Package 'PSAgraphics'

March 21, 2023

Title Propensity Score Analysis Graphics

Version 2.1.2

Date 2023-03-20

Maintainer Jason Bryer <jason@bryer.org>

Description A collection of functions that primarily produce graphics to aid in a Propensity Score Analysis (PSA). Functions include: cat.psa and box.psa to test balance within strata of categorical and quantitative covariates, circ.psa for a representation of the estimated effect size by stratum, loess.psa that provides a graphic and loess based effect size estimate, and various balance functions that provide measures of the balance achieved via a PSA in a categorical covariate.

License GPL (>= 2)

URL https://jbryer.github.io/PSAgraphics/

BugReports https://github.com/jbryer/PSAgraphics/issues

Depends R (>= 2.14.0), rpart

Encoding UTF-8

RoxygenNote 7.2.3

Imports graphics, grDevices, stats

NeedsCompilation no

Author James Helmreich [aut],
Robert Pruzek [aut],
Jason Bryer [ctb, cre] (<https://orcid.org/0000-0002-2454-0402>),
KuangNan Xiong [ctb]

Repository CRAN

Date/Publication 2023-03-21 12:30:02 UTC

R topics documented:

`granova-package` *Graphical Analysis of Variance*

Description

A collection of functions that primarily produce graphics to aid in a Propensity Score Analysis (PSA). Functions include: cat.psa and box.psa to test balance within strata of categorical and quantitative covariates, circ.psa for a representation of the estimated effect size by stratum, loess.psa that provides a graphic and loess based effect size estimate, and various balance functions that provide measures of the balance achieved via a PSA in a categorical covariate.

A collection of functions that primarily produce graphics to aid in a Propensity Score Analysis (PSA). Functions include: cat.psa and box.psa to test balance within strata of categorical and quantitative covariates, circ.psa for a representation of the estimated effect size by stratum, loess.psa that provides a graphic and loess based effect size estimate, and various balance functions that provide measures of the balance achieved via a PSA in a categorical covariate.

Details

Package:	PSAgraphics
Version:	2.0
License:	GPL (>= 2)

Package:	PSAgraphics
Version:	2.0
License:	GPL (>= 2)

Author(s)

James E. Helmreich <James.Helmreich@Marist.edu> and Robert M. Pruzek <RMPruzek@yahoo.com>

Maintainer: Jason Bryer <jason@bryer.org>

James E. Helmreich <James.Helmreich@Marist.edu> and

Robert M. Pruzek <RMPruzek@yahoo.com>

Maintainer: James E. Helmreich <James.Helmreich@Marist.edu>

See Also

box.psa cat.psa circ.psa loess.psa bal.ks.psa bal.ms.psa

bal.fe.psa cstrata.psa cv.trans.psa cv.bal.psa

box.psa cat.psa circ.psa loess.psa bal.ks.psa bal.ms.psa

bal.fe.psa cstrata.psa cv.trans.psa cv.bal.psa

`bal.cs.psa` *Balance for Categorical Covariate: Random Strata as part of a PSA*

Description

Function provides a measure of the balance achieved between control and treatment groups for a categorical covariate from user defined strata. This statistic is compared to the same measure for randomly permuted strata.

Usage

```
bal.cs.psa(
  categorical,
  treatment = NULL,
  strata = NULL,
  B = 1000,
  eps = 0.02,
  main = NULL,
  ...
)
```

Arguments

`categorical`	Categorical covariate that is being balanced within strata in a PSA. If `categorical` has three columns, then the second and third are assumed to be the treatment and strata respectively. Missing values are not allowed. May be factor or numeric.
`treatment`	Binary variable of same length as `categorical`; generally 0 for 'control,' 1 for 'treatment.'

strata	Integer variable; a vector of same length as categorical indicating the derived strata from estimated propensity scores.
B	Numeric; number of randomly generated iterations of the balance measure are created for the comparison distribution.
eps	Numeric; ensures that weighting is reasonable for small categories.
main	Title passed to histogram.
...	Other graphical parameters passed to histogram.

Details

This function measures the balance achieved across K strata for a categorical covariate with J categories. If p_{ijk} is the proportion of cases in stratum k, category j, and treatment i, then the statistic is the sum over all K, J of $|\sqrt{p_{0jk}+\epsilon} - \sqrt{p_{1jk}+\epsilon}|$. A permutation distribution is generated by randomly assigning cases to strata, thus generating B permuted stratifications and the associated B permutation statistics. The permutation stratifications are generated under a fixed marginals model to retain comparability with the original stratification. A histogram of the permutation statistics is produced with the original statistic referenced as a red dot.

Value

In addition to the histogram, a list with the following components is returned:

balance.orig	Balance measure of user defined strata.
rank.orig	Rank of original balance measure in comparison with the B randomly generated values.

Author(s)

James E. Helmreich < James.Helmreich@Marist.edu>

Robert M. Pruzek <RMPruzek@yahoo.com>

See Also

bal.cws.psa, bal.ms.psa, bal.ks.psa

Examples

```
#Everything random
categorical<-sample(4,1000,replace=TRUE)
treatment<-sample(c(0,1),1000,replace=TRUE)
strata<-sample(5,1000,replace=TRUE)
bal.cs.psa(categorical,treatment,strata)

#Perfect balance on 80%, random on last 20%
categorical<-rep(sample(5,1000,replace=TRUE),2)
treatment<-c(rep(0,1000),rep(1,1000))
strat<-sample(6,1200,replace=TRUE)
strat<-c(strat[1:1000],strat[1:800],strat[1001:1200])
bal.cs.psa(categorical,treatment,strat,B=200)
```

bal.fe.psa *Fisher's Exact Test for Independence of Treatments within Strata*

Description

Simple function that calls fisher.test repeatedly for each strata, testing the independence of treatements for the given covariate within strata.

Usage

```
bal.fe.psa(categorical, treatment = NULL, strata = NULL, FB = 2000)
```

Arguments

categorical	Categorical covariate that is being balanced within strata in a PSA. If categorical has three columns, then the second and third are assumed to be the treatment and strata respectively. Missing values are not allowed. May be factor or numeric.
treatment	Binary variable of same length as categorical; generally 0 for 'control,' 1 for 'treatment.'
strata	Integer variable; a vector of same length as categorical indicating the derived strata from estimated propensity scores.
FB	Numeric; number of replications sent to fisher.test.

Details

This function makes repeated calls to fisher.test, Fisher's Exact test, to test whether the distribution of the covariate categorical is independent of treatment within each stratum; a list of p-values for the test for each stratum are returned.

Value

Returns list of the same lenght as the number of strata containing p-values for the indpendence of treatment within each stratum derived from Fisher's Exact test.

Author(s)

James E. Helmreich < James.Helmreich@Marist.edu>

Robert M. Pruzek <RMPruzek@yahoo.com>

See Also

bal.cs.psa, bal.ms.psa, bal.ks.psa

Examples

```
#Everything random
categorical<-sample(4, 1000, replace = TRUE)
treatment<-sample(c(0,1), 1000, replace = TRUE)
strata<-sample(5, 1000, replace = TRUE)
bal.fe.psa(categorical, treatment, strata)

#Perfect balance on 80%, random on last 20%
categorical<-rep(sample(5,1000, replace=TRUE), 2)
treatment<-c(rep(0,1000), rep(1,1000))
strata<-sample(6, 1200, replace=TRUE)
strata<-c(strata[1:1000], strata[1:800], strata[1001:1200])
bal.fe.psa(categorical, treatment, strata)
```

---

bal.ks.psa *Kolgomorov-Smirnov 2 sample tests for multiple strata*

Description

Automates the Kolgomorov-Smirnov 2-sample nonparametric test of equivalence of two distrbutions across multiple pairs of sample distributions.

Usage

```
bal.ks.psa(continuous, treatment = NULL, strata = NULL)
```

Arguments

`continuous`	Quantitative covariate that is being balanced within strata in a PSA. If `continuous` has three columns, then the second and third are assumed to be the treatment and strata respectively. Missing values are not allowed.
`treatment`	Binary variable of same length as `continuous`; generally 0 for 'control,' 1 for 'treatment.'
`strata`	Integer variable (usually 1 - 5); A vector of same length as continuous indicating the derived strata from estimated propensity scores. Generally 5 or 6 strata are used, but graph works reasonably well at least up to 10 strata.

Details

Makes multiple calls to `ks.test`, returning a vector of p-values associated with strata from a Propensity Score Analysis.

Value

Returns a vector of same length as the number of strata containing the p-values from the KS-test of equivalence of distributions for each stratum-treatment pair.

Author(s)

James E. Helmreich <James.Helmreich@Marist.edu>

Robert M. Pruzek <RMPruzek@yahoo.com>

See Also

bal.ms.psa, bal.cs.psa, bal.cws.psa

Examples

```
continuous<-rnorm(1000)
treatment<-sample(c(0,1),1000,replace=TRUE)
strata<-sample(5,1000,replace=TRUE)
bal.ks.psa(continuous,treatment,strata)
```

---

bal.ms.psa *Balance for Continuous Covariate: Random Strata as part of a PSA*

Description

Function provides a measure (based on the trimmed mean) of the balance achieved between control and treatment groups for a continuous covariate from user defined strata. This statistic is compared to the same measure for randomly permuted strata.

Usage

```
bal.ms.psa(
  continuous,
  treatment = NULL,
  strata = NULL,
  trim = 0,
  B = 1000,
  main = NULL
)
```

Arguments

continuous	Quantitative covariate that is being balanced within strata in a PSA. If continuous has three columns, then the second and third are assumed to be the treatment and strata respectively. Missing values are not allowed.
treatment	Binary variable of same length as continuous; generally 0 for 'control,' 1 for 'treatment.'
strata	Integer variable; a vector of same length as continuous indicating the derived strata from estimated propensity scores.

trim	Fraction (0 to 0.5) of observations to be trimmed from each end of stratum-treatment level before the mean is computed. See mean.
B	Numeric; number of randomly generated iterations of the balance measure are created for the comparison distribution.
main	Title passed to histogram.

Details

This function measures the balance achieved across K strata for a continuous covariate. If μ_{ik} is the covariate trimmed (as specified by user) mean of cases in stratum k, treatment i, then the statistic is the sum over all K of $|\mu_{0k} - \mu_{1k}|$. A permutation distribution is generated by randomly assigning cases to strata, thus generating B permuted stratifications and the associated B permutation statistics. The permutation stratifications are generated under a fixed marginals model to retain comparability with the original stratification. A histogram of the permutation statistics is produced with the original statistic referenced as a red dot.

Value

In addition to the histogram, a list with the following components is returned:

balance.orig	Balance measure of user defined strata.
rank.orig	Rank of original balance measure in comparison with the B randomly generated values.

Author(s)

James E. Helmreich < James.Helmreich@Marist.edu>

Robert M. Pruzek <RMPruzek@yahoo.com>

See Also

bal.ks.psa, bal.cws.psa, bal.cs.psa

Examples

```
#Balance stat should be close to zero
meas<-rnorm(500)
continuous<-c(meas,meas+rnorm(500,0,.1))
treatment<-c(rep(0,500),rep(1,500))
strata<-rep(c(rep(1,100),rep(2,100),rep(3,100),rep(4,100),rep(5,100)),2)
bal.ms.psa(continuous,treatment,strata)


#Balance stat should be close to .4
meas<-rnorm(500)
continuous<-c(meas, meas[1:250] + runif(250,0,.2),
   meas[251:500]-runif(250,0,.2))
treatment<-c(rep(0,500),rep(1,500))
strata<-rep(c(rep(1,100), rep(2,100), rep(3,100),
   rep(4,100),rep(5,100)),2)
```

```
bal.ms.psa(continuous, treatment, strata, B=200)
```

---

box.psa *Compare balance graphically of a continuous covariate as part of a PSA*

Description

Given predefined strata and two level treatment for a continuous covariate from a propensity score analysis, box.psa draws pairs of side by side boxplots corresponding to control and treatment for each stratum.

Usage

```
box.psa(
  continuous,
  treatment = NULL,
  strata = NULL,
  boxwex = 0.17,
  offset = 0.17,
  col = c("yellow", "orange", "black", "red", "darkorange3"),
  xlab = "Stratum",
  legend.xy = NULL,
  legend.labels = NULL,
  pts = TRUE,
  balance = FALSE,
  trim = 0,
  B = 1000,
  ...
)
```

Arguments

continuous	Vector or N X 3 dataframe or matrix. If a vector, then represents the quantitative covariate that is being balanced within strata in a PSA. If continuous has three columns, then the second and third are assumed to be the treatment and strata respectively. Missing values are not allowed.
treatment	Binary vector of same length as continuous representing the two treatments; can be a character vector or factor.
strata	A vector or factor of same length as continuous indicating the derived strata from estimated propensity scores. May be numeric or character vector, or factor. Strata are ordered lexicographically in plot.
boxwex	Numeric; controls width of boxes. Default = 0.17
offset	Numeric; controls distance between the two boxes in each stratum. Default = 0.17

col	Default = c("yellow", "orange", "black", "red","darkorange3"). Color vector for the control boxes, treatment boxes, and line connecting their means.
xlab	Label for the x-axis; default = "Stratum". Other standard labels may be used as well.
legend.xy	Binary vector giving coordinates of the legend. By default the legend is placed to the top left.
legend.labels	Vector of labels for the legend; default is essentially c("Treatment (first)", "Treatment (second)", "Treatment Means Compared", "KS p-values", "Strata-Treatment Size") where treatment names are taken from treatment. Vector has four elements if balance = FALSE, ommitting "KS p-values".
pts	Logical; if TRUE then (jittered) points are added on top of the boxplots.
balance	Logical; if TRUE then bal.ms.psa provides a histogram of a permutation distribution and reference statstic to assess balance across strata; bal.ks.psa adds p-values to the graph derived from 2-sample Kologmorov-Smirnov tests of equivalence of control/treatment distributions within each stratum.
trim	If balance=TRUE, defines fraction (0 to 0.5) of observations to be trimmed from each end of stratum-treatment level before the mean is computed. See mean, bal.ms.psa.
B	Passed to bal.ms.psa if necessary, determines number of randomly generated comparison statistics. Default =1000.
...	Other graphical parameters passed to boxplot.

Details

Draws a pair of side by side boxplots for each stratum of a propensity score analysis. This allows visual comparisons within strata of the distribution of the given continuous covariate, and comparisons between strata as well. The number of observations in each boxplot are given below each box, and the means of paired treatment and control groups are connected.

Author(s)

James E. Helmreich <James.Helmreich@Marist.edu>

Robert M. Pruzek <RMPruzek@yahoo.com>

See Also

bal.ks.psa, bal.ms.psa, cat.psa

Examples

```
continuous<-rnorm(1000)
treatment<-sample(c(0,1),1000,replace=TRUE)
strata<-sample(5,1000,replace=TRUE)
box.psa(continuous, treatment, strata)

data(lindner)
```

```
attach(lindner)
lindner.ps <- glm(abcix ~ stent + height + female +
      diabetic + acutemi + ejecfrac + ves1proc,
      data = lindner, family = binomial)
ps<-lindner.ps$fitted
lindner.s5 <- as.numeric(cut(ps, quantile(ps, seq(0, 1, 1/5)),
      include.lowest = TRUE, labels = FALSE))
box.psa(ejecfrac, abcix, lindner.s5, xlab = "ejecfrac",
      legend.xy = c(3.5,110))

lindner.s10 <- as.numeric(cut(ps, quantile(ps, seq(0, 1, 1/5)),
      include.lowest = TRUE, labels = FALSE))
box.psa(height, abcix, lindner.s10, xlab="height",
      boxwex = .15, offset = .15, legend.xy = c(2,130), balance = TRUE)
```

`cat.psa` *Compare balance graphically of a categorical covariate as part of a PSA*

Description

Given predefined strata and two level treatment for a categorical covariate from a propensity score analysis, `cat.psa` draws pairs of side by side barplots corresponding to control and treatment for each stratum.

Usage

```
cat.psa(
  categorical,
  treatment = NULL,
  strata = NULL,
  catnames = NULL,
  catcol = "terrain.colors",
  width = 0.25,
  barlab = c("A", "B"),
  barnames = NULL,
  rtmar = 1.5,
  balance = FALSE,
  B = 1000,
  tbl = TRUE,
  cex.leg = 1,
  ...
)
```

Arguments

`categorical`	Vector or N X 3 dataframe or matrix. If a vector, then represents a categorical covariate that is being balanced within strata in a PSA. If `categorical` has

	three columns, then the second and third are assumed to be the `treatment` and `strata` respectively. Missing values are not allowed. May be factor or numeric.
`treatment`	Binary vector or factor of same length as `continuous` representing the two treatments.
`strata`	A vector or factor of same length as `continuous` indicating the derived strata from estimated propensity scores. Strata are ordered lexicographically in plot.
`catnames`	List of names in order of the categories; used in the plot legend. Default is `1:n`.
`catcol`	List of colors used for the categories, default is `terrain.colors`.
`width`	Controls width of bars, default = 0.25.
`barlab`	Binary list of single `treatment` character labels for the bars, default is `c("A", "B")`. These are defined in a legend by `barnames`.
`barnames`	Binary list of treatment names used in the legend; by default names are taken from `treatment`.
`rtmar`	Numeric. Governs size of right margin allocated for legend. Default = 1.5
`balance`	Logical. If `TRUE` a call is made to functions `bal.cs.psa` and `bal.cws.psa`. The former provides a reference histogram and ad hoc balance statistic, the second provides bootstrapped p-values for the two-way table formed in each statum. Default is `FALSE`.
`B`	Numeric; passed to `bal.cs.psa` governing size of reference histogram generated. Default is 100.
`tbl`	Logical; if `TRUE`, then a matrix of the proportions used in the creation of the bargraph is returned.
`cex.leg`	Numeric; value of `cex` (governing font size) passed to legend. Default = 1.
`...`	Other graphical parameters passed to plot.

Details

Pairs of bars are graphed side by side so that comparisons may be made within each stratum and across strata. If `balance` is `TRUE`, then the histogram represents an ad hoc balance measure of the given strata as compared to randomly generated strata. The p-values provided on the bargraph are bootstrapped in a standard fashion via randomly generated treatment divisions within given strata. For continuous covariates use `box.psa`.

Value

If `tbl` is `TRUE`, then a matrix is returned containing the proportions of each category, and in each treatment level and stratum that were used to draw the bargraph.

Author(s)

James E. Helmreich < James.Helmreich@Marist.edu>

Robert M. Pruzek <RMPruzek@yahoo.com>

See Also

`bal.cs.psa`, `bal.cws.psa`, `box.psa`

Examples

```
categorical<-sample(1:7,1000,replace=TRUE)
treatment<-sample(c(0,1),1000,replace=TRUE)
strata<-sample(5,1000,replace=TRUE)
cat.psa(categorical,treatment,strata)

data(lindner)
attach(lindner)
lindner.ps <- glm(abcix ~ stent + height + female +
      diabetic + acutemi + ejecfrac + ves1proc,
      data = lindner, family = binomial)
ps<-lindner.ps$fitted
lindner.s5 <- as.numeric(cut(ps, quantile(ps, seq(0, 1, 1/5)),
      include.lowest = TRUE, labels = FALSE))
cat.psa(stent, abcix, lindner.s5, xlab = "stent")

lindner.s10 <- as.numeric(cut(ps, quantile(ps, seq(0, 1, 1/10)),
      include.lowest = TRUE, labels = FALSE))
cat.psa(ves1proc,abcix, lindner.s10, balance = TRUE, xlab = "ves1proc")

#Using a rpart tree for strata
library(rpart)
lindner.rpart<-rpart(abcix ~ stent + height + female + diabetic +
      acutemi + ejecfrac + ves1proc, data=lindner, method="class")
lindner.tree<-factor(lindner.rpart$where, labels = 1:6)
cat.psa(stent, abcix, lindner.tree, xlab = "stent")
cat.psa(ves1proc, abcix, lindner.tree, xlab = "ves1proc")
```

---

circ.psa *Generates a Propensity Score Assessment Plot*

Description

Displays a graphic that summarizes outcomes in a propensity score analysis, based on strata that have been defined in the first Phase of a propensity score analysis (PSA). The graphic displays contributions of individual strata to the overall effect, weighing contributions of individual strata according to the relative sizes of the respective strata. The overall effect is plotted as a heavy dashed diagonal line that runs parallel to the identity diagonal.

Usage

```
circ.psa(
  response,
  treatment = NULL,
  strata = NULL,
  summary = FALSE,
  statistic = "mean",
```

```
  trim = 0,
  revc = FALSE,
  confint = TRUE,
  sw = 0.4,
  ne = 0.5,
  inc = 0.25,
  pw = 0.4,
  lab = TRUE,
  labcex = 1,
  xlab = NULL,
  ylab = NULL,
  main = NULL
)
```

Arguments

response	Either a numeric vector containing the response of interest in a propensity score analysis, or a three column array containing response, treatment and strata.
treatment	Binary variable of same length as response; generally 0 for 'control,' 1 for 'treatment'. A character vector with two labels or factor with two levels are also accepted.
strata	Generally integer variable; a vector of same length as response indicating the derived strata from estimated propensity scores. Generally 5 or 6 strata used, but function is effective for more strata. In the case when strata are defined via unique propensity scores (as from a tree), user may wish to define strata using factor.
summary	Logical (default FALSE). If TRUE then response must have rows corresponding to number of strata; the first two columns should contain treatment and control group sizes for each stratum, and the pair of columns should contain the appropriate summary statistics for each statum. For example, the four summary columns might have been generated by the strata.summary output of loess.psa.
statistic	A scalar summary of the center of the response distribution. Seen next item below. Default = "mean". Note that to generate this statistic the full vector of responses must have been input, not summaries.
trim	Allows for a trimmed mean as outcome measure, where trim is from 0 to .5 (.5 implying median).
revc	Logical; if TRUE then X and Y axes are interchanged in plot.
confint	Logical; if TRUE adds an approximate 95% confidence interval for the mean. The interval may not be realistic it the trim argument exceeds zero.
sw	Numerical argument (default = 0.4); extends axes on lower ends, effectively moving circles to lower left.
ne	Numerical argument (default = 0.5); extends axes on upper ends, effectively moving circles to upper right.
inc	Numerical argument (default = 0.35); controls circle sizes, but relative circle sizes are controlled via pw. In general one wants circle areas to appear subjectively to be sized in accordance with strata sizes.

pw	numerical argument (default = 0.4); controls relative circle sizes. pw denotes power or exponent for radius of circle.
lab	Logical (default TRUE); labels circles with stratum labels.
labcex	numerical argument (default = 1); controls the size of the circle labels.
xlab	Label for horizontal axis, by default taken from treatment.
ylab	Label for vertical axis, by default taken from treatment.
main	Main label for graph.

Details

A circle is plotted for each stratum, centered on the means for the treatment and control groups (for the X and Y axes) respectively. The sizes of the circles correspond to the relative sizes of the strata. A diagonal line (lower left to upper right) shows the identity, X=Y, so that circles on, say, the lower side of this line show that the corresponding X mean is larger than the Y mean for that stratum, and vice-versa. Parallel projections are made from the centers of the strata-cum-circles to difference scores that are plotted on a line segment on the lower-left corner of the graphic; the average difference, which corresponds to the average treatment effect (ATE) for the overall treatment effect, is plotted as a heavy (dark blue) dashed line parallel to the identity diagonal. Rug plots are shown on the upper and right margins of the graphic, for the X and Y marginal distributions. A 95% confidence interval for the overall effect is plotted to the left of the distribution of the stratum difference scores, centered on the ATE. Trimmed means can replace the conventional mean for both the ATE and the marginal distributions (however, the confidence interval calculations are likely to become less trustworthy as larger values of the trim argument are used).

Value

Generate a Propensity Assessment Plot, as well as numerical data for

summary.strata	An array with rows corresponding to strata and four columns; these show counts for control and treatment groups, as well as (possibly trimmed) mean response values for control and treatment.
wtd.Mn.(Name1)	Weighted mean of response for (Name1) group. Name taken from treatment.
wtd.Mn.(Name2)	Weighted mean of response for (Name2) group. Name taken from treatment.
ATE	Average Treatment Effect.
se.wtd	Weighted standard error for ATE
approx.t	Ratio of the average treatement effect and a standard error based on weighting of stratum variances.
df	Estimate of degree of freedom; response vector length minus twice number of strata.
CI.95	Approximate 95% confidence interval for overall effect size.

Author(s)

James E. Helmreich <James.Helmreich@Marist.edu>

Robert M. Pruzek <RMPruzek@yahoo.com>

See Also

loess.psa

Examples

```
##Random data with effect size 0
response <- rnorm(1000)
treatment <- sample(c(0,1), 1000, replace = TRUE)
strata <- sample(1:6, 1000, replace = TRUE)
circ.psa(response, treatment, strata)

##Random data with effect size -.2
response <- c(rnorm(500, 0, 12), rnorm(500, 6, 12))
treatment <- c(rep(0, 500), rep(1,500))
strata <- sample(1:5, 1000, replace = TRUE)
aaa <- cbind(response, treatment, strata)
circ.psa(aaa)

##Random data with effect size -2
response <- c(rt(100,3) * 2 + 20, rt(100,12) * 2 + 18)
treatment <- rep(c("A","B"), each = 100)
strata <- sample(c("X","Y","Z","U","V"), 200, replace = TRUE)
circ.psa(response, treatment, strata)

##Tree derived strata
library(rpart)
data(lindner)
attach(lindner)
lindner.rpart <- rpart(abcix ~ stent + height + female + diabetic +
     acutemi + ejecfrac + ves1proc, data = lindner, method = "class")
lindner.tree<-factor(lindner.rpart$where, labels = 1:6)
circ.psa(log(cardbill), abcix, lindner.tree)

##Loess derived strata
lindner.ps <- glm(abcix ~ stent + height + female +
      diabetic + acutemi + ejecfrac + ves1proc,
      data = lindner, family = binomial)
ps<-lindner.ps$fitted
lindner.loess<-loess.psa(log(cardbill), abcix, ps)
circ.psa(lindner.loess$summary.strata[, 1:4], summary = TRUE,
      inc = .1, labcex = .7)
```

---

cstrata.psa *Supports Multiple Methods for Defining and Visualizing (PS) Strata*

Description

Given propensity scores, allows strata to be directly user defined, possibly to: equalize sizes of strata, equalize the ranges of propensity scores, or to specify cut points on the unit interval. Once

strata are created, a simple graphic is generated to visualize or judge strata for overlap and appropriateness. If a regression tree has been used, propensity scores are defined for each leaf of the tree.

Usage

```
cstrata.psa(
  treatment,
  propensity,
  strata = NULL,
  int = NULL,
  tree = FALSE,
  minsize = 2,
  graphic = TRUE,
  colors = c("dark blue", "dark green"),
  xlab = "Estimated Propensity Scores with Random Heights",
  pch = c(16, 16)
)
```

Arguments

treatment	Binary vector or factor defining the two treatments
propensity	Vector of same length as treatment containing estimated propensity scores.
strata	Either a vector of same length as treatment of predefined stratum number, or one integer n used to assign rows to n strata propensity scores, each of approximately the same number of cases. If relatively few unique propensity scores have been defined (as from a classification tree) then the logical tree should be set equal to TRUE.
int	Either a number m used to divide [0,1] into m equal length subintervals, or a vector containing cut points between 0 and 1 that define subintervals (perhaps as suggested by loess.psa). In either case the subintervals define strata, for which sizes can differ.
tree	Logical, default FALSE. If there are few unique propensity scores, say from a recursively partitioned tree, then TRUE forces strata to be defined by the unique propensity scores.
minsize	Smallest allowable stratum-treatment size. If violated, rows in the stratum are removed. User may wish to redefine strata.
graphic	Logical, default TRUE. If set to FALSE the graphic is not provided.
colors	2-ary color vector. Sets the colors of the points in the graphic. Default = c("blue", "orange")
xlab	Label for the x axis; default = "Estimated Propensity Scores with Random Heights".
pch	2-ary vector; determines the shape of points in the graphic. Default = c(16, 16).

Value

`Original.Strata`
: Table of strata-treatment sizes before `minsize` evaluation.

`Used.Strata`
: Table of strata-treatment sizes after `minsize` evaluation.

`strata`
: Vector of the same length as `treatment`, indicating either the strata input by user or those created by the function.

Author(s)

James E. Helmreich < James.Helmreich@Marist.edu>

Robert M. Pruzek <RMPruzek@yahoo.com>

KuangNan Xiong <harryxkn@yahoo.com>

See Also

`cv.bal.psa`, `loess.psa`

Examples

```
data(lindner)
attach(lindner)
lindner.ps <- glm(abcix ~ stent + height + female +
      diabetic + acutemi + ejecfrac + ves1proc,
      data = lindner, family = binomial)
ps <- lindner.ps$fitted
cstrata.psa(abcix, ps, strata = 5)
cstrata.psa(abcix, ps, strata = 10)
cstrata.psa(abcix, ps, int = c(.37, .56, .87, 1))
```

---

cv.bal.psa *Multiple Covariate Balance Assessment Plot*

Description

Provides a graphic that depicts covarite effect size differences between treatment groups both before and after stratification. Function will create stata internally if desired, and returns numerical output used to create graphic.

Usage

```
cv.bal.psa(
  covariates,
  treatment,
  propensity,
  strata = NULL,
```

```
  int = NULL,
  tree = FALSE,
  minsize = 2,
  universal.psd = TRUE,
  trM = 0,
  absolute.es = TRUE,
  trt.value = NULL,
  use.trt.var = FALSE,
  verbose = FALSE,
  xlim = NULL,
  plot.strata = TRUE,
  ...
)
```

Arguments

covariates	Dataframe of covariates. Factors should be recoded using cv.trans.psa
treatment	Binary vector or factor defining the two treatments
propensity	Vector of same length as treatment containing estimated propensity scores.
strata	Either a vector of same length as treatment of predefined stratum number, or one integer n used to assign rows to n strata propensity scores, each of approximately the same number of cases. If relatively few unique propensity scores have been defined (as from a classification tree) then the logical tree should be set equal to TRUE.
int	Either a number m used to divide [0,1] into m equal length subintervals, or a vector containing cut points between 0 and 1 that define subintervals (perhaps as suggested by loess.psa). In either case the subintervals define strata, for which sizes can differ.
tree	Logical, default FALSE. If there are few unique propensity scores, say from a recursively partitioned tree, then TRUE forces strata to be defined by the unique propensity scores.
minsize	Smallest allowable stratum-treatment size. If violated, rows in the stratum are removed. User may wish to redefine strata.
universal.psd	Logical, default = TRUE. Forces standard deviations used to be unadjusted for stratification.
trM	Numeric, default = 0; passed to mean for trimming purposes.
absolute.es	Logical, default TRUE. If TRUE, graphic depicts absolute values of all effect sizes. Note that the adjusted effect size plotted is the absolute value of weighted averages of the signed by-stratum effect size values when absolute.es is TRUE.
trt.value	Character string; if desired allows the name of an active treatment to be given. Should be a level (value) of the treatment factor (vector).
use.trt.var	Logical, default FALSE. If TRUE, uses just active treatment standard deviations for effect size, as per a suggestion of Rubin and Stuart (see reference below).
verbose	Logical, default FALSE. Numerical output is returned invisibly.
xlim	Binary vector passed to plot for overriding default choices. Default NULL.

plot.strata	Logical, default TRUE. Adds effect size values for individual strata to graphic.
...	Other graphical parameters passed to plot.

Details

Effect sizes between treatments for each covariate are presented in one graphic, both before and after stratification.

Value

Graphic plots covariate balance before and after stratication on propensity scores. The default version (absolute.es = TRUE) plots the absolute values of effect sizes for each stratum, though the overall estimate is the weighted mean before taking the absolute values. Numerical output consists of seven addressable objects. If verbose is FALSE (default), output is not printed.

original.strata	Matrix of strata-treatment counts as originally input.
strata.used	Matrix of strata-treatment counts used in effectsize calculations after any minsize reductions.
mean.diff.strata.wtd	Matrix of strata by covariate weighted (by strata size) average differences.
mean.diff.unadj	Matrix of covariate effects sizes before stratification.
effect.sizes	Matrix of effect sizes by covariate and statum.
treatment.levels	Names of treatments.
effects.strata.treatment	Matrix of standard deviations and stratum-treatment covariate means used to calculate the effect.sizes.

Author(s)

Robert M. Pruzek <RMPruzek@yahoo.com>

James E. Helmreich <James.Helmreich@Marist.edu>

KuangNan Xiong <harryxkn@yahoo.com>

References

“Matching Methods for Causal Inference: A review and a look forward." Forthcoming in Statistical Science.

See Also

cv.bal.psa, loess.psa, cstrata.psa, cv.trans.psa

Examples

```
data(lindner)
attach(lindner)
lindner.ps <- glm(abcix ~ stent + height + female +
      diabetic + acutemi + ejecfrac + ves1proc,
      data = lindner, family = binomial)
ps<-lindner.ps$fitted
lindner.cv <- lindner[,4:10]
cv.bal.psa(lindner.cv, abcix, ps, strata = 5)
cv.bal.psa(lindner.cv, abcix, ps, strata = 10)
cv.bal.psa(lindner.cv, abcix, ps, int = c(.2, .5, .6, .75, .8))
```

---

cv.trans.psa *Transformation of Factors to Individual Levels*

Description

The function cv.trans.psa takes a covariate data frame and replaces each categorical covariate of n >=3 levels with n new binary covariate columns, one for each level. Transforms covariate dataframe for use with the function cv.bal.psa.

Usage

```
cv.trans.psa(covariates, fcol = NULL)
```

Arguments

covariates	A dataframe of covariates, presumably some factors.
fcol	An optional vector containing the factor columns in the covariate dataframe. In NULL (default) routine to identfy factors internally.

Value

Returns a dataframe covariates.transformed containing new columns for each level of more than binary factors. The rest of the covariate dataframe stays unchanged.

Author(s)

James E. Helmreich <James.Helmreich@Marist.edu>

Robert M. Pruzek <RMPruzek@yahoo.com>

KuangNan Xiong <harryxkn@yahoo.com>

See Also

cv.bal.psa, loess.psa, cstrata.psa, cv.trans.psa

Examples

```
#Note reordering of columns, binary factor and numeric column are unchanged.
f2 <- factor(sample(c(0, 1), 20, replace = TRUE))
f4 <- factor(sample(c("a", "b", "c", "d"), 20, replace = TRUE))
cv <- rnorm(20)
X <- data.frame(f2, f4, cv)
cv.trans.psa(X)
#
f2 <- factor(sample(c('c', 'C'), 20, replace = TRUE))
f4 <- factor(sample(c("b", "A", "d", "CC"), 20, replace = TRUE))
cv <- rnorm(20)
X <- data.frame(f2, f4, cv)
cv.trans.psa(X)
```

`lindner` *Data on 996 initial Percutaneous Coronary Interventions (PCIs) performed in 1997 at the Lindner Center, Christ Hospital, Cincinnati.*

Description

Data from an observational study of 996 patients receiving a PCI at Ohio Heart Health in 1997 and followed for at least 6 months by the staff of the Lindner Center. This is a landmark dataset in the literature on propensity score adjustment for treatment selection bias due to practice of evidence based medicine; patients receiving abciximab tended to be more severely diseased than those who did not receive a IIb/IIIa cascade blocker.

Data from an observational study of 996 patients receiving a PCI at Ohio Heart Health in 1997 and followed for at least 6 months by the staff of the Lindner Center. This is a landmark dataset in the literature on propensity score adjustment for treatment selection bias due to practice of evidence based medicine; patients receiving abciximab tended to be more severely diseased than those who did not receive a IIb/IIIa cascade blocker.

Format

A data frame with 996 observations on the following 10 variables, 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.

ejecfrac 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.

A data frame with 996 observations on the following 10 variables, no NAs.

list("lifepres") Mean life years preserved due to survival for at least 6 months following PCI; numeric value of either 11.4 or 0.

list("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.

list("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.

list("stent") Coronary stent deployment; numeric, with 1 meaning YES and 0 meaning NO.

list("height") Height in centimeters; numeric integer from 108 to 196.

list("female") Female gender; numeric, with 1 meaning YES and 0 meaning NO.

list("diabetic") Diabetes mellitus diagnosis; numeric, with 1 meaning YES and 0 meaning NO.

list("acutemi") Acute myocardial infarction within the previous 7 days; numeric, with 1 meaning YES and 0 meaning NO.

list("ejecfrac") Left ejection fraction; numeric value from 0 percent to 90 percent.

list("ves1proc") Number of vessels involved in the patient's initial PCI procedure; numeric integer from 0 to 5.

Source

Package USPS, by R. L. Obenchain.

Package USPS, by R. L. Obenchain.

`loess.psa` *Graphic for data and loess-based estimate of effect size after propensity score adjustment*

Description

Plots data points using propesity scores vs. the response, separately for treatment and control groups; points are distinguished by both type and color for the two groups. Also shows (non-linear, loess-based) regression curves for both groups. The loess regresion curves are then used to derive an overall estimate of effect size (based on number and/or location of strata as set by the user). Several other statistics are also provided, for both description and inference. Graphic motivated by a suggestion of R. L. Obenchain.

Usage

```
loess.psa(
  response,
  treatment = NULL,
  propensity = NULL,
  family = "gaussian",
  span = 0.7,
  degree = 1,
  minsize = 5,
  xlim = c(0, 1),
  colors = c("dark blue", "dark green", "blue", "dark green"),
  legend.xy = "topleft",
  legend = NULL,
  int = 10,
  lines = TRUE,
  strata.lines = TRUE,
  rg = TRUE,
  xlab = "Estimated Propensity Scores",
  ylab = "Response",
  pch = c(16, 1),
  ...
)
```

Arguments

response	Either a numeric vector containing the response of interest in a propensity score analysis, or a three column array containing response, treatment and strata.
treatment	Binary variable of same length as response; 0 for 'control,' 1 for 'treatment.'
propensity	Numeric vector of estimated propensity scores.
family	Passed to loess. Either "gaussian" (default) or "symmetric".
span	Parameter passed to loess governing degree of smoothing. Default = 0.7.
degree	Parameter passed to loess governing degree of polynomials used. Default = 1
minsize	Integer. Determines the minimum number of observations in each stratum treatment group allowed. If one of the treatment groups in a given statum does not meet this minsize, then all observations in this stratum are ignored as far as the effect size calculation is concerned.
xlim	Binary vector (min, max) providing the horizontal axis minimum and maximum. Default is c(0, 1).
colors	List of four colors used for control points, treatment points, control loess line, treatment loess line respectively. Default = c("seagreen3", "goldenrod1", "seagreen4", "goldenrod3").
legend.xy	Coordinates for legend box, see legend. Default = "topleft".
legend	Binary character vector containing the text of the legend. Default is taken from treatment.

int	Integer or ordered vector. If an integer is used, it represents the maximum number of equally sized strata. Alternatively, it may be a vector of cuts of the unit interval. Lower and upper ends need not be included. See examples. Default = 10.
lines	Logical; fitted loess values are plotted by default as points. If true, values are plotted as two lines.
strata.lines	Logical; default = TRUE. Creates light vertical lines that delineate strata.
rg	Logical; if TRUE (default) then rug plots are given for treatment and control propensity score and response distributions.
xlab	X axis label, default = "Estimated Propensity Scores".
ylab	Y axis label, default = "Response".
pch	Character types for plotted points, default = c(16, 1). Note: must be of length 2 to allow different plotting points for each treatment.
...	Optional parameters passed to points command.

Value

In addition to the plot, the function returns a list with the following components:

ATE	Estimated effect size based upon (number of) strata defined by int; that is, this is the Average Treatment Effect, after propensity-based adjustment.
se.wtd	Weighted standard error based on pooling of within-strata variance estimates.
CI.95	Approximate 95% confidence interval for the overall effect size (conditional on the specification of int).
summary.strata	A table with rows corresponding to strata; first two columns show counts (by statum) for both control and treatment; followed by mean differences for all strata. for control and treatment, followed by mean differences for all strata. The weighted average difference yields the effect size noted above.

Author(s)

James E. Helmreich <James.Helmreich@Marist.edu>

Robert M. Pruzek <RMPruzek@yahoo.com>

See Also

circ.psa

Examples

```
#Artificial example where ATE should be 1 over all of (0,1).
response1 <- c(rep(1, 100), rep(2, 100), rep(3, 100)) + rnorm(300, 0, .5)
response0 <- c(rep(0, 100), rep(1, 100), rep(2, 100)) + rnorm(300, 0, .5)
response <- c(response1, response0)
treatment <- c(rep(1, 300), rep(0, 300))
propensity <- rep(seq(.01, .99, (.98/299)), 2)
```

```
a <- data.frame(response, treatment, propensity)
loess.psa(a, span = .15, degree = 1, int = c(0, .33, .67, 1))


#Artificial example where estimates are unstable with varying
#numbers of strata. Note: sometimes get empty treatment/strata error.
rr <- c(rnorm(150, 3, .75), rnorm(700, 0, .75), rnorm(150, 3, .75),
     rnorm(150, -3, .75), rnorm(700, 0, .75), rnorm(150, -3, .75))
tt <- c(rep(1, 1000),rep(0, 1000))
pp <- NULL
for(i in 1:1000){pp <- c(pp, rnorm(1, 0, .05) + .00045*i + .25)}
for(i in 1:1000){pp <- c(pp, rnorm(1, 0, .05) + .00045*i + .4)}
a <- data.frame(rr, tt, pp)
loess.psa(a, span=.5, cex = .6)

#Using strata of possible interest as determined by loess lines.
data(lindner)
attach(lindner)
lindner.ps <- glm(abcix ~ stent + height + female +
      diabetic + acutemi + ejecfrac + ves1proc,
      data = lindner, family = binomial)
loess.psa(log(cardbill), abcix, lindner.ps$fitted,
         int = c(.37, .56, .87, 1), lines = TRUE)
         abline(v=c(.37, 56, .87))
```

Index