Package ‘MatchIt’

February 14, 2012

Version 2.4-20

Date 2011-10-24

Title MatchIt

Author Daniel Ho <daniel.e.ho@gmail.com>, Elizabeth Stuart <stuart@stat.harvard.edu>, Kosuke Imai <kimai@Princeton.Edu>,Gary King <king@harvard.edu>

Maintainer Kosuke Imai <kimai@Princeton.Edu>

Depends R (>= 2.7), MASS

Description MatchIt preprocesses data by selecting approximate matched samples of the treated and control groups with similar covariate distributions, drawing on a large variety of matching methods. After preprocessing data with MatchIt, whatever standard parametric technique one might have used without preprocessing can be used, but the results will be far less model dependent.

License GPL (>= 2)

URL http://gking.harvard.edu/matchit/

Suggests cem, optmatch, Matching, WhatIf, nnet, rpart, mgcv

Repository CRAN

Date/Publication 2011-10-24 19:02:13

R topics documented:

help.matchit 2
lalonde 2
match.data 3
matchit 4
user.prompt 7

Index **8**

help.matchit — *HTML Help for Matchit Commands and Models*

Description

The `help.matchit` command launches html help for Matchit commands and supported methods. The full manual is available online at http://gking.harvard.edu/matchit.

Usage

```
help.matchit(object)
```

Arguments

`object`	a character string representing a Matchit command or model. `help.matchit("command")` will take you to an index of Matchit commands and `help.matchit("method")` will take you to a list of matching methods. The following inputs are currently available: `exact`, `nearest`, `subclass`, `full`, `optimal`.

Author(s)

Daniel Ho <<daniel.ho@yale.edu>>; Kosuke Imai <<kimai@princeton.edu>>; Gary King <<king@harvard.edu>>; Elizabeth Stuart<<estuart@jhsph.edu>>

See Also

The complete document is available online at http://gking.harvard.edu/matchit.

lalonde — *Data from National Supported Work Demonstration and PSID, as analyzed by Dehejia and Wahba (1999).*

Description

This is a subsample of the data from the treated group in the National Supported Work Demonstration (NSW) and the comparison sample from the Current Population Survey (CPS). This data was previously analyzed extensively by Lalonde (1986) and Dehejia and Wahba (1999). The full dataset is available at http://www.columbia.edu/~rd247/nswdata.html.

Usage

```
data(lalonde)
```

Format

A data frame with 313 observations (185 treated, 429 control). There are 10 variables measured for each individual. "treat" is the treatment assignment (1=treated, 0=control). "age" is age in years. "educ" is education in number of years of schooling. "black" is an indicator for African-American (1=African-American, 0=not). "hispan" is an indicator for being of Hispanic origin (1=Hispanic, 0=not). "married" is an indicator for married (1=married, 0=not married). "nodegree" is an indicator for whether the individual has a high school degree (1=no degree, 0=degree). "re74" is income in 1974, in U.S. dollars. "re75" is income in 1975, in U.S. dollars. "re78" is income in 1978, in U.S. dollars.

Source

http://www.columbia.edu/~rd247/nswdata.html

References

Lalonde, R. (1986). Evaluating the econometric evaluations of training programs with experimental data. American Economic Review 76: 604-620.

Dehejia, R.H. and Wahba, S. (1999). Causal Effects in Nonexperimental Studies: Re-Evaluating the Evaluation of Training Programs. Journal of the American Statistical Association 94: 1053-1062.

`match.data` *Output Matched Data Sets*

Description

`match.data` outputs matched data sets from `matchit()`.

Usage

```
  match.data(object, group="all", distance = "distance",
weights = "weights", subclass = "subclass")
```

Arguments

`object`	The output object from `matchit`. This is a required input.
`group`	This argument specifies for which matched group the user wants to extract the data. Available options are `"all"` (all matched units), `"treat"` (matched units in the treatment group), and `"control"` (matched units in the control group). The default is `"all"`.
`distance`	This argument specifies the variable name used to store the distance measure. The default is `"distance"`.
`weights`	This argument specifies the variable name used to store the resulting weights from matching. The default is `"weights"`.
`subclass`	This argument specifies the variable name used to store the subclass indicator. The default is `"subclass"`.

Value

Returns a subset of the original data set sent to `matchit()`, with just the matched units. The data set also contains the additional variables `distance`, `weights`, and `subclass`. The variable `distance` gives the estimated distance measure, and `weights` gives the weights for each unit, generated in the matching procedure. The variable `subclass` gives the subclass index for each unit (if applicable). See the http://gking.harvard.edu/matchit/ for the complete documentation and type `demo(match.data)` at the R prompt to see a demonstration of the code.

Author(s)

Daniel Ho <daniel.ho@yale.edu>; Kosuke Imai <kimai@princeton.edu>; Gary King <king@harvard.edu>; Elizabeth Stuart <estuart@jhsph.edu>

See Also

Please use `help.matchit` to access the matchit reference manual. The complete document is available online at http://gking.harvard.edu/matchit.

---

matchit — *MatchIt: Matching Software for Causal Inference*

Description

`matchit` is the main command of the package *MatchIt*, which enables parametric models for causal inference to work better by selecting well-matched subsets of the original treated and control groups. MatchIt implements the suggestions of Ho, Imai, King, and Stuart (2004) for improving parametric statistical models by preprocessing data with nonparametric matching methods. MatchIt implements a wide range of sophisticated matching methods, making it possible to greatly reduce the dependence of causal inferences on hard-to-justify, but commonly made, statistical modeling assumptions. The software also easily fits into existing research practices since, after preprocessing with MatchIt, researchers can use whatever parametric model they would have used without MatchIt, but produce inferences with substantially more robustness and less sensitivity to modeling assumptions. Matched data sets created by MatchIt can be entered easily in Zelig (http://gking.harvard.edu/zelig) for subsequent parametric analyses. Full documentation is available online at http://gking.harvard.edu/matchit, and help for specific commands is available through `help.matchit`.

Usage

```
matchit(formula, data, method = "nearest", distance = "logit",
               distance.options = list(), discard = "none",
               reestimate = FALSE, ...)
```

Arguments

`formula` This argument takes the usual syntax of R formula, `treat ~ x1 + x2`, where `treat` is a binary treatment indicator and `x1` and `x2` are the pre-treatment covariates. Both the treatment indicator and pre-treatment covariates must be contained in the same data frame, which is specified as `data` (see below). All of the usual R syntax for formula works. For example, `x1:x2` represents the first order interaction term between `x1` and `x2`, and `I(x1^2)` represents the square term of `x1`. See `help(formula)` for details.

`data` This argument specifies the data frame containing the variables called in `formula`.

`method` This argument specifies a matching method. Currently, `"exact"` (exact matching), `"full"` (full matching), `"genetic"` (genetic matching), `"nearest"` (nearest neighbor matching), `"optimal"` (optimal matching), and `"subclass"` (subclassification) are available. The default is `"nearest"`. Note that within each of these matching methods, *MatchIt* offers a variety of options.

`distance` This argument specifies the method used to estimate the distance measure. The default is logistic regression, `"logit"`. A variety of other methods are available.

`distance.options`
This optional argument specifies the optional arguments that are passed to the model for estimating the distance measure. The input to this argument should be a list.

`discard` This argument specifies whether to discard units that fall outside some measure of support of the distance score before matching, and not allow them to be used at all in the matching procedure. Note that discarding units may change the quantity of interest being estimated. The options are: `"none"` (default), which discards no units before matching, `"both"`, which discards all units (treated and control) that are outside the support of the distance measure, `"control"`, which discards only control units outside the support of the distance measure of the treated units, and `"treat"`, which discards only treated units outside the support of the distance measure of the control units.

`reestimate` This argument specifies whether the model for distance measure should be reestimated after units are discarded. The input must be a logical value. The default is `FALSE`.

`...` Additional arguments to be passed to a variety of matching methods.

Details

The matching is done using the `matchit(treat ~ X, ...)` command, where `treat` is the vector of treatment assignments and `X` are the covariates to be used in the matching. There are a number of matching options, detailed below. The full syntax is `matchit(formula, data=NULL, discard=0, exact=FALSE, replace=FALSE, ratio=1, model="logit", reestimate=FALSE, nearest=TRUE, m.order=2, caliper=0, calclosest=FALSE, mahvars=NULL, subclass=0, sub.by="treat", counter=TRUE, full=FALSE, full.options=list(), ...)` A summary of the results can be seen graphically using `plot(matchitobject)`, or numerically using `summary(matchitobject)`. `print(matchitobject)` also prints out the output.

Value

call	The original matchit call.
formula	The formula used to specify the model for estimating the distance measure.
model	The output of the model used to estimate the distance measure. summary(m.out$model) will give the summary of the model where m.out is the output object from matchit.
match.matrix	An n_1 by ratio matrix where the row names, which can be obtained through row.names(match.matrix), represent the names of the treatment units, which come from the data frame specified in data. Each column stores the name(s) of the control unit(s) matched to the treatment unit of that row. For example, when the ratio input for nearest neighbor or optimal matching is specified as 3, the three columns of match.matrix represent the three control units matched to one treatment unit). NA indicates that the treatment unit was not matched.
discarded	A vector of length n that displays whether the units were ineligible for matching due to common support restrictions. It equals TRUE if unit i was discarded, and it is set to FALSE otherwise.
distance	A vector of length n with the estimated distance measure for each unit.
weights	A vector of length n that provides the weights assigned to each unit in the matching process. Unmatched units have weights equal to 0. Matched treated units have weight 1. Each matched control unit has weight proportional to the number of treatment units to which it was matched, and the sum of the control weights is equal to the number of uniquely matched control units.
subclass	The subclass index in an ordinal scale from 1 to the total number of subclasses as specified in subclass (or the total number of subclasses from full or exact matching). Unmatched units have NA.
q.cut	The subclass cut-points that classify the distance measure.
treat	The treatment indicator from data (the left-hand side of formula).
X	The covariates used for estimating the distance measure (the right-hand side of formula).
nn	A basic summary table of matched data (e.g., the number of matched units)

Author(s)

Daniel Ho <daniel.ho@yale.edu>; Kosuke Imai <kimai@princeton.edu>; Gary King <king@harvard.edu>; Elizabeth Stuart<estuart@jhsph.edu>

References

Daniel Ho, Kosuke Imai, Gary King, and Elizabeth Stuart (2007). Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. Political Analysis 15(3): 199-236. http://gking.harvard.edu/files/abs/matchp-abs.shtml

See Also

Please use help.matchit to access the matchit reference manual. The complete document is available online at http://gking.harvard.edu/matchit.

`user.prompt` *Pause in demo files*

Description

Use `user.prompt` while writing demo files to force users to hit return before continuing.

Usage

```
user.prompt()
```

Author(s)

Olivia Lau <olau@fas.harvard.edu>

See Also

`readline`

Examples

```
## Not run:
user.prompt()

## End(Not run)
```

Index

∗Topic **datasets**
lalonde, 2
∗Topic **documentation**
help.matchit, 2
∗Topic **environment**
matchit, 4
∗Topic **file**
user.prompt, 7
∗Topic **methods**
match.data, 3

help.matchit, 2

lalonde, 2

match.data, 3
MatchIt *(*matchit*)*, 4
Matchit *(*matchit*)*, 4
matchit, 4

user.prompt, 7