Package ‘nearfar’

August 31, 2018

Type Package
Title Near-Far Matching
Version 1.2
Date 2018-08-30
Author Joseph Rigdon, Michael Baiocchi, Sanjay Basu
Maintainer Joseph Rigdon <jrigdon@stanford.edu>
Description Near-far matching is a study design technique for preprocessing observational data to mimic a pair-randomized trial. Individuals are matched to be near on measured confounders and far on levels of an instrumental variable.
License GPL-3
Depends nbpMatching
Imports GenSA, MASS, car, stats
NeedsCompilation no
Repository CRAN
Date/Publication 2018-08-31 03:09:31

R topics documented:

nearfar-package ................................................................. 2
angrist ................................................................. 2
calipers ............................................................... 3
eff_ratio ............................................................... 4
matches ............................................................... 5
opt_nearfar ............................................................. 6
smahal ............................................................... 8
summary.nf ............................................................ 9
summ_matches ................................. 9

Index 11
**Description**

Near-far matching is a study design technique for preprocessing observational data to mimic a pair-randomized trial. Individuals are matched to be near on measured confounders and far on levels of an instrumental variable.

**Details**

- **Package:** nearfar
- **Type:** Package
- **Version:** 1.1
- **Date:** 2017-06-01
- **License:** GPL-3

**Author(s)**

Joseph Rigdon <jrigdon@stanford.edu>

**References**


---

**angrist**

*Angrist data set for education and wages*

**Description**

A random sample of 1000 observations from the data set used by Angrist and Krueger in their investigation of the impact of education on future wages.
**Format**

A data frame with 1000 observations on the following 7 variables.

- `wage` a numeric vector
- `educ` a numeric vector
- `qob` a numeric vector
- `IV` a numeric vector
- `age` a numeric vector
- `married` a numeric vector
- `race` a numeric vector

**Details**

This data set is a random sample of 1000 observations from the URL listed below.

**Source**


**References**


**Examples**

```r
library(nearfar)
str(angrist)
## maybe str(angrist) ; plot(angrist) ...
```

---

**Description**

Updates given distance matrix to prioritize specified measured confounders in a pair match. Used in consort with `matches` function to prioritize specific measured confounders in a near-far match in the `opt_nearfar` function.

**Usage**

```
calipers(distmat, variable, tolerance = 0.2)
```
eff_ratio

Arguments

distmat An object of class distance matrix
variable Named variable from list of measured confounders
tolerance Penalty to apply to mismatched observations; values near 0 penalize mismatches more

Value

Returns an updated distance matrix

See Also

matches, opt_nearfar

Examples

dd = mtcars[1:4, 2:3]
c = calipers(distmat=smaHdd, variable=dd$cyl, tolerance=0.2)
c

eff_ratio

Inference for effect ratio

Description

Conducts inference on effect ratio as described in Section 3.3 of Baiocchi (2010), resulting in an estimate and a permutation based confidence interval for the effect ratio.

Usage

eff_ratio(dta, match, outc, trt, alpha)

Arguments

dta The name of the data frame object
match Data frame where first column contains indices for those individuals encouraged into treatment by instrumental variable and second column contains indices for those individuals discouraged from treatment by instrumental variable; returned by both opt_nearfar and matches
outc The name of the outcome variable in quotes, e.g., “wages”
trt The name of the treatment variable, e.g., “educ”
alpha Level of confidence interval
**Value**

<table>
<thead>
<tr>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>est.emp</td>
<td>Empirical estimate of effect ratio</td>
</tr>
<tr>
<td>est.HL</td>
<td>Hodges-Lehmann type estimate of effect ratio</td>
</tr>
<tr>
<td>lower</td>
<td>Lower limit to 1-alpha/2 confidence interval for effect ratio</td>
</tr>
<tr>
<td>upper</td>
<td>Upper limit to 1-alpha/2 confidence interval for effect ratio</td>
</tr>
</tbody>
</table>

**Author(s)**

Joseph Rigdon <jrigdon@stanford.edu>

**References**


**Examples**

```r
k2 = matches(dta=mtcars, covs=c("cyl", "disp"), sinks=0.2, iv="carb",
             cutpoint=2, imp.var=c("cyl"), tol.var=0.03)

eff_ratio(dta=mtcars, match=k2, outc="wt", trt="gear", alpha=0.05)
```

**Description**

Given values of percent sinks and cutpoint, this function will find the corresponding near-far match.

**Usage**

```r
matches(dta, covs, iv = NA, imp.var = NA, tol.var = NA, sinks = 0,
cutpoint = NA)
```

**Arguments**

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>dta</td>
<td>The name of the data frame on which to do the matching</td>
</tr>
<tr>
<td>covs</td>
<td>A vector of the names of the covariates to make “near”, e.g., covs=c(&quot;age&quot;, &quot;sex&quot;, &quot;race&quot;)</td>
</tr>
<tr>
<td>iv</td>
<td>The name of the instrumental variable, e.g., iv=&quot;QOB&quot;</td>
</tr>
<tr>
<td>imp.var</td>
<td>A list of (up to 5) named variables to prioritize in the “near” matching</td>
</tr>
<tr>
<td>tol.var</td>
<td>A list of (up to 5) tolerances attached to the prioritized variables where 0 is highest penalty for mismatch</td>
</tr>
</tbody>
</table>
sinks Percentage of the data to match to sinks (and thus remove) if desired; default is 0

cutpoint Value below which individuals are too similar on iv; increase to make individuals more “far” in match

Details
Default settings yield a "near" match on only observed confounders in X; add IV, sinks, and cutpoint to get near-far match.

Value
A two-column matrix of row indices of paired matches

Author(s)
Joseph Rigdon <jrigdon@stanford.edu>

References

See Also
opt_nearfar

Examples
k2 = matches(dta=mtcars, covs=c("cyl", "disp"), sinks=0.2, iv="carb",
cutpoint=2, imp.var=c("cyl"), tol.var=0.03)
k2[1:5, ]
Arguments

dta  The name of the data frame on which matching was performed
trt  The name of the treatment variable, e.g., “educ”
iv  The name of the instrumental variable, e.g., iv="QOB"
covs A vector of the names of the covariates to make “near”, e.g., covs=c("age", "sex", "race")
trt.type  Treatment variable type: “cont” for continuous, or “bin” for binary
imp.var  A list of (up to 5) named variables to prioritize in the “near” matching
tol.var  A list of (up to 5) tolerances attached to the prioritized variables where 0 is highest penalty for mismatch
adjust.IV  if TRUE, include measured confounders in treatment~IV model that is optimized; if FALSE, exclude
sink.range  A two element vector of (min, max) for range of sinks over which to optimize in the near-far match; default (0, 0.5) such that maximally 50% of observations can be removed
cutp.range  a two element vector of (min, max) for range of cutpoints (how far apart the IV will become) over which to optimize in the near-far match; default is (one SD of IV, range of IV)
max.time.seconds  How long to let the optimization algorithm run; default is 300 seconds = 5 minutes

Value

n.calls  Number of calls made to the objective function
sink.range  A two element vector of (min, max) for range of sinks over which to optimize in the near-far match; default (0, 0.5) such that maximally 50% of observations can be removed
cutp.range  a two element vector of (min, max) for range of cutpoints (how far apart the IV will become) over which to optimize in the near-far match; default is (one SD of IV, range of IV)
pct.sink  Optimal percent sinks
cutp  Optimal cutpoint
maxF  Highest value of partial F-statistic (continuous treatment) or residual deviance (binary treatment) found by simulated annealing optimizer
match  A two column matrix where the first column is the index of an “encouraged” individual and the second column is the index of the corresponding “discouraged” individual from the pair matching
summ  A table of mean variable values for both the “encouraged” and “discouraged” groups across all variables plus absolute standardized differences for each variable
smahal

Author(s)
Joseph Rigdon <jrigdon@stanford.edu>

References

Examples
k = opt_nearfar(dta=mtcars, trt="drat", covs=c("cyl", "disp"),
  trt.type="cont", iv="carb", imp.var=NA, tol.var=NA, adjust.IV=TRUE,
  max.time.seconds=2)
summary(k)

smahal

Compute rank-based Mahalanobis distance matrix between each pair

Description
This function computes the rank-based Mahalanobis distance matrix between each pair of observations in the data set. Called by matches (and ultimately opt_nearfar) function to set up a distance matrix used to create pair matches.

Usage
smahal(X)

Arguments
X A matrix of observed confounders with n rows (observations) and p columns (variables)

Value
Returns the rank-based Mahalanobis distance matrix between every pair of observations

Examples
smahal(mtcars[1:4, 2:3])
**summary.nf**

*Summary method for object of class “nf”*

### Description

Displays key information, e.g., number of matches tried, and post-match balance, for `opt_nearfar` function.

### Usage

```r
## S3 method for class 'nf'
summary(object, ...)
```

### Arguments

- `object` Object of class “nf” returned by `opt_nearfar`
- `...` additional arguments affecting the summary produced

### Author(s)

Joseph Rigdon <jrigdon@stanford.edu>

### See Also

- `opt_nearfar`

### Examples

```r
k = opt_nearfar(dta=mtcars, trt="drat", covs=c("cyl", "disp"),
                trt.type="cont", iv="carb", imp.var=NA, tol.var=NA, adjust.IV=TRUE,
                max.time.seconds=1)
summary(k)
```

**summ_matches**

*Computes table of absolute standardized differences*

### Description

Computes absolute standardized differences for both continuous and binary variables. Called by `opt_nearfar` to summarize results of near-far match.

### Usage

```r
summ_matches(dta, iv, covs, match)
```
Arguments

- `dta` The name of the data frame on which matching was performed
- `iv` The name of the instrumental variable, e.g., `iv="QOB"`
- `covs` A vector of the names of the covariates to make "near", e.g., `covs=c("age", "sex", "race")`
- `match` A two-column matrix of row indices of paired matches

Value

A table of mean variable values for both the “encouraged” and “discouraged” groups across all variables plus absolute standardized differences for each variable

Author(s)

Joseph Rigdon <jrigdon@stanford.edu>

See Also

- `opt_nearfar`

Examples

```r
k2 = matches(dta=mtcars, covs=c("cyl", "disp"), sinks=0.2, iv="carb", cutpoint=2, imp.var=c("cyl"), tol.var=0.03)
summ_matches(dta=mtcars, iv="carb", covs=c("cyl", "disp"), match=k2)
```
Index

*Topic datasets
  angrist, 2

angrist, 2

calipers, 3

eff_ratio, 4

matches, 3, 4, 5, 8

nearfar-package, 2

opt_nearfar, 3–6, 6, 8–10

smahal, 8

summ_matches, 9

summary.nf, 9