Package ‘SelectionBias’

February 16, 2023

Title Calculates Bounds for the Selection Bias for Binary Treatment and Outcome Variables

Version 1.0.0


License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.2.3

Imports arm, stats

Suggests knitr, table1, rmarkdown, testthat (>= 3.0.0)

Config/testthat/edition 3

Depends R (>= 3.5.0)

LazyData true

VignetteBuilder knitr

NeedsCompilation no

Author Stina Zetterstrom [aut, cre] (<https://orcid.org/0000-0003-0730-9425>), Ingeborg Waernbaum [aut] (<https://orcid.org/0000-0002-4457-5311>)

Maintainer Stina Zetterstrom <stina.zetterstrom@statistics.uu.se>

Repository CRAN

Date/Publication 2023-02-16 15:20:09 UTC

R topics documented:

AFbound ................................................................. 2
**AFbound**

Description

AFbound() returns the assumption free bound for a dataset that consists of an outcome, a treatment and a selection variable. If the bias is negative, the recoding of the treatment has to be done manually.

Usage

AFbound(whichEst, outcome, treatment, selection)

Arguments

- **whichEst**: Input string. Defining the population parameter of interest. Available options are as follows. (1) Relative risk in the total population: "RR_tot", (2) Risk difference in the total population: "RD_tot", (3) Relative risk in the subpopulation: "RR_sub", (4) Risk difference in the subpopulation: "RD_sub".
- **outcome**: Input vector. A binary outcome variable.
- **treatment**: Input vector. A binary treatment variable.
- **selection**: Input vector or input scalar. A binary selection variable or a selection probability.

Value

A list with the assumption free bound.

References


Examples

# Example with selection indicator variable.
y = c(0, 0, 0, 1, 1, 1, 1)
tr = c(0, 0, 1, 1, 0, 0, 1)
sel = c(0, 1, 0, 1, 0, 1, 0)
AFbound(whichEst = "RR_tot", outcome = y, treatment = tr, selection = sel)
# Example with selection probability.
selprob = mean(sel)
AFbound(whichEst = "RR_tot", outcome = y[sel==1], treatment = tr[sel==1],
        selection = selprob)

# Example with simulated data.
n = 1000
tr = rbinom(n, 1, 0.5)
y = rbinom(n, 1, 0.2 + 0.05 * tr)
set = rbinom(n, 1, 0.4 + 0.1 * tr + 0.3 * y)
AFbound(whichEst = "RD_tot", outcome = y, treatment = tr, selection = sel)

---

**Smith and VanderWeele bound**

**Description**

`SVbound()` returns a list with the SV bound. All sensitivity parameters for the population of interest must be set to numbers, and the rest can be left as NULL. The sensitivity parameters can be inserted directly or as output from `SVboundparametersM()`. If the bias is negative, the recoding of the treatment has to be done manually.

**Usage**

```r
SVbound(
    whichEst,
    RR_UY_T1 = NULL,
    RR_UY_T0 = NULL,
    RR_SU_T1 = NULL,
    RR_SU_T0 = NULL,
    RR_UY_S1 = NULL,
    RR_TU_S1 = NULL,
    pY1_T1_S1 = NULL,
    pY1_T0_S1 = NULL
)
```

**Arguments**

- **whichEst**
  - Input string. Defining the causal estimand of interest. Available options are as follows. (1) Relative risk in the total population: "RR_tot", (2) Risk difference in the total population: "RD_tot", (3) Relative risk in the subpopulation: "RR_sub", (4) Risk difference in the subpopulation: "RD_sub".

- **RR_UY_T1**
  - Input value. The sensitivity parameter RR_UY|T=1. Must be greater than or equal to 1. Used in the bounds for the total population.

- **RR_UY_T0**
  - Input value. The sensitivity parameter RR_UY|T=0. Must be greater than or equal to 1. Used in the bounds for the total population.
RR_SU_T1  Input value. The sensitivity parameter RR_SU|T=1. Must be greater than or equal to 1. Used in the bounds for the total population.

RR_SU_T0  Input value. The sensitivity parameter RR_SU|T=0. Must be greater than or equal to 1. Used in the bounds for the total population.

RR_UY_S1  Input value. The sensitivity parameter RR_UY|S=1. Must be greater than or equal to 1. Used in the bounds for the subpopulation.

RR_TU_S1  Input value. The sensitivity parameter RR_TU|S=1. Must be greater than or equal to 1. Used in the bounds for the subpopulation.

pY1_T1_S1  Input value. The probability P(Y=1|T=1,I_S=1). Must be between 0 and 1. Used in the bounds for the risk difference estimands.

pY1_T0_S1  Input value. The probability P(Y=1|T=0,I_S=1). Must be between 0 and 1. Used in the bounds for the risk difference estimands.

Value

A list containing the Smith and VanderWeele bound.

References


Examples

# Example for relative risk in the total population.
SVbound(whichEst = "RR_tot", RR_UY_T1 = 2, RR_UY_T0 = 2,
RR_SU_T1 = 1.7, RR_SU_T0 = 1.5)

# Example for risk difference in the total population.
SVbound(whichEst = "RD_tot", RR_UY_T1 = 2, RR_UY_T0 = 2,
RR_SU_T1 = 1.7, RR_SU_T0 = 1.5, pY1_T1_S1 = 0.05, pY1_T0_S1 = 0.01)

# Example for relative risk in the subpopulation.
SVbound(whichEst = "RR_sub", RR_UY_S1 = 2.71, RR_TU_S1 = 2.33)

# Example for risk difference in the subpopulation.
SVbound(whichEst = "RD_sub", RR_UY_S1 = 2.71, RR_TU_S1 = 2.33,
pY1_T1_S1 = 0.05, pY1_T0_S1 = 0.01)
Sensitivity parameters for the Smith and VanderWeele bound

Description

SVboundparametersM() returns a list with the sensitivity parameters and an indicator if bias is negative and the treatment coding is reversed for an assumed model.

Usage

SVboundparametersM(
	whichEst,
	Vval,
	Uval,
	Tcoef,
	Ycoef,
	Scoef,
	Mmodel,
	pY1_T1_S1,
	pY1_T0_S1
)

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>whichEst</td>
<td>Input string. Defining the causal estimand of interest. Available options are as follows. (1) Relative risk in the total population: &quot;RR_tot&quot;, (2) Risk difference in the total population: &quot;RD_tot&quot;, (3) Relative risk in the subpopulation: &quot;RR_sub&quot;, (4) Risk difference in the subpopulation: &quot;RD_sub&quot;.</td>
</tr>
<tr>
<td>Vval</td>
<td>Input matrix. The first column is the values of the categories of V. The second column is the probabilities of the categories of V. If V is continuous, use a fine grid of values and probabilities.</td>
</tr>
<tr>
<td>Uval</td>
<td>Input matrix. The first column is the values of the categories of U. The second column is the probabilities of the categories of U. If U is continuous, use a fine grid of values and probabilities.</td>
</tr>
<tr>
<td>Tcoef</td>
<td>Input vector. Two numerical elements. The first element is the intercept in the model for the treatment. The second element is the slope in the model for the treatment.</td>
</tr>
<tr>
<td>Ycoef</td>
<td>Input vector. Three numerical elements. The first element is the intercept in the model for the outcome. The second element is the slope for T in the model for the outcome. The third element is the slope for U in the model for the outcome.</td>
</tr>
<tr>
<td>Scoef</td>
<td>Input matrix. Numerical matrix of size K by 4, where K is the number of selection variables. Each row is the coefficients for one selection variable. The first column is the intercepts in the models for the selection variables. The second column is the slopes for V in the models for the selection variables. The third column is the slopes for U in the models for the selection variables. The fourth column is the slopes for T in the models for the selection variables.</td>
</tr>
</tbody>
</table>
Mmodel  |  Input string. Defining the models for the variables in the M structure. If "P", the probit model is used. If "L", the logit model is
pY1_T1_S1 |  Input scalar. The observed probability P(Y=1|T=1,I_S=1).
pY1_T0_S1 |  Input scalar. The observed probability P(Y=1|T=0,I_S=1). used.

Value

A list containing the sensitivity parameters and an indicator if the treatment has been reversed.

References


Examples

# Example with no selection bias.
V = matrix(c(1, 0, 0.1, 0.9), ncol = 2)
U = matrix(c(1, 0, 0.1, 0.9), ncol = 2)
Tr = c(0, 1)
Y = c(0, 0, 1)
S = matrix(c(1, 0, 0, 0, 1, 0, 0, 0), nrow = 2, byrow = TRUE)
probT1 = 0.534
probT0 = 0.534
SVboundparametersM(whichEst = "RR_tot", Vval = V, Uval = U, Tcoef = Tr,
                    Ycoef = Y, Scoef = S, Mmodel = "P", pY1_T1_S1 = probT1, pY1_T0_S1 = probT0)

# Example with selection bias. DGP from the zika example.
V = matrix(c(1, 0, 0.85, 0.15), ncol = 2)
U = matrix(c(1, 0, 0.5, 0.5), ncol = 2)
Tr = c(-6.2, 1.75)
Y = c(-5.2, 5.0, -1.0)
S = matrix(c(1.2, 2.2, 0.0, 0.5, 2.0, -2.75, -4.0, 0.0), ncol = 4)
probT1 = 0.286
probT0 = 0.004
SVboundparametersM(whichEst = "RR_sub", Vval = V, Uval = U, Tcoef = Tr,
                    Ycoef = Y, Scoef = S, Mmodel = "L", pY1_T1_S1 = probT1, pY1_T0_S1 = probT0)

SVboundsharp  |  Check if the Smith and VanderWeele bound in the subpopulation is sharp
**SVboundsharp**

**Description**

`SVboundsharp()` returns a string that indicates if the SV bound is sharp, if it’s inconclusive or if it’s not sharp. If the bias is negative, the recoding of the treatment has to be done manually.

**Usage**

```r
SVboundsharp(BF_U, pY1_T0_S1, SVbound = NULL, AFbound = NULL)
```

**Arguments**

- **BF_U**
  - Input scalar. The bounding factor for the SV bounds in the subpopulation. Must be equal to or above 1. Can be inserted directly or as output from `SVboundparametersM()`.

- **pY1_T0_S1**
  - Input scalar. The probability $P(Y=1|T=0,I_S=1)$.

- **SVbound**
  - Optional input scalar. The SV bound, can be inserted directly or as output from `SVbound()`. Only necessary if one wants to know if the SV bound is not sharp.

- **AFbound**
  - Optional input scalar. The AF bound, can be inserted directly or as output from `AFbound()`. Only necessary if one wants to know if the SV bound is not sharp.

**Value**

A string stating if the SV bound is sharp, inconclusive or not sharp.

**References**


Zetterstrom, Stina and Waernbaum, Ingeborg. MANUSCRIPT XXX

**Examples**

```r
# Example where the SV bound is sharp.
SVboundsharp(BF_U = 1.56, pY1_T0_S1 = 0.33, SVbound = 1.56, AFbound = 3.0)

# Example where the SV bound is not sharp.
SVboundsharp(BF_U = 2, pY1_T0_S1 = 0.9, SVbound = 2, AFbound = 1.8)

# Example where the SV bound is inconclusive.
SVboundsharp(BF_U = 2, pY1_T0_S1 = 0.8, SVbound = 2, AFbound = 3)
```
**Description**

The data set is simulated to mimic real data. For the data generating process, see the vignette.

**Usage**

```r
data(zika_learner)
```

**Format**

A data frame with 5,000 observations on the following 7 binary variables:

- **mic_ceph**  Indication if the baby has microcephaly (1=microcephaly, 0=not microcephaly)
- **zika**  Indication if the mother is infected by zika (1=infected, 0=not infected)
- **urban**  Indication of the living area of the subject (1=urban, 0=rural)
- **SES**  Indication of the socioeconomic status of the subject (1=high, 0=low)
- **birth**  First selection variable. Indication if the baby is born (1=birth, 0=terminated birth)
- **hospital**  Second selection variable. Indication if the delivery is in a public hospital (1=public, 0=private)
- **sel_ind**  Selection indicator variable. Indication if the subject is included in the study (1=included, 0=not included)

**Details**

The data set is created to use in examples of selection bias. A similar example has previously been used in articles that construct bounds for selection bias (Smith and VanderWeele, 2019; Zetterstrom and Waernbaum, 2022).

**References**


https://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS?locations=BR
https://www.angloinfo.com/how-to/brazil/healthcare/health-system
Index

* datasets
  zika_learner, 8

AFbound, 2

SVbound, 3
SVboundparametersM, 5
SVboundsharp, 6

zika_learner, 8