Package ‘SAMUR’

March 27, 2015

Type Package
Title Stochastic Augmentation of Matched Data Using Restriction Methods
Version 0.6
Date 2015-03-26
Author Mansour T.A. Sharabiani, Alireza S. Mahani
Maintainer Alireza S. Mahani <alireza.s.mahani@gmail.com>
Description Augmenting a matched data set by generating multiple stochastic, matched samples from the data using a multi-dimensional histogram constructed from dropping the input matched data into a multi-dimensional grid built on the full data set. The resulting stochastic, matched sets will likely provide a collectively higher coverage of the full data set compared to the single matched set. Each stochastic match is without duplication, thus allowing downstream validation techniques such as cross-validation to be applied to each set without concern for overfitting.
License GPL (>= 2)
Imports Matching
NeedsCompilation no
Repository CRAN
Date/Publication 2015-03-27 07:15:54

R topics documented:

  samur ................................................................. 2
  summary.samur ................................................... 3

Index 5
Description

This function generates multiple subsets of the data in which the distribution of covariates is balanced across treatment groups. It works by binning the output of a base matching algorithm into a multidimensional histogram, and drawing - without replacement - from the full data set according to the histogram. This leads to higher data coverage across multiple matched subsets without duplication of cases within each subset.

Usage

```r
samur(formula, data, matched.subset, nsmp = 100
      , use.quantile = TRUE, breaks = 10)
## S3 method for class 'samur'
print(x, ...)
```

Arguments

- `formula`: Formula expression used to describe the treatment variable (lhs) and covariates used during matching (rhs).
- `data`: Data frame containing the treatment variables and matched covariates as specified in the formula.
- `matched.subset`: An integer vector representing the indexes of a subset of data that is the output of a base matching algorithm. It cannot contain duplicate values.
- `nsmp`: Number of stochastically matched subsets to generate.
- `use.quantile`: Should numeric covariates be binned using quantiles (TRUE) or not.
- `breaks`: Number of breaks to use in binning numeric covariates.
- `x`: An object of class samur, typically the output of function samur.
- `...`: Arguments passed to/from other methods.

Value

An object of class samur, a matrix of size length(matched.subset) by nsmp, where each column is a matched subset without case duplication. It also has the following attributes:

- `call`: Copy of function call.
- `formula`: Formula passed to the function.
- `mdg`: Multi-dimensional grid used for binning the matched data subsets.
- `mdh`: Multi-dimensional histogram resulting from binning data[matched.subset, ] according to the grid specified in mdg.
- `data`: Copy of data frame passed to the function.
Author(s)
Mansour T.A. Sharabiani, Alireza S. Mahani

See Also
summary.samur

Examples

```r
## not run:
library(samur)
library(matching)
data(lalonde)
myformula <- treat ~ age + educ
myglm <- glm(myformula, lalonde, family="binomial")
X <- myglm$fitted.values
# using M=1 and replace=F to ensure no duplication
bimatch <- Match(Tr = lalonde$treat, X = myglm$fitted.values,
                 , M = 1, replace = F, caliper = 0.25)
idx <- c(bimatch$index.control, bimatch$index.treated)
my.samur <- samur(formula = myformula, data = lalonde
                  , matched.subset = idx, nsmp = 100
                  , breaks = 10, use.quantile = TRUE)
summary(my.samur, nboots = 500)
```

## End(not run)

summary.samur  

**Summarizing Output of SAMUR Augmentation Function**

Description
summary method for class "samur".

Usage

```r
## S3 method for class 'samur'
summary(object, ...)
## S3 method for class 'summary.samur'
print(x, ...)
```

Arguments

- `object`  
  An object of class "samur", usually the result of a call to `samur`.
- `x`  
  An object of class "summary.samur", usually the result of a call to `summary.samur`.
- `...`  
  Further arguments to be passed to/from other methods. Current implementation of `summary.samur` passes arguments to `MatchBalance` function from `Matching` package.
Value

A list with the following elements:

- **min.pval.new**: A vector of length equal to number of samples (nsmp) generated by `samur`, each representing the minimum p-value from all univariate tests performed by the underlying function `MatchBalance`. It also has an attributed named `min.pval.orig`, containing a similar number for the original matched subset, i.e. `data[matched.subset, ]`.

- **min.pval.orig**: Same number as above, but for original matched subset.

- **coverage.new**: Percent of cases from full data set covered among all stochastic, matched samples.

- **coverage.orig**: Same as above, calculated for the original matched subset.

Note

All t-tests used for p-value calculations are "not" paired, since the philosophy of stochastic augmentation relaxes the notion of one-to-one matching.

Author(s)

Alireza S. Mahani, Mansour T.A. Sharabiani

See Also

- `samur`, `MatchBalance`
Index

MatchBalance, 4

print.samur(samur), 2
print.summary.samur(summary.samur), 3

samur, 2, 3, 4
summary.samur, 3, 3