Package: DRMR (via r-universe)

September 28, 2024

Type Package

Title Doubly-Ranked Stratification in Mendelian Randomization

Version 0.1.0

Author Haodong Tian

Maintainer Haodong Tian <haodong.tian@mrc-bsu.cam.ac.uk>

Description Doubly-ranked and residual stratification for instrumental
variable and Mendelian randomization studies and further
stratification-based analysis.

License GPL (>= 2)

Encoding UTF-8

LazyData true

Imports dplyr, fda, fda.usc, metafor, ggplot2, MendelianRandomization

Repository https://mrcieu.r-universe.dev

RemoteUrl https://github.com/HDTian/DRMR

RemoteRef HEAD

RemoteSha 3aa25c967e6429aac0b3f30afc088cd7f7a02732

Contents

---

getDat   *Get a Toy Data*

Description

Creat a simulated data set for stratification

Usage

```
getDat(N=10000,
        IVtype='cont',
        ZXmodel='A',
        XYmodel='1',
        printRR=FALSE
                    )
```

Arguments

N	a value indicates the sample size
IVtype	a character indicates which type of IV is used. Three IV types are available: the binary/dichotomous instrument (IVtype='bi'), the continuous instrument (IVtype='cont'), and the high-dimensional instruments (IVtype='high-dim')
ZXmodel	a character indicates which instrument-exposure model is used. The available chioces include: 'A', 'B', 'C', 'D', 'E', 'F', 'G' and 'H'. See the reference paper for the model details
XYmodel	a character indicates which exposure-outcome model is used. The available chioces include: 'A', 'B', 'C', 'D', 'E', 'F', 'G' and 'H'. See the reference paper for the model details
printRR	Logic value indicates whether the coefficient of determination values for the instrument strength to be returned.

Value

getDat returns a data frame consisting of the instrument, the exposure, the outcome information.

References

Tian, H., Mason, A. M., Liu, C., & Burgess, S. (2022). "Relaxing parametric assumptions for non-linear Mendelian randomization using a doubly-ranked stratification method". bioRxiv, 2022-06. (Doubly-ranked stratification)

Examples

```
dat<-getDat( IVtype='cont', ZXmodel='A',XYmodel='1' ) #get a toy data
```

---

getDatU *Get a Toy Data*

Description

Creat a simulated data set for stratification. `getDatU` contains different scenarios of the confounder effects on the outcome.

Usage

```
getDatU(N = 10000,
        IVtype = "cont",
        UYmodel = "U1",
        XYmodel = "1",
        printRR = FALSE)
```

Arguments

N	a value indicates the sample size
IVtype	a character indicates which type of IV is used. Two IV types are available: the binary/dichotomous instrument (`IVtype='bi'`) and the continuous instrument (`IVtype='cont'`).
UYmodel	a character indicates which confounder-outcome model is used. The available chioces include: `'U1'`, `'U2'` and `'U3'`. See the reference paper for the model details
XYmodel	a character indicates which exposure-outcome model is used. The available chioces include: `'A'`, `'B'`, `'C'`, `'D'`. See the reference paper for the model details
printRR	Logic value indicates whether the coefficient of determination values for the instrument strength to be returned.

Value

`getDatU` returns a data frame consisting of the instrument, the exposure, the outcome information.

References

Tian, H., Mason, A. M., Liu, C., & Burgess, S. (2022). "Relaxing parametric assumptions for nonlinear Mendelian randomization using a doubly-ranked stratification method". bioRxiv, 2022-06. (Doubly-ranked stratification)

Examples

```
dat<-getDatU(IVtype='cont', XYmodel='1',UYmodel='U1',printRR=TRUE )
```

---

getGRstats   *Get Gelman-Rubin statistics*

Description

Get the Gelman-Rubin uniformity statistics for each stratum. This is used to check the degree of coarsenness when the exposure is coarsened.

Usage

```
getGRstats(rdat, Nc = 2, roundnum = 3)
```

Arguments

rdat	a data containing the stratification information. rdat is the result of Stratify.
Nc	integer value indicates how many chain used for calculating the GR statistic. No absolutely optimal choice. Default value is Nc=2
roundnum	a digit indicates how many decimal places the result shoul be retained

Details

See supplementary Text S1 of the original paper for more details.

Value

getGRstats gives the GR statistic values for each stratum. Small values (<1.02, the heuristic threshold) indicates a good degree of the coarsenness.

References

Tian, H., Mason, A. M., Liu, C., & Burgess, S. (2022). "Relaxing parametric assumptions for non-linear Mendelian randomization using a doubly-ranked stratification method". bioRxiv, 2022-06.

Examples

```
dat<-getDat( IVtype='cont', ZXmodel='C',XYmodel='1' ) #get a toy data
rdat<-Stratify(dat)  #Do stratification on the data
getGRstats(rdat,Nc=2,roundnum=3)
getGRstats(rdat,Nc=5,roundnum=3)
getGRstats(rdat,Nc=10,roundnum=3)
getGRstats(rdat,Nc=100,roundnum=3)
```

---

getSummaryInf *Get the summary information*

Description

Return the summary information of each stratum

Usage

```
getSummaryInf(rdat,
              family_used='gaussian',
              covariate=FALSE,
              target = FALSE,
              XYmodel = "1",
              bxthre = 1e-05,
              getHeterQ = TRUE,
              onlyDR = FALSE)
```

Arguments

rdat	a data containing the stratification information. `rdat` is the result of `Stratify`.
family_used	a character indicates the type of the outcome. Currently support the exponential family that can be recognized by `glm` (e.g. `'gaussian'`, `'binomial'`, `'poisson'` etc) and the Cox PH model in survival analysis (`'coxph'`). The default value is `family_used='gaussian'` (used for continous outcome). Note for `family_used='coxph'` the outcome must be `Surv` objective (i.e. `Surv(time, time2, event)`).
covariate	logic value indicates whether or not adjust covariates.
target	logic value indicates whether or not calculate the target effect values fo each stratum. Target effects can only be known when the true causal effect is known. If you have real data, you may not need this argument. The default value is `target = FALSE`.
XYmodel	a character indicating the exposure-outcome model on which the target effect is calculated and based. Only applicable when `target=TRUE`.
bxthre	a threshold value for the instrument-exposure assocations. When absolute instrument-exposure association is less than the threshold value, the MR ests will not be calculated for this stratum to avoid extreme MR results.
getHeterQ	logic value indicating whether generates the heterogeneity testing results. The default value is `getHeterQ=TRUE`.
onlyDR	logic vlaue indicating whether only get the summary information for the doubly-ranked stratification. Default value is `onlyDR=FALSE`

Details

The heterogeneity statistic can be found in Supplementary Text S2. The details of the target effects can be found in Supplementary Text S3.

If covariates are to be adjusted, they should be named as C1, C2, etc in rdat.

Value

getSummaryInf returns a list consisting of the following elements: Rres, RHeterQ, DRres, DRHeterQ. Rres and DRres contain the summry information for each stratum by the residual and doubly-ranked stratification method. RHeterQ and DRHeterQ contains the heterogeneity information.

The summary information for each stratum include (all variables are stratum-specific)

size	sample size
min	minimal exposure (or covariate) value
1q	the 1st quartile exposure (or covariate) value
3q	the 3rd quartile exposure (or covariate) value
max	maximal exposure (or covariate) value
Fvalue	F statistic value of the instrument (ie the instrument strength)
bx	instrument-exposure associations
bxse	standard error of bx
by	instrument-outcome associations
byse	standard error of by
est	MR estimate
se	standard error of est
target	target effect values

Note that if the stratificaiton is for the covariate, the values of min, 1q, 3q and max are for the covariate as well. If you are confused about the variable, these values are for rdat$M (generally rdat$M==rdat$X).

The heterogeneity test information includes

Q statistic	heterogeneity statistic value
df	the degree of freedom
p-value	the p-value

References

Tian, H., Mason, A. M., Liu, C., & Burgess, S. (2022). "Relaxing parametric assumptions for non-linear Mendelian randomization using a doubly-ranked stratification method". bioRxiv, 2022-06.

Examples

```
dat<-getDat( IVtype='cont', ZXmodel='C',XYmodel='1' ) #get a toy data
rdat<-Stratify(dat)  #Do stratification on the data
RES<-getSummaryInf( rdat, target=TRUE, bxthre=1e-5, XYmodel='1',getHeterQ=TRUE)
```

Smooth	*Smoothing stratification results*

Description

Smooth helps to smooth the stratification results based on the stratum-specifc results.

Usage

```
Smooth(RES,
       StraMet='DR',
       Rall = NA,
       baseline = NA,
       splinestyle = 'Bspline',
       Norder = 1,
       XYmodel = '0',
       Knots = NA,
       Lambda = 0,
       random_effect = TRUE,
       getHeterQ = TRUE,
       Plot = FALSE,
       ylim_used = NA)
```

Arguments

RES	the summary information data derived by getSummaryInf(). One example is RES<-getSummaryInf(rdat,XYmodel='2')
StraMet	a character variable indicates which stratification method result will be used for further smoothing. Default is 'DR', representing the doubly-ranked stratification results.
Rall	a vector indicates the exposure range for smoothing and visualization. If Rall is not defined by user, the default exposure range is the mean exposure of the first stratum minus one and the mean exposure of the end stratum plus one.
baseline	a value representing the baseline exposure value for visualizing the causal effect shape. If baseline is not defined by user, the default baseline value is the mean exposure of the first stratum.
splinestyle	a character variable indicates the spline style used for smoothing. Default is 'Bspline', representing the B-spline style.
Norder	a positive integer value indicates the order to be used for smoothing. The default value is Norder=1.
XYmodel	a character variable indicates the index of the true underlying causal model.
Knots	either a vector or a character indicating the internal knots used for smoothing. If Knots is a vector, the vector is the internal knots. If Knots is a number character, the unform internal knots with number Knots over Rall will be used. If Knots is not defined by user, no internal knots will be considered.

Lambda	a positive value indicates the tuning parameter for smoothing roughness. Default value is `lambda=0`, which means no roughness penalty for smoothing.
random_effect	a logic value indicates whether to use random-effect for smoothing models. Default value is `random_effect=TRUE`
getHeterQ	a logic value indicates whether generates the heterogeneity testing results for the visulization results. The default value is `getHeterQ=TRUE`.
Plot	a logic value indicates whether print the visulization results immediately. Default is `Plot=FALSE`
ylim_used	a vector indicates the y axis limit for visulization of h'(x).

Details

The smoothing is based on B-spline system. Relevant papers will be appear to provide further details. The present smoothing methods for stratification results are fractional polynomial method and the piecewise linear method (see Reference below). Both of them can be achieved by `Smooth` (see Examples below).

Value

`Smooth()` gives a list containing the following elements:

thetahat	the estimated parameters for the B-spline basisfunctions
var.matrix	variance-covariance matrix of `thetahat`
summary	summary information of the smoothing results, consisting of the estimates, standard errors and p-values for the B-spline basisfunctions.
p	visulization plot for the derivatives of the causal effect shape
hp	visulization plot for the causal effect shape

References

Staley J R, Burgess S. Semiparametric methods for estimation of a nonlinear exposure-outcome relationship using instrumental variables with application to Mendelian randomization[J]. Genetic epidemiology, 2017, 41(4): 341-352.

Examples

```
dat<-getDat( IVtype='cont', ZXmodel='C',XYmodel='2' )
rdat<-Stratify(dat)
RES<-getSummaryInf( rdat,target=FALSE)

library(metafor)


#the fractional polynomial method (e.g. with degree 2)
smooth_res<-Smooth(RES,Norder=3,baseline=0)

#the piecewise linear method
cutting_values<-(RES$DRres$mean[-1] + head( RES$DRres$mean,-1) )/2
smooth_res<-Smooth(RES,Norder=1,baseline=0,Knots=cutting_values)
```

Stratify   *Stratification*

Description

Do doubly-ranked and residual stratification on a data

Usage

```
Stratify(dat, onExposure = TRUE, Ns = 10, SoP = NA, seed = NA)
```

Arguments

dat	a dataset to be stratified.
onExposure	Logic value to determine whether stratification is done on exposure or not. If onExposure=FALSE, the covariate rather than the exposure will be stratified on. Default value is onExposure==TRUE.
Ns	a value indicates the number of strata to be built.
SoP	a value indicates the size of pre-strata. Default size of pre-stratum is equal to the number of strata.
seed	a value of seed for reproducibility. Default value is seed=NA.

Details

Stratify does not require the pre-stratum or stratum has the equal size, so no need to drop individuals. Even if nrow(dat)/Ns or nrow(dat)/SoP is not an integer, stratification still works. One can let SoP>Ns, which will help to increase the stability of the doubly-ranked stratification results.

Note that the Stratify function will not induce internal randomness autonomously, in the sense that the same inputted data will always return the same output, so no worries about the reproducibility even you forget to fix seed. The argument seed help to induce internal randomness when the data contains the fixed instrument or exposure values.

Value

Stratify returns the same data as dat with augmented stratificaiton information. The new columns Rstratum, pre_stratum and DRstratum represents the residual stratification results, the pre-stratum, and the doubly-ranked stratification results rexpectively.

References

Burgess, S., Davies, N. M., & Thompson, S. G. (2014). "Instrumental variable analysis with a nonlinear exposure-outcome relationship". Epidemiology (Cambridge, Mass.), 25(6), 877. (Residual stratification)

Tian, H., Mason, A. M., Liu, C., & Burgess, S. (2022). "Relaxing parametric assumptions for nonlinear Mendelian randomization using a doubly-ranked stratification method". bioRxiv, 2022-06. (Doubly-ranked stratification)

Examples

```
dat<-getDat( IVtype='cont', ZXmodel='C',XYmodel='2' )
rdat<-Stratify(dat)
```

Index