Package ‘MedZIsc’

July 17, 2025

Type Package

Title Statistical Framework for Co-Mediators of Zero-Inflated Single-Cell Data

Version 0.0.4

Description A causal mediation framework for single-cell data that incorporates two key features ('MedZIsc', pronounced Magics): (1) zero-inflation using beta regression and (2) overdispersed expression counts using negative binomial regression. This approach also includes a screening step based on penalized and marginal models to handle high-dimensionality. Full methodological details are available in our recent preprint by Ahn S and Li Z (2025) <doi:10.48550/arXiv.2505.22986>.

License GPL-3

Encoding UTF-8

RoxygenNote 7.3.2

Depends R (>= 3.5.0)

Imports MASS, betareg, glmnet

Suggests knitr, rmarkdown, testthat (>= 3.0.0)

LazyData true

Config/testthat/edition 3

NeedsCompilation no

Author Seungjun Ahn [cre, aut] (ORCID: <`https://orcid.org/0000-0002-4816-8924`>), Zhigang Li [ctb]

Maintainer Seungjun Ahn <`seungjun.ahn@mountsinai.org`>

Repository CRAN

Date/Publication 2025-07-16 15:40:21 UTC

Contents

`adjust_Fg` *adjust_Fg*

Description

A function that adjusts zero proportion values to meet the requirements of beta regression by bounding values of 0 and 1 to 0.001 and 0.999.

Usage

```
adjust_Fg(Fg)
```

Arguments

`Fg`	A numeric vector of length n, where each element represents the proportion of zero counts for a given gene g across cells for subject i.

Value

A vector of adjusted zero proportions, with values constrained between 0.001 and 0.999.

`Magics` *Magics*

Description

A main function for conducting causal mediation analysis with co-mediators derived from zero-inflated single-cell data.

Usage

```
Magics(data.name, n_genes, covariate.names)
```

Arguments

`data.name`	A data.frame or matrix with N x (2G + k), where N is the number of samples, G is the number of genes (each gene contributes two features: one for the zero component and one for the non-zero component), and K is the number of covariates.
`n_genes`	An interger value. The number of genes (G) represented in the data.
`covariate.names`	A character vector to specify the column name of covariates.

Value

A list containing the following elements: (1) estimated coefficients from the outcome and two mediation models (M and F models in methodology paper); (2) standard errors corresponding to (1); (3) logical vector indicating whether each gene's mediator component (M model) is statistically significant; (4) logical vector indicating whether each gene's zero-inflation component (F model) is statistically significant; (5) Adjusted p-values for M and F model (joint significance test).

References

Ahn S, Li Z. A Statistical Framework for Co-Mediators of Zero-Inflated Single-Cell RNA-Seq Data. ArXiv. 2025 July 8:arXiv:2507.06113v1. Available at: `https://arxiv.org/pdf/2507.06113`

Examples

```
data("simulated_data")
n_genes = ncol(simulated_data[, grep("^(M_)", colnames(simulated_data))])
Magics(data.name = simulated_data, n_genes = n_genes, covariate.names = c("Z1", "Z2", "Z3"))
```

`simulated_data` *Simulated Mediation Dataset*

Description

A simulated dataset created for zero-inflated single-cell mediation analysis.

Usage

```
simulated_data
```

Format

An object of class `data.frame` with 400 rows and 405 columns.

Details

A simulated dataset used to evaluate mediation methods for zero-inflated single-cell data. The dataset includes 300 samples with a continuous outcome (Y), a binary exposure (X), three covariates (Z1–Z3), 200 aggregated gene expression values (M_1–M_200), and corresponding zero proportions (F_1–F_200).

Source

Simulated using code in `inst/scripts/simulate_example_data.R`

Index