

Package ‘PSW’

January 19, 2018

Type Package

Title Propensity Score Weighting Methods for Dichotomous Treatments

Description Provides propensity score weighting methods to control for confounding in causal inference with dichotomous treatments and continuous/binary outcomes. It includes the following functional modules: (1) visualization of the propensity score distribution in both treatment groups with mirror histogram, (2) covariate balance diagnosis, (3) propensity score model specification test, (4) weighted estimation of treatment effect, and (5) augmented estimation of treatment effect with outcome regression. The weighting methods include the inverse probability weight (IPW) for estimating the average treatment effect (ATE), the IPW for average treatment effect of the treated (ATT), the IPW for the average treatment effect of the controls (ATC), the matching weight (MW), the overlap weight (OVERLAP), and the trapezoidal weight (TRAPEZOIDAL). Sandwich variance estimation is provided to adjust for the sampling variability of the estimated propensity score. These methods are discussed by Hirano et al (2003) <DOI:10.1111/1468-0262.00442>, Lunceford and Davidian (2004) <DOI:10.1002/sim.1903>, Li and Greene (2013) <DOI:10.1515/ijb-2012-0030>, and Li et al (2016) <DOI:10.1080/01621459.2016.1260466>.

Version 1.1-3

Author Huzhang Mao <huzhangmao@gmail.com>, Liang Li <LLi15@mdanderson.org>

Maintainer Huzhang Mao <huzhangmao@gmail.com>

License GPL (>= 2)

Depends R (>= 3.0)

Imports stats, Hmisc, gtools, graphics

Encoding UTF-8

LazyData true

LazyLoad true

RoxygenNote 6.0.1

NeedsCompilation no

Repository CRAN

Date/Publication 2018-01-19 09:30:08 UTC

R topics documented:

psw	2
psw.aug	6
psw.balance	8
psw.mirror.hist	9
psw.spec.test	10
psw.wt	12
test_data	13

Index	14
--------------	-----------

psw	<i>Propensity score weighting</i>
-----	-----------------------------------

Description

psw is the main function to perform propensity score weighting analysis for (1) visualization of the propensity score distribution in both treatment groups, (2) covariate balance diagnosis, (3) propensity score model specification test, (4) treatment effect estimation and inference, and (5) augmented estimation with outcome regression when applicable.

Usage

```
psw(data, form.ps, weight, std.diff = FALSE, V.name = NULL,
     mirror.hist = FALSE, add.weight = FALSE, nclass = 50, wt = FALSE,
     out.var = NULL, family = "gaussian", aug = FALSE, form.outcome = NULL,
     spec.test = F, trans.type = NULL, K = 4)
```

Arguments

data	data frame to be used.
form.ps	propensity score model.
weight	weighting method to be used. Available methods are "ATE", "ATT", "ATC", "MW", "OVERLAP", and "TRAPEZOIDAL".
std.diff	calculate standardized mean difference as a percentage, std.diff=FALSE by default.
V.name	a vector of covariates on which standardized mean difference is computed or the specification test is performed. If V.name = NULL, the covariates in propensity score model are used.
mirror.hist	mirror histogram showing the propensity score distributions in both treatment groups, mirror.hist=FALSE by default.
add.weight	add propensity score weights to the mirror histogram, add.weight=FALSE by default and it is not available for weight="ATE", "ATT", or "ATC".
nclass	number of breaks in the mirror histogram.
wt	estimate the weighted estimator, wt=FALSE by default.

out.var	outcome variable, needed when wt=TRUE.
family	outcome family, either "gaussian" or "binomial", family="gaussian" by default.
aug	estimate the augmented estimator, aug=FALSE by default.
form.outcome	outcome model, needed when aug=TRUE.
spec.test	propensity score model specification test, spec.test=FALSE by default. Note that specification test is not available for weight="OVERLAP".
trans.type	a vector of the same length as V.name that specifies the transformation type for each element in V.name. Available transformations are "identity", "log", "logit", "sqrt", "Fisher". Needed when spec.test=T, and no transformation is performed with "identity". See Details.
K	value of K in $\omega(e_i) = \min(1, K \min(e_i, 1 - e_i))$ for "TRAPEZOIDAL" weight. The estimand is closer to the average treatment effect (ATE) with larger value of K . $K=4$ by default.

Details

In package PSW, treatment indicator (left handside of form.ps) should be dummy coded such that a value of 1 indicates the treated and a value of 0 indicates the control. All categorical covariates need to be dummy coded too. If the outcome belongs to the "gaussian" family, causal estimation is based on the mean difference between treatment groups. If the outcome belongs to the "binomial" family, causal estimation is based on risk difference, risk ratio, odds ratio or log odds ratio. The Delta method is used for variance estimation when applicable.

Let Z_i be the treatment indicator of subject i , e_i be the corresponding propensity score. Then propensity score weight, W_i , is defined as

$$W_i = \frac{\omega(e_i)}{Z_i e_i + (1 - Z_i)(1 - e_i)},$$

where $\omega(e_i)$ is a function depending on e_i . For "ATE", $\omega(e_i) = 1$, which leads to estimating the average treatment effect. For "ATT", $\omega(e_i) = e_i$, which leads to estimating average treatment effect for the treated. For "ATC", $\omega(e_i) = 1 - e_i$, which leads to estimating average treatment effect for the controls. For "MW", $\omega(e_i) = \min(e_i, 1 - e_i)$. For "OVERLAP", $\omega(e_i) = e_i(1 - e_i)$. For "TRAPEZOIDAL", $\omega(e_i) = \min(1, K \min(e_i, 1 - e_i))$. This type of weights are studied by Hirano, Imbens and Ridder (2003) and Li et al (2016). The $\omega(e_i)$ function is specified by the weight argument.

The matching weight ("MW") was proposed by Li and Greene (2013). The overlap weight ("OVERLAP") was proposed by Li et al (2016). These methods down weight subjects with propensity score close to 0 or 1. and hence improve the stability of computation.

A mirror histogram is produced to visualize the propensity score distributions in both treatment groups. In the mirror histogram, above the horizontal line is the histogram of the propensity scores of the control group, below is that of the treated group. The vertical axis of the histogram is the frequency. When add.weight=TRUE, the height of the green bar added to mirror histogram is the summation of the weights of subjects within the corresponding propensity score stratum. For weighting methods of "ATE", "ATT", "ATC", add.weight is not recommended for visualization because weights may be larger than 1.

Standardized mean difference for a covariate is defined as

$$\frac{100(\bar{x}_1 - \bar{x}_0)}{\sqrt{\frac{s_1^2 + s_0^2}{2}}},$$

where \bar{x}_1 and s_1^2 are weighted mean and standard deviation for the treated group, and \bar{x}_0 and s_0^2 are defined similarly for the control group. A plot showing the standardized mean difference before and after weighting adjustment will be generated to facilitate covariate balance diagnosis. It is sometimes recommended that the absolute values of standardized mean differences of all covariates should be less than 10% in order to claim covariate balance.

For the propensity score model specification test (Li and Greene, 2013), the quantity of interest is

$$\hat{B} = g \left\{ \frac{\sum_{i=1}^n W_i Z_i \mathbf{V}_i}{\sum_{i=1}^n W_i Z_i} \right\} - g \left\{ \frac{\sum_{i=1}^n W_i (1 - Z_i) \mathbf{V}_i}{\sum_{i=1}^n W_i (1 - Z_i)} \right\},$$

where \mathbf{V}_i is a vector of covariates whose balance are examined, and $g(\cdot)$ is a vector of monotone smooth transformations for the input. Transformation type is specified by argument `trans.type`, and available transformation types are "identity", "log", "logit", "sqrt", "Fisher". These transformations are recommended to improve the finite sample performance of the specification test. Log transformation ("log") and square root transformation ("sqrt") are recommended for skewed data, logit transformation ("logit") for binary data, and Fisher z-transformation ("Fisher") for bounded data between $(-1, 1)$. The current version of model specification test is not available for `weight="OVERLAP"` because it results in zero standardized difference.

For estimation of mean difference ("gaussian" family) or risk difference ("binomial" family), the weighted estimator is

$$\hat{\Delta} = \frac{\sum_{i=1}^n W_i Z_i Y_i}{\sum_{i=1}^n W_i Z_i} - \frac{\sum_{i=1}^n W_i (1 - Z_i) Y_i}{\sum_{i=1}^n W_i (1 - Z_i)},$$

and the augmented estimator is

$$\hat{\Delta}_{aug} = \frac{\sum_{i=1}^n \omega(e_i) \{m_{1i} - m_{0i}\}}{\sum_{i=1}^n \omega(e_i)} + \frac{\sum_{i=1}^n W_i Z_i \{Y_i - m_{1i}\}}{\sum_{i=1}^n W_i Z_i} - \frac{\sum_{i=1}^n W_i (1 - Z_i) \{Y_i - m_{0i}\}}{\sum_{i=1}^n W_i (1 - Z_i)},$$

where $m_{1i} = E[Y_i | \mathbf{X}_i, Z_i = 1]$ is the conditional expectation of outcome when treated given covariates \mathbf{X}_i , and $m_{0i} = E[Y_i | \mathbf{X}_i, Z_i = 0]$ is the conditional expectation of outcome when control given covariates \mathbf{X}_i . When the outcome belongs to the "binomial" family, the marginal probability is used to estimate risk ratio, odds ratio and log odds ratio. Sandwich variance estimation is used to adjust for the sampling variability in the estimated propensity scores (Li and Greene, 2013).

The augmented estimator $\hat{\Delta}_{aug}$ incorporates regression models for the outcome variable and has similar properties as the doubly robust IPW estimator (Lunceford and Davidian, 2004), but with one difference. The estimand of IPW estimator does not depend on the propensity score because $\omega(e_i) = 1$, while the estimands of other weighting methods depend on propensity score specification. Nonetheless, the proposed augmented estimator converges to the estimand defined by the corresponding propensity score model.

Value

psw returns a list of elements depending on the supplied arguments.

weight weighting method.

ps.model	object returned by fitting the propensity score model using glm with "binomial" family.
ps.hat	estimated propensity score.
W	estimated propensity score weight.
std.diff.before	A data frame of weighed mean, variance, and standardized mean difference for covariates in V.name (see below) by treatment groups before weighting. V.name is the row name and "treated.mean", "treated.var", "control.mean", "control.var", "std.diff.pct" are column names.
std.diff.after	A data frame of weighed mean, variance, and standardized mean difference for covariates in V.name by treatment groups after weighting.
est.wt	weighted estimator for mean difference when wt=T and family = "gaussian".
std.wt	standard error for est.wt.
est.aug	augmented estimator for mean difference when aug=T and family = "gaussian".
std.aug	standard error for est.aug.
est.risk.wt	weighted estimator for risk difference when wt=T and family = "binomial".
std.risk.wt	standard error for est.risk.wt.
est.risk.aug	augmented estimator for risk difference when aug=T and family = "binomial".
std.risk.aug	standard error for est.risk.aug.
est.rr.wt	weighted estimator for relative risk when wt=T and family = "binomial".
std.rr.wt	standard error for est.rr.wt.
est.or.wt	weighted estimator for odds ratio when wt=T and family = "binomial".
std.or.wt	standard error for est.or.wt.
est.lor.wt	weighted estimator for log odds ratio when wt=T and family = "binomial".
std.lor.wt	standard error for est.lor.wt.
V.name	covariates for balance diagnosis and specification test.
test.stat	test statistic for the specification test, which follows the χ_{df}^2 distribution under the null, available when spec.test=T.
df	degree of freedom for the specification test, df=rank(var.B.hat), available when spec.test=T.
pvalue	pvalue of the specification test when spec.test=T.

References

- Hirano K, Imbens GW and Ridder G. "Efficient estimation of average treatment effects using the estimated propensity score." *Econometrica* 2003; 71(4): 1161-1189.
- Li F, Morgan KL and Zaslavsky AM. "Balancing covariates via propensity score weighting." *J Am Stat Assoc* 2016; DOI:10.1080/01621459.2016.1260466.
- Li L and Greene T. "A weighting analogue to pair matching in propensity score analysis." *Int J Biostat* 2013; 9(2):215-234.
- Lunceford JK and Davidian M. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Stat Med.* 2004; 23(19):2937-2960.

See Also

[psw.balance](#), [psw.spec.test](#), [psw.wt](#), [psw.aug](#), [psw.mirror.hist](#).

Examples

```
# Load the test data set
data(test_data);

# Propensity score model
form.ps <- "Z ~ X1 + X2 + X3 + X4";

# A vector of covariates
V.name <- c( "X1", "X2", "X3", "X4" );

#1. Standardized difference with "ATE"
tmp1 <- psw( data = test_data, form.ps = form.ps, weight = "ATE",
std.diff = TRUE, V.name = V.name );

#2. Mirror histogram and add estimated matching weight to it
tmp2 <- psw( data = test_data, form.ps = form.ps, weight = "MW",
mirror.hist = TRUE, add.weight = TRUE );

#3. Estimate average treatment effect with "ATE"
tmp3 <- psw( data = test_data, form.ps = form.ps, weight = "ATE", wt = TRUE,
out.var = "Y", family = "gaussian" );

#4. Augmented estimator with "OVERLAP"
# outcome model
form.out <- "Y ~ X1 + X2 + X3 + X4";
tmp4 <- psw( data = test_data, form.ps = form.ps, weight = "OVERLAP", aug = TRUE,
form.outcome = form.out, family = "gaussian" );

#5. Propensity score model specification test with "MW".
# A vector of transformation types for covariates in V.name.
trans.type <- c( "identity", "identity", "logit", "logit" );
tmp5 <- psw( data = test_data, form.ps = form.ps, weight = "MW", spec.test = TRUE,
V.name = V.name, trans.type = trans.type );
```

psw.aug

Propensity score weighting with augmented estimation

Description

psw.aug is the function to estimate the augmented estimator for mean difference (mean outcome difference for "gaussian" family and risk difference for "binomial" family). The augmented estimator is consistent for the estimand defined by the corresponding propensity score model.

Usage

```
psw.aug(data, form.ps, weight, form.outcome, family = "gaussian", K = 4)
```

Arguments

data	data frame to be used.
form.ps	propensity score model.
weight	weighting method to be used. Available methods are "ATE", "ATT", "ATC", "MW", "OVERLAP", and "TRAPEZOIDAL".
form.outcome	outcome model.
family	outcome family, either "gaussian" or "binomial". family="gaussian" by default.
K	value of K in $\omega(e_i) = \min(1, K \min(e_i, 1 - e_i))$ for "TRAPEZOIDAL" weight. The estimand is closer to the average treatment effect (ATE) with larger value of K . $K=4$ by default.

Details

psw.aug is used to estimate the augmented estimator, $\hat{\Delta}_{aug}$, and make inference using the sandwich variance that adjusts for the sampling variability in the estimated propensity score.

Value

A list of weighting method, fitted propensity score model, estimated propensity scores, estimated propensity score weights, augmented estimator and associated standard error.

weight	weighting method.
ps.model	object returned by fitting the propensity score model using glm with "binomial" family.
ps.hat	estimated propensity score.
W	estimated propensity score weight.
est.aug	augmented estimator for mean difference when family = "gaussian".
std.aug	standard error for est.aug.
est.risk.aug	augmented estimator for risk difference when family = "binomial".
std.risk.aug	standard error for est.risk.aug.

Examples

```
# Load the test data set
data(test_data);
# Propensity score model
form.ps <- "Z ~ X1 + X2 + X3 + X4";
# Outcome model
form.out <- "Y ~ X1 + X2 + X3 + X4";
tmp <- psw.aug( data = test_data, form.ps = form.ps, weight = "ATE",
form.outcome = form.out, family="gaussian" );
```

psw.balance

*Balance checking using standardized mean difference***Description**

psw.balance is used to compute the standardized mean difference (in percentage) for balance diagnosis.

Usage

```
psw.balance(data, form.ps, weight, V.name = NULL, K = 4)
```

Arguments

data	data frame to be used.
form.ps	propensity score model.
weight	weighting method to be used. Available methods are "ATE", "ATT", "ATC", "MW", "OVERLAP", and "TRAPEZOIDAL".
V.name	a vector of covariates on which standardized mean difference is computed. If V.name = NULL, the covariates in propensity score model are used.
K	value of K in $\omega(e_i) = \min(1, K \min(e_i, 1 - e_i))$ for "TRAPEZOIDAL" weight.

Value

A list of weighting method, fitted propensity score model, estimated propensity scores, estimated propensity score weights, standardized mean difference before and after weighting adjustment.

weight	weighting method.
ps.model	object returned by fitting the propensity score model using glm with "binomial" family.
ps.hat	estimated propensity score.
W	estimated propensity score weight.
std.diff.before	A data frame of weighed mean, variance, and standardized mean difference for covariates in V.name by treatment groups before weighting. V.name is the row name and "treated.mean", "treated.var", "control.mean", "control.var", "std.diff.pct" are column names.
std.diff.after	A data frame of weighed mean, variance, and standardized mean difference for covariates in V.name by treatment groups after weighting.

See Also

[psw](#), [psw.spec.test](#)

Examples

```
# Load the test data set
data(test_data);
# Propensity score model
form.ps <- "Z ~ X1 + X2 + X3 + X4";
# A vector of covariates
V.name <- c( "X1", "X2", "X3", "X4" );
tmp <- psw.balance( data = test_data, weight = "MW", form.ps = form.ps,
V.name = V.name );
```

psw.mirror.hist *Mirror histogram*

Description

psw.mirror.hist is used to plot the mirror histogram that visualizes the propensity score distributions in both treatment groups.

Usage

```
psw.mirror.hist(data, form.ps, weight, add.weight = FALSE, nclass = 50,
K = 4)
```

Arguments

data	data frame to be used.
form.ps	propensity score model.
weight	weighting method to be used. Available methods are "ATE", "ATT", "ATC", "MW", "OVERLAP", and "TRAPEZOIDAL".
add.weight	add propensity score weights to the mirror histogram, add.weight=FALSE by default and it is not available for weight="ATE", "ATT" or "ATC".
nclass	number of breaks in the mirror histogram.
K	value of K in $\omega(e_i) = \min(1, K \min(e_i, 1 - e_i))$ for "TRAPEZOIDAL" weight.

Details

See psw.

Value

NULL.

See Also

[psw](#)

Examples

```
# Load the test data set
data(test_data);
# Propensity score model
form.ps <- "Z ~ X1 + X2 + X3 + X4";
tmp <- psw.mirror.hist( data = test_data, weight = "MW", form.ps = form.ps,
add.weight = TRUE );
```

psw.spec.test

Propensity score model specification test

Description

psw.spec.test is used to test the sufficiency of propensity score model in balancing covariates between groups.

Usage

```
psw.spec.test(data, form.ps, weight, V.name, trans.type, K = 4)
```

Arguments

data	data frame to be used.
form.ps	propensity score model.
weight	weighting method to be used. Available methods are "ATE", "ATT", "ATC", "MW", and "TRAPEZOIDAL". Note that specification test is not available for weight = "OVERLAP".
V.name	a vector of covariates on which the specification test is performed.
trans.type	a vector of the same length as V.name that specifies the transformation type for each element in V.name. Available transformations are "identity", "log", "logit", "sqrt", and "Fisher". No transformation is performed with "identity".
K	value of K in $\omega(e_i) = \min(1, K \min(e_i, 1 - e_i))$ for "TRAPEZOIDAL" weight. K=4 by default.

Details

In the data set, treatment indicator should be numerically specified such that a value of 1 indicates the treated and a value of 0 indicates the control. The null hypothesis is that the propensity score model is correctly specified; the alternative is that the propensity score model is misspecified. Therefore, this test is a goodness-of-fit test of propensity score model, with the test statistic being a metric of covariate balance. # Rejection of the specification test implies current propensity score model is inadequate for balancing covariates between groups.

Value

A list of model specification test results.

weight	weighting method.
ps.model	object returned by fitting the propensity score model using glm with "binomial" family.
ps.hat	estimated propensity score.
W	estimated propensity score weight.
V.name	covariates in the specification test.
g.B1.hat	a vector of transformed weighted average for covariates in the treated group when spec.test=T.
g.B0.hat	a vector of transformed weighted average for covariates in the control group when spec.test=T.
B.hat	difference between eta.B1.hat and eta.B0.hat when spec.test=T.
var.B.hat	covariance matrix for B.hat when spec.test=T.
test.stat	test statistic for the specification test, which follows the χ_{df}^2 distribution under the null, available when spec.test=T.
df	degree of freedom for the specification test, $df = \text{rank}(\text{var.B.hat})$, available when spec.test=T.
pvalue	pvalue of the specification test when spec.test=T.

See Also

[psw](#), [psw.balance](#)

Examples

```
# Load the test data set
data(test_data);
# Propensity score model
form.ps <- "Z ~ X1 + X2 + X3 + X4";
# A vector of covariates
V.name <- c( "X1", "X2", "X3", "X4" );
# A vector of transformation types for covariates in V.name.
trans.type <- c( "identity", "identity", "logit", "logit" );
tmp <- psw.spec.test( data = test_data, form.ps = form.ps,
weight = "MW", V.name = V.name, trans.type = trans.type );
```


std.rr.wt	standard error for est.rr.wt.
est.or.wt	weighted estimator for odds ratio when family = "binomial".
std.or.wt	standard error for est.or.wt.
est.lor.wt	weighted estimator for log odds ratio when family = "binomial".
std.lor.wt	standard error for est.lor.wt.

See Also

[psw](#)

Examples

```
# Load the test data set
data(test_data);
# Propensity score model
form.ps <- "Z ~ X1 + X2 + X3 + X4";
tmp <- psw.wt( data = test_data, weight = "ATE", form.ps = form.ps,
out.var = "Y", family = "gaussian" );
```

test_data

Test data

Description

A simulated data frame for illustration. In the test data, X_1 and X_2 are continuous variables, X_3 and X_4 are binary variables, Y is the continuous outcome, and Z is the dichotomous treatment indicator.

Usage

```
test_data
```

Index

psw, [2](#), [8](#), [9](#), [11](#), [13](#)

psw.aug, [6](#), [6](#)

psw.balance, [6](#), [8](#), [11](#)

psw.mirror.hist, [6](#), [9](#)

psw.spec.test, [6](#), [8](#), [10](#)

psw.wt, [6](#), [12](#)

test_data, [13](#)