Package ‘PSweight’

September 15, 2020

Type Package

Title Propensity Score Weighting for Causal Inference with Observational Studies and Randomized Trials

Version 1.1.0

Date 2020-09-09

Maintainer Tianhui Zhou <tianhui.zhou@duke.edu>

Description Supports propensity score weighting analysis of observational studies and randomized trials. Enables the estimation and inference of average causal effects with binary and multiple treatments using overlap weights (ATO), inverse probability of treatment weights (ATE), average treatment effect among the treated weights (ATT), matching weights (ATM) and entropy weights (ATEN), with and without propensity score trimming. These weights are members of the family of balancing weights introduced in Li, Morgan and Zaslavsky (2018) <doi:10.1080/01621459.2016.1260466> and Li and Li (2019) <doi:10.1214/19-AOAS1282>.

Depends R (>= 3.5.0)

License GPL (>= 2)

URL https://github.com/thuizhou/PSweight

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

Suggests knitr, rmarkdown

Imports nnet, MASS, ggplot2, numDeriv

NeedsCompilation no

Author Tianhui Zhou [aut, cre], Guangyu Tong [aut], Fan Li [aut], Laine Thomas [aut], Fan Li [aut]

Repository CRAN

Date/Publication 2020-09-15 17:10:37 UTC
### plot.SumStat

#### Description

Summarize the SumStat object, generate histogram or density of estimated propensity scores and plot the balance statistics under weighting versus no weighting.

#### Usage

```r
## S3 method for class 'SumStat'
plot(
x, 
  type = "balance",
  weighted.var = TRUE,
  threshold = 0.1,
  metric = "ASD",
  breaks = 50,
  ...
)
```

#### Arguments

- `x`: a `SumStat` object obtained with `SumStat` function.
- `type`: a character indicating the type of plot to produce, including histogram of estimated propensity scores ("hist"), density of estimated propensity scores ("density"), and plot of balance statistics ("balance").
- `weighted.var`: logical. Indicating whether weighted variance should be used in calculating the balance statistics. Default is `TRUE`.
- `threshold`: an optional numeric value indicating the balance threshold for the balance plot. Default is `0.1`. Only valid when `type = "balance"`. 
metric: a character indicating the type of metric used in balance plot. Only "ASD" or "PSD" is allowed. If not specified, the default is "ASD". See `summary.SumStat` for additional details on balance metrics.

breaks: a single number giving the number of cells for the histogram. Default is 50.

... further arguments passed to or from other methods.

Details

For the balance plot, a vertical line at `threshold` is used to define balance on covariates. The default value is `threshold = 0.1` following Austin and Stuart (2015). If more than 2 treatments are considered, only density of the estimated generalized propensity scores will be produced, regardless of whether `type = "density"` or `type = "hist"`.

Value

Plot of the indicated type.

References


Examples

data("psdata")
ps.formula <- trt ~ cov1 + cov2 + cov3 + cov4 + cov5 + cov6
msstat <- SumStat(ps.formula, trtgrp = "2", data = subset(psdata, trt > 1), weight = c("IPW", "overlap", "treated", "entropy", "matching"))
plot(msstat, type = "hist")
plot(msstat, type = "balance", weighted.var = TRUE, threshold = 0.1, metric = "ASD")

Description

The `print` method for class "PSweight"

Usage

```r
## S3 method for class 'PSweight'
print(x, ...)
```
print.SumStat

Arguments

  x  an object used to select a method.
  ... further arguments passed to or from other methods.

Value

  The output from print

print.PSweightsum  Print the results of Summary.PSweight

Description

  The print method for class "PSweightsum"

Usage

  ## S3 method for class 'PSweightsum'
  print(x, ...)

Arguments

  x  an object used to select a method.
  ... further arguments passed to or from other methods.

Value

  The output from print

print.SumStat  Print the results of SumStat

Description

  The print method for class "SumStat"

Usage

  ## S3 method for class 'SumStat'
  print(x, ...)

Arguments

  x  an object used to select a method.
  ... further arguments passed to or from other methods.
print.SumSumStat

Value
The output from print

Description
The print method for class "SumSumStat"

Usage
## S3 method for class 'SumSumStat'
print(x, ...)

Arguments
x an object used to select a method.
... further arguments passed to or from other methods.

Value
The output from print

psdata Simulated dataset for PSweight

Description
This is a simulated observational study with three treatment groups to illustated the utility of PSweight.

Usage
data(psdata)

Format
A data frame with 1500 rows and 8 columns.

Details
The simulated dataset includes 1500 rows, with each row represents information recorded from each individual. There are 8 variables (columns). The treatment is the variable trt, which has three treatment arms. The outcome of interest is variable Y. cov1-cov6 are pre-treatment covariates among which cov1-cov5 are continuous, and cov6 is binary.
Examples

```r
data("psdata")
```

PStrim

*Trim the input data and propensity estimate*

Description

Trim the original data and propensity estimate according to symmetric propensity score trimming rules.

Usage

```r
PStrim(
  data,
  ps.formula = NULL,
  zname = NULL,
  ps.estimate = NULL,
  delta = 0,
  optimal = FALSE
)
```

Arguments

data an optional data frame containing the variables required by ps.formula.

ps.formula an object of class `formula` (or one that can be coerced to that class): a symbolic description of the propensity score model to be fitted. Additional details of model specification are given under "Details". This argument is optional if ps.estimate is not NULL.

zname an optional character specifying the name of the treatment variable in data. Unless ps.formula is specified, zname is required.

ps.estimate an optional matrix or data frame containing estimated (generalized) propensity scores for each observation. Typically, this is an N by J matrix, where N is the number of observations and J is the total number of treatment levels. Preferably, the column name of this matrix should match the name of treatment level, if column name is missing or there is a mismatch, the column names would be assigned according to alphabetic order of the treatment levels. A vector of propensity score estimates is also allowed in ps.estimate, in which case a binary treatment is implied and the input is regarded as the propensity to receive the last category of treatment by alphabetic order, unless otherwise stated by trtgrp.

delta trimming threshold for estimated (generalized) propensity scores. Should be no larger than 1 / number of treatment groups. Default is 0, corresponding to no trimming.

optimal an logical argument indicating if optimal trimming should be used. Default is FALSE.
Details
A typical form for \texttt{ps.formula} is \texttt{treatment \sim terms} where \texttt{treatment} is the treatment variable (identical to the variable name used to specify \texttt{zname}) and \texttt{terms} is a series of terms which specifies a linear predictor for \texttt{treatment}. \texttt{ps.formula} specifies generalized linear model for estimating the propensity scores, when \texttt{ps.estimate} is \texttt{NULL}. See \texttt{glm} for more details on generalized linear models.

When comparing two treatments, \texttt{ps.estimate} can either be a vector or a two-column matrix of estimated propensity scores. If a vector is supplied, it is assumed to be the propensity scores to receive the treatment, and the treatment group corresponds to the last group in the alphabetic order, unless otherwise specified by \texttt{trtgrp}. When comparing multiple (J>=3) treatments, \texttt{ps.estimate} needs to be specified as an N by J matrix, where N indicates the number of observations, and J indicates the total number of treatments. This matrix specifies the estimated generalized propensity scores to receive each of the J treatments.

With binary treatments, \texttt{delta} defines the symmetric propensity score trimming rule following Crump et al. (2009). With multiple treatments, \texttt{delta} defines the symmetric multinomial trimming rule introduced in Yoshida et al. (2019). With binary treatments and when \texttt{optimal} equals \texttt{TRUE}, the trimming function implements the optimal symmetric trimming rule in Crump et al. (2009). The optimal trimming threshold \texttt{delta} is then returned. With multiple treatments and \texttt{optimal} equals \texttt{TRUE}, the trimming function implements the optimal trimming rule in Yang et al. (2016). The optimal cutoff \texttt{lambda}, which defines the acceptable upper bound for the sum of inverse generalized propensity scores, is returned. See Yang et al. (2016) and Li and Li (2019) for details.

The argument \texttt{zname} is required when \texttt{ps.estimate} is not \texttt{NULL}.

Value
\texttt{PStrim} returns a list of the following values:

\begin{itemize}
\item \texttt{data} a data frame of trimmed data.
\item \texttt{trim_sum} a table summarizing the number of cases by treatment groups before and after trimming.
\item \texttt{ps.estimate} a data frame of propensity estimate after trimming.
\item \texttt{delta} an optional output of trimming threshold for symmetric trimming.
\item \texttt{lambda} an optional output trimming threshold for optimal trimming with multiple treatment groups.
\end{itemize}

References
Examples

data("psdata")

# the propensity model
ps.formula<-trt~cov1+cov2+cov3+cov4+cov5+cov6

# trim the original data by setting the threshold of propensity as 0.05
PStrim(data=psdata, ps.formula=ps.formula, delta=0.05)
PStrim(data=psdata, ps.formula=ps.formula, optimal=TRUE)

PSweight

Estimate average causal effects by propensity score weighting

Description

The function PSweight is used to estimate the average potential outcomes corresponding to each treatment group among the target population. The function currently implements the following types of weights: the inverse probability of treatment weights (IPW: target population is the combined population), average treatment effect among the treated weights (treated: target population is the population receiving a specified treatment), overlap weights (overlap: target population is the overlap population at clinical equipoise), matching weights (matching: target population is population obtained under 1:1 matching), entropy weights (entropy: target population is the population weighted by the entropy function). Augmented propensity score weighting estimators are also allowed, with propensity scores and outcome model estimates either estimated within the function, or supplied by external routines.

Usage

PSweight(
    ps.formula = NULL,
    ps.estimate = NULL,
    trtgrp = NULL,
    zname = NULL,
    yname,
    data,
    weight = "overlap",
    delta = 0,
    augmentation = FALSE,
    bootstrap = FALSE,
    R = 200,
    out.formula = NULL,
    out.estimate = NULL,
    family = "gaussian"
)
Arguments

ps.formula an object of class formula (or one that can be coerced to that class): a symbolic description of the propensity score model to be fitted. Additional details of model specification are given under "Details". This argument is optional if ps.estimate is not NULL.

ps.estimate an optional matrix or data frame containing estimated (generalized) propensity scores for each observation. Typically, this is an N by J matrix, where N is the number of observations and J is the total number of treatment levels. Preferably, the column name of this matrix should match the name of treatment level, if column name is missing or there is a mismatch, the column names would be assigned according to alaphabetic order of the treatment levels. A vector of propensity score estimates is also allowed in ps.estimate, in which case a binary treatment is implied and the input is regarded as the propensity to receive the last category of treatment by alphabetic order, unless otherwise stated by trtgrp.

trtgrp an optional character defining the "treated" population for estimating the average treatment effect among the treated (ATT). Only necessary if weight = "treated". This option can also be used to specify the treatment (in a two-treatment setting) when a vector argument is supplied for ps.estimate. Default value is the last group in the alphabetic order.

zname an optional character specifying the name of the treatment variable in data.

yname an optional character specifying the name of the outcome variable in data.

data an optional data frame containing the variables in the propensity score model and outcome model (if augmented estimator is used). If not found in data, the variables are taken from environment(formula).

weight a character or vector of characters including the types of weights to be used. "IPW" specifies the inverse probability of treatment weights for estimating the average treatment effect among the combined population. "treated" specifies the weights for estimating the average treatment effect among the treated. "overlap" specifies the (generalized) overlap weights for estimating the average treatment effect among the overlap population, or population at clinical equipoise. "matching" specifies the matching weights for estimating the average treatment effect among the matched population (ATM). "entropy" specifies the entropy weights for the average treatment effect of entropy weighted population (ATEN). Default is "overlap".

delta trimming threshold for estimated (generalized) propensity scores. Should be no larger than 1 / number of treatment groups. Default is 0, corresponding to no trimming.

augmentation logical. Indicate whether augmented weighting estimators should be used. Default is FALSE.

bootstrap logical. Indicate whether bootstrap is used to estimate the standard error of the point estimates. Default is FALSE.

R an optional integer indicating number of bootstrap replicates. Default is R = 200.
out.formula  an object of class `formula` (or one that can be coerced to that class): a symbolic
description of the outcome model to be fitted. Additional details of model spec-
ification are given under "Details". This argument is optional if `out.estimate`
is not NULL.

out.estimate  an optional matrix or data frame containing estimated potential outcomes for
each observation. Typically, this is an N by J matrix, where N is the number
of observations and J is the total number of treatment levels. Preferably, the
column name of this matrix should match the name of treatment level, if column
name is missing or there is a mismatch, the column names would be assigned
according to alphabetic order of the treatment levels, with a similar mechanism
as in `ps.estimate`.

family  a description of the error distribution and link function to be used in the out-
come model. Only required if `out.formula` is provided. Supported distribu-
tional families include "gaussian" (link = identity), "binomial" (link = logit) and "poisson" (link = log). See `family` in `glm` for more details. De-
fault is "gaussian".

Details

A typical form for `ps.formula` is treatment ~ terms where treatment is the treatment variable
(identical to the variable name used to specify `zname`) and terms is a series of terms which specifies
a linear predictor for treatment. Similarly, a typical form for `out.formula` is outcome ~ terms
where outcome is the outcome variable (identical to the variable name used to specify `yname`) and
terms is a series of terms which specifies a linear predictor for outcome. Both `ps.formula` and
`out.formula` specify generalized linear models when `ps.estimate` and/or `out.estimate` is NULL.
See `glm` for more details on generalized linear models.

When comparing two treatments, `ps.estimate` can either be a vector or a two-column matrix of
estimated propensity scores. If a vector is supplied, it is assumed to be the propensity scores to
receive the treatment, and the treatment group corresponds to the last group in the alphabetic order,
unless otherwise specified by `trtgrp`. When comparing multiple (J>=3) treatments, `ps.estimate`
needs to be specified as an N by J matrix, where N indicates the number of observations, and J
indicates the total number of treatments. This matrix specifies the estimated generalized propensity
scores to receive each of the J treatments. In general, `ps.estimate` should have column names that
indicate the level of the treatment variable, which should match the levels given in `Z`. If column
names are empty or there is a mismatch, the column names will be created following the alphabetic
order of values in `Z`, and the rightmost column of `ps.estimate` is assumed to be the treatment group,
when estimating ATT. `trtgrp` can also be used to specify the treatment group for estimating ATT.
The same mechanism applies to `out.estimate`, except that the input for `out.estimate` must be
an N by J matrix, where each row corresponds to the estimated potential outcomes (corresponding
to each treatment) for each observation.

The argument `zname` and/or `yname` is required when `ps.estimate` and/or `out.estimate` is not
NULL.

Current version of PSweight allows for five types of propensity score weights used to estimate ATE
(IPW), ATT (treated) and ATO (overlap), ATM (matching) and ATEN (entropy). These weights are
members of larger class of balancing weights defined in Li, Morgan, and Zaslavsky (2018). Spec-
cific definitions of these weights are provided in Li, Morgan, and Zaslavsky (2018), Li and Greene
(2013), Zhou, Matsouaka and Thomas (2020). When there is a practical violation of the positivity
assumption, `delta` defines the symmetric propensity score trimming rule following Crump et
al. (2009). With multiple treatments, delta defines the multinomial trimming rule introduced in Yoshida et al. (2019). The overlap weights can also be considered as a data-driven continuous trimming strategy without specifying trimming rules, see Li, Thomas and Li (2019). Additional details on balancing weights and generalized overlap weights for multiple treatment groups are provided in Li and Li (2019).

If augmentation = TRUE, an augmented weighting estimator will be implemented. For binary treatments, the augmented weighting estimator is presented in Mao, Li and Greene (2018). For multiple treatments, the augmented weighting estimator is mentioned in Li and Li (2019), and additional details will appear in our ongoing work (Zhou et al. 2020+). When weight = "ATE", the augmented estimator is also referred to as a doubly-robust (DR) estimator.

When bootstrap = TRUE, the variance will be calculated by nonparametric bootstrap, with R bootstrap replications. The default of R is 200. Otherwise, the variance will be calculated using the sandwich variance formula obtained in the M-estimation framework.

**Value**

PSweight returns a PSweight object containing a list of the following values: estimated propensity scores, average potential outcomes corresponding to each treatment, variance-covariance matrix of the point estimates, the label for each treatment group, and estimates in each bootstrap replicate if bootstrap = TRUE. A summary of PSweight can be obtained with summary.PSweight.

- **trtgrp** a character indicating the treatment group.
- **propensity** a data frame of estimated propensity scores.
- **muhat** average potential outcomes by treatment groups, with reference to specific target populations.
- **covmu** variance-covariance matrix of muhat.
- **muboot** an optional list of point estimates in each bootstrap replicate bootstrap = TRUE.
- **group** a table of treatment group labels corresponding to the output point estimates muhat.

**References**


summary.PSweight


Examples

data("psdata")
# the propensity and outcome models
ps.formula<-trt~cov1+cov2+cov3+cov4+cov5+cov6
out.formula<-Y~cov1+cov2+cov3+cov4+cov5+cov6

# without augmentation
ato1<-PSweight(ps.formula = ps.formula,yname = 'Y',data = psdata,weight = 'overlap')
summary(ato1)

# augmented weighting estimator
ato2<-PSweight(ps.formula = ps.formula,yname = 'Y',data = psdata,
augmentation = TRUE,out.formula = out.formula,family = 'gaussian',weight = 'overlap')
summary(ato2)

summary.PSweight

Summarize a PSweight object

Description

summary.PSweight is used to summarize the results from PSWeight. The output contains the average causal effects defined by specific contrasts, as well as their standard error estimates.

Usage

## S3 method for class 'PSweight'
summary(object, contrast = NULL, type = "DIF", ...)

Arguments

object a PSWeight object obtained from the PSWeight function.
contrast a vector or matrix specifying the causal contrast of interest. The average causal effects will be defined by such contrasts. For multiple treatments, the contrast parameters are explained in Li and Li (2019) for estimating general causal effects. Default is all pairwise contrasts between any two treatment groups.
type a character specifying the target estimand. The most commonly seen additive estimand is specified by type = "DIF", abbreviated for weighted difference-in-means. This is the usual pairwise average treatment effects as those defined in Li, Morgan, and Zaslavsky (2018) and Li and Li (2019). For binary (or count outcomes), we also allow two ratio estimands: causal relative risk (type = "RR") and causal odds ratio (type = "OR"). Estimates for these two ratio estimands
will be reported on the log scale (log relative risk and log odds ratio) to improve
the approximate for asymptotic normality. With binary outcomes, "DIF" is the
same as the average causal risk difference. Default is "DIF" if left empty.

... further arguments passed to or from other methods.

Details

For the contrast argument, one specifies the contrast of interest and thus defines the target esti-
mmand for comparing treatments. For example, if there are three treatment levels: A, B, and C, the
contrast A-C (i.e., E[Y(A)] - E[Y(C)]) can be specified by c(1,0,-1). The contrasts of A-C and
B-C can be jointly specified by rbind(c(1,0,-1),c(0,1,-1)).

For estimating the causal relative risk (type = "RR"), the contrast is specified at the log scale. For
example, the contrast A-C (specified by c(1,0,-1)) implies the estimation of log{E[Y(A)]} -
log{E[Y(C)]}. For estimating the causal odds ratio, the contrast is specified at the log odds scale.
For example, the contrast A-C (specified by c(1,0,-1)) implies the estimation of log{E[Y(A)]/E[1-
Y(A)]} - log{E[Y(C)]/E[1-Y(C)]}.

The variance of the contrasts will be estimated by the delta method (if sandwich variance is used,
or bootstrap = FALSE), or nonparametric bootstrap (if bootstrap = TRUE). Details will be given in
Zhou et al. (2020+).

The argument type takes one of three options: "DIF", "RR", or "OR", with "DIF" as the default
option. Typically, "RR" is relevant for binary or count outcomes, and "OR" is relevant only for
binary outcomes. "DIF" applies to all types of outcomes.

Value

A list of following values:

- trtgrp  a character indicating the treatment group, or target population under ATT weights.
- estimates  a matrix of point estimates, standard errors and 95 for contrasts of interest.
- bootestimates  a list of data frames containing estimated contrasts in each bootstrap replicate, if
  bootstrap is used to estimate standard errors.
- contrast  a table listing the specified contrasts of interest.
- group  a table of treatment group labels corresponding to the output point estimates, provided in
  results obtained from PSweight.

References


Li, F., Li, F. (2019). Propensity score weighting for causal inference with multiple treatments. The

Examples

## For examples, run: example(PSweight).
summary.SumStat

Summarize a SumStat object.

Description

summary.SumStat is used to summarize results obtained from function `SumStat`. The output includes effective sample sizes and tables for balance statistics.

Usage

```r
## S3 method for class 'SumStat'
summary(object, weighted.var = TRUE, metric = "ASD", ...)
```

Arguments

- `object`: a SumStat object obtained with the `SumStat` function.
- `weighted.var`: logical. Indicate whether the propensity score weighted variance should be used in calculating the balance metrics. Default is `TRUE`.
- `metric`: a character indicating the type of balance metrics. "ASD" refers to the pairwise absolute standardized difference and "PSD" refers to the population standardized difference. Default is "ASD".
- `...`: further arguments passed to or from other methods.

Details

For `metric`, the two options "ASD" and "PSD" are defined in Li and Li (2019) for the general family of balancing weights. Similar definitions are also given in McCaffrey et al. (2013) for inverse probability weighting. `weighted.var` specifies whether weighted or unweighted variance should be used in calculating ASD or PSD. An example of weighted variance with two treatment groups is given in Austin and Stuart (2015). For more than two treatment groups, the maximum of ASD (across all pairs of treatments) and maximum of PSD (across all treatments) are calculated, as explained in Li and Li (2019).

Value

A list of tables containing effective sample sizes and balance statistics on covariates for specified propensity score weighting schemes.

- `effective.sample.size`: a table of effective sample sizes. This serves as a conservative measure to characterize the variance inflation or precision loss due to weighting, see Li and Li (2019).
- `unweighted`: A table summarizing mean, variance by treatment groups, and standardized mean difference.
- `IPW`: If "IPW" is specified, this is a data table summarizing mean, variance by treatment groups, and standardized mean difference under inverse probability of treatment weighting.
- `treated`: If "treated" is specified, this is a data table summarizing mean, variance by treatment groups, and standardized mean difference under the ATT weights.
overlap If "overlap" is specified, this is a data table summarizing mean, variance by treatment groups, and standardized mean difference under the (generalized) overlap weights.

matching If "matching" is specified, this is a data table summarizing mean, variance by treatment groups, and standardized mean difference under the (generalized) matching weights.

entropy If "entropy" is specified, this is a data table summarizing mean, variance by treatment groups, and standardized mean difference under the (generalized) entropy weights.

References


Examples

```r
## For examples, run: example(SumStat).
```

| SumStat | Calculate summary statistics for propensity score weighting |

Description

SumStat is used to generate distributional plots of the estimated propensity scores and balance diagnostics after propensity score weighting.

Usage

```r
SumStat(
  ps.formula,
  ps.estimate = NULL,
  trtgrp = NULL,
  Z = NULL,
  covM = NULL,
  zname = NULL,
  xname = NULL,
  data = NULL,
  weight = "overlap",
  delta = 0
)
```
Arguments

ps.formula
an object of class \texttt{formula} (or one that can be coerced to that class): a symbolic description of the propensity score model to be fitted. Additional details of model specification are given under "Details". This argument is optional if \texttt{ps.estimate} is not \texttt{NULL}.

ps.estimate
an optional matrix or data frame containing estimated (generalized) propensity scores for each observation. Typically, this is an N by J matrix, where N is the number of observations and J is the total number of treatment levels. Preferably, the column names of this matrix should match the names of treatment level, if column names are missing or there is a mismatch, the column names would be assigned according to the alphabetic order of treatment levels. A vector of propensity score estimates is also allowed in \texttt{ps.estimate}, in which case a binary treatment is implied and the input is regarded as the propensity to receive the last category of treatment by alphabetic order, unless otherwise stated by \texttt{trtgrp}.

trtgrp
an optional character defining the "treated" population for estimating the average treatment effect among the treated (ATT). Only necessary if \texttt{weight = "treated"}. This option can also be used to specify the treatment (in a two-treatment setting) when a vector argument is supplied for \texttt{ps.estimate}. Default value is the last group in the alphabetic order.

Z
an optional vector specifying the values of treatment, only necessary when the covariate matrix \texttt{covM} is provided instead of \texttt{data}.

covM
an optional covariate matrix or data frame including covariates, their interactions and higher-order terms. When the covariate matrix \texttt{covM} is provided, the balance statistics are generated according to each column of this matrix.

zname
an optional character specifying the name of the treatment variable in \texttt{data}.

xname
an optional vector of characters including the names of covariates in \texttt{data}.

data
an optional data frame containing the variables in the propensity score model. If not found in \texttt{data}, the variables are taken from \texttt{environment(formula)}.

weight
a character or vector of characters including the types of weights to be used. "IPW" specifies the inverse probability weights for estimating the average treatment effect among the combined population (ATE). "treated" specifies the weights for estimating the average treatment effect among the treated (ATT). "overlap" specifies the (generalized) overlap weights for estimating the average treatment effect among the overlap population (ATO), or population at clinical equipoise. "matching" specifies the matching weights for estimating the average treatment effect among the matched population (ATM). "entropy" specifies the entropy weights for the average treatment effect of entropy weighted population (ATEN). Default is "overlap".

delta
trimming threshold for estimated (generalized) propensity scores. Should be no larger than 1 / number of treatment groups. Default is 0, corresponding to no trimming.

Details

A typical form for \texttt{ps.formula} is \texttt{treatment ~ terms} where \texttt{treatment} is the treatment variable (identical to the variable name used to specify \texttt{zname}) and \texttt{terms} is a series of terms which specifies
a linear predictor for treatment. \texttt{ps.formula} specifies logistic or multinomial logistic models for estimating the propensity scores, when \texttt{ps.estimate} is \texttt{NULL}.

When comparing two treatments, \texttt{ps.estimate} can either be a vector or a two-column matrix of estimated propensity scores. If a vector is supplied, it is assumed to be the propensity scores to receive the treatment, and the treatment group corresponds to the last group in the alphabetical order, unless otherwise specified by \texttt{trtgrp}. When comparing multiple (J>=3) treatments, \texttt{ps.estimate} needs to be specified as an N by J matrix, where N indicates the number of observations, and J indicates the total number of treatments. This matrix specifies the estimated generalized propensity scores to receive each of the J treatments. In general, \texttt{ps.estimate} should have column names that indicate the level of the treatment variable, which should match the levels given in \texttt{Z}. If column names are empty or there is a mismatch, the column names will be created following the alphabetical order of treatment levels. The rightmost column of \texttt{ps.estimate} is then assumed to be the treatment group when estimating ATT ("treated"). \texttt{trtgrp} can also be used to specify the treatment group for estimating ATT.

To generate balance statistics, one can directly specify \texttt{Z} and \texttt{covM} to indicate the treatment levels and covariate matrix. Alternatively, one can supply \texttt{data}, \texttt{zname}, and \texttt{xname} to indicate the same information. When both are specified, the function will prioritize inputs from \texttt{Z} and \texttt{covM}. When \texttt{ps.estimate} is not \texttt{NULL}, argument \texttt{zname}.

Current version of \texttt{PSweight} allows for five types of propensity score weights used to estimate ATE ("IPW"), ATT ("treated"), and ATO("overlap"), ATM "matching" and ATEN "entropy". These weights are members of a larger class of balancing weights defined in Li, Morgan, and Zaslavsky (2018). When there is a practical violation of the positivity assumption, \texttt{delta} defines the symmetric propensity score trimming rule following Crump et al. (2009). With multiple treatments, \texttt{delta} defines the multinomial trimming rule introduced in Yoshida et al. (2019). The overlap weights can also be considered as a data-driven continuous trimming strategy without specifying trimming rules, see Li, Thomas and Li (2019). Additional details on balancing weights and generalized overlap weights for multiple treatment groups are provided in Li and Li (2019). For details about matching weights and entropy weights, please refer to Li and Greene (2013) and Zhou, Matsouaka and Thomas (2020).

### Value

\texttt{SumStat} returns a \texttt{SumStat} object including a list of the following value: treatment group, propensity scores, propensity score weights, effective sample sizes, and balance statistics. A summary of \texttt{SumStat} can be obtained with \texttt{summary.SumStat}.

- \texttt{trtgrp} a character indicating the treatment group.
- \texttt{propensity} a data frame of estimated propensity scores.
- \texttt{ps.weights} a data frame of propensity score weights.
- \texttt{ess} a table of effective sample sizes. This serves as a conservative measure to characterize the variance inflation or precision loss due to weighting, see Li and Li (2019).
- \texttt{unweighted.sumstat} A list of tables including covariate means and variances by treatment group and standardized mean differences.
- \texttt{ATE.sumstat} If "IPW" is included in \texttt{weight}, this is a list of summary statistics using inverse probability weighting.
- \texttt{ATT.sumstat} If "treated" is included in \texttt{weight}, this is a list of summary statistics using the ATT weights.
AT0.sumstat  If "overlap" is included in weight, this is a list of summary statistics using the overlap weights.

ATM.sumstat  If "matching" is included in weight, this is a list of summary statistics using the matching weights.

ATEN.sumstat  If "entropy" is included in weight, this is a list of summary statistics using the entropy weights.

trim  If delta > 0, this is a table summarizing the number of observations before and after trimming.

References


Examples

data("psdata")
# the propensity model
ps.formula<-trt~cov1+cov2+cov3+cov4+cov5+cov6

# using SumStat to estimate propensity scores
msstat <- SumStat(ps.formula, trtgrp="2", data=psdata,
weight=c("IPW","overlap","treated","entropy","matching"))
summary(msstat)

# importing user-supplied propensity scores "e.h"
fit <- nnet::multinom(formula=ps.formula, data=psdata, maxit=500, trace=FALSE)
e.h <- fit$fitted.values
varname <- c("cov1","cov2","cov3","cov4","cov5","cov6")
msstat0 <- SumStat(zname="trt", xname=varname, data=psdata, ps.estimate=e.h,
trtgrp="2", weight=c("IPW","overlap","treated","entropy","matching"))
summary(msstat0)
Index

* dataset
  psdata, 5

family, 10
formula, 6, 9, 10, 16

glm, 10

plot.SumStat, 2
print, 3–5
print.PSweight, 3
print.PSweightsum, 4
print.SumStat, 4
print.SumSumStat, 5
psdata, 5
PStrim, 6
PSweight, 8, 12, 13

summary.PSweight, 11, 12
summary.SumStat, 3, 14, 17
SumStat, 2, 14, 15