Package ‘ipwErrorY’

May 2, 2019

Type Package

Title Inverse Probability Weighted Estimation of Average Treatment
Effect with Misclassified Binary Outcome

Version 2.1

Date 2019-05-02

Author Di Shu <shudi1991@gmail.com>, Grace Y. Yi <yyi@uwaterloo.ca>

Maintainer Di Shu <shudi1991@gmail.com>

Description An implementation of the correction methods pro-
posed by Shu and Yi (2017) <doi:10.1177/0962280217743777> for the inverse probabil-
ity weighted (IPW) estimation of average treatment effect (ATE) with misclassified binary out-
comes. Logistic regression model is assumed for treatment model for all implemented correc-
tion methods, and is assumed for the outcome model for the implemented doubly robust correc-
tion method. Misclassification probability given a true value of the outcome is as-
sumed to be the same for all individuals.

License GPL (>= 2)

Imports nleqslv, stats

RoxygenNote 6.0.1

NeedsCompilation no

Repository CRAN

Date/Publication 2019-05-02 10:00:03 UTC

R topics documented:

  ipwErrorY-package ........................................ 2
  Est2Replicates ........................................ 2
  EstValidation ........................................ 4
  KnownError ........................................... 5
  KnownErrorDR ........................................ 6

Index 8
ipwErrorY-package  

Inverse Probability Weighted (IPW) Estimation of Average Treatment Effect (ATE) with Misclassified Binary Outcome

Description

This package is an implementation of the correction methods proposed by Shu and Yi (2017) for the inverse probability weighted (IPW) estimation of average treatment effect (ATE) with misclassified binary outcomes. Logistic regression model is assumed for treatment model for all implemented correction methods, and is assumed for the outcome model for the implemented doubly robust correction method. Misclassification probability given a true value of the outcome is assumed to be the same for all individuals.

Details

The ipwErrorY package implements correction methods developed by Shu and Yi (2017) to adjust for misclassification in binary outcomes in the inverse probability weighted estimation of average treatment effect. The function KnownError implements the correction method with known outcome misclassification probabilities. The function EstValidation implements the optimal linear combination correction method when validation data are available. The function Est2Replicates implements the correction method when two independent replicates of the outcome are available. The function KnownErrorDR implements the doubly robust correction method with known outcome misclassification probabilities.

Author(s)

Di Shu and Grace Y. Yi
Maintainer: Di Shu <shudi1991@gmail.com>

References


Est2Replicates  

Estimation of ATE with Two Replicates

Description

Estimation of average treatment effect when misclassification probabilities are unknown but two independent replicates of the outcome are available
Est2Replicates

Usage

Est2Replicates(data, indA, indYerror, indx, constraint = c("sensitivity equals specificity", "known specificity", "known specificity", "known prevalence"), sensitivity = NULL, specificity = NULL, prevalence = NULL, confidence = 0.95)

Arguments

data The dataset to be analyzed in the form of R data frame without missing data
indA A column name indicating the binary treatment variable
indYerror A vector of two column names indicating replicates of the binary outcome variable
indx A vector of column names indicating the covariates included in the treatment model
constraint The constraint to be used; the default assumes sensitivity equals specificity
sensitivity The specified sensitivity between 0 and 1 when imposing the constraint that sensitivity is known, and the default is set to be NULL
specificity The specified specificity between 0 and 1 when imposing the constraint that specificity is known, and the default is set to be NULL
prevalence The specified prevalence between 0 and 1 when imposing the constraint that prevalence is known, and the default is set to be NULL
confidence The confidence level between 0 and 1; the default is 0.95 corresponding to a 95 per cent confidence interval

Value

A list of the estimate of average treatment effect, sandwich-variance-based standard error, confidence interval, imposed constraint, and the information on sensitivity and specificity

Examples

# create a dataset with sensitivity=0.95 and specificity=0.85
set.seed(100)
X1=rnorm(2000)
A=rbinom(2000, 1, 1/(1+exp(-0.2-X1)))
Y=rbinom(2000, 1, 1/(1+exp(-0.2-A-X1)))
y1=which(Y==1)
y0=which(Y==0)
Yast1=Y
Yast1[y1]=rbinom(length(y1), 1, 0.95)
Yast1[y0]=rbinom(length(y0), 1, 0.15)
Yast2=Y
Yast2[y1]=rbinom(length(y1), 1, 0.95)
Yast2[y0]=rbinom(length(y0), 1, 0.15)
da=data.frame(A=A, X1=X1, Yast1=Yast1, Yast2=Yast2)
head(da)
# apply the correction method assuming specificity=0.85
Est2Replicates(da,"A",c("Yast1","Yast2"),"X1","known specificity",NULL,0.95,0.95)
EstValidation

Estimation of ATE with Validation Data

Description

Estimation of average treatment effect using the optimal linear combination method when misclassification probabilities are unknown but validation data are available.

Usage

EstValidation(maindata, validationdata, indA, indYerror, indX, indY, confidence = 0.95)

Arguments

- `maindata`: The non-validation main data in the form of R data frame without missing data.
- `validationdata`: The validation data in the form of R data frame without missing data.
- `indA`: A column name indicating the binary treatment variable.
- `indYerror`: A column name indicating the misclassified binary outcome variable.
- `indX`: A vector of column names indicating the covariates included in the treatment model.
- `indY`: A column name indicating the true binary outcome variable.
- `confidence`: The confidence level between 0 and 1; the default is 0.95 corresponding to a 95 per cent confidence interval.

Value

A list of the estimate of average treatment effect, sandwich-variance-based standard error, confidence interval, and the estimated sensitivity and specificity.

Examples

# create main data and validation data with sensitivity=0.95 and specificity=0.85
set.seed(100)
X1=rnorm(1200)
A=rbinom(1200,1,1/(1+exp(-0.2*X1)))
Y=rbinom(1200,1,1/(1+exp(-0.2-A-X1)))
y1=which(Y==1)
y0=which(Y==0)
Yast=Y
Yast[y1]=rbinom(length(y1),1,0.95)
Yast[y0]=rbinom(length(y0),1,0.15)
mainda=data.frame(A=A,X1=X1,Yast=Yast)
X1=rnorm(800)
A=rbinom(800,1,1/(1+exp(-0.2-X1)))
Y=rbinom(800,1,1/(1+exp(-0.2-A-X1)))
y1=which(Y==1)
KnownError

\[ y_0 = \text{which}(Y == 0) \]
\[ Y = y \]
\[ Y_{y1} = \text{rbinom}(\text{length}(y), 1, 0.95) \]
\[ Y_{y0} = \text{rbinom}(\text{length}(y_0), 1, 0.15) \]
\[ \text{validationdata} = \text{data.frame}(A = A, X1 = X1, Y = Y, Yast = Yast) \]
\[ \text{head}(\text{maindata}) \]
\[ \text{head}(\text{validationdata}) \]

# apply the optimal linear combination correction method
\[ \text{EstValidation}(\text{maindata}, \text{validationdata}, \text{"A"}, \text{"Yast"}, \text{"X1"}, \text{"Y"}, 0.95) \]

---

### Description

Estimation of average treatment effect with known outcome misclassification probabilities, i.e., known sensitivity and specificity

### Usage

```
KnownError(data, indA, indYerror, indX, sensitivity, specificity, confidence = 0.95)
```

### Arguments

- **data**: The dataset to be analyzed in the form of R data frame without missing data
- **indA**: A column name indicating the binary treatment variable
- **indYerror**: A column name indicating the misclassified binary outcome variable
- **indX**: A vector of column names indicating the covariates included in the treatment model
- **sensitivity**: The specified sensitivity between 0 and 1
- **specificity**: The specified specificity between 0 and 1
- **confidence**: The confidence level between 0 and 1; the default is 0.95 corresponding to a 95 per cent confidence interval

### Value

A list of the estimate of average treatment effect, sandwich-variance-based standard error and confidence interval
Examples

```r
# create a dataset with sensitivity=0.95 and specificity=0.85
set.seed(100)
X1=rnorm(2000)
A=rbinom(2000,1,1/(1+exp(-0.2-X1)))
Y=rbinom(2000,1,1/(1+exp(-0.2-A-X1)))
y1=which(Y==1)
y0=which(Y==0)
Yast=Y
Yast[y1]=rbinom(length(y1),1,0.95)
Yast[y0]=rbinom(length(y0),1,0.15)
da=data.frame(X1=X1,A=A,Yast=Yast)
head(da)
# apply the correction method with sensitivity=0.95 and specificity=0.85
KnownError(da,"A","Yast","X1",0.95,0.85,0.95)
```

---

**KnownErrorDR**

**Doubly Robust Estimation of ATE with Known Error**

**Description**

Doubly robust estimation of average treatment effect with known outcome misclassification probabilities, i.e., known sensitivity and specificity

**Usage**

```r
KnownErrorDR(data, indA, indYerror, indXtrt, indXout, sensitivity, specificity, sharePara = FALSE, confidence = 0.95)
```

**Arguments**

- **data**: The dataset to be analyzed in the form of R data frame without missing data
- **indA**: A column name indicating the binary treatment variable
- **indYerror**: A column name indicating the misclassified binary outcome variable
- **indXtrt**: A vector of column names indicating the covariates included in the treatment model
- **indXout**: A vector of column names indicating the covariates included in the outcome model
- **sensitivity**: The specified sensitivity between 0 and 1
- **specificity**: The specified specificity between 0 and 1
- **sharePara**: if the treated and untreated groups share parameters for covariates in the logistic outcome model (i.e., assuming Y~ T+X), then set sharePara=TRUE; if not (i.e., modeling Y~ X for the treated and untreated groups separately), then set sharePara=FALSE. By default, sharePara=FALSE
- **confidence**: The confidence level between 0 and 1; the default is 0.95 corresponding to a 95 per cent confidence interval
Value

A list of the estimate of average treatment effect, sandwich-variance-based standard error and confidence interval

Examples

# create a dataset with sensitivity=0.95 and specificity=0.85
set.seed(100)
X=rnorm(2000)
xx=X^2
A=rbinom(2000,1,1/(1+exp(-0.1-X-0.2*xx)))
Y=rbinom(2000,1,1/(1+exp(1-A-0.5*X-xx)))
y1=which(Y==1)
y0=which(Y==0)
Y[y1]=rbinom(length(y1),1,0.95)
Y[y0]=rbinom(length(y0),1,0.15)
Yast=Y
da=data.frame(A=A,X=X,xx=xx,Yast=Yast)
head(da)

# apply the doubly robust correction method with sensitivity=0.95 and specificity=0.85
KnownErrorDR(da,"A","Yast",c("X","xx"),c("X","xx"),0.95,0.85,FALSE,0.95)
Index

*Topic package
  ipwErrorY-package, 2

Est2Replicates, 2, 2
EstValidation, 2, 4

ipwErrorY-package, 2

KnownError, 2, 5
KnownErrorDR, 2, 6