Package 'PolyGIM'

May 7, 2023

Type Package

Title Integrative Analysis of Individual-Level Data and Summary Statistics For Polytomous Disease

Version 1.0.0

Date 2023-5-7

Author Sheng Fu, Kai Yu

Maintainer Sheng Fu <fusheng1007@gmail.com>

Description A procedure that integrates individual-level and summary data from multiple external binary logistic regression models to fit the polytomous logistic regression (PLR) model, enabling more efficient statistical inference.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Imports VGAM, splines, stats4

RoxygenNote 7.2.3

Depends R (>= 2.10)

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

R topics documented:

PolyGIM-package *Improve the model of disease subtype heterogeneity by leveraging external summary data*

Description

This package proposes a generalized integration procedure that utilizes individual-level data and summary data from multiple binary logistic regression models obtained from external sources, in order to fit the polytomous logistic regression (PLR) model for more efficient statistical inference.

Details

Researchers are often interested in understanding the disease subtype heterogeneity by testing whether a risk exposure has the same level of effect on different disease subtypes. The polytomous logistic regression (PLR) model provides a flexible tool for such an evaluation. Disease subtype heterogeneity can also be investigated with a case-only study that uses a case-case comparison procedure to directly assess the difference between risk effects on two disease subtypes. We develop PolyGIM, a procedure to fit the PLR model by integrating individual-level data with summary data extracted from multiple studies under different designs. The summary data consist of coefficient estimates from working logistic regression models established by external studies.

Author(s)

Sheng Fu, Kai Yu

References

Fu, S., Purdue M. P., Zhang H., Qin J., Song L., Berndt S. I., Yu K. (2023). Improve the model of disease subtype heterogeneity by leveraging external summary data. Submitted.

Fu, S., Deng, L., Zhang, H., Qin, J., Yu, K. (2023). Integrative analysis of individual-level data and high-dimensional summary statistics. Bioinformatics, 39(4).

Zhang, H., Deng, L., Wheeler, W., Qin, J., Yu, K. (2022). Integrative analysis of multiple case-control studies. Biometrics, 78(3), 1080-1091.

Zhang, H., Deng, L., Schiffman, M., Qin, J., Yu, K. (2020). Generalized integration model for improved statistical inference by leveraging external summary data. Biometrika, 107(3), 689-703.

data *Data for example in PolyGIM*

Description

`data` is a Rdata file used in the example of `PolyGIM`.

Usage

```
data("data")
```

Format

A rdata file contains the following 4 objects.

`int` a data frame with 21 covairates (X1,..,X21), a PRS score covariate (score) and polytoumous outcomes (y)

`models` a list of 147 summary data and models from marginal binary logistic regression models

`nctrl` a matrix specifying the number of cases shared in datasets that are used to fit the working models given in `models`

`nctrl` a matrix specifying the number of controls shared in datasets that are used to fit the working models given in `models`

Details

This is a dataset from which internal and external data are extracted for the example.

`polygim_opt` *PolyGIM procedure with optimal choice of V*

Description

Integrating Summary Statistics and Individual-level Data Using the Optimal V

Usage

```
polygim_opt(formula, data, models, ncase, nctrl)
```

Arguments

`formula`	A formula.
`data`	A internal individual-level data frame containing all variables that are specified in `formula` and `models`.
`models`	A list of external model and summary statistics from multiple external studies.
`ncase`	A matrix specifying the number of cases shared in datasets that are used to fit the working models given in `models`.
`nctrl`	A matrix specifying the number of controls shared in datasets that are used to fit the working models given in `models`.

Examples

```
data(data, package="PolyGIM")
formula = "y~score"
fit = polygim_opt(formula, int, models, ncase, nctrl)
# estimate
fit$theta
# the corresponding standard errors
fit$se
```

polygim_v *PolyGIM procedure with a user defined V*

Description

Integrating Summary Statistics and Individual-level Data with Fixed V

Usage

```
polygim_v(formula, data, models, ncase, nctrl, V)
```

Arguments

formula	A formula.
data	A internal individual-level data frame containing all variables that are specified in formula and models.
models	A list of external model and summary statistics from multiple external studies.
ncase	A matrix specifying the number of cases shared in datasets that are used to fit the working models given in models.
nctrl	A matrix specifying the number of controls shared in datasets that are used to fit the working models given in models.
V	The variance-covariance matrix of external estimate summary data in models.

Examples

```
library("PolyGIM")
data(data, package = "PolyGIM")
formula = "y~score"
V = diag(length(models))
fit = polygim_v(formula, int, models, ncase, nctrl, V)
# estimate
fit$theta
# the corresponding standard errors
fit$se
```

Index