Package ‘twang’

July 2, 2017

Version 1.5
Date 2017-06-30
Title Toolkit for Weighting and Analysis of Nonequivalent Groups
Author Greg Ridgeway, Dan McCaffrey, Andrew Morral, Beth Ann Griffin, Lane Burgette
<burgette@rand.org>
Maintainer Lane Burgette <burgette@rand.org>
Depends R (>= 2.10), gbm (>= 1.5-3), survey, xtable, lattice, latticeExtra
Description Provides functions for propensity score estimating and weighting, nonresponse weighting, and diagnosis of the weights.
License GPL (>= 2)
NeedsCompilation yes
Repository CRAN
Repository/R-Forge/Project twang
Repository/R-Forge/Revision 75
Repository/R-Forge/DateTimeStamp 2017-07-01 01:59:38
Date/Publication 2017-07-02 00:01:17 UTC

R topics documented:

AOD ................................................................. 2
bal.stat ......................................................... 3
bal.table ......................................................... 4
boxplot.mnps .................................................. 5
boxplot.ps ..................................................... 6
desc.wts ....................................................... 7
dx.wts ......................................................... 9
degsingle ..................................................... 10
get.weights ................................................... 11
get.weights.num ............................................ 12
Subset of Alcohol and Other Drug treatment data

Description

A small subset of the data from McCaffrey et al. (2013).

Usage

data(AOD)

Format

A data frame with 600 observations on the following 10 variables.

treat Treatment that each study subject received. Either community, metcbt5, or scy.
suf12 outcome variable, substance use frequency at 12 month follow-up
illact covariate, illicit activities scale
crimjust covariate, criminal justice involvement
subprob covariate, substance use problem scale
subdep covariate, substance use dependence scale
white 1 if non-Hispanic white, 0 otherwise
References


bal.stat

Calculate weighted balance statistics

Description

bal.stat compares the treatment and control subjects by means, standard deviations, effect size, and KS statistics

Usage

bal.stat(data,
vars = NULL,
treat.var,
w.all,
sampw,
get.means = TRUE,
get.ks = TRUE,
na.action = "level",
estimand,
multinom, fillNAs = FALSE)

Arguments

data a data frame containing the data
vars a vector of character strings with the names of the variables on which the function will assess the balance
treat.var the name of the treatment variable
w.all observation weights (e.g. propensity score weights, sampling weights, or both)
sampw sampling weights. These are passed in addition to w.all because the "un-weighted" results should be adjusted for sample weights (though not propensity score weights).
get.means logical. If TRUE then bal.stat will compute means and variances
get.ks logical. If TRUE then bal.stat will compute KS statistics
na.action a character string indicating how bal.stat should handle missing values. Current options are "level", "exclude", or "lowest"
estimand either "ATT" or "ATE"
multinom TRUE if used for multinomial propensity scores.
fillNAs If TRUE, fills in zeros for missing values.
Details

`bal.table` calls auxiliary functions for each variable and assembles the results in a table.

Value

`getMeans` and `get.KS` manipulate the inclusion of certain columns in the returned result.

See Also

The example for `ps` contains an example of the use of `bal.table`.

## Description

Extract the balance table from `ps`, `dx.wts`, and `mnps` objects.

## Usage

```r
bal.table(x, digits = 3, collapse.to = c("pair","covariate","stop.method")[1],
          subset.var = NULL, subset.treat = NULL, subset.stop.method = NULL,
          es.cutoff = 0, ks.cutoff = 0, p.cutoff = 1, ks.p.cutoff = 1,
          timeperiods = NULL, ...)  
```

## Arguments

- **x**: a `ps` or `dx.wts` object.
- **digits**: The number of digits that the numerical entries should be rounded to.
- **collapse.to**: For `mnps` ATE objects, the comparisons can be given for all pairs (default), summarized by pre-treatment covariate and stop.method, or as a single summary for each stop.method.
- **subset.var**: Eliminate all but a specified subset of covariates.
- **subset.treat**: Subset to either all pairs that include a specified treatment or a single pair of treatments.
- **subset.stop.method**: Subset to a subset of stop.method\'s used to fit the ps object.
- **es.cutoff**: Subsets to comparisons with absolute ES values bigger than es.cutoff.
- **ks.cutoff**: Subsets to comparisons with KS values bigger than ks.cutoff.
- **p.cutoff**: Subsets to comparisons with t- or chi-squared p-values no bigger than p.cutoff.
- **ks.p.cutoff**: Subsets to comparisons with KS p-values no bigger than ks.p.cutoff.
- **timeperiods**: Used to subset times for iptw fits.
- **...**: Additional arugments.
Details

`bal.table` is a generic function for extracting balance tables from `ps` and `dx.wts` objects. These objects usually have several sets of candidate weights, one for an unweighted analysis and perhaps several stop.methods. `bal.table` will return a table for each set of weights combined into a list. Each list component will be named as given in the x, usually the name of the stop.method. The balance table labeled “unw” indicates the unweighted analysis.

Value

Returns a data frame containing the balance information.

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>tx.mn</td>
<td>The mean of the treatment group</td>
</tr>
<tr>
<td>tx.sd</td>
<td>The standard deviation of the treatment group</td>
</tr>
<tr>
<td>ct.mn</td>
<td>The mean of the control group</td>
</tr>
<tr>
<td>ct.sd</td>
<td>The standard deviation of the control group</td>
</tr>
<tr>
<td>std.eff.sz</td>
<td>The standardized effect size, ((\text{tx.mn-ct.mn})/\text{tx.sd}). If \text{tx.sd} is small or 0, the standardized effect size can be large or INF. Therefore standardized effect sizes greater than 500 are set to NA</td>
</tr>
<tr>
<td>stat</td>
<td>the t-statistic for numeric variables and the chi-square statistic for continuous variables</td>
</tr>
<tr>
<td>p</td>
<td>the p-value for the test associated with stat</td>
</tr>
<tr>
<td>ks</td>
<td>the KS statistic</td>
</tr>
<tr>
<td>ks.pval</td>
<td>the KS p-value computed using the analytic approximation, which does not necessarily work well with a lot of ties</td>
</tr>
</tbody>
</table>

See Also

The example for `ps` contains an example of the use of `bal.table`

---

**Description**

This function produces a collection of diagnostic plots for mnps objects.

**Usage**

```r
## S3 method for class 'mnps'
boxplot(x, stop.method = NULL, color = TRUE, figureRows = NULL,
        singlePlot = NULL, multiPage = FALSE, time = NULL, print = TRUE, ...)
```
Arguments

- **x**: A `ps` object
- **stop.method**: Only 1 `stop.method` can be presented at a time for `mnps` objects. Use a numeric indicator of which `stop.method` (among those specified when fitting the `mnps` object) should be used.
- **color**: If FALSE, a grayscale figure will be returned.
- **figureRows**: The number of rows in the figure. Defaults to the number of panels.
- **singlePlot**: If multiple sets of boxplots are produced, `singlePlot` can be used to select only one. For example, `singlePlot = 2` would return only the second set of boxplots.
- **multiPage**: When multiple frames of a figure are produced, `multiPage = TRUE` will print each frame on a different page. This is intended for situations where the graphical output is being saved to a file.
- **time**: For use with `iptw` fits.
- **print**: If FALSE, the figure is returned but not printed.
- **...**: Additional arguments that may be passed to the underlying `lattice` package plotting functions

Details

This function produces lattice-style graphics of diagnostic plots.

References


See Also

- `ps`

---

**boxplot.ps**  
*Boxplots for ps objects*

Description

This function produces a collection of diagnostic plots for `ps` objects.

Usage

```r
## S3 method for class 'ps'
boxplot(x, subset=NULL, color = TRUE, time = NULL, ...)
```
**Arguments**

- **x** A ps object
- **subset** If multiple `stop.method` rules were used in the `ps()` call, subset restricts the plots of a subset of the stopping rules that were employed. This argument expects a subset of the integers from 1 to `k`, if `k stop.method` rules were used.
- **color** If set to FALSE, grayscale figures will be produced
- **time** Used to specify a subset of times for use with the `iptw` function. Ignored for standard `ps` fits.
- **...** Additional arguments that may be passed to the underlying `lattice` package plotting functions

**Details**

This function produces lattice-style graphics of diagnostic plots.

**References**


**See Also**

- **ps**

---

**desc.wts**

*Diagnosis of weights*

**Description**

desc.wts assesses the quality of a set of weights on balancing a treatment and control group.

**Usage**

desc.wts(data, 
    w, 
    sampw = sampw, 
    vars = NULL, 
    treat.var, 
    tp, 
    na.action = "level", 
    perm.test.iters=0, 
    verbose=TRUE, 
    alerts.stack, 
    estimand, multinom = FALSE, fillNAs = FALSE)
Arguments

data a data frame containing the dataset
w a vector of weights equal to nrow(data)
sampw sampling weights, if provided
vars a vector of variable names corresponding to data
treat.var the name of the treatment variable
tp a title for the method “type” used to create the weights, used to label the results
na.action a string indicating the method for handling missing data
perm.test.iters an non-negative integer giving the number of iterations of the permutation test for the KS statistic. If perm.test.iters=0 then the function returns an analytic approximation to the p-value. This argument is ignored is x is a ps object. Setting perm.test.iters=200 will yield precision to within 3% if the true p-value is 0.05. Use perm.test.iters=500 to be within 2%
verbose if TRUE, lots of information will be printed to monitor the the progress of the fitting
alerts.stack an object for collecting warnings issued during the analyses
estimand the estimand of interest: either "ATT" or "ATE"
multinom Indicator that weights are from a propensity score analysis with 3 or more treatment groups.
fillNAs If TRUE fills NAs with zeros.

Details
desc.wts calls bal.stat to assess covariate balance. If perm.test.iters>0 it will call bal.stat multiple times to compute Monte Carlo p-values for the KS statistics and the maximum KS statistic. It assembles the results into a list object, which usually becomes the desc component of ps objects that ps returns.

Value

See the description of the desc component of the ps object that ps returns

See Also

ps
Description

donx.wts takes a ps object or a set of propensity scores and computes diagnostics assessing covariates balance.

Usage

dx.wts(x,
  data,
estimand,
  vars=NULL,
treat.var,
  x.as.weights=TRUE,
sampw=NULL,
  perm.test.iters=0)

Arguments

x a data frame, matrix, or vector of propensity score weights or a ps object. x can also be a data frame, matrix, or vector of propensity scores if x.as.weights=FALSE
data a data frame
estimand the estimand of interest: either "ATT" or "ATE"
vars a vector of character strings naming variables in data on which to assess balance	
treat.var a character string indicating which variable in data contains the 0/1 treatment group indicator
x.as.weights TRUE or FALSE indicating whether x specifies propensity score weights or propensity scores. Ignored if x is a ps object
sampw optional sampling weights. If x is a ps object then the sampling weights should have been passed to ps and not specified here. dx.wts will issue a warning if x is a ps object and sampw is also specified
perm.test.iters an non-negative integer giving the number of iterations of the permutation test for the KS statistic. If perm.test.iters=0 then the function returns an analytic approximation to the p-value. This argument is ignored is x is a ps object. Setting perm.test.iters=200 will yield precision to within 3% if the true p-value is 0.05. Use perm.test.iters=500 to be within 2%

Details

Creates a balance table that compares unweighted and weighted means and standard deviations, computes effect sizes, and KS statistics to assess the ability of the propensity scores to balance the treatment and control groups.
Value

Returns a list containing

- `treat` the vector of 0/1 treatment assignment indicators
- `desc` a nested list containing detailed diagnostic information on the weights. This includes the number of treatment and control subjects, the effective sample size, the largest KS statistic, the average absolute effect size, and the complete balance table
- `summary.tab` a data frame showing balance information
- `ps` the given propensity scores
- `w` the given weights
- `datestamp` the date and time of the call to `dx.wts`
- `parameters` the parameters used when calling `dx.wts`
- `alerts` text containing any warnings accumulated during the estimation
- `varNames` the variable names

See Also

The example for `ps` contains an example of the use of `dx.wts`

---

egsingle

US Sustaining Effects study

Description

A subset of the mathematics scores from the U.S. Sustaining Effects Study. The subset consists of information on 1721 students from 60 schools. This dataset is available in the `mlmrev` package.

Usage

data(egsingle)

Format

A data frame with 7230 observations on the following 12 variables.

- `schoolid` a factor of school identifiers
- `childid` a factor of student identifiers
- `year` a numeric vector indicating the year of the test
- `grade` a numeric vector indicating the student's grade
- `math` a numeric vector of test scores on the IRT scale score metric
- `retained` a factor with levels 0 1 indicating if the student has been retained in a grade.
- `female` a factor with levels Female Male
black  a factor with levels 0/1 indicating if the student is Black
hispanic a factor with levels 0/1 indicating if the student is Hispanic
size  a numeric vector indicating the number of students enrolled in the school
lowinc a numeric vector giving the percentage of low-income students in the school
mobility a numeric vector

Source
Reproduced from the mlmRev package for use in the section on nonresponse weighting in the twang package vignette. These data are distributed with the HLM software package (Bryk, Raudenbush, and Congdon, 1996). Conversion to the R format is described in Doran and Lockwood (2006).

References

get.weights

Extract propensity score weights

Description
Extracts propensity score weights from a ps or mnps object.

Usage
get.weights(ps1,
stop.method = NULL,
estimand = NULL,
withSampW = TRUE)

Arguments
ps1             a ps or mnps object
stop.method     indicates which set of weights to retrieve from the ps object
estimand        indicates whether the weights are for the average treatment effect on the treated (ATT) or the average treatment effect on the population (ATE). By default, get.weights will use the estimand used to fit the ps object.
withSampW       Returns weights with sample weights multiplied in, if they were provided in the original ps or mnps call.

Details
Weights for ATT are 1 for the treatment cases and p/(1-p) for the control cases.
Weights for ATE are 1/p for the treatment cases and 1/(1-p) for the control cases.
get.weights.num

Value

A vector of weights

See Also

ps

get.weights.num

Get numerators to stabilize propensity score weights for iptw fits.

Description

Forms numerators to stabilize weights for an iptw object.

Usage

get.weights.num(iptw,
fitlist)

Arguments

iptw An iptw object.

fitList A list containing objects with an associated "fitted" function.

Details

Returns numerator of stabilized weights to be used in conjunction with get.weights.unstab

Value

A vector of stabilizing factors for weights.

See Also

iptw
get.weights.unstab

Extract unstabilized propensity score weights for iptw fits.

Description

Extracts propensity score weights from an iptw or mniptw object.

Usage

get.weights.unstab(x, stop.method = NULL, withSampW = TRUE)

Arguments

x A iptw or mniptw object.
stop.method The stop method used for the fit of interest.
withSampW Returns weights with sample weights multiplied in, if they were provided in the original iptw call.

Details

Weights are the reciprocal of the product of the probability of receiving the treatment received.

Value

a data.frame of weights

See Also

iptw

iptw

Inverse probability of treatment weighting for marginal structural models

Description

iptw uses gbm to estimate propensity scores for sequential treatments.
Usage

iptw(formula, data, timeInVariant = NULL, n.trees = 10000, stop.method = "es.max", cumulative = TRUE, timeIndicators = NULL, ID = NULL, priorTreatment = TRUE, ...)

Arguments

formula Either a single formula (long format) or a list with formulas
data The dataset, includes treatment assignment as well as covariates
timeInvariant An optional formula (with no left-hand variable) specifying time-invariant characteristics.
n.trees number of gbm iterations passed on to gbm
stop.method A method or methods of measuring and summarizing balance across pretreatment variables. Current options are ks.mean, ks.max, es.mean, and es.max. ks refers to the Kolmogorov-Smirnov statistic and es refers to standardized effect size. These are summarized across the pretreatment variables by either the maximum (.max) or the mean (.mean).
cumulative If TRUE, the time t model includes time-varying characteristics from times 1 through t-1.
timeIndicators For long format fits, a vector of times for each observation.
ID For long format fits, a vector of numeric identifiers for unique analytic units.
priorTreatment For long format fits, includes treatment levels from previous times if TRUE. This argument is ignored for wide format fits.
... Additional arguments that are passed to ps function.

Details

This function uses generalized boosted models to estimate inverse probability of treatment weights for sequential treatments.

Value

Returns an object of class iptw, a list containing

psList A list of ps objects with length equal to the number of time periods.
estimand The specified estimand.
stop.methods The stopping rules used to optimize iptw balance.
nFits The number of ps objects (i.e., the number of distinct time points.)
uniqueTimes The unique times in the specified model.
iptwExLong

See Also

ps

iptwExLong

Example data for iptw function (long version)

Description

These data are simulated to demonstrate the `iptw` function in the "long" data format.

Usage

data(lindner)

Format

A list with a covariate matrix and outcomes.

covariates  Time-invariant covariates are gender and age. The time-varying covariate is use. The treatment indicator is given by tx. An individual level identifier is given in ID, and the time period is time.

outcome  Vector of post-treatment outcomes.

iptwExWide

Example data for iptw function (wide version)

Description

These data are simulated to demonstrate the `iptw` function in the "wide" data format.

Usage

data(lindner)

Format

A list with a covariate matrix and outcomes.

gender  Gender.
age  Age.
use0  Baseline substance use.
use1  Use following first time period treatment.
use2  Use following second time period treatment.
tx1  Treatment indicator (first time period).
tx2  Treatment indicator (second time period).
tx3  Treatment indicator (third time period).
covariates  Time-invariant covariates are gender and age. The time-varying covariate is use. The treatment indicator is given by tx. An individual level identifier is given in ID, and the time period is time.
outcome  Post-treatment outcomes.

lalonde  Lalonde’s National Supported Work Demonstration data

Description
One of the datasets used by Dehejia and Wahba in their paper ”Causal Effects in Non-Experimental Studies: Reevaluating the Evaluation of Training Programs.” Also used as an example dataset in the MatchIt package.

Usage
data(lalonde)

Format
A data frame with 614 observations on the following 10 variables.
treat  1 if treated in the National Supported Work Demonstration, 0 if from the Current Population Survey
age  age
educ  years of education
black  1 if black, 0 otherwise
hispan  1 if Hispanic, 0 otherwise
married  1 if married, 0 otherwise
nodegree  1 if no degree, 0 otherwise
reW74  earnings in 1974 (pretreatment)
reW75  earnings in 1975 (pretreatment)
reW78  earnings in 1978 (outcome)

Source

References
Description

These data are adapted from the lindner dataset in the USPS package. The description comes from that package, except for the variable sixMonthSurvive, which is a recode of lifepres

Data from an observational study of 996 patients receiving an initial Percutaneous Coronary Intervention (PCI) at Ohio Heart Health, Christ Hospital, Cincinnati in 1997 and followed for at least 6 months by the staff of the Lindner Center. The patients thought to be more severely diseased were assigned to treatment with abciximab (an expensive, high-molecular-weight IIb/IIIa cascade blocker); in fact, only 298 (29.9 percent) of patients received usual-care-alone with their initial PCI.

Usage

data(lindner)

Format

A data frame of 10 variables collected on 996 patients; no NAs.

- **lifepres**: Mean life years preserved due to survival for at least 6 months following PCI; numeric value of either 11.4 or 0.
- **cardbill**: Cardiac related costs incurred within 6 months of patient’s initial PCI; numeric value in 1998 dollars; costs were truncated by death for the 26 patients with lifepres == 0.
- **abcix**: Numeric treatment selection indicator; 0 implies usual PCI care alone; 1 implies usual PCI care deliberately augmented by either planned or rescue treatment with abciximab.
- **stent**: Coronary stent deployment; numeric, with 1 meaning YES and 0 meaning NO.
- **height**: Height in centimeters; numeric integer from 108 to 196.
- **female**: Female gender; numeric, with 1 meaning YES and 0 meaning NO.
- **diabetic**: Diabetes mellitus diagnosis; numeric, with 1 meaning YES and 0 meaning NO.
- **acutemi**: Acute myocardial infarction within the previous 7 days; numeric, with 1 meaning YES and 0 meaning NO.
- **ejectfrac**: Left ejection fraction; numeric value from 0 percent to 90 percent.
- **ves1proc**: Number of vessels involved in the patient’s initial PCI procedure; numeric integer from 0 to 5.
- **sixMonthSurvive**: Survival at six months — a recoded version of lifepres.

References


Extract table of means from an mnps object

Description
Extracts table of means from an mnps object.

Usage
means.table(mnps, stop.method = 1, includeSD = FALSE, digits = NULL)

Arguments
- **mnps**: An mnps object.
- **stop.method**: Indicates which set of weights to retrieve from the ps object. Either the name of the stop.method used, or a natural number with 1, for example, indicating the first stop.method specified.
- **includeSD**: Indicates whether standard deviations as well as means are to be displayed. By default, they are not displayed.
- **digits**: If not NULL, results will be rounded to the specified number of digits.

Details
Displays a table with weighted and unweighted means and standardized effect sizes, and – if requested – standard deviations.

Value
A table of means, standardized effect sizes, and perhaps standard deviations, by treatment group.

See Also
mnps
Example data for iptw function (long version, more than two treatments).

Description
These data are simulated to demonstrate the iptw function in the "long" data format.

Usage
data(lindner)

Format
A list with a covariate matrix and outcomes.

covariates Time-invariant covariates are gender and age. The time-varying covariate is use. The treatment indicator is given by tx. An individual level identifier is given in ID, and the time period is time.
outcome Vector of post-treatment outcomes.

Example data for iptw function (wide version, more than two treatments)

Description
These data are simulated to demonstrate the iptw function in the "wide" data format.

Usage
data(lindner)

Format
A list with a covariate matrix and outcomes.

gender Gender.
age Age.
use0 Baseline substance use.
use1 Use following first time period treatment.
use2 Use following second time period treatment.
tx1 Treatment indicator (first time period).
tx2 Treatment indicator (second time period).
tx3 Treatment indicator (third time period).

covariates Time-invariant covariates are gender and age. The time-varying covariate is use. The treatment indicator is given by tx. An individual level identifier is given in ID, and the time period is time.

outcome Post-treatment outcomes.

---

**mnps**

*Propensity score estimation*

---

**Description**

*mnps* calculates propensity scores and diagnoses them using a variety of methods, but centered on using boosted logistic regression as implemented in *gbm*.

**Usage**

```r
mnps(formula = formula(data),
     data,
     n.trees = 10000,
     interaction.depth = 3,
     shrinkage = 0.01,
     bag.fraction = 1.0,
     perm.test.iters=0,
     print.level = 2,
     iterlim = 1000,
     verbose = TRUE,
     estimand = "ATE",
     stop.method = "es.max",
     sampw = NULL,
     treatATT = NULL, ...)
```

**Arguments**

- **formula** A formula for the propensity score model with the treatment indicator on the left side of the formula and the potential confounding variables on the right side.
- **data** The dataset, includes treatment assignment as well as covariates
- **n.trees** number of gbm iterations passed on to *gbm*
- **interaction.depth** interaction.depth passed on to *gbm*
- **shrinkage** shrinkage passed on to *gbm*
- **bag.fraction** bag.fraction passed on to *gbm*
perm.test.iters

a non-negative integer giving the number of iterations of the permutation test for the KS statistic. If perm.test.iters=0 then the function returns an analytic approximation to the p-value. Setting perm.test.iters=200 will yield precision to within 3% if the true p-value is 0.05. Use perm.test.iters=500 to be within 2%

print.level

the amount of detail to print to the screen

iterlim

maximum number of iterations for the direct optimization

verbose

if TRUE, lots of information will be printed to monitor the the progress of the fitting

estimand

The causal effect of interest. Options are "ATE" (average treatment effect), which attempts to estimate the change in the outcome if the treatment were applied to the entire population versus if the control were applied to the entire population, or "ATT" (average treatment effect on the treated) which attempts to estimate the analogous effect, averaging only over the treated population.

stop.method

A method or methods of measuring and summarizing balance across pretreatment variables. Current options are ks.mean, ks.max, es.mean, and es.max. ks refers to the Kolmogorov-Smirnov statistic and es refers to standardized effect size. These are summarized across the pretreatment variables by either the maximum (.max) or the mean (.mean).

sampw

Optional sampling weights.

treatATT

If the estimand is specified to be ATT, this argument is used to specify which treatment condition is considered 'the treated'. It must be one of the levels of the treatment variable. It is ignored for ATE analyses.

... Additional arguments.

Details

formula should be something like "treatment ~ X1 + X2 + X3". The treatment variable should be a variable with three or more levels. There is no need to specify interaction terms in the formula. interaction.depth controls the level of interactions to allow in the propensity score model.

Note that — unlike earlier versions of twang — plotting functions are no longer included in the ps() function. See plot for details of the plots.

Value

Returns an object of class mnp, which consists of the following.

psList

A list of ps objects.

nFits

The number of calls to ps that were used to form the mnp object.

estimand

The estimand — either ATT or ATE — that was specified in the call to mnp.

treatATT

For ATT fits, the treatment category that is considered "the treated"

treatLev

The levels of the treatment variable.

levExceptTreatAtt

The levels of the treatment variable, excluding the treatATT level.
data The data used to fit the model.
treatVar The vector of treatment indicators
stopMethods The stop.method vector specified in the call to mnps.
sampw Sampling weights provided to mnps, if any.

Author(s)
Lane Burgette <burgette@rand.org>, Beth Ann Griffin <bethg@rand.org>, Dan McCaffrey <danielm@rand.org>

References

See Also
ps

plot.mnps Plots for mnps objects

Description
This function produces a collection of diagnostic plots for ps objects.

Usage

## S3 method for class 'mnps'
plot(x, plots = "optimize", pairwiseMax = TRUE, figureRows = NULL, color = TRUE, subset = NULL, treatments = NULL, singlePlot = NULL, multipage = FALSE, time = NULL, print = TRUE, ...)  

Arguments

x An mnps object.
plots An indicator of which type of plot is desired. The options are
"optimize" or 1 A plot of the balance criteria as a function of the GBM iteration
"boxplot" or 2 Boxplots of the propensity scores for the treatment and control cases
"es" or 3 Plots of the standardized effect size of the pre-treatment variables before and after reweighing
"t" or 4 Plots of the p-values from t-statistics comparing means of treated and control subjects for pretreatment variables, before and after weighting.
"ks" or 5 Plots of the p-values from Kolmogorov-Smirnov statistics comparing distributions of pretreatment variables of treated and control subjects, before and after weighting.

pairwiseMax If FALSE, the plots for the underlying ps fits will be returned. Otherwise, pairwise maxima will be returned.

figureRows The number of rows of figures that should be used. If left as NULL, twang tries to find a reasonable value.

color If color = FALSE figures will be gray scale.

subset Used to restrict which of the stop.methods will be used in the figure. For example subset = c(1, 3) would indicate that the first and third stop.methods (in alphabetical order of those specified in the original call to mnp) should be included in the figure.

treatments Only applicable when pairwiseMax is FALSE and plots 3, 4, and 5. If left at NULL, panels for all treatment pairs are created. If one level of the treatment variable is specified, plots comparing that treatment to all others are produced. If two levels are specified, a comparison for that single pair is produced.

singlePlot For plot calls that produce multiple plots, specifying an integer value of singlePlot will return only the corresponding plot. E.g., specifying singlePlot = 2 will return the second plot.

multiPage When multiple frames of a figure are produced, multiPage = TRUE will print each frame on a different page. This is intended for situations where the graphical output is being saved to a file.

time For use with iptw.

print If FALSE, the figure is returned but not printed.

... Additional arguments.

Details
This function produces lattice-style graphics of diagnostic plots.

References

See Also
mnp
plot.ps  

Plots for ps objects

Description

This function produces a collection of diagnostic plots for ps objects.

Usage

## S3 method for class 'ps'
plot(x, plots = "optimize", subset=NA, color = TRUE, ...)

Arguments

x  
A ps object

plots  
An indicator of which type of plot is desired. The options are

"optimize" or 1 A plot of the balance criteria as a function of the GBM iteration

"boxplot" or 2 Boxplots of the propensity scores for the treatment and control cases

"es" or 3 Plots of the standardized effect size of the pre-treatment variables before and after reweighing

"t" or 4 Plots of the p-values from t-statistics comparing means of treated and control subjects for pretreatment variables, before and after weighting.

"ks" or 5 Plots of the p-values from Kolmogorov-Smirnov statistics comparing distributions of pretreatment variables of treated and control subjects, before and after weighting.

"histogram" or 6 Histogram of weights for treated and control subjects.

subset  
If multiple stop.method rules were used in the ps() call, subset restricts the plots of a subset of the stopping rules that were employed. This argument expects a subset of the integers from 1 to k, if k stop.methods were used.

color  
If set to FALSE, grayscale figures will be produced

...  
Additional arguments that may be passed to the underlying lattice package plotting functions

Details

This function produces lattice-style graphics of diagnostic plots.

References

See Also

ps

print.dxwts

Print a diagnosis of the weights

Description

Prints a diagnosis of the weights. Extracts summary.tab from the dx.wts object

Usage

## S3 method for class 'dxwts'
p(x, ...)

Arguments

x          a dx.wts object
...

... further arguments passed to or from other methods

Value

See ps for a description of the components of the table

ps

Propensity score estimation

Description

ps calculates propensity scores and diagnoses them using a variety of methods, but centered on using boosted logistic regression as implemented in gbm

Usage

ps(formula = formula(data),
data,
n.trees = 10000,
interaction.depth = 3,
shrinkage = 0.01,
bag.fraction = 1.0,
perm.test.iters=0,
print.level = 2,
iterlim = 1000,
verbose = TRUE,
estimand = "ATE",
stop.method = c("ks.mean", "es.mean"),
sampw = NULL,
multinom = FALSE, ...)

Arguments

formula A formula for the propensity score model with the treatment indicator on the left side of the formula and the potential confounding variables on the right side.
data The dataset, includes treatment assignment as well as covariates
n.trees number of gbm iterations passed on to gbm
interaction.depth interaction.depth passed on to gbm
shrinkage shrinkage passed on to gbm
bag.fraction bag.fraction passed on to gbm
perm.test.iters a non-negative integer giving the number of iterations of the permutation test for the KS statistic. If perm.test.iters=0 then the function returns an analytic approximation to the p-value. Setting perm.test.iters=200 will yield precision to within 3% if the true p-value is 0.05. Use perm.test.iters=500 to be within 2%
print.level the amount of detail to print to the screen
iterlim maximum number of iterations for the direct optimization
verbose if TRUE, lots of information will be printed to monitor the progress of the fitting
estimand The causal effect of interest. Options are "ATE" (average treatment effect), which attempts to estimate the change in the outcome if the treatment were applied to the entire population versus if the control were applied to the entire population, or "ATT" (average treatment effect on the treated) which attempts to estimate the analogous effect, averaging only over the treated population.
stop.method A method or methods of measuring and summarizing balance across pretreatment variables. Current options are ks.mean, ks.max, es.mean, and es.max. ks refers to the Kolmogorov-Smirnov statistic and es refers to standardized effect size. These are summarized across the pretreatment variables by either the maximum (.max) or the mean (.mean).
sampw Optional sampling weights.
multinom Set to true only when called from mnps function.
... Additional arguments.

Details

formula should be something like "treatment ~ X1 + X2 + X3". The treatment variable should be a 0/1 indicator. There is no need to specify interaction terms in the formula. interaction.depth controls the level of interactions to allow in the propensity score model.

Note that — unlike earlier versions of twang — plotting functions are no longer included in the ps() function. See plot for details of the plots.
Value

Returns an object of class ps, a list containing

- **gbm.obj**: The returned gbm object
- **treat**: The treatment variable.
- **desc**: A list containing balance tables for each method selected in stop.methods. Includes a component for the unweighted analysis names “unw”. Each desc component includes a list with the following components:
  - **ess**: The effective sample size of the control group
  - **n.treat**: The number of subjects in the treatment group
  - **n.ctrl**: The number of subjects in the control group
  - **max.es**: The largest effect size across the covariates
  - **mean.es**: The mean absolute effect size
  - **max.ks**: The largest KS statistic across the covariates
  - **mean.ks**: The average KS statistic across the covariates
  - **bal.tab**: A (potentially large) table summarizing the quality of the weights for equalizing the distribution of features across the two groups. This table is best extracted using the bal.table method. See the help for bal.table for details on the table’s contents
  - **n.trees**: The estimated optimal number of gbm iterations to optimize the loss function for the associated stop.methods
  - **ps**: A data frame containing the estimated propensity scores. Each column is associated with one of the methods selected in stop.methods
  - **w**: A data frame containing the propensity score weights. Each column is associated with one of the methods selected in stop.methods. If sampling weights are given then these are incorporated into these weights.
  - **estimand**: The estimand of interest (ATT or ATE).
  - **datestamp**: Records the date of the analysis
  - **parameters**: Saves the ps call
  - **alerts**: Text containing any warnings accumulated during the estimation
  - **iters**: A sequence of iterations used in the GBM fits used by plot function.
  - **balance**: The balance measures for the pretreatment covariates, with a column for each stop.method.
  - **n.trees**: Maximum number of trees considered in GBM fit.
  - **data**: Data as specified in the data argument.

Author(s)

Greg Ridgeway <gregr@rand.org>, Dan McCaffrey <danielm@rand.org>, Andrew Morral <morr@rand.org>, Lane Burgette <burgette@rand.org>

References

See Also

gbm

Description

Simulated example data for assessing race bias in traffic stop outcomes

Usage

data(raceprofiling)

Format

A data frame with 5000 observations on the following 10 variables.

- `id`: an ID for each traffic stop
- `nhood`: a factor indicating the neighborhood in which the stop occurred.
- `reason`: The reason for the stop, mechanical/registration violations, dangerous moving violation, non-dangerous moving violation
- `resident`: an indicator whether the driver is a resident of the city
- `age`: driver's age
- `male`: an indicator whether the driver was male
- `race`: the race of the driver, with levels A, B, H, W
- `hour`: the hour of the stop (24-hour clock)
- `month`: and ordered factor indicating in which month the stop took place
- `citation`: an indicator of whether the driver received a citation

Source

This is simulated data to demonstrate how to use twang to adjust estimates of racial bias for important factors. This dataset does not represent real data from any real law enforcement agency.

References


http://www.i-pensieri.com/gregr/rp.shtml

Examples

data(raceprofiling)

# the first five lines of the dataset
raceprofiling[1:5,]
sensitivity analyses for propensity score analyses

Description
sensitivity

Usage

sensitivity(ps1,
  data,
  outcome,
  order.by.importance = TRUE,
  verbose = TRUE)

Arguments
ps1 A ps object.
data The dataset including the outcomes
outcome The outcome of interest.
order.by.importance Orders the output by relative importance of covariates.
verbose If TRUE, extra information will be printed.

Details
Performs the sensitivity analyses described in Ridgeway (2006). This is a beta version of this functionality. Please let the developers know if you have problems with it.

Value
tx Summary for treated observations.
ctrl Summary for control observations.

References
stop.methods

Object only used for backward compatibility

Description

In older versions of twang, the ps function specified the stop.method in a different manner. This stop.methods object is used to ensure backward compatibility; new twang users should not make use of it.

Details

This is merely a vector with the names of the stopping rules.

See Also

ps

summary.mnps

Summarize an mnps object

Description

Computes summary information about a stored mnps object

Usage

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

Arguments

object a ps object

... additional arguments affecting the summary produced

Details

Compresses the information in the desc component of the ps object into a short summary table describing the size of the dataset and the quality of the propensity score weights.

Value

See ps for details on the returned table

See Also

ps, mnps
Summary

Summarize a ps object

**Description**

Computes summary information about a stored ps object

**Usage**

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

**Arguments**

- `object`: a ps object
- `...`: additional arguments affecting the summary produced

**Details**

Compresses the information in the desc component of the ps object into a short summary table describing the size of the dataset and the quality of the propensity score weights.

**Value**

See `ps` for details on the returned table

**See Also**

`ps`
Index

*Topic **datasets**
  AOD, 2
  egsingle, 10
  iptwExLong, 15
  iptwExWide, 15
  lalonde, 16
  lindner, 17
  mnIptwExLong, 19
  mnIptwExWide, 19
  raceprofiling, 28

*Topic **models**
  bal.table, 4
  desc.wts, 7
  dx.wts, 9
  iptw, 13
  mnp, 20
  plot.mnps, 22
  plot.ps, 24
  ps, 25
  summary.mnps, 30
  summary.ps, 31

*Topic **multivariate**
  bal.stat, 3
  mnp, 20
  plot.mnps, 22
  plot.ps, 24
  ps, 25

*Topic **print**
  print.dwts, 25

*Topic **utilities**
  get.weights, 11
  get.weights.num, 12
  get.weights.unst, 13
  means.table, 18
  stop.methods, 30

AOD, 2
bal.stat, 3, 8
bal.table, 4, 27
boxplot.mnps, 5
boxplot.ps, 6
desc.wts, 7
dx.wts, 4, 5, 9, 25
egsingle, 10

*gbm*, 13, 14, 20, 25–28
get.weights, 11
get.weights.num, 12
get.weights.unst, 13

iptw, 12, 13, 13
iptwExLong, 15
iptwExWide, 15

lalonde, 16
lindner, 17

means.table, 18
mnIptwExLong, 19
mnIptwExWide, 19

mnp, 4, 18, 20, 23, 30

plot, 21, 26
plot.mnps, 22
plot.ps, 24

print.dwts, 25

ps, 4–10, 12, 15, 25, 25, 30, 31

raceprofiling, 28

sensitivity, 29
stop.methods, 30

summary.mnps, 30
summary.ps, 31