Package ‘MatchingFrontier’

February 19, 2015

Version 0.3.17
Date 2014-09-05
Type Package
Title Computation of the Balance - Sample Size Frontier in Matching Methods for Causal Inference
Author Gary King, Christopher Lucas, and Richard Nielsen
Maintainer Christopher Lucas <clucas@fas.harvard.edu>
Description MatchingFrontier returns the subset of the data with the minimum imbalance for every possible subset size (N - 1, N - 2, ...), down to the data set with the minimum possible imbalance. The package also includes tools for the estimation of causal effects for each subset size, as well as functions for visualization and data export.
Imports igraph, MASS
LazyData true
License GPL-3
NeedsCompilation no
Repository CRAN
Date/Publication 2014-09-16 23:58:00

R topics documented:

estimateEffects .................................................. 2
generateDataset .................................................. 3
lalonde ............................................................. 3
makeFrontier ...................................................... 4
parallelPlot ....................................................... 5
plotEstimates ....................................................... 6
plotFrontier ....................................................... 7

Index 9
estimateEffects  

Estimate Effects on the Frontier

Description

estimateEffects() is used to estimate the effect of the treatment along the entire frontier.

Usage

estimateEffects(frontier.object, formula, prop.estimated = 1)

Arguments

frontier.object  
An object generated by makeFrontier().

formula  
An object of class formula (or one that can be coerced to that class). This will be passed to lm().

prop.estimated  
The proportion of points on the frontier to estimate. By default, 100% of the points on the frontier are estimated. To estimate less than 100% of the points, pass the proportion to be estimated to prop.estimated (for example, .6 to estimate 60% of the points).

Examples

data(lalonde)

# set seed for sampling a small data set for this example
set.seed(02139)
lalonde <- lalonde[c(which(lalonde$treat == 1),
                  sample(which(lalonde$treat == 0), 500),]
match.on <- colnames(lalonde)[!(colnames(lalonde) %in% c('re78',
                 'treat'))]

my.frontier <- makeFrontier(dataset = lalonde,
                          treatment = 'treat',
                          outcome = 're78',
                          match.on = match.on)
my.estimates <- estimateEffects(my.frontier, 're78 ~ treat',
                               prop.estimated = .05)
generateDataset  
Generate a data set that is on the balance - sample size frontier.

Description

generateDataset() allows the user to export a data set that sits on the frontier.

Usage

generateDataset(frontier.object, N)

Arguments

  frontier.object
      An object generated by makeFrontier().
  N
      The number of observations left in the exported data set. If the user selects an undefined point, generateDataset returns a dataset from the nearest defined point on the frontier.

Examples

data(lalonde)

  # set seed for sampling a small data set for this example
  set.seed(02139)
  lalonde <- lalonde[c(which(lalonde$treat == 1),
                     sample(which(lalonde$treat == 0), 500),]

  match.on <- colnames(lalonde)[!(colnames(lalonde) %in% c('re78', 'treat'))]
  my.frontier <- makeFrontier(dataset = lalonde,
                             treatment = 'treat',
                             outcome = 're78',
                             match.on = match.on)
  n <- 300  # Identify the point from which to select the data
  matched.data <- generateDataset(my.frontier, N = n)

lalonde  
Modified Lalonde dataset

Description

This is a modified version of the Lalonde experimental dataset used for explanatory purposes only.

Usage

data(lalonde)
makeFrontier

Format
A data frame with 16437 observations on the following 11 variables.

treat treatment variable indicator
age age
education years of education
black race indicator variable
married marital status indicator variable
nodegree indicator variable for not possessing a degree
re74 real earnings in 1974
re75 real earnings in 1975
re78 real earnings in 1978 (post-treatment outcome)
hispanic ethnic indicator variable

makeFrontier Compute the balance - sample size frontier.

Description
makeFrontier() computes the balance - frontier sample size and can be used with estimateEffects to estimate effects along the balance - sample size frontier.

Usage
makeFrontier(dataset, treatment, outcome, match.on, QOI = "FSATT",
metric = "Mahal", ratio = "variable", breaks = NULL)

Arguments
dataset The data set contain containing the treatment, outcome, and variable to match on.
treatment The name of the treatment.
outcome The name of the outcome.
match.on A vector of colnames indicating which variables are to be matched on.
QOI The quantity of interest to be estimated. By default, feasible sample average treatment effect on the treated or FSATT. The other option is SATT (sample average treatment effect on the treated).
metric The metric used to measure imbalance. Defaults to average mahalanobis distance to nearest match. The other option is L1.
ratio Variable or fixed ratio. See King, Lucas, and Nielsen for details.
breaks Can be used with L1 to provide user-specified breaks.
Examples

data(lalonde)

# set seed for sampling a small data set for this example
set.seed(02139)
lalonde <- lalonde[c(which(lalonde$ treatment == 1),
                    sample(which(lalonde$ treatment == 0), 500)), ]

match.on <- colnames(lalonde)[!(colnames(lalonde) %in% c('re78', 'treat'))]
my.frontier <- makeFrontier(dataset = lalonde,
                             treatment = 'treat',
                             outcome = 're78',
                             match.on = match.on)

parallelPlot

Create a parallel plot for a specified point on the frontier.

Description

parallelPlot() creates a parallel plot for a specified point on the frontier. Wraps parcoord() from MASS.

Usage

parallelPlot(frontier.object, N, variables, treated.col = 'grey', control.col = 'black')

Arguments

frontier.object

An object generated by makeFrontier().

N

The number of observations left in the exported data set. If the user selects an undefined point, generateDataset returns a dataset from the nearest defined point on the frontier.

variables

The variables to be included in the parallel plot.

treated.col

The color of the lines corresponding to observations assigned to the treatment. Grey by default.

control.col

The color of the lines corresponding to observations assigned to the control. Black by default.

References

plotEstimates

Examples

```r
data(lalonde)

# set seed for sampling a small data set for this example
set.seed(02139)
lalonde <- lalonde[c(which(lalonde$treat == 1),
                      sample(which(lalonde$treat == 0), 500), ]

match.on <- colnames(lalonde)[!(colnames(lalonde) %in% c('re78', 'treat'))]
mahal.frontier <- makefrontier(dataset = lalonde,
                                treatment = 'treat',
                                outcome = 're78',
                                match.on = match.on)

parallelPlot(mahal.frontier,
             N = 300,
             variables = c('age',
                           're74',
                           're75',
                           'black'),
             treated.col = 'grey',
             control.col = 'blue'
)
```

plotEstimates

Plot estimates along the frontier.

Description

plotEstimates() plots estimates along the frontier.

Usage

```r
plotEstimates(estimates.object, xlab = "Number of Observations Pruned",
              ylab = "Estimate", main = "Effects Plot", ...)
```

Arguments

- `estimates.object`
  An object generated by `estimateEffects()`
- `xlab`
  The label for the x-axis. Defaults to 'Number of Observations Pruned'.
- `ylab`
  The label for the y-axis. Defaults to Estimate'.
- `main`
  The main label. Defaults to 'Effects Plot'.
- `...`
  Additional arguments to be passed to `plot`.

Details

plotEstimates() wraps `plot` and uses ...to pass additional arguments to the base `plot()` function, like color, axis range, etc.
Examples

data(lalonde)

# set seed for sampling a small data set for this example
set.seed(02139)
lalonde <- lalonde[c(which(lalonde$dresp == 1),
                      sample(which(lalonde$dresp == 0), 500)), ]

match.on <- colnames(lalonde)[!(colnames(lalonde) %in% c('re78', 'treat'))]
my.frontier <- makeFrontier(dataset = lalonde,
                          treatment = 'treat',
                          outcome = 're78',
                          match.on = match.on)
my.estimates <- estimateEffects(my.frontier, 're78 ~ treat',
                                prop.estimated = .05)
plotEstimates(my.estimates)

plotFrontier

**Plot the balance - sample size frontier.**

Description

plotFrontier() plots the balance - sample size frontier.

Usage

plotFrontier(frontier.object, xlab = "Number of Observations Pruned",
             ylab = frontier.object$metric, main = "Frontier Plot", ...)

Arguments

- **frontier.object**: An object generated by makeFrontier()
- **xlab**: The label for the x-axis. Defaults to 'Number of Observations Pruned'.
- **ylab**: The label for the y-axis. Defaults to the selected metric.
- **main**: The main label. Defaults to 'Effects Plot'.
- **...**: Additional arguments to be passed to plot.

Details

plotEstimates() wraps plot and uses ... to pass additional arguments to the base plot() function, like color, axis range, etc.
Examples

data(lalonde)

# set seed for sampling a small data set for this example
set.seed(02139)
lalonde <- lalonde[which(lalonde$treat == 1),
                   sample(which(lalonde$treat == 0), 500), ]

match.on <- colnames(lalonde)[!(colnames(lalonde) %in% c('re78', 'treat'))]
my.frontier <- makeFrontier(dataset = lalonde,
                             treatment = 'treat',
                             outcome = 're78',
                             match.on = match.on)

plotFrontier(my.frontier)
Index

estimateEffects, 2
generateDataset, 3
lalonde, 3
makeFrontier, 4
parallelPlot, 5
plotEstimates, 6
plotFrontier, 7