

Package ‘USPS’

April 17, 2009

Title Unsupervised and Supervised methods of Propensity Score Adjustment for Bias

Version 1.2-0

Date 2009-02-10

Author Bob Obenchain <wizbob@att.net>

Maintainer Bob Obenchain <wizbob@att.net>

Depends R (>= 1.8.0), cluster, lattice, gss

Description Unsupervised methods: identify and display the distribution of Local Treatment Differences (LTDs) and Local Average Treatment Effects (LATEs) across Clusters of patients chosen so as to be relatively well matched on patient baseline X-covariates. Supervised methods: estimate and validate Propensity Scores and use them to either sub-group or smooth observed patient outcomes over the common support of alternative treatment cohorts.

License GPL (>= 2)

URL <http://www.r-project.org>, <http://members.iquest.net/~softrx/>

Repository CRAN

Date/Publication 2009-02-15 09:26:37

R topics documented:

USPS-package	2
lindner	3
plot.SPSloess	4
plot.SPSsmoot	6
plot.UPSnnltd	8
SPSbalan	9
SPSloess	11
SPSlogit	13
SPSnbins	14
SPSoutco	16
SPSsmoot	18

UPSaccum	20
UPSaltdd	22
UPSgraph	24
UPShclus	25
UPSivadj	27
UPSnnltd	29

Index	32
--------------	-----------

USPS-package	<i>Unsupervised and Supervised Propensity Scoring adjustments for Bias and Confounding</i>
--------------	--

Description

Identify and display the distribution of Local Treatment Differences (LTDs) and Local Average Treatment Effects (LATEs) across Clusters of patients chosen so as to be relatively well matched within Clusters defined by patient baseline X-covariates.

Details

Package: USPS
 Type: Package
 Version: 1.2-0
 Date: 2009-02-10
 License: GNU GENERAL PUBLIC LICENSE, Version 2, June 1991

SUPERVISED OUTCOME BINNING AND SMOOTHING WITH ESTIMATED PROPENSITY SCORES:

Once one has fitted a somewhat smooth curve through scatters of observed outcomes, Y, versus 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?"

UNSUPERVISED NEAREST NEIGHBORS / LOCAL TREATMENT DIFFERENCES:

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.

UNSUPERVISED INSTRUMENTAL VARIABLES / LOCAL AVERAGE TREATMENT EFFECTS:

Multiple calls to UPSivadj(n) for varying numbers of clusters, n, yield alternative linear smoothes of LATE estimates plotted versus within cluster propensity score (observed treatment fraction) percentages.

Author(s)

Bob Obenchain <wizbob@att.net>

References

Green PJ, Silverman BW. (1994) **Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach**. *Chapman and Hall*.

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.

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

Obenchain RL. **USPSinR.pdf** ../R_HOME/library/USPS 2009; 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.

Examples

```
demo(abcix)
```

```
lindner
```

Lindner Center data on 996 PCI patients analyzed by Kereiakes et al.(2000)

Description

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

Usage

```
data(lindner)
```

Format

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

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

cardbill Cardiac related costs incurred within 6 months of patient's initial PCI; numeric value in 1998 dollars; costs were truncated by death for the 26 patients with lifepres == 0.

abcix Numeric treatment selection indicator; 0 implies usual PCI care alone; 1 implies usual PCI care deliberately augmented by either planned or rescue treatment with abciximab.

stent Coronary stent deployment; numeric, with 1 meaning YES and 0 meaning NO.

height Height in centimeters; numeric integer from 108 to 196.

female Female gender; numeric, with 1 meaning YES and 0 meaning NO.

diabetic Diabetes mellitus diagnosis; numeric, with 1 meaning YES and 0 meaning NO.

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

ejectfrac Left ejection fraction; numeric value from 0 percent to 90 percent.

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

References

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

Obenchain RL. (2009) **USPSinR.pdf** ../R_HOME/library/USPS 40 pages.

Examples

```
# Demo of USPS functionality on the lindner dataset...
demo(abcix)
```

plot.SPSloess	<i>Display LOESS Smooth of Outcome by Treatment in Supervised Propensity Scoring</i>
---------------	--

Description

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

Usage

```
## S3 method for class 'SPSloess':
plot(x, tcol="blue", ucol="red", dcol="green3", ...)
```

Arguments

x	output list object of class SPSloess.
tcol	optional; quoted name of color for treated patient smooth.
ucol	optional; quoted name of color for untreated patient smooth.
dcol	optional: quoted name of color for combined patient density.
...	optional; argument(s) passed on to plot().

Details

Plots of SPSloess objects display each patient's propensity score versus his/her observed (continuous) outcome. Patients receiving the "standard" treatment (trtm=0) are represented by cyan circles, while patients receiving the "new" treatment (trtm=1) are represented by magenta triangles. The smooth fits of outcome to propensity score within treatment cohorts are shown as cyan (trtm=0) and magenta (trtm=1) curves, respectively, superimposed upon the scatter.

Because smooth fits can be difficult to see when the scatters contain many points, a second plot rescaled to show only the two smooth (lowess or spline) fits, again using cyan (trtm=0) and magenta (trtm=1) curves. For details, see the returned lofit data frame.

Finally, a third plot shows total patient frequencies (black circles) within a 100-cell histogram along the propensity score axis as well as the corresponding density() smooth in red. For details, see the returned logrid data frame.

Winsorizing Cost data: `PSframeTRIMBILL < - pmin(PSframecardbill, 50000)`

The `fam="symmetric"` default option of SPSloess tends to be fairly robust to outlying outcomes, at least when the loess span is wide enough. Thus reducing (Winsorizing) outlying cardbill values to \$50K (as illustrated above) should have little effect on a fitted loess smooth with an appropriate span. Looking for the effects of Winsorizing on SPSloess() or SPSsmoot() constitutes "sensitivity analysis."

The original lowess() function of Cleveland and Devlin (1988) could be used here because only one X variable (namely, fitted propensity score) is involved, but I choose loess() instead to give users flexibility to choose between `fam="gaussian"` and `fam="symmetric"` option, which provides some resistance to outlying outcome values.

SPSloess() fits can tend to look rather "rough" compared to SPSsmoot() fits. Cubic spline smoothing appears to give answers that are interpretable as smoothed mean values for highly skewed distributions. Loess smoothing, at least when `fam="symmetric"`, tends to give answers more easily interpretable as modes or medians of highly skewed distributions. This median versus mean analogy may help explain why the weighted average signed treatment differences from SPSloess() tend to seem more precise than those from SPSsmoot() for highly skewed distributions.

Value

NULL

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. (2009) **USPSinR.pdf** ../R_HOME/library/USPS 40 pages.

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

See Also

[SPSlogit](#), [SPSsmoot](#) and [SPSoutco](#).

Examples

```
data(lindner)
PStreat <- abcix~stent+height+female+diabetic+acutemi+ejecfrac+veslproc
logtSPS <- SPSlogit(lindner, PStreat, PSfit, PSrnk, PSbin, appn="lindSPS")

SPScbls5 <- SPSloess(lindSPS, abcix, PSfit, cardbill, span=.5)
SPScbls5
plot(SPScbls5)
```

plot.SPSsmoot	<i>Display Spline Smooth of Outcome by Treatment in Supervised Propensity Scoring</i>
---------------	---

Description

Express Expected Outcome by Treatment as Spline Functions of Fitted Propensity Scores.

Usage

```
## S3 method for class 'SPSsmoot':
plot(x, tcol = "blue", ucol = "red", dcol = "green3", ...)
```

Arguments

x	output list object of class SPSsmoot.
tcol	optional; quoted name of color for treated patient smooth.
ucol	optional; quoted name of color for untreated patient smooth.
dcol	optional; quoted name of color for combined patient density.
...	optional; argument(s) passed on to plot().

Details

Plots of SPSsmoot objects display each patient's propensity score versus his/her observed (continuous) outcome. Patients receiving the "standard" treatment (trtm=0) are represented by cyan circles, while patients receiving the "new" treatment (trtm=1) are represented by magenta triangles. The smooth fits of outcome to propensity score within treatment cohorts are shown as cyan (trtm=0) and magenta (trtm=1) curves, respectively, superimposed upon the scatter.

Because smooth fits can be difficult to see when the scatters contain many points, a second plot rescaled to show only the two smooth (lowess or spline) fits, again using cyan (trtm=0) and magenta (trtm=1) curves. For details, see the returned ssfit data frame.

Finally, a third plot shows total patient frequencies (black circles) within a 100-cell histogram along the propensity score axis as well as the corresponding density() smooth in red. For details, see the returned ssgrid data frame.

SPSloess() fits can tend to look rather "rough" compared to SPSsmoot() fits. Cubic spline smoothing appears to give answers that are interpretable as smoothed mean values for highly skewed distributions. Loess smoothing, at least when fam="symmetric," tends to give answers more easily interpretable as modes or medians of highly skewed distributions. This median versus mean analogy may help explain why the weighted average signed treatment differences from SPSloess() tend to seem more precise than those from SPSsmoot() for highly skewed distributions.

Value

NULL

Author(s)

Bob Obenchain <wizbob@att.net>

References

Chambers JM, Hastie T. (1992) **Statistical Models in S** Wadsworth & Brooks/Cole.

Green PJ, Silverman BW. (1994) **Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach.** Chapman and Hall.

Hastie TJ, Tibshirani RJ. (1990) **Generalized Additive Models.** Chapman and Hall.

Obenchain RL. (2009) **USPSinR.pdf** ../R_HOME/library/USPS 40 pages.

Sheather SJ, Jones MC. (1991) A reliable data-based bandwidth selection method for kernel density estimation. *J Roy Statist Soc B* **53**: 683-690.

R implementation of smooth.spline() by Ripley BD and Maechler M. ('spar/lambda', etc).

See Also

[SPSloess](#), [SPSbalan](#) and [SPSoutco](#).

Examples

```
data(lindner)
PStreat <- abcix~stent+height+female+diabetic+acutemi+ejecfrac+veslproc
logtSPS <- SPSlogit(lindner, PStreat, PSfit, PSrnk, PSbin, appn="lindSPS")

SPScbss7 <- SPSsmoot(lindSPS, abcix, PSfit, cardbill, df=7)
SPScbss7
plot(SPScbss7)
```

plot.UPSnnltd	<i>Display plots of the NN/LTD Distribution in Unsupervised Propensity Scoring</i>
---------------	--

Description

Make a Snow Ball ("snob"), Probability Density ("dens") and/or Cumulative Distribution Function ("cdf") plot of the Local Treatment Difference (LTD) Distribution across all Informative Clusters within a requested total number clusters in patient baseline X-covariate space.

Usage

```
## S3 method for class 'UPSnnltd':
plot(x, pballs = TRUE, nnplot = "snob", nalpha = 1.4, ...)
```

Arguments

x	Object of class UPSnnltd.
pballs	Logical; Display cluster-size weighted (red) and inverse-variance weighted (green) "Power Balls" to represent across cluster LTD averages within the "Snowball" plot.
nnplot	String; Display "all" plots together, "seq" = one after the other, "snob" = only Snow Balls, "dens" = only fitted gss probability density, or "cdf" = only the gss cumulative distribution.
nalpha	Numeric; alpha argument to ssden() function.
...	Optional argument(s) to pass on to plot().

Details

Plots of UPSnnltd objects describe the LTD Distribution across all informative clusters. The generalized smoothing spline (gss) displays corresponding to the nnplot = "dens" and "cdf" options may fail to exist when the number of informative clusters is less than 60.

Value

NULL

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. (2009) **USPSinR.pdf** ../R_HOME/library/USPS 40 pages.

See Also

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

Examples

```
data(lindner)
UPSxvars <- c("stent", "height", "female", "diabetic", "acutemi", "ejecfrac", "veslproc")
UPSharch <- UPShclus(lindner, UPSxvars)
UPSaccum(UPSharch, lindner, abcix, lifepres, faclev=1, scedas="homo", accobj="ABClife")

lif070nn <- UPSnnltd(70)
lif070nn
plot(lif070nn)
```

SPSbalan

Test for Within-Bin X-covariate Balance in Supervised Propensity Scoring

Description

Test for Conditional Independence of X-covariate Distributions from Treatment Selection within Given, Adjacent PS Bins.

Usage

```
xcvobj <- SPSbalan(dframe, trtm, qbin, xvar, faclev=3)
```

Arguments

dframe	Name of augmented data.frame written to the appn="" argument of SPSlogit().
trtm	Name of the two-level treatment factor 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.

Details

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.

Value

An output list object of class SPSbalan.

"contin"uous xvar => only the following four outputs...

aovdiff	ANOVA output for marginal test.
form2	Formula for differences in X due to bins and to treatment nested within bins.
bindiff	ANOVA output for the nested within bin model.
df3	Output data.frame containing 3 variables: X-covariate, treatment and bin.
factab	Marginal table of counts by X-factor level and treatment.
tab	Three-way table of counts by X-factor level, treatment and bin.
cumchi	Cumulative Chi-Square statistic for interaction in the three-way, nested table.
cumdf	Degrees 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. (2009) **USPSinR.pdf** ../R_HOME/library/USPS 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](#), [SPSnbins](#) and [SPSoutco](#).

Examples

```
data(lindner)
PStreat <- abcix~stent+height+female+diabetic+acutemi+ejecfrac+veslproc
logtSPS <- SPSlogit(lindner, PStreat, PSfit, PSrnk, PSbin, appn="lindSPS")

SPSbalvs <- SPSbalan(lindSPS, abcix, PSbin, veslproc)
SPSbalvs
plot(SPSbalvs)
```

SPSloess	<i>LOESS Smoothing of Outcome by Treatment in Supervised Propensity Scoring</i>
----------	---

Description

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

Usage

```
losobj <- SPSloess(dframe, trtm, pscr, yvar, faclev=3, deg=2, span=0.75, fam="sym")
```

Arguments

<code>dframe</code>	data.frame of the form returned by SPSlogit().
<code>trtm</code>	the two-level factor on the left-hand-side in the formula argument to SPSlogit().
<code>pscr</code>	fitted propensity scores of the form returned by SPSlogit().
<code>yvar</code>	continuous outcome measure or result unknown at the time patient was assigned (possibly non-randomly) to treatment; "NA"s are allowed in yvar.
<code>faclev</code>	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.
<code>deg</code>	optional; degree (1=linear or 2=quadratic) of the local fit.
<code>span</code>	optional; span (0 to 2) argument for the loess() function.
<code>fam</code>	optional; "gaussian" or "symmetric".

Details

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.

Value

An output list object of class SPSloess:

logrid	loess 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 (F1F0), 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 loess fits could not be extrapolated because no observed outcomes were available.
losub0, losub1	loess 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.)
span	loess span setting.
lotdif	outcome treatment difference mean.
lotsde	outcome 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. (2009) **USPSinR.pdf** ../R_HOME/library/USPS 40 pages.
- Ripley BD, loess() based on the 'cloess' package of Cleveland, Grosse and Shyu.

See Also

[SPSlogit](#), [SPSsmoot](#) and [SPSoutco](#).

Examples

```
data(lindner)
PStreat <- abcix~stent+height+female+diabetic+acutemi+ejecfrac+veslproc
logtSPS <- SPSlogit(lindner, PStreat, PSfit, PSrnk, PSbin, appn="lindSPS")

SPScbls5 <- SPSloess(lindSPS, abcix, PSfit, cardbill, span=.5)
SPScbls5
plot(SPScbls5)
```

SPSlogit	<i>Propensity Score prediction of Treatment Selection from Patient Baseline X-covariates</i>
----------	--

Description

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

Usage

```
SPSlobj <- SPSlogit(dframe, form, pfit, prnk, qbin, bins=5, appn="")
```

Arguments

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

<code>dframe</code>	Name of input data.frame containing X, t & Y variables.
<code>dfoutnam</code>	Name of output data.frame augmented by <code>pfit</code> , <code>prnk</code> and <code>qbin</code> variables.
<code>trtm</code>	Name of two-level treatment factor variable.
<code>form</code>	<code>glm()</code> formula for logistic regression.
<code>pfit</code>	Name of predicted PS variable.
<code>prnk</code>	Name of variable containing PS tied-ranks.
<code>qbin</code>	Name of variable containing assigned PS bin number for each patient.
<code>bins</code>	Number of adjacent PS bins desired.
<code>glmobj</code>	Output object from invocation of <code>glm()</code> 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. (2009) **USPSinR.pdf** ../R_HOME/library/USPS 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](#).

Examples

```
data(lindner)
PStreat <- abcix~stent+height+female+diabetic+acutemi+ejecfrac+veslproc
logtSPS <- SPSlogit(lindner, PStreat, PSfit, PSrnk, PSbin, appn="lindSPS")
logtSPS
```

SPSnbins

Change the Number of Bins in Supervised Propensity Scoring

Description

Change the Number of Bins in Supervised Propensity Scoring

Usage

```
dframe2 <- SPSnbins(dframe1, prnk, qbin, bins=8)
```

Arguments

dframe1	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:

`dframe2` Modified 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. (2009) **USPSinR.pdf** `../R_HOME/library/USPS` 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](#).

Examples

```
data(lindner)
PStreat <- abcix~stent+height+female+diabetic+acutemi+ejecfrac+veslproc
logtSPS <- SPSlogit(lindner, PStreat, PSfit, PSrnk, PSbin, appn="lindSPS")
logtSPS

# If imbalance within the SPSlogit() default of bins=5 was detected, then ...

lindSPS <- SPSnbins(lindSPS, PSrnk, PSbin6, bins=6)
table(lindSPS$PSbin,lindSPS$PSbin6)
```

SPSoutco

Examine Treatment Differences on an Outcome Measure in Supervised Propensity Scoring

Description

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

Usage

```
outobj <- SPSoutco(dframe, trtm, qbin, yvar, faclev=3)
```

Arguments

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:

dframe	Name of augmented data.frame written to the appn="" argument of SPSlogit().
trtm	Name of the two-level treatment factor variable.
yvar	Name of an outcome Y variable.
bins	Number of variable containing bin numbers.
PStdif	Character string describing the treatment difference.
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.
wkbdif	Across cluster average difference, inverse variance weights.
wwsde	Standard error of wkbdif.
form	Formula for overall, marginal treatment difference on X-covariate.
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.
youtype	"contin"uous => only next six outputs; "factor" => only last four outputs.
aovdiff	ANOVA output for marginal test.
form2	Formula for differences in X due to bins and to treatment nested within bins.
bindiff	ANOVA summary for treatment nested within bin.
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.
factab	Marginal table of counts by Y-factor level and treatment.
tab	Three-way table of counts by Y-factor level, treatment and bin.
cumchi	Cumulative Chi-Square statistic for interaction in the three-way, nested table.
cumdf	Degrees 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. (2009) **USPSinR.pdf** ../R_HOME/library/USPS 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](#).

Examples

```

data(lindner)
PStreat <- abcix~stent+height+female+diabetic+acutemi+ejecfrac+veslproc
logtSPS <- SPSlogit(lindner, PStreat, PSfit, PSrnk, PSbin, appn="lindSPS")

SPSlifeo <- SPSoutco(lindSPS, abcix, PSbin, lifepres, faclev=1)
SPSlifeo
plot(SPSlifeo)

```

SPSsmoot

Spline Smoothing of Outcome by Treatment in Supervised Propensity Scoring

Description

Express Expected Outcome by Treatment as Spline Functions of Fitted Propensity Scores.

Usage

```
smoobj <- SPSsmoot(dframe, trtm, pscr, yvar, faclev=3, df=5, spar=NULL, cv=F, pen
```

Arguments

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.
cv	optional; ordinary cross-validation (T) or generalized cross-validation, GCV (F).
df	optional; degrees-of-freedom of B-spline fit.
spar	spar argument for smooth.spline() function.
penalty	coefficient of penalty for df in the GCV criterion.

Details

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 SPSsmoot() output object contains two data frames, ssgrid and ssfit, useful in further computations.

Value

An output list object of class SPSsmoot:

ssgrid	spline 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.
spsub0, spsub1	spline fit data.frames containing 4 variables for each distinct PS value in ssfit. These 4 variables are named PS, YAVG, TRT==0 and 1, respectively, and FIT = spline prediction.
df	smooth.spline() degrees-of-freedom
sptdif	outcome treatment difference mean.
sptsde	outcome treatment difference standard deviation.

Author(s)

Bob Obenchain <wizbob@att.net>

References

- Chambers JM, Hastie T. (1992) **Statistical Models in S** *Wadsworth & Brooks/Cole*.
- Green PJ, Silverman BW. (1994) **Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach**. *Chapman and Hall*.
- Hastie TJ, Tibshirani RJ. (1990) **Generalized Additive Models**. *Chapman and Hall*.
- Obenchain RL. (2009) **USPSinR.pdf** ../R_HOME/library/USPS 40 pages.
- Sheather SJ, Jones MC. (1991) A reliable data-based bandwidth selection method for kernel density estimation. *J Roy Statist Soc B* **53**: 683-690.
- R implementation of smooth.spline() by Ripley BD and Maechler M. ('spar/lambda', etc).

See Also

[SPSloess](#), [SPSbalan](#) and [SPSoutco](#).

Examples

```

data(lindner)
PStreat <- abcix~stent+height+female+diabetic+acutemi+ejecfrac+veslproc
logtSPS <- SPSlogit(lindner, PStreat, PSfit, PSrnk, PSbin, appn="lindSPS")

SPScbss7 <- SPSsmoot(lindSPS, abcix, PSfit, cardbill, df=7)
SPScbss7
plot(SPScbss7)

```

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(hiclus, dframe, trtm, yvar, faclev=3, scedas="hete", accobj="UPSframe")
```

Arguments

hiclus	Name of UPSclus() output object created using the diana, agnes or hclust method.
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"
accobj	Name of the object for accumulation of I-plots to be ultimately displayed using UPSgraph().

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

The output object will automatically be named UPSaccum.pars:

hiclus	Name of a diana, agnes or hclust object created by UPSclus().
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 a proportion.
scedas	Scedasticity assumption: "homo" or "hete"
accobj	Name of the object for accumulation of I-plots to be ultimately displayed using UPSgraph().
nnymax	Maximum NN LTD Standard Error observed; Upper NN plot limit; initialized to zero.
nnxmin	Minimum NN LTD observed; Left NN plot limit; initialized to zero.
nnxmax	Maximum 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. (2009) **USPSinR.pdf** ../R_HOME/library/USPS 40 pages.

See Also

[UPSnnltd](#), [UPSivadj](#) and [UPShclus](#).

Examples

```
data(lindner)
UPSxvars <- c("stent", "height", "female", "diabetic", "acutemi", "ejecfrac", "veslproc")
UPSharch <- UPShclus(lindner, UPSxvars)

UPSaccum(UPSharch, lindner, abcix, lifepres, faclev=1, scedas="homo", accobj="ABClife")
UPSaccum.pars

lif001nn <- UPSnnltd(1)
lif020nn <- UPSnnltd(20)
lif070nn <- UPSnnltd(70)
UPSgraph()

ABClife
```

 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(dframe, trtm, yvar, faclev=3, scedas="homo", NNobj=NA, clus=50, reps=10, s
```

Arguments

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

An output list object of class UPSaltdd:

dframe	Name of data.frame containing 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"
NNobj	Name of an existing UPSnnltd object or NA.
clus	Number of Random Clusters requested per Replication.
reps	Number of overall Replications, each with the same number of requested clusters.
pats	Number of patients with no NAs in their yvar outcome and trtm factor.
seed	Seed for Monte Carlo random number generator.
altd	Matrix of LTDs and relative weights from artificial clusters.
alxmin	Minimum artificial LTD value.
alxmax	Maximum artificial LTD value.
alymax	Maximum weight among artificial LTDs.
altdcdf	Vector of artificial LTD x-coordinates for smoothed CDF.
qq	Vector of equally spaced CDF values from 0.0 to 1.0.
nnltd	Optional matrix of relevant NN/LTDs and relative weights.
nnlxmin	Optional minimum NN/LTD value.
nnlxmax	Optional maximum NN/LTD value.
nnlymax	Optional maximum weight among NN/LTDs.
nnltdcdf	Optional vector of NN/LTD x-coordinates for smoothed CDF.
nq	Optional 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. (2009) **USPSinR.pdf** ../R_HOME/library/USPS 41 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](#).

Examples

```
data(lindner)
abcdf <- UPSaltd(lindner, abcix, lifepres, faclev=1)
abcdf
plot(abcdf)
```

UPSgraph

Display Sensitivity Analysis Graphic in Unsupervised Propensity 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(nncol = "red", nwcol = "green3", ivcol = "blue", ...)
```

Arguments

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().
...	Optional parameter(s) passed on to plot().

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.

Value

NULL

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. (2009) **USPSinR.pdf** ../R_HOME/library/USPS 40 pages.
- Rubin DB. (1980) Bias reduction using Mahalanobis metric matching. *Biometrics* **36**: 293-298.

See Also

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

Examples

```
data(lindner)
UPSxvars <- c("stent", "height", "female", "diabetic", "acutemi", "ejefrac", "veslproc")
UPSharch <- UPShclus(lindner, UPSxvars)
UPSaccum(UPSharch, lindner, abcix, lifepres, faclev=1, scedas="homo", accobj="ABClife")
lif001nn <- UPSnnltd(1)
lif020nn <- UPSnnltd(20)
lif070nn <- UPSnnltd(70)
lif120nn <- UPSnnltd(120)

UPSgraph()

ABClife
```

UPShclus	<i>Hierarchical Clustering of Patients on X-covariates for Unsupervised Propensity Scoring</i>
----------	--

Description

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

Usage

```
hclobj <- UPShclus(dframe, xvars, method="diana")
```

Arguments

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".

Details

The first step in an Unsupervised Propensity Scoring analysis 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:

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".
upshcl	Hierarchical 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. (2009) **USPSinR.pdf** ../R_HOME/library/USPS 40 pages.
- Rubin DB. (1980) Bias reduction using Mahalanobis metric matching. *Biometrics* **36**: 293-298.

See Also

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

Examples

```
data(lindner)
UPSxvars <- c("stent", "height", "female", "diabetic", "acutemi", "ejefrac", "veslproc")
UPSharch <- UPShclus(lindner, UPSxvars)
plot(UPSharch)
```

UPSivadj	<i>Instrumental Variable LATE Linear Fitting in Unsupervised Propensity Scoring</i>
----------	---

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

```
ivobj <- UPSivadj(numclust)
```

Arguments

`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:

<code>hiclus</code>	Name of clustering object created by UPShclus().
<code>dframe</code>	Name of data.frame containing X, t & Y variables.
<code>trtm</code>	Name of treatment factor variable.
<code>yvar</code>	Name of outcome Y variable.
<code>numclust</code>	Number of clusters requested.
<code>actclust</code>	Number of clusters actually produced.
<code>scedas</code>	Scedasticity assumption: "homo" or "hete"
<code>PStdif</code>	Character string describing the treatment difference.
<code>ivhbindf</code>	Vector containing cluster number for each patient.
<code>rawmean</code>	Unadjusted outcome mean by treatment group.
<code>rawvars</code>	Unadjusted outcome variance by treatment group.
<code>rawfreq</code>	Number of patients by treatment group.
<code>ratdif</code>	Unadjusted mean outcome difference between treatments.
<code>ratsde</code>	Standard error of unadjusted mean treatment difference.

binmean	Unadjusted mean outcome by cluster and treatment.
binfreq	Number of patients by bin and treatment.
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" => next eleven outputs; "factor" => no additional output items.
pbinout	LATE regardless of treatment by cluster.
pbinpsp	Within-Cluster Treatment Percentage = non-parametric Propensity Score.
pbinsiz	Cluster radii measure: square root of total number of patients.
symsiz	Symbol size of largest possible Snowball in a UPSivadj() plot with 1 cluster.
ivfit	lm() output for linear smooth across clusters.
ivtzero	Predicted outcome at PS percentage zero.
ivtxsde	Standard deviation of outcome prediction at PS percentage zero.
ivtdiff	Predicted outcome difference for PS percentage 100 minus that at zero.
ivtdsde	Standard deviation of outcome difference.
ivt100p	Predicted outcome at PS percentage 100.
ivt1pse	Standard 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. (2009) **USPSinR.pdf** ../R_HOME/library/USPS 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](#).

Examples

```

data(lindner)
UPSxvars <- c("stent", "height", "female", "diabetic", "acutemi", "ejecfrac", "veslproc")
UPSharch <- UPShclus(lindner, UPSxvars)
UPSaccum(UPSharch, lindner, abcix, lifepres, faclev=1, scedas="homo", accobj="ABClife")

lif100iv <- UPSivadj(100)
lif100iv
plot(lif100iv)

```

UPSnnltd	<i>Nearest Neighbor Distribution of LTDs in Unsupervised Propensity Scoring</i>
----------	---

Description

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

Usage

```
nnobj <- UPSnnltd(numclust)
```

Arguments

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:

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	"contin"uous => 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.

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. (2009) **USPSinR.pdf** ../R_HOME/library/USPS 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](#).

Examples

```
data(lindner)
UPSxvars <- c("stent", "height", "female", "diabetic", "acutemi", "ejecfrac", "veslproc")
UPSharch <- UPShclus(lindner, UPSxvars)
UPSaccum(UPSharch, lindner, abcix, lifepres, faclev=1, scedas="homo", accobj="ABClife")

lif070nn <- UPSnnltd(70)
lif070nn
plot(lif070nn)
```

Index

- *Topic **cluster**
 - UPShclus, 25
 - *Topic **datasets**
 - lindner, 3
 - *Topic **design**
 - SPSnbins, 14
 - UPSaccum, 19
 - UPShclus, 25
 - *Topic **hplot**
 - plot.SPSloess, 4
 - plot.SPSsmoot, 5
 - plot.UPSnnltd, 7
 - SPSloess, 10
 - SPSoutco, 15
 - UPSgraph, 23
 - UPSivadj, 26
 - *Topic **htest**
 - SPSbalan, 8
 - *Topic **methods**
 - plot.SPSloess, 4
 - plot.SPSsmoot, 5
 - plot.UPSnnltd, 7
 - *Topic **models**
 - SPSlogit, 12
 - *Topic **nonparametric**
 - SPSoutco, 15
 - UPSaltdd, 21
 - UPSivadj, 26
 - UPSnnltd, 28
 - *Topic **package**
 - USPS-package, 1
 - *Topic **smooth**
 - SPSloess, 10
 - SPSsmoot, 17
 - *Topic **univar**
 - UPSaccum, 19
- lindner, 3
- plot.SPSloess, 4
- plot.SPSsmoot, 5
- plot.UPSnnltd, 7
- SPSbalan, 7, 8, 13, 14, 17, 19
- SPSloess, 7, 10, 19
- SPSlogit, 5, 9, 11, 12, 14, 17
- SPSnbins, 9, 13, 14, 17
- SPSoutco, 5, 7, 9, 11, 13, 14, 15, 19
- SPSsmoot, 5, 11, 17
- UPSaccum, 8, 19, 23, 24, 26, 28, 30
- UPSaltdd, 21
- UPSgraph, 8, 23, 23, 26, 28, 30
- UPShclus, 21, 25
- UPSivadj, 8, 21, 24, 26, 30
- UPSnnltd, 21, 23, 24, 26, 28, 28
- USPS-package, 1