

# Package ‘PSAboot’

December 5, 2016

**Type** Package

**Title** Bootstrapping for Propensity Score Analysis

**Version** 1.1.4

**Date** 2016-02-25

**Author** Jason Bryer <jason@bryer.org>

**Maintainer** Jason Bryer <jason@bryer.org>

**URL** <http://jason.bryer.org/PSAboot>, <http://github.com/jbryer/PSAboot>

**BugReports** <https://github.com/jbryer/PSAboot/issues>

**Description** Bootstrapping for propensity score analysis and matching.

**License** GPL

**LazyLoad** yes

**VignetteBuilder** knitr

**Depends** R (>= 3.0),graphics,ggplot2,PSAgraphics

**Imports**

utils,Matching,MatchIt,modeltools,parallel,party,psych,reshape2,rpart,TriMatch,ggthemes,stats

**Suggests** devtools,knitr,multilevelPSA

**RoxygenNote** 5.0.1

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2016-12-05 18:28:46

## R topics documented:

PSAboot-package . . . . .	2
as.data.frame.PSAbootSummary . . . . .	2
balance . . . . .	3
balance.matching . . . . .	4
boot.ctree . . . . .	4
boot.matching . . . . .	5

boot.matchit . . . . .	5
boot.rpart . . . . .	6
boot.strata . . . . .	6
boxplot.PSABoot . . . . .	7
boxplot.PSABoot.balance . . . . .	8
getPSABootMethods . . . . .	8
hist.PSABoot . . . . .	9
matrixplot . . . . .	9
pisa.psa.cols . . . . .	10
pisalux . . . . .	10
pisausa . . . . .	12
plot.PSABoot . . . . .	14
plot.PSABoot.balance . . . . .	14
print.PSABoot . . . . .	15
print.PSABoot.balance . . . . .	15
print.PSABootSummary . . . . .	16
psa.strata . . . . .	16
PSABoot . . . . .	17
q25 . . . . .	18
q75 . . . . .	19
summary.PSABoot . . . . .	19

**Index** 20

---

PSABoot-package      *Bootstrapping for Propensity Score Analysis*

---

**Description**

Bootstrapping procedures for Propensity Score Analysis.

---

as.data.frame.PSABootSummary

*Convert the results of PSABoot summary to a data frame.*

---

**Description**

Convert the results of PSABoot summary to a data frame.

**Usage**

```
## S3 method for class 'PSABootSummary'
as.data.frame(x, row.names = NULL,
  optional = FALSE, ...)
```

**Arguments**

x	results of <a href="#">summary.PSAbboot</a>
row.names	row names.
optional	unused.
...	unused.

---

balance	<i>Returns a summary of the balance for all bootstrap samples.</i>
---------	--

---

**Description**

This method provides some crude overall measures of balance.

**Usage**

```
balance(psaboot, na.rm = TRUE, pool.fun = mean)
```

**Arguments**

psaboot	results from <a href="#">PSAbboot</a> .
na.rm	should NAs be removed. NAs generally occur when there is insufficient sample for a particular covariate or an unused level.
pool.fun	a function specifying how the effect sizes across all covariates should be combined. Possible values include mean (default), <a href="#">q25</a> , <a href="#">q75</a> , median, max, or any function that takes a vector of numeric values.

**Value**

a list with three elements:

**unadjusted** named numeric vector with unadjusted effect size before adjustment for each covariate

**complete** a matrix with adjusted effect size for each covariate (columns) for each method (rows).

**pooled** a matrix with mean adjusted effect size for all covariates for each method (columns) and each bootstrap sample (rows).

**balances** a list with an M x n covariates matrix for each method.

---

balance.matching	Returns balance for each covariate from propensity score matching.
------------------	--

---

### Description

Returns balance for each covariate from propensity score matching.

### Usage

```
balance.matching(index.treated, index.control, covs)
```

### Arguments

index.treated	a vector with the index of treated rows in covs.
index.control	a vector with the index of control rows in covs.
covs	data frame or matrix of covariates. Factors should already be recoded. See <a href="#">cv.trans.psa</a>

### Value

a named vector with one element per covariate.

---

boot.ctree	Stratification using classification trees for bootstrapping.
------------	--

---

### Description

Stratification using classification trees for bootstrapping.

### Usage

```
boot.ctree(Tr, Y, X, X.trans, formu, minStrata = 5, ...)
```

### Arguments

Tr	vector indicating treatment assignment.
Y	vector of outcome.
X	matrix or data frame of covariates.
X.trans	a data frame of X with factors recoded. See <a href="#">cv.trans.psa</a>
formu	the formula to use to estimate propensity scores. Note that the dependent variable (i.e. treatment variable) name will be updated using the Tr vector.
minStrata	minimum number of treatment or control units within a strata to include that strata.
...	other parameters passed from <a href="#">PSAboot</a>

---

boot.matching	<i>Matching package implementation for bootstrapping.</i>
---------------	---

---

**Description**

Matching package implementation for bootstrapping.

**Usage**

```
boot.matching(Tr, Y, X, X.trans, formu, estimand = "ATE", ...)
```

**Arguments**

Tr	vector indicating treatment assignment.
Y	vector of outcome.
X	matrix or data frame of covariates.
X.trans	a data frame of X with factors recoded. See <a href="#">cv.trans.psa</a>
formu	the formula to use to estimate propensity scores. Note that the dependent variable (i.e. treatment variable) name will be updated using the Tr vector.
estimand	character string for estimand, either ATE, ATT, or ATC. See <a href="#">Match</a> for more details.
...	other parameters passed to <a href="#">Match</a> .

---

boot.matchit	<i>MatchIt package implementation for bootstrapping.</i>
--------------	--

---

**Description**

MatchIt package implementation for bootstrapping.

**Usage**

```
boot.matchit(Tr, Y, X, X.trans, formu, ...)
```

**Arguments**

Tr	vector indicating treatment assignment.
Y	vector of outcome.
X	matrix or data frame of covariates.
X.trans	a data frame of X with factors recoded. See <a href="#">cv.trans.psa</a>
formu	the formula to use to estimate propensity scores. Note that the dependent variable (i.e. treatment variable) name will be updated using the Tr vector.
...	other parameters passed from <a href="#">PSAboot</a>

---

boot.rpart	<i>Stratification using classification trees for bootstrapping.</i>
------------	---

---

**Description**

Stratification using classification trees for bootstrapping.

**Usage**

```
boot.rpart(Tr, Y, X, X.trans, formu, minStrata = 5, ...)
```

**Arguments**

Tr	vector indicating treatment assignment.
Y	vector of outcome.
X	matrix or data frame of covariates.
X.trans	a data frame of X with factors recoded. See <a href="#">cv.trans.psa</a>
formu	the formula to use to estimate propensity scores. Note that the dependent variable (i.e. treatment variable) name will be updated using the Tr vector.
minStrata	minimum number of treatment or control units within a strata to include that strata.
...	other parameters passed from <a href="#">PSAboot</a>

---

boot.strata	<i>Stratification implementation for bootstrapping.</i>
-------------	---

---

**Description**

Stratification implementation for bootstrapping.

**Usage**

```
boot.strata(Tr, Y, X, X.trans, formu, nstrata = 5, ...)
```

**Arguments**

Tr	vector indicating treatment assignment.
Y	vector of outcome.
X	matrix or data frame of covariates.
X.trans	a data frame of X with factors recoded. See <a href="#">cv.trans.psa</a>
formu	the formula to use to estimate propensity scores. Note that the dependent variable (i.e. treatment variable) name will be updated using the Tr vector.
nstrata	number of strata to divide the propensity scores.
...	other parameters passed from <a href="#">PSAboot</a>

**Value**

a list with three elements:

`summary` a named numeric vector (with at minimum `estimate`, `ci.min`, and `ci.max` but other values allowed)

`balance` a named numeric vector with one element per covariate listed in `X.trans` representing a balance statistic (usually standardized effect size after adjustment)

`details` an arbitrary object that contains the full results of the analysis

---

boxplot.PSAboot	<i>Boxplot of PSA bootstrap results.</i>
-----------------	--

---

**Description**

Boxplot of PSA bootstrap results.

**Usage**

```
## S3 method for class 'PSAboot'
boxplot(x, bootstrap.mean.color = "blue",
        bootstrap.ci.color = "green", bootstrap.ci.width = 0.5,
        bootstrap.ci.size = 3, overall.mean.color = "red", tufte = FALSE,
        coord.flip = TRUE, ...)
```

**Arguments**

<code>x</code>	result of <a href="#">PSAboot</a> .
<code>bootstrap.mean.color</code>	the color of the point for the bootstrap mean, or NA to omit.
<code>bootstrap.ci.color</code>	the color of the confidence intervals of the bootstrap samples, or NA to omit.
<code>bootstrap.ci.width</code>	the width of the confidence interval lines at the end.
<code>bootstrap.ci.size</code>	the size of the confidence interval lines.
<code>overall.mean.color</code>	the color of the point for the overall (before bootstrapping) mean, or NA to omit.
<code>tufte</code>	use Tufte's boxplot style. Requires the <code>ggthemes</code> package.
<code>coord.flip</code>	Whether to flip the coordinates.
<code>...</code>	unused

---

```
boxplot.PSAboot.balance
```

*Boxplot of the balance statistics for bootstrapped samples.*

---

### Description

Boxplot of the balance statistics for bootstrapped samples.

### Usage

```
## S3 method for class 'PSAboot.balance'
boxplot(x, unadjusted.color = "red",
        pooled.color = "blue", point.size = 3, point.alpha = 0.5, ...)
```

### Arguments

<code>x</code>	results of <a href="#">balance</a>
<code>unadjusted.color</code>	the color used for the unadjusted effect size.
<code>pooled.color</code>	the color used for the mean bootstrap effect size.
<code>point.size</code>	the size of the points.
<code>point.alpha</code>	the transparency level for the points.
<code>...</code>	other parameters passed to <a href="#">facet_wrap</a>

### Value

a ggplot2 expression.

---

```
getPSAbootMethods
```

*Returns a vector with the default methods used by PSAboot.*

---

### Description

The current default methods are:

**Stratification** [boot.strata](#)

**ctree** [boot.ctree](#)

**rpart** [boot.rpart](#)

**Matching** [boot.matching](#)

**MatchIt** [boot.matchit](#)

### Usage

```
getPSAbootMethods()
```



**Details**

The default methods can be changed by setting the `PSAboot.methods` option using `options('PSAboot.methods'=c(...))` where `...` is a named list of functions.

**Value**

a vector of methods for use by [PSAboot](#)

---

hist.PSAboot	<i>Histogram of PSA bootstrap results</i>
--------------	---

---

**Description**

Histogram of PSA bootstrap results

**Usage**

```
## S3 method for class 'PSAboot'
hist(x, ...)
```

**Arguments**

x	result of <a href="#">PSAboot</a> .
...	other parameters passed to <a href="#">geom_histogram</a>

---

matrixplot	<i>Matrix Plot of Boostrapped Propensity Score Analysis</i>
------------	---

---

**Description**

Matrix Plot of Boostrapped Propensity Score Analysis

**Usage**

```
matrixplot(bm)
```

**Arguments**

bm	result from <a href="#">PSAboot</a> .
----	---------------------------------------

---

pisa.psa.cols	<i>Character vector representing the list of covariates used for estimating propensity scores.</i>
---------------	--

---

**Description**

Character vector representing the list of covariates used for estimating propensity scores.

**Format**

a character vector with covariate names for estimating propensity scores.

---

pisalux	<i>Programme of International Student Assessment (PISA) results from the Luxembourg in 2009.</i>
---------	--

---

**Description**

Student results from the 2009 Programme of International Student Assessment (PISA) as provided by the Organization for Economic Co-operation and Development (OECD). See <http://www.pisa.oecd.org/> for more information including the code book.

**Format**

a data frame with 4,622 rows and 65 columns.

CNT Country

SCHOOLID SchoolID

ST01Q01 Grade

ST04Q01 Sex

ST05Q01 Attend

ST06Q01 Age

ST07Q01 Repeat

ST08Q01 At home mother

ST08Q02 At home father

ST08Q03 At home brothers

ST08Q04 At home sisters

ST08Q05 At home grandparents

ST08Q06 At home others

ST10Q01 Mother highest schooling

ST12Q01 Mother current job status

ST14Q01 Father highest schooling  
ST16Q01 Father current job status  
ST19Q01 Language at home  
ST20Q01 Desk  
ST20Q02 Own room  
ST20Q03 Study place  
ST20Q04 Computer  
ST20Q05 Software  
ST20Q06 Internet  
ST20Q07 Literature  
ST20Q08 Poetry  
ST20Q09 Art  
ST20Q10 Textbooks  
ST20Q12 Dictionary  
ST20Q13 Dishwasher  
ST20Q14 DVD  
ST21Q01 How many cellphones  
ST21Q02 How many TVs  
ST21Q03 How many computers  
ST21Q04 How many cars  
ST21Q05 How many rooms bath or shower  
ST22Q01 How many books  
ST23Q01 Reading enjoyment time  
ST31Q01 Enrich in test language  
ST31Q02 Enrich in mathematics  
ST31Q03 Enrich in science  
ST31Q05 Remedial in test language  
ST31Q06 Remedial in mathematics  
ST31Q07 Remedial in science  
ST32Q01 Out of school lessons in test language  
ST32Q02 Out of school lessons maths  
ST32Q03 Out of school lessons in science  
PUBPRIV Public or private school  
STRATIO Student to teacher ratio in school

**Details**

Note that missing values have been imputed using the <http://cran.r-project.org/web/packages/mice/index.html> package. Details on the specific procedure are in the `pisa.impute` function in the `pisa` package.

## References

Organisation for Economic Co-operation and Development (2009). Programme for International Student Assessment (PISA). <http://www.pisa.oecd.org/>

---

pisausa

*Programme of International Student Assessment (PISA) results from the United States in 2009.*

---

## Description

Student results from the 2009 Programme of International Student Assessment (PISA) as provided by the Organization for Economic Co-operation and Development (OECD). See <http://www.pisa.oecd.org/> for more information including the code book.

## Format

a data frame with 5,233 rows and 65 columns.

CNT Country  
 SCHOOLID SchoolID  
 ST01Q01 Grade  
 ST04Q01 Sex  
 ST05Q01 Attend  
 ST06Q01 Age  
 ST07Q01 Repeat  
 ST08Q01 At home mother  
 ST08Q02 At home father  
 ST08Q03 At home brothers  
 ST08Q04 At home sisters  
 ST08Q05 At home grandparents  
 ST08Q06 At home others  
 ST10Q01 Mother highest schooling  
 ST12Q01 Mother current job status  
 ST14Q01 Father highest schooling  
 ST16Q01 Father current job status  
 ST19Q01 Language at home  
 ST20Q01 Desk  
 ST20Q02 Own room  
 ST20Q03 Study place  
 ST20Q04 Computer

ST20Q05 Software  
ST20Q06 Internet  
ST20Q07 Literature  
ST20Q08 Poetry  
ST20Q09 Art  
ST20Q10 Textbooks  
ST20Q12 Dictionary  
ST20Q13 Dishwasher  
ST20Q14 DVD  
ST21Q01 How many cellphones  
ST21Q02 How many TVs  
ST21Q03 How many computers  
ST21Q04 How many cars  
ST21Q05 How many rooms bath or shower  
ST22Q01 How many books  
ST23Q01 Reading enjoyment time  
ST31Q01 Enrich in test language  
ST31Q02 Enrich in mathematics  
ST31Q03 Enrich in science  
ST31Q05 Remedial in test language  
ST31Q06 Remedial in mathematics  
ST31Q07 Remedial in science  
ST32Q01 Out of school lessons in test language  
ST32Q02 Out of school lessons maths  
ST32Q03 Out of school lessons in science  
PUBPRIV Public or private school  
STRATIO Student to teacher ratio in school

### Details

Note that missing values have been imputed using the <http://cran.r-project.org/web/packages/mice/index.html> package. Details on the specific procedure are in the `pisa.impute` function in the `pisa` package.

### References

Organisation for Economic Co-operation and Development (2009). Programme for International Student Assessment (PISA). <http://www.pisa.oecd.org/>

---

plot.PSABoot                    *Plot the results of PSABoot*

---

### Description

Plot the results of PSABoot

### Usage

```
## S3 method for class 'PSABoot'
plot(x, sort = "all", ci.sig.color = "red",
     plot.overall = FALSE, plot.bootstrap = TRUE, ...)
```

### Arguments

x	result of <a href="#">PSABoot</a> .
sort	how the sort the rows by mean difference. Options are to sort using the mean difference from matching, stratification, both individually, or no sorting.
ci.sig.color	the color used for confidence intervals that do not span zero.
plot.overall	whether to plot vertical lines for the overall (non-bootstrapped) estimate and confidence interval.
plot.bootstrap	whether to plot vertical lines for the bootstrap pooled estimate and confidence interval.
...	currently unused.

---

plot.PSABoot.balance    *Plot method for balance.*

---

### Description

Plot method for balance.

### Usage

```
## S3 method for class 'PSABoot.balance'
plot(x, unadjusted.color = "red",
     complete.color = "blue", pooled.color = "black", ...)
```

**Arguments**

x	results from <a href="#">balance</a>
unadjusted.color	color of the vertical line representing the mean unadjusted effect size for all covariates.
complete.color	color of the vertical line representing the mean adjusted effect size for all covariates using the complete dataset.
pooled.color	color of the vertical line representing the mean adjusted effect size for all covariates across all bootstrapped samples.
...	currently unused.

---

print.PSAboot      *Print results of PSAboot*

---

**Description**

Print results of PSAboot

**Usage**

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

**Arguments**

x	result of <a href="#">PSAboot</a> .
...	currently unused.

---

print.PSAboot.balance      *Print method for balance.*

---

**Description**

This is a crude measure of overall balance. Absolute value of the standardized effect sizes are calculated for each covariate. Overall balance statistics are the mean of those effect sizes after adjustment for each method across all bootstrap samples.

**Usage**

```
## S3 method for class 'PSAboot.balance'
print(x, na.rm = TRUE, ...)
```

**Arguments**

x	results from <a href="#">balance</a> .
na.rm	whether NA balance statistics should be removed before averaging them.
...	currently unused.

---

```
print.PSABootSummary Print method for PSABoot Summary.
```

---

### Description

Print method for PSABoot Summary.

### Usage

```
## S3 method for class 'PSABootSummary'
print(x, digits = 3, ...)
```

### Arguments

x	result of <a href="#">summary.PSABoot</a>
digits	desired number of digits after the decimal point.
...	unused.

---

```
psa.strata Propensity Score Analysis using Stratification
```

---

### Description

Propensity Score Analysis using Stratification

### Usage

```
psa.strata(Y, Tr, strata, trim = 0, minStrata = 5)
```

### Arguments

Y	response variable.
Tr	treatment variable.
strata	strata identifier.
trim	allows for a trimmed mean as outcome measure, where trim is from 0 to .5 (.5 implying median).
minStrata	minimum number of treatment or control units within a strata to include that strata.



**Description**

Bootstrapping has become a popular resampling method for estimating sampling distributions. And propensity score analysis (PSA) has become popular for estimating causal effects in observational studies. This function implements bootstrapping specifically for PSA. Like typical bootstrapping methods, this function estimates treatment effects for  $M$  random samples. However, unlike typical bootstrap methods, this function allows for separate sample sizes for treatment and control units. That is, under certain circumstances (e.g. when the ratio of treatment-to-control units is large) bootstrapping only the control units may be desirable. Additionally, this function provides a framework to use multiple PSA methods for each bootstrap sample.

**Usage**

```
PSAboot(Tr, Y, X, M = 100, formu = as.formula(paste0("treat ~ ",
  paste0(names(X), collapse = " + "))), control.ratio = 5,
  control.sample.size = min(control.ratio * min(table(Tr)), max(table(Tr))),
  control.replace = TRUE, treated.sample.size = min(table(Tr)),
  treated.replace = TRUE, methods = getPSAbootMethods(), parallel = TRUE,
  seed = NULL, ...)
```

**Arguments**

Tr	numeric (0 or 1) or logical vector of treatment indicators.
Y	vector of outcome variable.
X	matrix or data frame of covariates used to estimate the propensity scores.
M	number of bootstrap samples to generate.
formu	formula used for estimating propensity scores. The default is to use all covariates in X.
control.ratio	the ratio of control units to sample relative to the treatment units.
control.sample.size	the size of each bootstrap sample of control units.
control.replace	whether to use replacement when sampling from control units.
treated.sample.size	the size of each bootstrap sample of treatment units. The default uses all treatment units for each bootstrap sample.
treated.replace	whether to use replacement when sampling from treated units.
methods	a named vector of functions for each PSA method to use.
parallel	whether to run the bootstrap samples in parallel.
seed	random seed. Each iteration, $i$ , will use a seed of $seed + i$ .
...	other parameters passed to <a href="#">Match</a> and <a href="#">psa.strata</a>

**Value**

a list with following elements:

**overall.summary** Data frame with the results using the complete dataset (i.e. unbootstrapped results).

**overall.details** Objects returned from each method for complete dataset.

**pooled.summary** Data frame with results of each bootstrap sample.

**pooled.details** List of objects returned from each method for each bootstrap sample.

**control.sample.size** sample size used for control units.

**treated.sample.size** sample size used for treated units.

**control.replace** whether control units were sampled with replacement.

**treated.replace** whether treated units were sampled with replacement.

**Tr** vector of treatment assignment.

**Y** vector out outcome.

**X** matrix or data frame of covariates.

**M** number of bootstrap samples.

**See Also**

`getPSAbootMethods`

---

q25

*Return the 25th percentile.*

---

**Description**

Return the 25th percentile.

**Usage**

```
q25(x, na.rm = FALSE, ...)
```

**Arguments**

<code>x</code>	numeric vector.
<code>na.rm</code>	logical; if true, any NA and NaN's are removed from <code>x</code> before the quantiles are computed
<code>...</code>	other parameters passed to <a href="#">quantile</a> .

**Value**

the 25th percetile.

---

q75	<i>Returns the 75th percentile.</i>
-----	-------------------------------------

---

**Description**

Returns the 75th percentile.

**Usage**

```
q75(x, na.rm = FALSE, ...)
```

**Arguments**

x	numeric vector.
na.rm	logical; if true, any NA and NaN's are removed from x before the quantiles are computed
...	other parameters passed to <a href="#">quantile</a> .

**Value**

the 75th percentile.

---

summary.PSAbboot	<i>Summary of pooled results from PSAbboot</i>
------------------	--

---

**Description**

Summary of pooled results from PSAbboot

**Usage**

```
## S3 method for class 'PSAbboot'  
summary(object, ...)
```

**Arguments**

object	result of <a href="#">PSAbboot</a> .
...	currently unused.

**Value**

a list with pooled summary statistics.

# Index

## \*Topic **datasets**

- `pisa.psa.cols`, 10
- `as.data.frame.PSABootSummary`, 2
- `balance`, 3, 8, 15
- `balance.matching`, 4
- `boot.ctree`, 4, 8
- `boot.matching`, 5, 8
- `boot.matchit`, 5, 8
- `boot.rpart`, 6, 8
- `boot.strata`, 6, 8
- `boxplot.PSABoot`, 7
- `boxplot.PSABoot.balance`, 8
- `cv.trans.psa`, 4–6
- `facet_wrap`, 8
- `geom_histogram`, 9
- `getPSABootMethods`, 8
- `hist.PSABoot`, 9
- `Match`, 5, 17
- `matrixplot`, 9
- `pisa.psa.cols`, 10
- `pisalux`, 10
- `pisausa`, 12
- `plot.PSABoot`, 14
- `plot.PSABoot.balance`, 14
- `print.PSABoot`, 15
- `print.PSABoot.balance`, 15
- `print.PSABootSummary`, 16
- `psa.strata`, 16, 17
- `PSABoot`, 3–9, 14, 15, 17, 19
- `PSABoot-package`, 2
- `q25`, 3, 18
- `q75`, 3, 19
- `quantile`, 18, 19
- `summary.PSABoot`, 3, 16, 19