALSE, # Disabled only for simulation data
  verbose = TRUE
) # Parallel computing feature is recommended
```

```
summary(e3)

# Longitudinal mediator + survival HIMA:
data(SurvivalLongData)
pheno_data <- SurvivalLongData$PhenoData
mediator_data <- SurvivalLongData$Mediator

e3long <- hima(Surv(Tstart, Tstop, Status) ~ Treatment + Sex + Age,
  data.pheno = pheno_data,
  data.M = mediator_data,
  mediator.type = "gaussian",
  penalty = "lasso",
  longitudinal = TRUE,
  id.var = "ID",
  scale = FALSE, # Disabled only for simulation data
  verbose = TRUE
) # Parallel computing feature is recommended
summary(e3long)

# Example 4 (compositional data as mediator, e.g., microbiome):
data(MicrobiomeData)
pheno_data <- MicrobiomeData$PhenoData
mediator_data <- MicrobiomeData$Mediator

e4 <- hima(Outcome ~ Treatment + Sex + Age,
  data.pheno = pheno_data,
  data.M = mediator_data,
  mediator.type = "compositional",
  penalty = "DBlasso",
  verbose = TRUE
) # Scaling is always enabled internally for hima_microbiome
summary(e4)

#' # Example 5 (quantile mediation analysis - quantile HIMA):
data(QuantileData)
pheno_data <- QuantileData$PhenoData
mediator_data <- QuantileData$Mediator

# Note that the function will prompt input for quantile level.
e5 <- hima(Outcome ~ Treatment + Sex + Age,
  data.pheno = pheno_data,
  data.M = mediator_data,
  mediator.type = "gaussian",
  quantile = TRUE,
  penalty = "MCP", # Quantile HIMA does not support DBlasso
  scale = FALSE, # Disabled only for simulation data
  tau = c(0.3, 0.5, 0.7),
  verbose = TRUE
) # Specify multiple quantile level
summary(e5)

## End(Not run)
```

`hima_classic` *Classic high-dimensional mediation analysis*

Description

`hima_classic` is used to estimate and test classic high-dimensional mediation effects (linear & logistic regression).

Usage

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

Arguments

`X`	a vector of exposure. Do not use `data.frame` or `matrix`.
`M`	a `data.frame` or `matrix` of high-dimensional mediators. Rows represent samples, columns represent variables.
`Y`	a vector of outcome. Can be either continuous or binary (0-1). Do not use `data.frame` or `matrix`.
`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 (i.e., `COV.XM`. Using different sets of covariates is allowed but this needs to be handled carefully.
`Y.type`	data type of outcome (`Y`). Either `'continuous'` (default) or `'binary'`.
`M.type`	data type of mediator (`M`). Either `'gaussian'` (default) or `'negbin'` (i.e., negative binomial).

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)) for continuous outcome, or ceiling(n/(2*log(n))) for binary outcome, 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.
Bonfcut	Bonferroni-corrected p value cutoff applied to select significant mediators. Default = 0.05.
verbose	logical. Should the function be verbose? Default = FALSE.
parallel	logical. Enable parallel computing feature? Default = FALSE.
ncore	number of cores to run parallel computing Valid when parallel = TRUE.
...	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.

# When Y is continuous and normally distributed
# Example 1 (continuous outcome):
data(ContinuousOutcome)
pheno_data <- ContinuousOutcome$PhenoData
mediator_data <- ContinuousOutcome$Mediator

hima.fit <- hima_classic(
  X = pheno_data$Treatment,
```

```
  Y = pheno_data$Outcome,
  M = mediator_data,
  COV.XM = pheno_data[, c("Sex", "Age")],
  Y.type = "continuous",
  scale = FALSE, # Disabled only for simulation data
  verbose = TRUE
)
hima.fit

# When Y is binary
# Example 2 (binary outcome):
data(BinaryOutcome)
pheno_data <- BinaryOutcome$PhenoData
mediator_data <- BinaryOutcome$Mediator

hima.logistic.fit <- hima_classic(
  X = pheno_data$Treatment,
  Y = pheno_data$Disease,
  M = mediator_data,
  COV.XM = pheno_data[, c("Sex", "Age")],
  Y.type = "binary",
  scale = FALSE, # Disabled only for simulation data
  verbose = TRUE
)
hima.logistic.fit

## End(Not run)
```

`hima_dblasso` *High-dimensional mediation analysis with de-biased lasso penalty*

Description

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

Usage

```
hima_dblasso(
  X,
  M,
  Y,
  COV = NULL,
  topN = NULL,
  scale = TRUE,
  FDRcut = 0.05,
  verbose = FALSE,
  parallel = FALSE,
```

```
  ncore = 1
)
```

Arguments

X	a vector of exposure. Do not use data.frame or matrix.
M	a data.frame or matrix of high-dimensional mediators. Rows represent samples, columns represent variables.
Y	a vector of outcome. Can be either continuous or binary (0-1). Do not use data.frame or matrix.
COV	a data.frame or matrix of covariates dataset for testing the association M ~ X and Y ~ M.
topN	an integer specifying the number of top markers from sure independent screening. Default = NULL. If NULL, topN will be 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	HDMT pointwise FDR cutoff applied to select significant mediators. Default = 0.05.
verbose	logical. Should the function be verbose? Default = FALSE.
parallel	logical. Enable parallel computing feature? Default = FALSE.
ncore	number of cores to run parallel computing Valid when parallel = TRUE.

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.

# Y is continuous and normally distributed
# Example:
data(ContinuousOutcome)
pheno_data <- ContinuousOutcome$PhenoData
mediator_data <- ContinuousOutcome$Mediator

hima_dblasso.fit <- hima_dblasso(
  X = pheno_data$Treatment,
  Y = pheno_data$Outcome,
  M = mediator_data,
  COV = pheno_data[, c("Sex", "Age")],
  scale = FALSE, # Disabled only for simulation data
  FDRcut = 0.05,
  verbose = TRUE
)
hima_dblasso.fit

## End(Not run)
```

`hima_efficient` *Efficient high-dimensional mediation analysis*

Description

`hima_efficient` 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

```
hima_efficient(
  X,
  M,
  Y,
  COV = NULL,
  topN = NULL,
  scale = TRUE,
  FDRcut = 0.05,
  verbose = FALSE,
  parallel = FALSE,
  ncore = 1
)
```

Arguments

X	a vector of exposure. Do not use data.frame or matrix.
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.
parallel	logical. Enable parallel computing feature? Default = FALSE.
ncore	number of cores to run parallel computing Valid when parallel = TRUE.

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.

# Y is continuous and normally distributed
# Example (continuous outcome):
data(ContinuousOutcome)
pheno_data <- ContinuousOutcome$PhenoData
mediator_data <- ContinuousOutcome$Mediator

hima_efficient.fit <- hima_efficient(
  X = pheno_data$Treatment,
  Y = pheno_data$Outcome,
  M = mediator_data,
  COV = pheno_data[, c("Sex", "Age")],
  scale = FALSE, # Disabled only for simulation data
  FDRcut = 0.05,
  verbose = TRUE
)
hima_efficient.fit

## End(Not run)
```

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

Description

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

Usage

```
hima_microbiome(
  X,
  OTU,
  Y,
  COV = NULL,
  FDRcut = 0.05,
  verbose = FALSE,
  parallel = FALSE,
  ncore = 1
)
```

Arguments

X	a vector of exposure. Do not use data.frame or matrix.
OTU	a data.frame or matrix of high-dimensional Operational Taxonomic Unit (OTU) data (mediators). Rows represent samples, columns represent variables.
Y	a vector of continuous outcome. Binary outcome is not allowed. Do not use data.frame or matrix.
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.
parallel	logical. Enable parallel computing feature? Default = FALSE.
ncore	number of cores to run parallel computing Valid when parallel = TRUE.

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(MicrobiomeData)
pheno_data <- MicrobiomeData$PhenoData
mediator_data <- MicrobiomeData$Mediator
```

```
hima_microbiome.fit <- hima_microbiome(
  X = pheno_data$Treatment,
  Y = pheno_data$Outcome,
  OTU = mediator_data,
  COV = pheno_data[, c("Sex", "Age")],
  FDRcut = 0.05,
  verbose = TRUE
)
hima_microbiome.fit

## End(Not run)
```

`hima_quantile` *High-dimensional quantile mediation analysis*

Description

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

Usage

```
hima_quantile(
  X,
  M,
  Y,
  COV = NULL,
  penalty = c("MCP", "SCAD", "lasso"),
  topN = NULL,
  tau = 0.5,
  scale = TRUE,
  Bonfcut = 0.05,
  verbose = FALSE,
  parallel = FALSE,
  ncore = 1,
  ...
)
```

Arguments

`X`	a vector of exposure. Do not use `data.frame` or `matrix`.
`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.
parallel	logical. Enable parallel computing feature? Default = FALSE.
ncore	number of cores to run parallel computing Valid when parallel = TRUE.
...	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(QuantileData)
pheno_data <- QuantileData$PhenoData
mediator_data <- QuantileData$Mediator
```

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

## End(Not run)
```

`hima_survival` *High-dimensional mediation analysis for survival outcome data*

Description

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

Usage

```
hima_survival(
  X,
  M,
  OT,
  status,
  COV = NULL,
  topN = NULL,
  scale = TRUE,
  FDRcut = 0.05,
  verbose = FALSE,
  parallel = FALSE,
  ncore = 1
)
```

Arguments

`X`	a vector of exposure. Do not use `data.frame` or `matrix`.
`M`	a `data.frame` or `matrix` of high-dimensional mediators. Rows represent samples, columns represent mediator variables.
`OT`	a vector of observed failure times.
`status`	a vector of censoring indicator (`status` = `1`: uncensored; `status` = `0`: censored)
`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 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	HDMT pointwise FDR cutoff applied to select significant mediators. Default = 0.05.
verbose	logical. Should the function be verbose? Default = FALSE.
parallel	logical. Enable parallel computing feature? Default = FALSE.
ncore	number of cores to run parallel computing Valid when parallel = TRUE.

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(SurvivalData)
pheno_data <- SurvivalData$PhenoData
mediator_data <- SurvivalData$Mediator

hima_survival.fit <- hima_survival(
  X = pheno_data$Treatment,
  OT = pheno_data$Time,
  status = pheno_data$Status,
  M = mediator_data,
  COV = pheno_data[, c("Sex", "Age")],
  scale = FALSE, # Disabled only for simulation data
```

```
  FDRcut = 0.05,
  verbose = TRUE
)
hima_survival.fit

## End(Not run)
```

`hima_survival_long` *High-dimensional mediation analysis for longitudinal mediator and survival outcome data*

Description

`hima_survival_long` estimates and tests high-dimensional longitudinal mediation effects for survival data in a counting process framework.

Usage

```
hima_survival_long(
  X,
  M,
  tstart,
  tstop,
  status,
  id,
  COV = NULL,
  topN = NULL,
  scale = TRUE,
  Bonfcut = 0.05,
  verbose = FALSE,
  parallel = FALSE,
  ncore = 1
)
```

Arguments

`X`	A numeric vector of exposure values (do not use `data.frame` or `matrix`).
`M`	A `data.frame` or `matrix` of high-dimensional mediators (rows = observations/intervals, columns = mediators).
`tstart`	A numeric vector of starting times for each observation/interval (e.g., entry time in a counting-process setup).
`tstop`	A numeric vector of stopping times for each observation/interval (e.g., event/censoring time in a counting-process setup).
`status`	A numeric vector of censoring indicators (`1` = `event`, `0` = `censored`).
`id`	A vector of subject identifiers (used for clustering/random effects).

COV	A matrix or data.frame of adjusting covariates. Rows represent samples, columns represent variables. Can be NULL.
topN	Integer specifying the number of top mediators retained after sure independent screening (SIS). If NULL (default), topN = ceiling(n/log(n)), where n is the number of unique subjects. When topN exceeds the total number of mediators, all mediators are kept (i.e., the low-dimensional scenario).
scale	Logical. Should the function scale the exposure, mediators, and covariates? Default = TRUE.
Bonfcut	Bonferroni-corrected p value cutoff applied to select significant mediators. Default = 0.05.
verbose	Logical. Should progress messages be printed? Default = FALSE.
parallel	Logical. Enable parallel computing for SIS? Default = FALSE.
ncore	Integer specifying the number of cores to use when parallel = TRUE.

Value

A data.frame containing mediation testing results of significant mediators (joint p-value < Bonfcut).

Index Mediator name of the selected significant mediator.

alpha_hat Coefficient estimates for the exposure (X) –> mediator (M) model (adjusted for covariates).

alpha_se Standard error for alpha_hat.

beta_hat Coefficient estimates for the mediator (M) –> outcome (Y) model (adjusted for covariates and exposure).

beta_se Standard error for beta_hat.

IDE Indirect (mediation) effect estimate, i.e., alpha_hat * beta_hat.

rimp Relative importance of the mediator.

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

References

Liu L, Zhang H, Zheng Y, Gao T, Zheng C, Zhang K, Hou L, Liu L. High-dimensional mediation analysis for longitudinal mediators and survival outcomes. Briefings in Bioinformatics. 2025. DOI: 10.1093/bib/bbaf206. PMID: 40350699 PMCID: PMC12066418

Examples

```
## Not run:
data(SurvivalLongData)
pheno_data <- SurvivalLongData$PhenoData
mediator_data <- SurvivalLongData$Mediator

hima_survival_long.fit <- hima_survival_long(
  X = pheno_data$Treatment,
  M = mediator_data,
  tstart = pheno_data$Tstart,
```

```
  tstop = pheno_data$Tstop,
  status = pheno_data$Status,
  id = pheno_data$ID,
  COV = pheno_data[, c("Sex", "Age")],
  verbose = TRUE
)
hima_survival_long.fit

## End(Not run)
```

`MicrobiomeData` *Compositional Mediator Dataset for HIMA Demo*

Description

A dataset containing phenotype data and high-dimensional compositional mediators (e.g., microbiome). The dataset was simulated using parameters generated from real data.

Usage

```
MicrobiomeData
```

Format

A list with the following components:

PhenoData A data frame containing:

- **Treatment** treated (value = 1) or not treated (value = 0).
- **Outcome** a normally distributed continuous outcome variable.
- **Sex** female (value = 1) or male (value = 0).
- **Age** age of the participant.

Mediator A matrix of high-dimensional compositional mediators (rows: samples, columns: variables).

Examples

```
data(MicrobiomeData)
head(MicrobiomeData$PhenoData)
```

QuantileData *Quantile Mediation Dataset for HIMA Demo*

Description

A dataset containing phenotype data and high-dimensional mediators for quantile mediation analysis. The dataset was simulated using parameters generated from real data.

Usage

```
QuantileData
```

Format

A list with the following components:

PhenoData A data frame containing:

- **Treatment** treated (value = 1) or not treated (value = 0).
- **Outcome** an abnormally distributed continuous outcome variable.
- **Sex** female (value = 1) or male (value = 0).
- **Age** age of the participant.

Mediator A matrix of high-dimensional mediators (rows: samples, columns: variables).

Examples

```
data(QuantileData)
head(QuantileData$PhenoData)
```

SurvivalData *Survival Outcome Dataset for HIMA Demo*

Description

A dataset containing phenotype data and high-dimensional mediators for survival outcome analysis. The dataset was simulated using parameters generated from real data.

Usage

```
SurvivalData
```

Format

A list with the following components:

PhenoData A data frame containing:

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

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

Time time to the event or censoring.

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

Age age of the participant.

Mediator A matrix of high-dimensional mediators (rows: samples, columns: variables).

Examples

```
data(SurvivalData)
head(SurvivalData$PhenoData)
```

SurvivalLongData *Longitudinal Mediators with Survival Outcome Dataset for HIMA Demo*

Description

A simulated dataset for demonstrating high-dimensional and longitudinal mediation analysis with survival outcomes in a counting-process framework. The data were generated under a longitudinal mediator model and a piecewise-constant Weibull survival model, mimicking real-world analysis settings.

Usage

```
SurvivalLongData
```

Format

A list with the following components:

PhenoData A data frame where each row represents one interval (`tstart`, `tstop`) for a subject in counting-process format. It contains:

ID Subject identifier (may appear multiple times due to interval splitting).

Tstart Start time of the interval.

Tstop Stop time of the interval (event or censoring time).

Status Event indicator for the interval (`1` = `event`, `0` = `no event`).

Treatment Exposure variable for each subject.

Sex Binary covariate: `1` = `male`, `0` = `female`.

Age Age of the subject in years.

Mediator A numeric matrix of high-dimensional longitudinal mediators aligned with the rows of `PhenoData`. Columns correspond to mediator variables (`M1`, `M2`, ...), and rows correspond to observation intervals in the counting-process setup.

Examples

```
data(SurvivalLongData)
head(SurvivalLongData$PhenoData)
```

Index

∗ **datasets**
BinaryOutcome, 3
ContinuousOutcome, 4
MicrobiomeData, 24
QuantileData, 25
SurvivalData, 25
SurvivalLongData, 26
∗ **package**
HIMA-package, 2
~, *5*

BinaryOutcome, 3

ContinuousOutcome, 4

hima, 5
HIMA-package, 2
hima_classic, 10
hima_dblasso, 12
hima_efficient, 14
hima_microbiome, 16
hima_quantile, 18
hima_survival, 20
hima_survival_long, 22

MicrobiomeData, 24

QuantileData, 25

SurvivalData, 25
SurvivalLongData, 26