Package ‘TwoPhaseInd’

March 17, 2016

Type Package

Title Estimate Gene-Treatment Interaction Exploiting Randomization

Version 1.1.1

Author James Dai [aut, cre],
Xiaoyu Wang [aut]

Maintainer James Dai <jdai@fredhutch.org>

Description Estimation of gene-treatment interactions in randomized clinical trials exploiting gene-
treatment independence.

License GPL (>= 2)

LazyLoad no

NeedsCompilation yes

Imports survival

Repository CRAN

Date/Publication 2016-03-17 23:13:20

R topics documented:

aco1arm .................................................. 2
aco2arm .................................................. 4
acoarm .................................................. 5
acodata .................................................. 8
caseonly .................................................. 9
char2num .................................................. 10
data .................................................. mele 10
remove_missingdata ........................................ 12
remove_rarevariants ..................................... 13
spml .................................................. 13
whiBioMarker ............................................ 15

Index 17
A function to estimate parameters in augmented case-only designs, the genotype is ascertained for a random subcohort from the active treatment arm or the placebo arm

Description

This function estimates parameters of proportional hazards model with gene-treatment interaction. It employs case-cohort estimation incorporating the case-only estimators. The method was published in Dai et al. (2015) Biometrics.

Usage

```r
colarm(data, svtime, event, treatment, BaselineMarker, id, subcohort, esttype = 1, augment = 1, extra)
```

Arguments

- `data` A data frame used to access the following data.
- `svtime` A character string of column name, corresponds to one column of the data frame, which is used to store the failure time variable (numeric).
- `event` A character string of column name, corresponds to one column of the data frame, which is used to store the indicator of failure event (1: failure, 0: not failure).
- `treatment` A character string of column name, corresponds to one column of the data frame, which is used to store the binary vector of treatment variable (1: treatment, 0: placebo).
- `BaselineMarker` A character string of column name, corresponds to one column of the data frame, which is used to store a vector of biomarker.
- `id` A character string of column name, corresponds to one column of the data frame, which is used to store the sample identifier.
- `subcohort` A character string of column name, corresponds to one column of the data frame, which is used to store the indicator of sub-cohort (1: sample belong to the sub-cohort, 0: not belong to the sub-cohort).
- `esttype` The option of estimation methods (1: Self-Prentice estimator, 0: Lin-Ying estimator).
- `augment` The indicator of whether subcohort was drawn from the active treatment arm (augment=1) or from the placebo arm (augment=0).
- `extra` A string vector of column name(s), corresponds to more or more column(s) of the data frame, which is/are used to store the extra baseline covariate(s) to be adjusted for in addition to treatment and biomarker.

Details

The function returns estimates of the proportional hazards model, and variance of the estimates. The method was published in Dai et al. (2015) Biometrics.
Value

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>beta</td>
<td>Estimated parameter</td>
</tr>
<tr>
<td>stder</td>
<td>Standard error</td>
</tr>
<tr>
<td>pVal</td>
<td>p value</td>
</tr>
</tbody>
</table>

Author(s)

James Y. Dai

References


See Also

aco2arm

Examples

```r
## Load the example data
data(acodata)
## ACO in active arm
rfit1 <- acolarm(data=acodata,
                 svtime="vacc1_evinf",
                 event="f_evinf",
                 treatment="f_treat",
                 BaselineMarker="fcgr2a.3",
                 id="ptid",
                 subcohort="subcoh",
                 esttype=1,
                 augment=1,
                 extra=c("f_agele30","f_hsv_2","f_ad5gt18","f_crcm",
                         "any_drug","num_male_part_cat","uias","uras"))

rfit1

## ACO in placebo arm
rfit2 <- acolarm(data=acodata,
                 svtime="vacc1_evinf",
                 event="f_evinf",
                 treatment="f_treat",
                 BaselineMarker="fcgr2a.3",
                 id="ptid",
                 subcohort="subcoh",
                 esttype=1,
                 augment=0,
                 extra=c("f_agele30","f_hsv_2","f_ad5gt18","f_crcm",
                         "any_drug","num_male_part_cat","uias","uras"))

rfit2
```
aco2arm

A function to estimate parameters in augmented case-only designs, the genotype is ascertained for a random subcohort from both the active treatment arm and the placebo arm

Description

This function estimates parameters of proportional hazards model with gene-treatment interaction. It employs case-cohort estimation incorporating the case-only estimators. The method was published in Dai et al. (2015) Biometrics.

Usage

aco2arm(data, svtime, event, treatment, BaselineMarker, id, subcohort, esttype = 1, extra)

Arguments

data A data frame used to access the following data.
svtime A character string of column name, corresponds to one column of the data frame, which is used to store the failure time variable (numeric).
event A character string of column name, corresponds to one column of the data frame, which is used to store the indicator of failure event (1: failure, 0: not failure).
treatment A character string of column name, corresponds to one column of the data frame, which is used to store the binary vector of treatment variable (1: treatment, 0: placebo).
BaselineMarker A character string of column name, corresponds to one column of the data frame, which is used to store a vector of biomarker.
id A character string of column name, corresponds to one column of the data frame, which is used to store the sample identifier.
subcohort A character string of column name, corresponds to one column of the data frame, which is used to store the indicator of sub-cohort (1: sample belong to the sub-cohort, 0: not belong to the sub-cohort)
esttype The option of estimation methods (1: Self-Prentice estimator, 0: Lin-Ying estimator).
extra A string vector of column name(s), corresponds to more or more column(s) of the data frame, which is/are used to store the extra baseline covariate(s) to be adjusted for in addition to treatment and biomarker.

Details

The function returns estimates of the proportional hazards model, and variance of the estimates. The method was published in Dai et al. (2015) Biometrics.
acoarm

Value

<table>
<thead>
<tr>
<th>beta</th>
<th>Estimated parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>stder</td>
<td>Standard error</td>
</tr>
<tr>
<td>pVal</td>
<td>p value</td>
</tr>
</tbody>
</table>

Author(s)

James Y. Dai

References


See Also

aco1arm

Examples

```
## Load the example data
data(acodata)
## Case-cohort + case-only estimators
rfit1 <- aco2arm(data=acodata,
                 svttime="vacc1_evinf",
                 event="f_evinf",
                 treatment="f_treat",
                 BaselineMarker="fcgr2a.3",
                 id="ptid",
                 subcohort="subcoh",
                 esttype=1,
                 extra=c("f_agele30","f_hsv_2","f_ad5gt18","f_crcm",
                         "any_drug","num_male_part_cat","uias","uras"))
rfit1
```

acoarm

A function to estimate parameters in Cox proportional hazard models by augmented case-only designs for randomized clinical trials with failure time endpoints

Description

This function estimates parameters of proportional hazards models with gene-treatment interactions. It employs classical case-cohort estimation methods, incorporating the case-only estimators. The method was published in Dai et al. (2015) Biometrics.
Usage

acoarm(data, svtime, event, treatment, BaselineMarker, id, subcohort, esttype = 1, augment = 1, extra = NULL)

Arguments

data A data frame used to access the following data.
svtime A character string of column name, corresponds to one column of the data frame, which is used to store the failure time variable (numeric).
event A character string of column name, corresponds to one column of the data frame, which is used to store the indicator of failure event (1: failure, 0: not failure).
treatment A character string of column name, corresponds to one column of the data frame, which is used to store the binary vector of treatment variable (1: treatment, 0: placebo).
BaselineMarker A character string of column name, corresponds to one column of the data frame, which is used to store a vector of baseline biomarker that is under investigation for interaction with treatment. The BaselineMarker variable is missing for those who are not sampled in the case-cohort.
id A character string of column name, corresponds to one column of the data frame, which is used to store the sample identifier.
subcohort A character string of column name, corresponds to one column of the data frame, which is used to store the indicator of sub-cohort (1: sample belong to the sub-cohort, 0: not belong to the sub-cohort)
esttype The option of estimation methods (1: Self-Prentice estimator, 0: Lin-Ying estimator).
augment The indicator of whether subcohort was drawn from the placebo arm (augment=0), from the active treatment arm (augment=1), or from both arms (augment=2).
extra A string vector of column name(s), corresponds to more or more column(s) of the data frame, which is/are used to store the extra baseline covariate(s) to be adjusted for in addition to treatment and biomarker.

Details

The function returns point estimates and standard error estimates of parameters in the proportional hazards model. The method was published in Dai et al. (2015) Biometrics.

Value

beta Estimated parameter
stder Estimated standard error of parameter estimates
pVal p value
Author(s)
James Y. Dai

References

Examples

```r
## Load the example data
data(acodata)

## ACO in placebo arm
rfit0 <- acoarm(data = acodata, 
svtime = "vacci1_evinf", 
etvent = "f_evinf", 
treatment = "f_treat", 
BaselineMarker = "fcgr2a.3", 
id = "ptid", 
subcohort = "subcoh", 
esttype = 1, 
augment = 0, 
extra = c("f_agele30", "f_hsv_2", "f_ad5gt18", "f_crcm", "any_drug", "num_male_part_cat", "uias", "uras"))

rfit0

## ACO in active arm
rfit1 <- acoarm(data = acodata, 
svtime = "vacci1_evinf", 
etvent = "f_evinf", 
treatment = "f_treat", 
BaselineMarker = "fcgr2a.3", 
id = "ptid", 
subcohort = "subcoh", 
esttype = 1, 
augment = 1, 
extra = c("f_agele30", "f_hsv_2", "f_ad5gt18", "f_crcm", "any_drug", "num_male_part_cat", "uias", "uras"))

rfit1

## ACO in both arms
rfit2 <- acoarm(data = acodata, 
svtime = "vacci1_evinf", 
etvent = "f_evinf", 
treatment = "f_treat", 
BaselineMarker = "fcgr2a.3", 
id = "ptid", 
subcohort = "subcoh", 
esttype = 1, 
augment = 2, 
extra = c("f_agele30", "f_hsv_2", "f_ad5gt18", "f_crcm", "any_drug", "num_male_part_cat", "uias", "uras"))
```
RfitAcodata

RfitAcodata

Acodata

A dataset from the STEP trial to study the interactions between gene and vaccine on HIV infection

Description

A dataset from the STEP trial to study the interactions between gene and vaccine on HIV infection

Usage

data("acodata")

Format

A data frame with 907 observations on the following 14 variables.

vacc1_evinf  the time to HIV infection, a numeric vector
f_evinf    the indicator variable for HIV infection, a numeric vector
subcoh    the indicator of whether the participant was selected into the sub-cohort for genotyping, a logical vector
ptid    participant identifier, a numeric vector
f_treat    vaccine assignment variable, a numeric vector
f_cgr2a.3  the genotype of Fcr receptor FcrRIIIa, the biomarker of interest here, a numeric vector
f_agele30  a numeric vector
f_hsv_2    a numeric vector
f_adUgt18   a numeric vector
f_crcm     a numeric vector
any_drug   a numeric vector
num_male_part_cat     a numeric vector
uias     a numeric vector
uras     a numeric vector

Details

A dataset from the STEP trial to study the interactions between gene and vaccine on HIV infection

References


A function to deal with case-only designs

Description

This function estimates parameters of case-only designs.

Usage

```r
caseonly(data, treatment, BaselineMarker, extra = NULL, fraction = 0.5)
```

Arguments

data A data frame used to access the following data.
treatment A character string of column name, corresponds to one column of the data frame, which is used to store the binary vector of treatment variable (1: treatment, 0: placebo).
BaselineMarker A character string of column name, corresponds to one column of the data frame, which is used to store a vector of biomarker.
extra A string vector of column name(s), corresponds to more or more column(s) of the data frame, which is/are used to store the extra baseline covariate(s) to be adjusted for in addition to treatment and biomarker.
fraction The randomization fraction of active treatment assignment.

Details

This function estimates parameters of case-only designs. It estimates two parameters for "treatment effect when baselineMarker=0" and treatment+baselineMarker interaction".

Value

For each parameter, it returns:

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>beta</td>
<td>Estimated parameter</td>
</tr>
<tr>
<td>stder</td>
<td>Standard error</td>
</tr>
<tr>
<td>pVal</td>
<td>p value</td>
</tr>
</tbody>
</table>

Author(s)

James Y. Dai
References


Examples

```r
#form the data
data(acodata)
cdata=acodata[acodata[,2]==1,]
cfit=caseonly(data=cdata,
  treatment="f_treat",
  BaselineMarker="fcgr2a.3",
  extra=c("f_agele30","f_hsv_2","f_ad5gt18","f_crcm",
    "any_drug","num_male_part_cat","uias","uras"))
cfit
```

---

**char2num**

A function used in acoarm to transform categorical variable to integers

---

**Description**

Transform category data to integers 0..levels(data)-1. The numeric variable can be then used in acoarm models.

**Usage**

```r
char2num(data)
```

**Arguments**

data  
data is a dataframe composed of categorical variables.

---

**mele**

function to compute the maximum estimated likelihood estimator

---

**Description**

This function computes the maximum estimated likelihood estimator (MELE) of regression parameters, which assess treatment-biomarker interactions in studies with two-phase sampling in randomized clinical trials. The function has an option to incorporate the independence between a randomized treatment and the baseline markers.
Usage

```r
mele(data, response, treatment, BaselineMarker, extra = NULL, phase,
ind = TRUE, maxit=2000)
```

Arguments

- **data**: A data frame used to access the following data. Each row contains the response and predictors of a study participant. All variables are numerical.
- **response**: A character string of column name, corresponds to one column of the data frame, which is used to store a numeric vector of response. The response variable should be coded as 1 for cases and 0 for controls.
- **treatment**: A character string of column name, corresponds to one column of the data frame, which is used to store a binary vector of the treatment. The treatment variable should be coded as 1 for treatment and 0 for placebo.
- **BaselineMarker**: A character string of column name, corresponds to one column of the data frame, which is used to store a vector of biomarker that is assessed for interaction with the treatment. The BaselineMarker variable is missing for those who are not sampled in the second phase.
- **extra**: A string vector of column name(s), corresponds to one or more column(s) of the data frame, which are used to store the extra covariate(s) to be adjusted for in addition to treatment and biomarker. All extra variables are missing for those who are not sampled in the second phase.
- **phase**: A character string of column name, correspond to one column of the data frame, which is used to store the indicator of two-phase sampling (1: not being sampled for measuring biomarker; 2: being sampled for measuring biomarker).
- **ind**: A logical flag. TRUE indicates incorporating the independence between the randomized treatment and the baseline markers.
- **maxit**: A integer number of the maximal number of iteration.

Details

The function returns estimates, standard errors, and p values for MELE of a regression model for treatment-biomarker interaction studies with two-phase sampling in randomized trials, response ~ treatment + biomarker + treatment*biomarker + other covariates. Treatment and response are available for all the samples, while baseline biomarker data are available for a subset of samples. The mele can incorporate the independence between the treatment and baseline biomarkers ascertained in the phase-two sample.

Value

- **beta**: Estimated parameter
- **stder**: Standard error
- **pVal**: p value

Author(s)

James Y. Dai
References


See Also

spmle

Examples

```r
## Load the example data
data(whiBioMarker)
## Here is an example of MELE with exploiting independent and with confounding factors:

melIndExtra <- mele(data=whiBioMarker, # dataset
response="stroke", # response variable
treatment="hrtdisp", # treatment variable
BaselineMarker="papbl", # environment variable
extra=c( # extra variable(s)
  "age" # age
  ,"dias" # diabetes
  ,"hyp" # hypertension
  ,"syst" # systolic
  ,"diabtrt" # diastolic BP
  ,"lmsepi" # waist:hip ratio
), # extra variable(s)
phase="phase", # phase indicator
ind=TRUE # independent or non-independent
)
```

remove_missingdata A function used in acoarm to remove missing data

Description

It is used to remove samples which have NA/missing data in covariates.

Usage

```r
remove_missingdata(data)
```

Arguments

data data is a dataframe.
remove_rarevariants

A function used in spmle and acoarm to remove rare-variant covariates

Description

It is used to remove a rare-variant covariates, which can cause divergence problem.

Usage

remove_rarevariants(data, cutoff = 0.02)

Arguments

data A dataframe composed of covariates.
cutoff Proportion cutoff. If data composed of more than (1-cutoff) proportion of a constant value, we call it rare-variant.

spmle

function to compute the semiparametric maximum likelihood estimator

Description

This function computes the semiparametric maximum likelihood estimator (SPMLE) of regression parameters, which assess treatment-biomarker interactions in studies with two-phase sampling in randomized clinical trials. The function has an option to incorporate the independence between a randomized treatment and the baseline markers.

Usage

spmle(data, response, treatment, BaselineMarker, extra = NULL, phase, ind = TRUE, difffactor = 0.001, maxit = 1000)

Arguments

data A data frame used to access the following data. Each row contains the response and predictors of a study participant. All variables are numerical.
response A character string of column name, corresponds to one column of the data frame, which is used to store a numeric vector of response. The response variable should be coded as 1 for cases and 0 for controls.
treatment A character string of column name, corresponds to one column of the data frame, which is used to store a binary vector of the treatment. The treatment variable should be coded as 1 for treatment and 0 for placebo.
BaselineMarker: A character string of column name, corresponds to one column of the data frame, which is used to store a vector of biomarker that is assessed for interaction with the treatment. The BaselineMarker variable is missing for those who are not sampled in the second phase.

extra: A string vector of column name(s), corresponds to one or more column(s) of the data frame, which are used to store the extra covariate(s) to be adjusted for in addition to treatment and biomarker. All extra variables are missing for those who are not sampled in the second phase.

phase: A character string of column name, correspond to one column of the data frame, which is used to store the indicator of two-phase sampling (1: not being sampled for measuring biomarker; 2: being sampled for measuring biomarker).

ind: A logical flag. TRUE indicates incorporating the independence between the randomized treatment and the baseline markers.

diff factor: A decimal number of the differentiation factor, used to control the step of numerical differentiation.

maxit: A integer number of the maximal number of numerical differentiation iteration.

Details

The function returns estimates, standard errors, and p values for SPMLE for parameters of a regression model for treatment-biomarker interaction studies with two-phase sampling in randomized trials, response ~ treatment + biomarker + treatment*biomarker + other covariates. Treatment and response are available for all the samples, while biomarker data are available for a subset of samples. The SPMLE can incorporate the independence between the treatment and baseline biomarkers ascertained in the phase-two sample. A profile likelihood based Newton-Raphson algorithm is used to compute SPMLE.

Value

- beta: Estimated parameter
- stder: Standard error
- pVal: p value

Author(s)

James Y. Dai

References


See Also

mele
Examples

```r
## Load the example data
data(whiBioMarker)

## Here is an example of SPMLE with exploiting independent and with confounding factors:
```

Description

A dataset from a Women's Health Initiative (WHI) hormone trial to study the interaction between biomarker and hormone therapy on stroke.

Usage

```r
data("whiBioMarker")
```

Format

A data frame consisting of 10 observations, with the following columns:

- `stroke`: a binary indicator vector of stroke; 1=has stroke
- `hrtdisp`: a binary indicator vector of treatment in the Estrogen Plus Progestin Trial; 1="Estrogen Plus Progestin", 0="placebo"
- `papbl`: a numeric vector of Biomarker PAP (plasmin-antiplasmin complex) in logarithmic scale (base 10)
- `age`: an integer vector of age
- `dias`: A binary indicator vector of Diastolic BP; 1="Yes"
- `hyp`: a vector of hypertension with levels Missing, No, Yes
- `syst`: an integer vector of Systolic BP
diabtrt A vector of Diabetes with levels: Missing, No, Yes
lmsepi A vector of episodes per week of moderate and strenuous recreational physical activity of
  >= 20 minutes duration with levels 2 ~ <4 episodes per week, 4+ episodes per week, Missing, No activity, Some activity
phase a numeric vector of phase; 1: phase 1, 2:phase 2

Details
It is an two-phase sampling example dataset adapted from Kooperberg et al. (2007) to demonstrate the usage of MELE and SPMLE algorithms in Dai et al. (2009).

Source

References

Examples
data(whiBioMarker)
str(whiBioMarker)
colnames(whiBioMarker)
Index

*Topic case-cohort designs
   aco1arm, 2
   aco2arm, 4
   acoarm, 5

*Topic case-only designs
   aco1arm, 2
   aco2arm, 4
   acoarm, 5
   caseonly, 9

*Topic datasets
   acodata, 8
   whiBioMarker, 15

*Topic semiparametric maximum likelihood estimate
   mele, 10
   spmle, 13

*Topic two-phase sampling
   mele, 10
   spmle, 13

aco1arm, 2, 5
aco2arm, 3, 4
acoarm, 5
acodata, 8
caseonly, 9
char2num, 10
mele, 10, 14
remove_missingdata, 12
remove_rarevariants, 13

spmle, 12, 13
whiBioMarker, 15