Package 'LocalControl'

May 20, 2022

Type Package

Title Nonparametric Methods for Generating High Quality Comparative
Effectiveness Evidence

Version 1.1.3

Date 2022-05-20

Description Implements novel nonparametric approaches to address
biases and confounding when comparing treatments or exposures in
observational studies of outcomes. While designed and appropriate for use
in studies involving medicine and the life sciences, the package can be
used in other situations involving outcomes with multiple confounders.
The package implements a family of methods for non-parametric bias correction
when comparing treatments in observational studies, including survival
analysis settings, where competing risks and/or censoring may be present.
The approach extends to bias-corrected personalized predictions of treatment
outcome differences, and analysis of heterogeneity of treatment effect-sizes
across patient subgroups. For further details, please see:
Lauve NR, Nelson SJ, Young SS, Obenchain RL, Lambert CG. LocalControl:
An R Package for Comparative Safety and Effectiveness Research.
Journal of Statistical Software. 2020. p. 1–32. Available from <doi:10.18637/jss.v096.i04>.

License Apache License 2.0 | file LICENSE

SystemRequirements C++11

LazyData TRUE

LinkingTo Rcpp

Imports Rcpp, gss, cluster, lattice, stats, graphics

Suggests colorspace, RColorBrewer, TeachingDemos, data.table, ggplot2,
gridExtra, rpart, rpart.plot, xtable

RoxygenNote 7.1.2

Encoding UTF-8

NeedsCompilation yes

Depends R (>= 3.0.0)

Repository CRAN

Maintainer Christophe G. Lambert <cglambert@unm.edu>

Author Nicolas R. Lauve [aut] (<<https://orcid.org/0000-0002-9348-0319>>),
      Stuart J. Nelson [aut] (<<https://orcid.org/0000-0002-8756-0179>>),
      S. Stanley Young [aut] (<<https://orcid.org/0000-0001-9449-5478>>),
      Robert L. Obenchain [aut] (<<https://orcid.org/0000-0002-8395-1666>>),
      Melania Pintilie [ctb],
      Martin Kutz [ctb],
      Christophe G. Lambert [aut, cre]
        (<<https://orcid.org/0000-0003-1994-2893>>)

R **topics documented:**

---

cardSim	*Simulated cardiac medication data for survival analysis*

Description

This dataset was created to demonstrate the effects of Local Control on correcting bias within a set
of data.

Format

A data frame with 1000 rows and 6 columns:

id Unique identifier for each row.

time Time in years to the outcome specified by status.

status 1 if the patient experienced cardiac arrest. 0 if censored before that.

drug Medication the patient received for cardiac health (drug 1 or drug 0).

age Age of the patient, ranges from 18 to 65 years.

bmi Patient body mass index. Majority of observations fall between 22 and 30.

Author(s)

Lauve NR, Lambert CG

---

framingham *Framingham heart study data extract on smoking and hypertension.*

Description

Data collected over a 24 year study suitable for competing risks survival analysis of hypertension and death as a function of smoking.

Format

A data frame with 2316 rows and 11 columns:

female Sex of the patient. 1=female, 0=male.

totchol Total cholesterol of patient at study entry.

age Age of the patient at study entry.

bmi Patient body mass index.

BPVar Average units of systolic and diastolic blood pressure above normal: ((SystolicBP-120)/2) + (DiasystolicBP-80)

heartrte Patient heartrate taken at study entry.

glucose Patient blood glucose level.

cursmoke Whether or not the patient was a smoker at the time of study entry.

outcome Did the patient die, experience hypertension, or leave the study without experiencing either event.

time_outcome The time at which the patient experienced outcome.

cigpday Number of cigarettes smoked per day at time of study entry.

References

- Dawber TR, Meadors GF, Moore FE Jr. Epidemiological approaches to heart disease: the Framingham Study. Am J Public Health Nations Health. 1951;41(3):279-281.
- Teaching Datasets - Public Use Datasets. https://biolincc.nhlbi.nih.gov/teaching/.

---

lindner	*Lindner Center for Research and Education study on Abciximab cost-effectiveness and survival*

Description

The effects of Abciximab use on both survival and cardiac billing.

Format

A data frame with 996 rows and 10 columns:

lifepres Life years preserved post treatment: 0 (died) vs. 11.6 (survived).

cardbill Cardiac related billing in dollars within 12 months.

abcix Indicates whether the patient received Abciximab treatment: 1=yes 0=no.

stent Was a stent depolyed? 1=yes, 0=no.

height Patient height in centimeters.

female Patient sex: 1=female, 0=male.

diabetic Was the patient diabetic? 1=yes, 0=no.

acutemi Had the patient suffered an acute myocardial infarction witih the last seven days? 1=yes, 0=no.

ejecfrac Left ventricular ejection fraction.

ves1proc Number of vessels involved in the first PCI procedure.

References

Kereiakes DJ, Obenchain RL, Barber BL, Smith A, McDonald M, Broderick TM, Runyon JP, Shimshak TM, Schneider JF, Hattemer CR, Roth EM, Whang DD, Cocks D, Abbottsmith CW. Abciximab provides cost-effective survival advantage in high-volume interventional practice. Am Heart J. 2000;140(4):603-610.

LocalControl	*Local Control*

Description

Implements a non-parametric methodology for correcting biases when comparing the outcomes of two treatments in a cross-sectional or case control observational study. This implementation of Local Control uses nearest neighbors to each point within a given radius to compare treatment outcomes. Local Control matches along a continuum of similarity (radii), clustering the near neighbors to a given observation by variables thought to be sources of bias and confounding. This is analogous to combining a host of smaller studies that are each homogeneous within themselves, but represent the spectrum of variability of observations across diverse subpopulations. As the clusters

get smaller, some of them can become noninformative, whereby all cluster members contain only one treatment, and there is no basis for comparison. Each observation has a unique set of near-neighbors, and the approach becomes more akin to a non-parametric density estimate using similar observations within a covariate hypersphere of a given radius. The global treatment difference is taken as the average of the treatment differences of the neighborhood around each observation.

While `LocalControlClassic` uses the number of clusters as a varying parameter to visualize treatment differences as a function of similarity of observations, this function instead uses a varying radius. The maximum radius enclosing all observations corresponds to the biased estimate which compares the outcome of all those with treatment A versus all those with treatment B. An easily interpretable graph can be created to illustrate the change in estimated outcome difference between two treatments, on average, across all clusters, as a function of using smaller and more homogenous clusters. The `LocalControlNearestNeighborsConfidence` procedure statistically resamples this Local Control process to generate confidence estimates. It is also helpful to plot a box-plot of the local treatment difference at a radius of zero, requiring that every observation has at least one perfect match on the other treatment. When perfect matches exist, one can estimate the treatment difference without making assumptions about the relative importance of the clustering variables. The `plot.LocalControlCS` function will plot both visualizations in a single graph.

Usage

```
LocalControl(
  data,
  modelForm = NULL,
  outcomeType = "default",
  treatmentColName,
  outcomeColName,
  cenCode = 0,
  clusterVars,
  timeColName = "",
  treatmentCode,
  labelColName = "",
  radStepType = "exp",
  radDecayRate = 0.8,
  radMinFract = 0.01,
  radiusLevels = numeric(),
  normalize = TRUE,
  verbose = FALSE,
  numThreads = 1
)
```

Arguments

`data`	DataFrame containing all variables which will be used for the analysis.
`modelForm`	A formula containing the necessary variables for Local Control analysis. This can be used as an alternative to the primary interface for cross-sectional studies. The formula should be in the following format: "outcome ~ treatment | clusterVar1 ... clusterVarN".
`outcomeType`	Specifys the outcome type for the analysis.

treatmentColName

A string containing the name of a column in data. The column contains the treatment variable specifying the treatment groups.

outcomeColName   A string containing the name of a column in data. The column contains the outcome variable to be compared between the treatment groups.

cenCode   A value specifying which of the outcome values corresponds to a censored observation.

clusterVars   A character vector containing column names in data. Each column contains an X-variable, or covariate which will be used to form patient clusters.

timeColName   A string containing the name of a column in data. The column contains the time to outcome for each of the observations in data.

treatmentCode   (optional) A string containing one of the factor levels from the treatment column. If provided, the corresponding treatment will be considered "Treatment 1". Otherwise, the first "level" of the column will be considered the primary treatment.

labelColName   (optional) A string containing the name of a column from data. The column contains labels for each of the observations in data, defaults to the row indices.

radStepType   (optional) Used in the generation of correction radii. The step type used to generate each correction radius after the maximum. Currently accepts "unif" and "exp" (default). "unif" for uniform decay ex: (radDecayRate = 0.1) (1, 0.9, 0.8, 0.7, ..., ~minRadFract, 0) "exp" for exponential decay ex: (radDecayRate = 0.9) (1, 0.9, 0.81, 0.729, ..., ~minRadFract, 0)

radDecayRate   (optional) Used in the generation of correction radii. The size of the "step" between each of the generated correction radii. If radStepType == "exp", radDecayRate must be a value between (0,1). This value defaults to 0.8.

radMinFract   (optional) Used in the generation of correction radii. A floating point number representing the smallest fraction of the maximum radius to use as a correction radius.

radiusLevels   (optional) By default, Local Control builds a set of radii to fit data. The radiusLevels parameter allows users to override the construction by explicitly providing a set of radii.

normalize   (optional) Logical value. Tells local control if it should or should not normalize the covariates. Default is TRUE.

verbose   (optional) Logical value. Display or suppress the console output during the call to Local Control. Default is FALSE.

numThreads   (optional) An integer value specifying the number of threads which will be assigned to the analysis. The maximum number of threads varies depending on the system hardware. Defaults to 1 thread.

Value

A list containing the results from the call to LocalControl.

- outcomes List containing two dataframes for the average T1 and T0 outcomes within each cluster at each radius.

- counts List containing two dataframes which hold the number of T1 and T0 patients within each cluster at each radius.

- ltds Dataframe containing the average LTD within each cluster at each radius.

- summary Dataframe containing summary statistics about the analysis for each radius.

- params List containing the parameters used to call LocalControl.

References

- Lauve NR, Nelson SJ, Young SS, Obenchain RL, Lambert CG. LocalControl: An R Package for Comparative Safety and Effectiveness Research. Journal of Statistical Software. 2020. p. 1–32. Available from: http://dx.doi.org/10.18637/jss.v096.i04

- Fischer K, Gartner B, Kutz M. Fast Smallest-Enclosing-Ball Computation in High Dimensions. In: Algorithms - ESA 2003. Springer, Berlin, Heidelberg; 2003:630-641.

- Martin Kutz, Kaspar Fischer, Bernd Gartner. miniball-1.0.3. https://github.com/hbf/miniball.

Examples

```
# cross-sectional

data(lindner)
linVars <- c("stent", "height", "female", "diabetic", "acutemi",
             "ejecfrac", "ves1proc")
csresults = LocalControl(data = lindner,
                         clusterVars = linVars,
                         treatmentColName = "abcix",
                         outcomeColName = "cardbill",
                         treatmentCode = 1)
plot(csresults)


# survival / competing risks example

data(cardSim)
crresults = LocalControl(data = cardSim, outcomeType = "survival",
                         outcomeColName = "status",
                         timeColName = "time",
                         treatmentColName = "drug",
                         treatmentCode = 1,
                         clusterVars = c("age", "bmi"))
plot(crresults)
```

```
LocalControl-deprecated
```
Deprecated LocalControl functions

Description

These functions are provided for compatibility with previous versions of LocalControl. They may eventually be completely removed.

Details

localControlNearestNeighbors	Now called using LocalControl with the outcomeType = "cross-sectional".
localControlCompetingRisks	Now called using LocalControl with the outcomeType = "survival".
plotLocalControlCIF	Now called using plot.LocalControlCR.
plotLocalControlLTD	Now called using plot.LocalControlCS.

LocalControlClassic   *Local Control Classic*

Description

LocalControlClassic was originally contained in the deprecated CRAN package USPS, this function is a combination of three of the original USPS functions, UPShclus, UPSaccum, and UP-Snnltd. This replicates the original implementation of the Local Control functionality in Robert Obenchain's USPS package. Some of the features have been removed due to deprecation of R packages distributed through CRAN. For a given number of patient clusters in baseline X-covariate space, LocalControlClassic() characterizes the distribution of Nearest Neighbor "Local Treatement Differences" (LTDs) on a specified Y-outcome variable.

Usage

```
LocalControlClassic(
  data,
  clusterVars,
  treatmentColName,
  outcomeColName,
  faclev = 3,
  scedas = "homo",
  clusterMethod = "ward",
  clusterDist = "euclidean",
  clusterCounts = c(50, 100, 200)
)
```

Arguments

data	The data frame containing all baseline X covariates.
clusterVars	List of names of X variable(s).

treatmentColName	
	Name of treatment factor variable.
outcomeColName	Name of outcome Y variable.
faclev	Maximum number of different numerical values an outcome variable can assume without automatically being converted into a "factor" variable; faclev=1 causes a binary indicator to be treated as a continuous variable determining an average or proportion.
scedas	Scedasticity assumption: "homo" or "hete".
clusterMethod	Type of clustering method, defaults to "complete". Currently implemented methods: "ward", "single", "complete" or "average".
clusterDist	Distance type to use, defaults to "euclidean". Currently implemented: "euclidean", "manhattan", "maximum", or "minkowski".
clusterCounts	A vector containing different number of clusters in baseline X-covariate space which Local Control will iterate over.

Value

Returns a list containing several elements.

hiclus	Name of clustering object created by UPShclus().
dframe	Name of data.frame containing X, t & Y variables.
trtm	Name of treatment factor variable.
yvar	Name of outcome Y variable.
numclust	Number of clusters requested.
actclust	Number of clusters actually produced.
scedas	Scedasticity assumption: "homo" or "hete"
PStdif	Character string describing the treatment difference.
nnhbindf	Vector containing cluster number for each patient.
rawmean	Unadjusted outcome mean by treatment group.
rawvars	Unadjusted outcome variance by treatment group.
rawfreq	Number of patients by treatment group.
ratdif	Unadjusted mean outcome difference between treatments.
ratsde	Standard error of unadjusted mean treatment difference.
binmean	Unadjusted mean outcome by cluster and treatment.
binvars	Unadjusted variance by cluster and treatment.
binfreq	Number of patients by bin and treatment.
awbdif	Across cluster average difference with cluster size weights.
awbsde	Standard error of awbdif.
wwbdif	Across cluster average difference, inverse variance weights.
wwbsde	Standard error of wwbdif.

faclev	Maximum number of different numerical values an outcome variable can assume without automatically being converted into a "factor" variable; faclev=1 causes a binary indicator to be treated as a continuous variable determining an average or proportion.
youtype	"continuous" => only next eight outputs; "factor" => only last three outputs.
aovdiff	ANOVA summary for treatment main effect only.
form2	Formula for outcome differences due to bins and to treatment nested within bins.
bindiff	ANOVA summary for treatment nested within cluster.
sig2	Estimate of error mean square in nested model.
pbindif	Unadjusted treatment difference by cluster.
pbinsde	Standard error of the unadjusted difference by cluster.
pbinsiz	Cluster radii measure: square root of total number of patients.
symsiz	Symbol size of largest possible Snowball in a UPSnnltd() plot with 1 cluster.
factab	Marginal table of counts by Y-factor level and treatment.
cumchi	Cumulative Chi-Square statistic for interaction in the three-way, nested table.
cumdf	Degrees of-Freedom for the Cumulative Chi-Squared.

References

- Obenchain, RL. *USPS package: Unsupervised and Supervised Propensity Scoring in R.* https://cran.r-project.org/src/contrib/Archive/USPS/ 2005.

- Obenchain, RL. *The "Local Control" Approach to Adjustment for Treatment Selection Bias and Confounding (illustrated with JMP Scripts).* Observational Studies. Cary, NC: SAS Press. 2009.

- Obenchain RL. The local control approach using JMP. In: Faries D, Leon AC, Haro JM, Obenchain RL, eds. Analysis of Observational Health Care Data Using SAS. Cary, NC: SAS Institute; 2010:151-194.

- Obenchain RL, Young SS. Advancing statistical thinking in observational health care research. J Stat Theory Pract. 2013;7(2):456-506.

- Faries DE, Chen Y, Lipkovich I, Zagar A, Liu X, Obenchain RL. Local control for identifying subgroups of interest in observational research: persistence of treatment for major depressive disorder. Int J Methods Psychiatr Res. 2013;22(3):185-194.

- Lopiano KK, Obenchain RL, Young SS. Fair treatment comparisons in observational research. Stat Anal Data Min. 2014;7(5):376-384.

- Young SS, Obenchain RL, Lambert CG (2016) A problem of bias and response heterogeneity. In: Alan Moghissi A, Ross G (eds) Standing with giants: A collection of public health essays in memoriam to Dr. Elizabeth M. Whelan. American Council on Science and Health, New York, NY, pp 153-169.

Examples

```
data(lindner)

cvars <- c("stent","height","female","diabetic","acutemi",
           "ejecfrac","ves1proc")
numClusters <- c(1, 2, 10, 15, 20, 25, 30, 35, 40, 45, 50)
results <- LocalControlClassic( data = lindner,
                                clusterVars = cvars,
                                treatmentColName = "abcix",
                                outcomeColName = "cardbill",
                                clusterCounts = numClusters)
UPSLTDdist(results,ylim=c(-15000,15000))
```

LocalControlCompetingRisksConfidence

Calculate confidence intervals around the cumulative incidence functions (CIFs) generated by LocalControl when outcomeType = "survival".

Description

Given the output of `LocalControl`, this function produces pointwise standard error estimates for the cumulative incidence functions (CIFs) using a modified version of Choudhury's approach (2002). This function currently supports the creation of 90%, 95%, 98%, and 99% confidence intervals with linear, log(-log), and arcsine transformations of the estimates.

Usage

```
LocalControlCompetingRisksConfidence(
  LCCompRisk,
  confLevel = "95%",
  confTransform = "asin"
)
```

Arguments

LCCompRisk	Output from a successful call to LocalControl with outcomeType = "survival".
confLevel	Level of confidence with which the confidence intervals will be formed. Choices are: "90%", "95%", "98%", "99%".
confTransform	Transformation of the confidence intervals, defaults to arcsin ("asin"). "log" and "linear" are also implemented.

References

- Lauve NR, Nelson SJ, Young SS, Obenchain RL, Lambert CG. LocalControl: An R Package for Comparative Safety and Effectiveness Research. Journal of Statistical Software. 2020. p. 1–32. Available from: http://dx.doi.org/10.18637/jss.v096.i04
- Choudhury JB (2002) Non-parametric confidence interval estimation for competing risks analysis: application to contraceptive data. Stat Med 21:1129-1144. doi: 10.1002/sim.1070

Examples

```
data(cardSim)
results = LocalControl(data = cardSim,
                          outcomeType = "survival",
                          outcomeColName = "status",
                          timeColName = "time",
                          treatmentColName = "drug",
                          treatmentCode = 1,
                          clusterVars = c("age", "bmi"))

conf = LocalControlCompetingRisksConfidence(results)
```

LocalControlNearestNeighborsConfidence

Provides a bootstrapped confidence interval estimate for LocalControl LTDs.

Description

Given a number of bootstrap iterations and the params used to call [LocalControl](#) with outcome-Type = "default", this function calls LocalControl nBootstrap times. The 50% and 95% quantiles are drawn from the distribution of results to produce the LTD confidence intervals.

Usage

```
LocalControlNearestNeighborsConfidence(
  data,
  nBootstrap,
  randSeed,
  treatmentColName,
  treatmentCode = "",
  outcomeColName,
  clusterVars,
  labelColName = "",
  numThreads = 1,
  radiusLevels = numeric(),
  radStepType = "exp",
  radDecayRate = 0.8,
```

```
    radMinFract = 0.01,
    normalize = TRUE,
    verbose = FALSE
)
```

Arguments

data	DataFrame containing all variables which will be used for the analysis.
nBootstrap	The number of times to resample and run LocalControl for the confidence intervals.
randSeed	The seed used to set random number generator state prior to resampling. No default value, provide one for reproducible results.
treatmentColName	
	A string containing the name of a column in data. The column contains the treatment variable specifying the treatment groups.
treatmentCode	(optional) A string containing one of the factor levels from the treatment column. If provided, the corresponding treatment will be considered "Treatment 1". Otherwise, the first "level" of the column will be considered the primary treatment.
outcomeColName	A string containing the name of a column in data. The column contains the outcome variable to be compared between the treatment groups. If outcomeType = "survival", the outcome column holds the failure/censor assignments.
clusterVars	A character vector containing column names in data. Each column contains an X-variable, or covariate which will be used to form patient clusters.
labelColName	(optional) A string containing the name of a column from data. The column contains labels for each of the observations in data, defaults to the row indices.
numThreads	(optional) An integer value specifying the number of threads which will be assigned to the analysis. The maximum number of threads varies depending on the system hardware. Defaults to 1 thread.
radiusLevels	(optional) By default, Local Control builds a set of radii to fit data. The radiusLevels parameter allows users to override the construction by explicitly providing a set of radii.
radStepType	(optional) Used in the generation of correction radii. The step type used to generate each correction radius after the maximum. Currently accepts "unif" and "exp" (default). "unif" for uniform decay ex: (radDecayRate = 0.1) (1, 0.9, 0.8, 0.7, ..., ~minRadFract, 0) "exp" for exponential decay ex: (radDecayRate = 0.9) (1, 0.9, 0.81, 0.729, ..., ~minRadFract, 0)
radDecayRate	(optional) Used in the generation of correction radii. The size of the "step" between each of the generated correction radii. If radStepType == "exp", radDecayRate must be a value between (0,1). This value defaults to 0.8.
radMinFract	(optional) Used in the generation of correction radii. A floating point number representing the smallest fraction of the maximum radius to use as a correction radius.
normalize	(optional) Logical value. Tells local control if it should or should not normalize the covariates. Default is TRUE.

verbose   (optional) Logical value. Display or suppress the console output during the call
                   to Local Control. Default is FALSE.

References

- Lauve NR, Nelson SJ, Young SS, Obenchain RL, Lambert CG. LocalControl: An R Package
  for Comparative Safety and Effectiveness Research. Journal of Statistical Software. 2020. p.
  1–32. Available from: http://dx.doi.org/10.18637/jss.v096.i04

- Kereiakes DJ, Obenchain RL, Barber BL, Smith A, McDonald M, Broderick TM, Runyon JP,
  Shimshak TM, Schneider JF, Hattemer CR, Roth EM, Whang DD, Cocks D, Abbottsmith CW.
  Abciximab provides cost-effective survival advantage in high-volume interventional practice.
  Am Heart J. 2000 Oct;140(4):603–610. PMID: 11011333

Examples

```
## Not run:
#input the abciximab study data of Kereiakes et al. (2000).
data(lindner)

linVars <- c("stent", "height", "female", "diabetic", "acutemi",
             "ejecfrac", "ves1proc")
results <- LocalControl(data = lindner,
                        clusterVars = linVars,
                        treatmentColName = "abcix",
                        outcomeColName = "cardbill",
                        treatmentCode = 1)

#Calculate the confidence intervals via resampling.
confResults = LocalControlNearestNeighborsConfidence(
                                 data = lindner,
                                 clusterVars = linVars,
                                 treatmentColName = "abcix",
                                 outcomeColName = "cardbill",
                                 treatmentCode = 1, nBootstrap = 20)

# Plot the local treatment difference with confidence intervals.
plot(results, confResults)

## End(Not run)
```

---

plot.LocalControlCR   *Plot cumulative incidence functions (CIFs) from Local Control.*

Description

Given the results from LocalControl with outcomeType = "survival", plot a corrected and uncor-
rected cumulative incidence function (CIF) for both groups.

Usage

```
## S3 method for class 'LocalControlCR'
plot(
  x,
  ...,
  rad2plot,
  xlim,
  ylim = c(0, 1),
  col1 = "blue",
  col0 = "red",
  xlab = "Time",
  ylab = "Cumulative incidence",
  legendLocation = "topleft",
  main = "",
  group1 = "Treatment 1",
  group0 = "Treatment 0"
)
```

Arguments

x	Return object from LocalControl with outcomeType = "survival".
...	Arguments passed on to `graphics::plot`
rad2plot	The index or name ("rad_#") of the radius to plot. By default, the radius with pct_informative closest to 0.8 will be selected.
xlim	The x axis bounds. Defaults to c(0, max(lccrResults$Failtimes)).
ylim	The y axis bounds. Defaults to c(0,1).
col1	The plot color for group 1.
col0	The plot color for group 0.
xlab	The x axis label. Defaults to "Time".
ylab	The y axis label. Defaults to "Cumulative incidence".
legendLocation	The location to place the legend. Default "topleft".
main	The main plot title. Default is empty.
group1	The name of the primary group (Treatment 1).
group0	The name of the secondary group (Treatment 0).

References

- Lauve NR, Nelson SJ, Young SS, Obenchain RL, Lambert CG. LocalControl: An R Package for Comparative Safety and Effectiveness Research. Journal of Statistical Software. 2020. p. 1–32. Available from: http://dx.doi.org/10.18637/jss.v096.i04

Examples

```
data("cardSim")
results = LocalControl(data = cardSim,
                       outcomeType = "survival",
                       outcomeColName = "status",
                       timeColName = "time",
                       treatmentColName = "drug",
                       treatmentCode = 1,
                       clusterVars = c("age", "bmi"))
plot(results)
```

---

plot.LocalControlCS	*Plots the local treatment difference as a function of radius for Local-Control.*

Description

Creates a plot where the y axis represents the local treatment difference, while the x axis represents the percentage of the maximum radius. If the confidence summary (nnConfidence) is provided, the 50% and 95% confidence estimates are also plotted.

Usage

```
## S3 method for class 'LocalControlCS'
plot(
  x,
  ...,
  nnConfidence,
  ylim,
  legendLocation = "bottomleft",
  ylab = "LTD",
  xlab = "Fraction of maximum radius",
  main = ""
)
```

Arguments

x	Return object from LocalControl with "default" outcomeType.
...	Arguments passed on to graphics::plot
nnConfidence	Return object from LocalControlNearestNeighborsConfidence
ylim	The y axis bounds. Defaults to c(0,1).
legendLocation	The location to place the legend. Default "topleft".
ylab	The y axis label. Defaults to "LTD".
xlab	The x axis label. Defaults to "Fraction of maximum radius".
main	The main plot title. Default is empty.

References

- Lauve NR, Nelson SJ, Young SS, Obenchain RL, Lambert CG. LocalControl: An R Package for Comparative Safety and Effectiveness Research. Journal of Statistical Software. 2020. p. 1–32. Available from: http://dx.doi.org/10.18637/jss.v096.i04

Examples

```
data(lindner)
# Specify clustering variables.
linVars <- c("stent", "height", "female", "diabetic",
             "acutemi", "ejecfrac", "ves1proc")

# Call Local Control once.
linRes <- LocalControl(data = lindner,
                       clusterVars = linVars,
                       treatmentColName = "abcix",
                       outcomeColName = "cardbill",
                       treatmentCode = 1)

# Plot the local treatment differences from Local Control without
# confidence intervals.
plot(linRes, ylim =  c(-6000, 3600))

#If the confidence intervals are calculated:
#linConfidence = LocalControlNearestNeighborsConfidence(
#                                    data = lindner,
#                                    clusterVars = linVars,
#                                    treatmentColName = "abcix",
#                                    outcomeColName = "cardbill",
#                                    treatmentCode = 1, nBootstrap = 100)

# Plot the local treatment difference with confidence intervals.
#plot(linRes, linConfidence)
```

SPSbalan | *Test for Within-Bin X-covariate Balance in Supervised Propensiy Scoring*

Description

Test for Conditional Independence of X-covariate Distributions from Treatment Selection within Given, Adjacent PS Bins. The second step in Supervised Propensity Scoring analyses is to verify that baseline X-covariates have the same distribution, regardless of treatment, within each fitted PS bin.

Usage

```
SPSbalan(envir, dframe, trtm, yvar, qbin, xvar, faclev = 3)
```

Arguments

`envir`	The local control environment
`dframe`	Name of augmented data.frame written to the appn="" argument of SPSlogit().
`trtm`	Name of the two-level treatment factor variable.
`yvar`	The outcome variable.
`qbin`	Name of variable containing bin numbers.
`xvar`	Name of one baseline covariate X variable used in the SPSlogit() PS model.
`faclev`	Maximum number of different numerical values an X-covariate can assume without automatically being converted into a "factor" variable; faclev=1 causes a binary indicator to be treated as a continuous variable determining a proportion.

Value

An output list object of class SPSbalan. The first four are returned with a continuous x-variable. The next 4 are used if it is a factor variable.

- aovdiffANOVA output for marginal test.

- form2Formula for differences in X due to bins and to treatment nested within bins.

- bindiffANOVA output for the nested within bin model.

- df3Output data.frame containing 3 variables: X-covariate, treatment and bin.

- factabMarginal table of counts by X-factor level and treatment.

- tabThree-way table of counts by X-factor level, treatment and bin.

- cumchiCumulative Chi-Square statistic for interaction in the three-way, nested table.

- cumdfDegrees of-Freedom for the Cumulative Chi-Squared.

Author(s)

Bob Obenchain <wizbob@att.net>

References

- Cochran WG. (1968) The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics* **24**: 205-213.

- Obenchain RL. (2011) **USPSinR.pdf** USPS R-package vignette, 40 pages.

- Rosenbaum PR, Rubin RB. (1983) The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* **70**: 41-55.

- Rosenbaum PR, Rubin DB. (1984) Reducing Bias in Observational Studies Using Subclassification on a Propensity Score. *J Amer Stat Assoc* **79**: 516-524.

SPSloess	*LOESS Smoothing of Outcome by Treatment in Supervised Propensiy Scoring*

Description

Express Expected Outcome by Treatment as LOESS Smooths of Fitted Propensity Scores.

Usage

```
SPSloess(
  envir,
  dframe,
  trtm,
  pscr,
  yvar,
  faclev = 3,
  deg = 2,
  span = 0.75,
  fam = "symmetric"
)
```

Arguments

envir	Local control classic environment.
dframe	data.frame of the form returned by SPSlogit().
trtm	the two-level factor on the left-hand-side in the formula argument to SPSlogit().
pscr	fitted propensity scores of the form returned by SPSlogit().
yvar	continuous outcome measure or result unknown at the time patient was assigned (possibly non-randomly) to treatment; "NA"s are allowed in yvar.
faclev	optional; maximum number of distinct numerical values a variable can assume and yet still be converted into a factor variable; faclev=1 causes a binary indicator to be treated as a continuous variable determining a proportion.
deg	optional; degree (1=linear or 2=quadratic) of the local fit.
span	optional; span (0 to 2) argument for the loess() function.
fam	optional; "gaussian" or "symmetric".

Details

SPSloess

Once one has fitted a somewhat smooth curve through scatters of observed outcomes, Y, versus the fitted propensity scores, X, for the patients in each of the two treatment groups, one can consider the question: "Over the range where both smooth curves are defined (i.e. their common support), what is the (weighted) average signed difference between these two curves?"

If the distribution of patients (either treated or untreated) were UNIFORM over this range, the (unweighted) average signed difference (treated minus untreated) would be an appropriate estimate of the overall difference in outcome due to choice of treatment.

Histogram patient counts within 100 cells of width 0.01 provide a naive "non-parametric density estimate" for the distribution of total patients (treated or untreated) along the propensity score axis. The weighted average difference (and standard error) displayed by SPSsmoot() are based on an R density() smooth of these counts.

In situations where the propensity scoring distribution for all patients in a therapeutic class is known to differ from that of the patients within the current study, that population weighted average would also be of interest. Thus the SPSloess() output object contains two data frames, logrid and lofit, useful in further computations.

- logridloess grid data.frame containing 11 variables and 100 observations. The PS variable contains propensity score "cell means" of 0.005 to 0.995 in steps of 0.010. Variables F0, S0 and C0 for treatment 0 and variables F1, S1 and C1 for treatment 1 contain fitted smooth spline values, standard error estimates and patient counts, respectively. The DIF variable is simply (F1\-F0), the SED variable is sqrt(S1\^2+S0\^2), the HST variable is proportional to (C0+C1), and the DEN variable is the estimated probability density of patients along the PS axis. Observations with "NA" for variables F0, S0, F1 or S1 represent "extremes" where the lowess fits could not be extrapolated because no observed outcomes were available.

- losub0, losub1loess fit data.frame contains 4 variables for each distinct PS value in lofit. These 4 variables are named PS, YAVG, TRT==0 and 1, respectively, and FIT = spline prediction for the specified degrees-of-freedom (default df=1.)

- spanloess span setting.

- lotdifoutcome treatment difference mean.

- lotsdeoutcome treatment difference standard deviation.

Author(s)

Bob Obenchain <wizbob@att.net>

References

Cleveland WS, Devlin SJ. (1988) Locally-weighted regression: an approach to regression analysis by local fitting. *J Amer Stat Assoc* **83**: 596-610.

Cleveland WS, Grosse E, Shyu WM. (1992) Local regression models. Chapter 8 of **Statistical Models in S** eds Chambers JM and Hastie TJ. *Wadsworth & Brooks/Cole*.

Obenchain RL. (2011) **USPSinR.pdf** USPS R-package vignette, 40 pages.

Ripley BD, loess() based on the 'cloess' package of Cleveland, Grosse and Shyu.

SPSlogit	*Propensity Score prediction of Treatment Selection from Patient Baseline X-covariates*

Description

Use a logistic regression model to predict Treatment Selection from Patient Baseline X-covariates in Supervised Propensity Scoring.

Usage

```
SPSlogit(envir, dframe, form, pfit, prnk, qbin, bins = 5, appn = "")
```

Arguments

envir	name of the working local control classic environment.
dframe	data.frame containing X, t and Y variables.
form	Valid formula for glm()with family = binomial(), with the two-level treatment factor variable as the left-hand-side of the formula.
pfit	Name of variable to store PS predictions.
prnk	Name of variable to store tied-ranks of PS predictions.
qbin	Name of variable to store the assigned bin number for each patient.
bins	optional; number of adjacent PS bins desired; default to 5.
appn	optional; append the pfit, prank and qbin variables to the input dfname when appn=="", else save augmented data.frame to name specified within a non-blank appn string.

Details

The first phase of Supervised Propensity Scoring is to develop a logit (or probit) model predicting treatment choice from patient baseline X characteristics. SPSlogit uses a call to glm()with family = binomial() to fit a logistic regression.

Value

An output list object of class SPSlogit:

- dframeName of input data.frame containing X, t & Y variables.
- dfoutnamName of output data.frame augmented by pfit, prank and qbin variables.
- trtmName of two-level treatment factor variable.
- formglm() formula for logistic regression.
- pfitName of predicted PS variable.
- prankName of variable containing PS tied-ranks.
- qbinName of variable containing assigned PS bin number for each patient.
- binsNumber of adjacent PS bins desired.
- glmobjOutput object from invocation of glm() with family = binomial().

Author(s)

Bob Obenchain <wizbob@att.net>

References

Cochran WG. (1968) The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics* **24**: 205-213.

Kereiakes DJ, Obenchain RL, Barber BL, et al. (2000) Abciximab provides cost effective survival advantage in high volume interventional practice. *Am Heart J* **140**: 603-610.

Obenchain RL. (2011) **USPSinR.pdf** USPS R-package vignette, 40 pages.

Rosenbaum PR, Rubin RB. (1983) The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* **70**: 41-55.

Rosenbaum PR, Rubin DB. (1984) Reducing Bias in Observational Studies Using Subclassification on a Propensity Score. *J Amer Stat Assoc* **79**: 516-524.

See Also

SPSbalan, SPSnbins and SPSoutco.

SPSnbins	*Change the Number of Bins in Supervised Propensiy Scoring*

Description

Change the Number of Bins in Supervised Propensiy Scoring

Usage

```
SPSnbins(envir, dframe, prnk, qbin, bins = 8)
```

Arguments

envir	name of the working local control classic environment.
dframe	Name of data.frame of the form output by SPSlogit().
prnk	Name of PS tied-rank variable from previous call to SPSlogit().
qbin	Name of variable to contain the re-assigned bin number for each patient.
bins	Number of PS bins desired.

Details

Part or all of the first phase of Supervised Propensity Scoring will need to be redone if SPSbalan() detects dependence of within-bin X-covariate distributions upon treatment choice. Use SPSnbins() to change (increase) the number of adjacent PS bins. If this does not achieve balance, invoke SPSlogit() again to modify the form of your PS logistic model, typically by adding interaction and/or curvature terms in continuous X-covariates.

Value

An output data.frame with new variables inserted:

- dframe2Modified version of the data.frame specified as the first argument to SPSnbins().

Author(s)

Bob Obenchain <wizbob@att.net>

References

Cochran WG. (1968) The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics* **24**: 205-213.

Obenchain RL. (2011) **USPSinR.pdf** USPS R-package vignette, 40 pages.

Rosenbaum PR, Rubin DB. (1984) Reducing Bias in Observational Studies Using Subclassification on a Propensity Score. *J Amer Stat Assoc* **79**: 516-524.

See Also

SPSlogit, SPSbalan and SPSoutco.

SPSoutco	*Examine Treatment Differences on an Outcome Measure in Supervised Propensiy Scoring*

Description

Examine Within-Bin Treatment Differences on an Outcome Measure and Average these Differences across Bins.

Usage

```
SPSoutco(envir, dframe, trtm, qbin, yvar, faclev = 3)
```

Arguments

envir	name of the working local control classic environment.
dframe	Name of augmented data.frame written to the appn="" argument of SPSlogit().
trtm	Name of treatment factor variable.
qbin	Name of variable containing the PS bin number for each patient.
yvar	Name of an outcome Y variable.
faclev	Maximum number of different numerical values an X-covariate can assume without automatically being converted into a "factor" variable; faclev=1 causes a binary indicator to be treated as a continuous variable determining an average or proportion.

Details

Once the second phase of Supervised Propensity Scoring confirms, using SPSbalan(), that X-covariate Distributions have been Balanced Within-Bins, the third phase can start: Examining Within-Bin Outcome Difference due to Treatment and Averaging these Differences across Bins. Graphical displays of SPSoutco() results feature R barplot() invocations.

Value

An output list object of class SPSoutco:

- dframeName of augmented data.frame written to the appn="" argument of SPSlogit().
- trtmName of the two-level treatment factor variable.
- yvarName of an outcome Y variable.
- binsNumber of variable containing bin numbers.
- PStdifCharacter string describing the treatment difference.
- rawmeanUnadjusted outcome mean by treatment group.
- rawvarsUnadjusted outcome variance by treatment group.
- rawfreqNumber of patients by treatment group.
- ratdifUnadjusted mean outcome difference between treatments.
- ratsdeStandard error of unadjusted mean treatment difference.
- binmeanUnadjusted mean outcome by cluster and treatment.
- binvarsUnadjusted variance by cluster and treatment.
- binfreqNumber of patients by bin and treatment.
- awbdifAcross cluster average difference with cluster size weights.
- awbsdeStandard error of awbdif.
- wwbdifAcross cluster average difference, inverse variance weights.
- wwbsdeStandard error of wwbdif.
- formFormula for overall, marginal treatment difference on X-covariate.
- faclevMaximum number of different numerical values an X-covariate can assume without automatically being converted into a "factor" variable; faclev=1 causes a binary indicator to be treated as a continuous variable determining an average or proportion.
- youtype"contin"uous => only next six outputs; "factor" => only last four outputs.
- aovdiffANOVA output for marginal test.
- form2Formula for differences in X due to bins and to treatment nested within bins.
- bindiffANOVA summary for treatment nested within bin.
- pbindifUnadjusted treatment difference by cluster.
- pbinsdeStandard error of the unadjusted difference by cluster.
- pbinsizCluster radii measure: square root of total number of patients.
- factabMarginal table of counts by Y-factor level and treatment.
- tabThree-way table of counts by Y-factor level, treatment and bin.
- cumchiCumulative Chi-Square statistic for interaction in the three-way, nested table.
- cumdfDegrees of-Freedom for the Cumulative Chi-Squared.

Author(s)

Bob Obenchain <wizbob@att.net>

References

Cochran WG. (1968) The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics* **24**: 205-213.

Obenchain RL. (2011) **USPSinR.pdf** USPS R-package vignette, 40 pages.

Rosenbaum PR, Rubin RB. (1983) The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* **70**: 41-55.

Rosenbaum PR, Rubin DB. (1984) Reducing Bias in Observational Studies Using Subclassification on a Propensity Score. *J Amer Stat Assoc* **79**: 516-524.

See Also

SPSlogit, SPSbalan and SPSnbins.

UPSaccum	*Prepare for Accumulation of (Outcome,Treatment) Results in Unsupervised Propensity Scoring*

Description

Specify key result accumulation parameters: Treatment t-Factor, Outcome Y-variable, faclev setting, scedasticity assumption, and name of the UPSgraph() data accumulation object.

Usage

```
UPSaccum(envir, dframe, trtm, yvar, faclev = 3, scedas = "homo")
```

Arguments

envir	name of the working local control classic environment.
dframe	Name of data.frame containing the X, t & Y variables.
trtm	Name of treatment factor variable.
yvar	Name of outcome Y variable.
faclev	Maximum number of different numerical values an outcome variable can assume without automatically being converted into a "factor" variable; faclev=1 causes a binary indicator to be treated as a continuous variable determining an average or proportion.
scedas	Scedasticity assumption: "homo" or "hete"

Details

The second phase in an Unsupervised Propensity Scoring analysis is to prepare to accumulate results over a wide range of values for "Number of Clusters." As the number of such clusters increases, individual clusters will tend to become smaller and smaller and, thus, more and more compact in covariate X-space.

Value

- hiclusName of a diana, agnes or hclust object created by UPShclus().

- dframeName of data.frame containing the X, t & Y variables.

- trtmName of treatment factor variable.

- yvarName of outcome Y variable.

- faclevMaximum number of different numerical values an outcome variable can assume without automatically being converted into a "factor" variable; faclev=1 causes a binary indicator to be treated as a continuous variable determining a proportion.

- scedasScedasticity assumption: "homo" or "hete"

- accobjName of the object for accumulation of I-plots to be ultimately displayed using UPS-graph().

- nnymaxMaximum NN LTD Standard Error observed; Upper NN plot limit; initialized to zero.

- nnxminMinimum NN LTD observed; Left NN plot limit; initialized to zero.

- nnxmaxMaximum NN LTD observed; Right NN plot limit; initialized to zero.

Author(s)

Bob Obenchain <wizbob@att.net>

References

Obenchain RL. (2004) Unsupervised Propensity Scoring: NN and IV Plots. *Proceedings of the American Statistical Association (on CD)* 8 pages.

Obenchain RL. (2011) **USPSinR.pdf** USPS R-package vignette, 40 pages.

See Also

UPSnnltd, UPSivadj and UPShclus.

UPSaltdd *Artificial Distribution of LTDs from Random Clusters*

Description

For a given number of clusters, UPSaltdd() characterizes the potentially biased distribution of "Local Treatment Differences" (LTDs) in a continuous outcome y-variable between two treatment groups due to Random Clusterings. When the NNobj argument is not NA and specifies an existing UPSnnltd() object, UPSaltdd() also computes a smoothed CDF for the NN/LTD distribution for direct comparison with the Artificial LTD distribution.

Usage

```
UPSaltdd(
  envir,
  dframe,
  trtm,
  yvar,
  faclev = 3,
  scedas = "homo",
  NNobj = NA,
  clus = 50,
  reps = 10,
  seed = 12345
)
```

Arguments

envir	name of the working local control classic environment.
dframe	Name of data.frame containing a treatment-factor and the outcome y-variable.
trtm	Name of treatment factor variable with two levels.
yvar	Name of continuous outcome variable.
faclev	Maximum number of different numerical values an outcome variable can assume without automatically being converted into a "factor" variable; faclev=1 causes a binary indicator to be treated as a continuous variable determining an average or proportion.
scedas	Scedasticity assumption: "homo" or "hete"
NNobj	Name of an existing UPSnnltd object or NA.
clus	Number of Random Clusters requested per Replication; ignored when NNobj is not NA.
reps	Number of overall Replications, each with the same number of requested clusters.
seed	Seed for Monte Carlo random number generator.

Details

Multiple calls to UPSaltdd() for different UPSnnltd objects or different numbers of clusters are typically made after first invoking UPSgraph().

Value

- dframeName of data.frame containing X, t & Y variables.
- trtmName of treatment factor variable.
- yvarName of outcome Y variable.
- faclevMaximum number of different numerical values an outcome variable can assume without automatically being converted into a "factor" variable; faclev=1 causes a binary indicator to be treated as a continuous variable determining an average or proportion.
- scedasScedasticity assumption: "homo" or "hete"
- NNobjName of an existing UPSnnltd object or NA.
- clusNumber of Random Clusters requested per Replication.
- repsNumber of overall Replications, each with the same number of requested clusters.
- patsNumber of patients with no NAs in their yvar outcome and trtm factor.
- seedSeed for Monte Carlo random number generator.
- altddMatrix of LTDs and relative weights from artificial clusters.
- alxminMinimum artificial LTD value.
- alxmaxMaximum artificial LTD value.
- alymaxMaximum weight among artificial LTDs.
- altdcdfVector of artificial LTD x-coordinates for smoothed CDF.
- qqVector of equally spaced CDF values from 0.0 to 1.0.
- nnltddOptional matrix of relevant NN/LTDs and relative weights.
- nnlxminOptional minimum NN/LTD value.
- nnlxmaxOptional maximum NN/LTD value.
- nnlymaxOptional maximum weight among NN/LTDs.
- nnltdcdfOptional vector of NN/LTD x-coordinates for smoothed CDF.
- nqOptional vector of equally spaced CDF values from 0.0 to 1.0.

Author(s)

Bob Obenchain <wizbob@att.net>

References

Obenchain RL. (2004) Unsupervised Propensity Scoring: NN and IV Plots. *Proceedings of the American Statistical Association (on CD)* 8 pages.

Obenchain RL. (2011) **USPSinR.pdf** USPS R-package vignette, 40 pages.

Rosenbaum PR, Rubin RB. (1983) The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* **70**: 41-55.

Rubin DB. (1980) Bias reduction using Mahalanobis metric matching. *Biometrics* **36**: 293-298.

See Also

[UPSnnltd](#), [UPSaccum](#) and [UPSgraph](#).

UPSboxplot	*Returns a series of boxplots comparing LTD distributions given different numbers of clusters.*

Description

Given the output of [LocalControlClassic](#), this function uses all or some of the UPSnnltd objects contained to create a series of boxplots of the local treatment difference at each of the different numbers of requested clusters.

Usage

```
UPSboxplot(envir, clusterSubset = c())
```

Arguments

envir   A LocalControlClassic environment containing UPSnnltd objects.

clusterSubset (optional) A vector containing requested cluster counts. If provided, the boxplot is created using only the UPSnnltd objects corresponding to the requested cluster counts.

Value

Returns the call to boxplot with the formula: "ltd ~ numclst".

Adds the "ltdds" object to the Local Control environment.

Examples

```
data(lindner)
cvars <- c("stent","height","female","diabetic","acutemi",
           "ejecfrac","ves1proc")
numClusters <- c(1, 5, 10, 20, 40, 50)

results <- LocalControlClassic(data = lindner,
                               clusterVars = cvars,
                               treatmentColName = "abcix",
                               outcomeColName = "cardbill",
                               clusterCounts = numClusters)

bxp <- UPSboxplot(results)
```

UPSgraph
Display Sensitivity Analysis Graphic in Unsupervised Propensiy Scoring

---

Description

Plot summary of results from multiple calls to UPSnnltd() and/or UPSivadj() after an initial setup call to UPSaccum(). The UPSgraph() plot displays any sensitivity of the LTD and LOA Distributions to choice of Number of Clusters in X-space.

Usage

```
UPSgraph(envir, nncol = "red", nwcol = "green3", ivcol = "blue", ...)
```

Arguments

envir	name of the working local control classic environment.
nncol	optional; string specifying color for display of the Mean of the LTD distribution when weighted by cluster size from any calls to UPSnnltd().
nwcol	optional; string specifying color for display of the Mean of the LTD distribution when weighted inversely proportional to variance from any calls to UPSnnltd().
ivcol	optional; string specifying color for display of the Difference in LOA predictions, at PS = 100% minus that at PS = 0%, from any calls to UPSivadj().
...	Additional arguments to pass to the plotting function.

Details

The third phase of Unsupervised Propensity Scoring is a graphical Sensitivity Analysis that depicts how the Overall Means of the LTD and LOA distributions change with the number of clusters.

Author(s)

Bob Obenchain <wizbob@att.net>

References

Kaufman L, Rousseeuw PJ. (1990) **Finding Groups in Data. An Introduction to Cluster Analysis**. *New York: John Wiley and Sons*.

Obenchain RL. (2004) Unsupervised Propensity Scoring: NN and IV Plots. *Proceedings of the American Statistical Association (on CD)* 8 pages.

Obenchain RL. (2011) **USPSinR.pdf** USPS R-package vignette, 40 pages.

Rubin DB. (1980) Bias reduction using Mahalanobis metric matching. *Biometrics* **36**: 293-298.

See Also

UPSnnltd, UPSivadj and UPSaccum.

UPShclus	*Hierarchical Clustering of Patients on X-covariates for Unsupervised Propensiy Scoring*

Description

Derive a full, hierarchical clustering tree (dendrogram) for all patients (regardless of treatment received) using Mahalonobis between-patient distances computed from specified baseline X-covariate characteristics.

Usage

```
UPShclus(envir, dframe, xvars, method, metric)
```

Arguments

envir	name of the working local control classic environment.
dframe	Name of data.frame containing baseline X covariates.
xvars	List of names of X variable(s).
method	Hierarchical Clustering Method: "diana", "agnes" or "hclus".
metric	A valid distance metric for clustering.

Details

The first step in an Unsupervised Propensity Scoring alalysis is always to hierarchically cluster patients in baseline X-covariate space. UPShclus uses a Mahalabobis metric and clustering methods from the R "cluster" library for this key initial step.

Value

An output list object of class UPShclus:

- dframeName of data.frame containing baseline X covariates.
- xvarsList of names of X variable(s).
- methodHierarchical Clustering Method: "diana", "agnes" or "hclus".
- upshclHierarchical clustering object created by choice between three possible methods.

Author(s)

Bob Obenchain <wizbob@att.net>

References

Kaufman L, Rousseeuw PJ. (1990) **Finding Groups in Data. An Introduction to Cluster Analysis**. New York: John Wiley and Sons.

Kereiakes DJ, Obenchain RL, Barber BL, et al. (2000) Abciximab provides cost effective survival advantage in high volume interventional practice. *Am Heart J* **140**: 603-610.

Obenchain RL. (2004) Unsupervised Propensity Scoring: NN and IV Plots. *Proceedings of the American Statistical Association (on CD)* 8 pages.

Obenchain RL. (2011) **USPSinR.pdf** USPS R-package vignette, 40 pages.

Rubin DB. (1980) Bias reduction using Mahalanobis metric matching. *Biometrics* **36**: 293-298.

See Also

UPSaccum, UPSnnltd and UPSgraph.

UPSivadj	*Instrumental Variable LATE Linear Fitting in Unsupervised Propensiy Scoring*

Description

For a given number of patient clusters in baseline X-covariate space and a specified Y-outcome variable, linearly smooth the distribution of Local Average Treatment Effects (LATEs) plotted versus Within-Cluster Treatment Selection (PS) Percentages.

Usage

```
UPSivadj(envir, numclust)
```

Arguments

envir	name of the working local control classic environment.
numclust	Number of clusters in baseline X-covariate space.

Details

Multiple calls to UPSivadj(n) for varying numbers of clusters n are made after first invoking UPShclus() to hierarchically cluster patients in X-space and then invoking UPSaccum() to specify a Y outcome variable and a two-level treatment factor t. UPSivadj(n) linearly smoothes the LATE distribution when plotted versus within cluster propensity score percentages.

Value

An output list object of class UPSivadj:

- hiclusName of clustering object created by UPShclus().
- dframeName of data.frame containing X, t & Y variables.
- trtmName of treatment factor variable.
- yvarName of outcome Y variable.
- numclustNumber of clusters requested.
- actclustNumber of clusters actually produced.
- scedasScedasticity assumption: "homo" or "hete"
- PStdifCharacter string describing the treatment difference.
- ivhbindfVector containing cluster number for each patient.
- rawmeanUnadjusted outcome mean by treatment group.
- rawvarsUnadjusted outcome variance by treatment group.
- rawfreqNumber of patients by treatment group.
- ratdifUnadjusted mean outcome difference between treatments.
- ratsdeStandard error of unadjusted mean treatment difference.
- binmeanUnadjusted mean outcome by cluster and treatment.
- binfreqNumber of patients by bin and treatment.
- faclevMaximum number of different numerical values an outcome variable can assume without automatically being converted into a "factor" variable; faclev=1 causes a binary indicator to be treated as a continuous variable determining an average or proportion.
- youtype"contin"uous => next eleven outputs; "factor" => no additional output items.
- pbinoutLATE regardless of treatment by cluster.
- pbinpspWithin-Cluster Treatment Percentage = non-parametric Propensity Score.
- pbinsizCluster radii measure: square root of total number of patients.
- symsizSymbol size of largest possible Snowball in a UPSivadj() plot with 1 cluster.
- ivfitlm() output for linear smooth across clusters.
- ivtzeroPredicted outcome at PS percentage zero.
- ivtxsdeStandard deviation of outcome prediction at PS percentage zero.
- ivtdiffPredicted outcome difference for PS percentage 100 minus that at zero.
- ivtdsdeStandard deviation of outcome difference.
- ivt100pPredicted outcome at PS percentage 100.
- ivt1pseStandard deviation of outcome prediction at PS percentage 100.

Author(s)

Bob Obenchain <wizbob@att.net>

References

Imbens GW, Angrist JD. (1994) Identification and Estimation of Local Average Treatment Effects (LATEs). *Econometrica* **62**: 467-475.

Obenchain RL. (2004) Unsupervised Propensity Scoring: NN and IV Plots. *Proceedings of the American Statistical Association (on CD)* 8 pages.

Obenchain RL. (2011) **USPSinR.pdf** USPS R-package vignette, 40 pages.-

McClellan M, McNeil BJ, Newhouse JP. (1994) Does More Intensive Treatment of Myocardial Infarction in the Elderly Reduce Mortality?: Analysis Using Instrumental Variables. *JAMA* **272**: 859-866.

Rosenbaum PR, Rubin RB. (1983) The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* **70**: 41-55.

See Also

UPSnnltd, UPSaccum and UPSgraph.

UPSLTDdist *Plot the LTD distribution as a function of the number of clusters.*

Description

This function creates a plot displaying the distribution of Local Treatment Differences (LTDs) as a function of the number of clusters created for all UPSnnltd objects in the provided environment. The hinges and whiskers are generated using boxplot.stats.

Usage

```
UPSLTDdist(envir, legloc = "bottomleft", ...)
```

Arguments

envir	A LocalControlClassic environment containing UPSnnltd objects.
legloc	Where to place the legend in the returned plot. Defaults to "bottomleft".
...	Arguments passed on to graphics::plot

Value

Returns the LTD distribution plot.

Adds the "ltdds" object to envir.

Examples

```
data(lindner)
cvars <- c("stent","height","female","diabetic","acutemi",
           "ejecfrac","ves1proc")
numClusters <- c(1, 2, 10, 15, 20, 25, 30, 35, 40, 45, 50)
results <- LocalControlClassic(data = lindner,
                               clusterVars = cvars,
                               treatmentColName = "abcix",
                               outcomeColName = "cardbill",
                               clusterCounts = numClusters)
UPSLTDdist(results,ylim=c(-15000,15000))
```

UPSnnltd	*Nearest Neighbor Distribution of LTDs in Unsupensiy Scoring*

Description

For a given number of patient clusters in baseline X-covariate space, UPSnnltd() characterizes the distribution of Nearest Neighbor "Local Treatemnt Differences" (LTDs) on a specified Y-outcome variable.

Usage

```
UPSnnltd(envir, numclust)
```

Arguments

envir   name of the working local control classic environment.

numclust   Number of clusters in baseline X-covariate space.

Details

Multiple calls to UPSnnltd(n) for varying numbers of clusters, n, are typically made after first invoking UPShclus() to hierarchically cluster patients in X-space and then invoking UPSaccum() to specify a Y outcome variable and a two-level treatment factor t. UPSnnltd(n) then determines the LTD Distribution corresponding to n clusters and, optionally, displays this distribution in a "Snowball" plot.

Value

An output list object of class UPSnnltd:

- hiclusName of clustering object created by UPShclus().
- dframeName of data.frame containing X, t & Y variables.

- trtmName of treatment factor variable.
- yvarName of outcome Y variable.
- numclustNumber of clusters requested.
- actclustNumber of clusters actually produced.
- scedasScedasticity assumption: "homo" or "hete"
- PStdifCharacter string describing the treatment difference.
- nnhbindfVector containing cluster number for each patient.
- rawmeanUnadjusted outcome mean by treatment group.
- rawvarsUnadjusted outcome variance by treatment group.
- rawfreqNumber of patients by treatment group.
- ratdifUnadjusted mean outcome difference between treatments.
- ratsdeStandard error of unadjusted mean treatment difference.
- binmeanUnadjusted mean outcome by cluster and treatment.
- binvarsUnadjusted variance by cluster and treatment.
- binfreqNumber of patients by bin and treatment.
- awbdifAcross cluster average difference with cluster size weights.
- awbsdeStandard error of awbdif.
- wwbdifAcross cluster average difference, inverse variance weights.
- wwbsdeStandard error of wwbdif.
- faclevMaximum number of different numerical values an outcome variable can assume without automatically being converted into a "factor" variable; faclev=1 causes a binary indicator to be treated as a continuous variable determining an average or proportion.
- youtype"contin"uous => only next eight outputs; "factor" => only last three outputs.
- aovdiffANOVA summary for treatment main effect only.
- form2Formula for outcome differences due to bins and to treatment nested within bins.
- bindiffANOVA summary for treatment nested within cluster.
- sig2Estimate of error mean square in nested model.
- pbindifUnadjusted treatment difference by cluster.
- pbinsdeStandard error of the unadjusted difference by cluster.
- pbinsizCluster radii measure: square root of total number of patients.
- symsizSymbol size of largest possible Snowball in a UPSnnltd() plot with 1 cluster.
- factabMarginal table of counts by Y-factor level and treatment.
- cumchiCumulative Chi-Square statistic for interaction in the three-way, nested table.
- cumdfDegrees of-Freedom for the Cumulative Chi-Squared.

Author(s)

Bob Obenchain <wizbob@att.net>

References

Obenchain RL. (2004) Unsupervised Propensity Scoring: NN and IV Plots. *Proceedings of the American Statistical Association (on CD)* 8 pages.

Obenchain RL. (2011) **USPSinR.pdf** USPS R-package vignette, 40 pages.

Rosenbaum PR, Rubin RB. (1983) The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* **70**: 41–55.

Rubin DB. (1980) Bias reduction using Mahalanobis metric matching. *Biometrics* **36**: 293-298.

See Also

UPSivadj, UPSaccum and UPSgraph.

Index