Package 'met'

August 14, 2020

Type Package

Title Evaluating and Improving Matched Samples in Observational Studies

Version 0.1.0

Author Ruoqi Yu

Maintainer Ruoqi Yu <ruoqiyu@wharton.upenn.edu>

Description Evaluate covariate balance of a matched sample, and suggest a new feature that the new match should balance if the current match is not satisfactory (only available with method being generalized Kolmogorov-Smirnov statistic on a finite set, GFKS for short).

License MIT+file LICENSE

Encoding UTF-8

LazyData true

Depends R (>= 3.5.0)

Imports stats, DiPs

Suggests optmatch

NeedsCompilation no

Repository CRAN

Date/Publication 2020-08-14 14:10:03 UTC

R topics documented:

gfks1_stat *Univariate GFKS statistics*

Description

Compute the GFKS statistics and identify its location for each column of x. This function and its use are discussed in Yu (2020).

Usage

```
gfks1_stat(x,z)
```

Arguments

x	A matrix with length(z) rows giving the discretized covariates.
z	A vector whose ith coordinate is 1 for a treated unit and is 0 for a control.

Value

gfks	A vector of GFKS statistics for each discrete covariate.
where	A vector of locations that the GFKS statistics occur, i.e., the location of the maximum difference of the empirical distribution function of each discrete covariate in the treated and control groups.

References

Yu, R. (2020) Evaluating and Improving a Matched Comparison of Antidepressants and Bone Density. Under revision.

Examples

```
library(optmatch)
data("SSRI")
attach(SSRI)
X<-cbind(female,black,education)
dist<-DiPs::maha_dense(z,X)
o<-DiPs::match(z, dist, SSRI)
M0<-o$data
Xm<-cbind(M0$female,M0$black,M0$education)
gfks1_stat(Xm,M0$z)
detach(SSRI)
```

gfks2_stat *Bivariate GFKS statistics*

Description

Compute the GFKS statistics and identify its location for all marginal distributions and bivariate distributions for x. This function and its use are discussed in Yu (2020).

Usage

```
gfks2_stat(x,z)
```

Arguments

x	A matrix with length(z) rows giving the discretized covariates.
z	A vector whose ith coordinate is 1 for a treated unit and is 0 for a control.

Value

gfks	A vector of GFKS statistics for all marginal and bivariate distributions.
where	A vector of locations v that the GFKS statistics occur, i.e., the location of the maximum difference of the empirical distribution function in the treated and control groups.
direction	A vector of directions indicating which quadrant the GFKS statistics occur – 1 for '<=' and '<=', 2 for '<=' and '>', 3 for '>' and '<=', 4 for '>' and '>'.

References

Yu, R. (2020) Evaluating and Improving a Matched Comparison of Antidepressants and Bone Density. Under revision.

Examples

```
library(optmatch)
data("SSRI")
attach(SSRI)
X<-cbind(female,black,education)
dist<-DiPs::maha_dense(z,X)
o<-DiPs::match(z, dist, SSRI)
M0<-o$data
Xm<-cbind(M0$female,M0$black,M0$education)
gfks2_stat(Xm,M0$z)
detach(SSRI)
```

ks_stat *Kolmogorov-Smirnov statistics*

Description

Compute the Kolmogorov-Smirnov statistics for each individual covariate in x.

Usage

```
ks_stat(x,z)
```

Arguments

x	A matrix with length(z) rows giving the covariates.
z	A vector whose ith coordinate is 1 for a treated unit and is 0 for a control.

Value

A vector of Kolmogorov-Smirnov statistics for each column of x.

Examples

```
library(optmatch)
data("SSRI")
attach(SSRI)
X<-cbind(female,black,education)
dist<-DiPs::maha_dense(z,X)
o<-DiPs::match(z, dist, SSRI)
M0<-o$data
Xm<-cbind(M0$female,M0$black,M0$education)
ks_stat(Xm,M0$z)
detach(SSRI)
```

met *Balance evaluation of matched samples*

Description

Evaluate a matched sample with minimum p-value of multiple measures of imbalance. This function and its use are discussed in Yu (2020).

Usage

```
met(x,z,method='GFKS-1',prob=c(0,0.11,0.35,0.65,0.89,1),
continuous=rep(FALSE,dim(x)[2]),nperm=1000,xf=NULL,zf=NULL)
```

Arguments

x	A matrix with length(z) rows giving the discretized covariates.
z	A vector whose ith coordinate is 1 for a treated unit and is 0 for a control.
method	Balance evaluation methods, can be chosen from 'GFKS-1', 'GFKS-2','SMD','KS','t','wilcoxon'.
prob	If method is chosen as 'GFKS-1'or 'GFKS-2', continuous variables in x are discretized based on prob, i.e., only the quantiles at prob are considered.
continuous	A vector of length dim(x)[2] indicating whether the covariates are discrete (F) or continuous (T). If continuous, GFKS statistics are evaluted at the quantiles at prob.
nperm	Number of simulated randomized experiments to approximate the p-values.
xf	A matrix with length(zf) rows giving the covariates before matching. xf should be specified if method is chosen as 'GFKS-1', 'GFKS-2', 'SMD'.
zf	A vector whose ith coordinate is 1 for a treated unit and is 0 for a control. zf should be specified if method is chosen as 'GFKS-1', 'GFKS-2', 'SMD'.

Value

statistic	A summary statistic to combine all statistics – the minimum p-value.
pvalue	A vector of locations v that the GFKS statistics occur, i.e., the location of the maximum difference of the empirical distribution function in the treated and control groups.
imbalance	A vector indicating which marginal distribution or joint distribution is most imbalanced. Not NULL if method is chosen as 'GFKS-1' or 'GFKS-2'.
location	Location that statistic occur. Not NULL if method is chosen as 'GFKS-1' or 'GFKS-2'.
direction	Direction indicating which quadrant statistic occur. 1 for '<=' and '<=', 2 for '<=' and '>', 3 for '>' and '<=', 4 for '>' and '>'. Not NULL if method is chosen as 'GFKS-1' or 'GFKS-2'.
ind.pvalue	P-values for the observational match.
null.pvalues	P-values for the simulated randomized experiments.
variable	A new binary variable that the next iteration match should balance. Not NULL if method is chosen as 'GFKS-1' or 'GFKS-2'.

References

Yu, R. (2020) Evaluating and Improving a Matched Comparison of Antidepressants and Bone Density. Under revision.

Examples

```
library(optmatch)
data("SSRI")
attach(SSRI)
X<-cbind(female,black,education,age,bmi)
dist<-DiPs::maha_dense(z,X)
```

```
o<-DiPs::match(z, dist, SSRI)
M0<-o$data
Xm<-cbind(M0$female,M0$black,M0$education,M0$age,M0$bmi)
met(Xm,M0$z,'GFKS-1',continuous=c(FALSE,FALSE,FALSE,TRUE,TRUE),nperm=100,xf=X,zf=z)
detach(SSRI)


#real data application in Yu(2020)
library(DiPs)
data("SSRI")
attach(SSRI)
X<-cbind(age,female,black,hispanic,povertyNA,povertyFill,education,height,weight,
bmi,cotinine,hd,diabetes,insurance,weighmore,weighless,weightchange,physicalact,dietsup)
XX<-cbind(age,female,black,hispanic,povertyNA,povertyFill,education,height,weight,
bmi,cotinine,hd,diabetes,insurance,weighmore,weighless,weightchange,physicalact,dietsup,pr)
detach(SSRI)

dat=SSRI
#basic match
dist<-maha_dense(dat$z,X)
dist<-addcaliper(dist, dat$z, dat$pr, c(-.2,.2), stdev = TRUE, penalty = 1000)
o<-match(dat$z,dist, dat,fine=factor(dat$education), ncontrol = 4)
M0<-o$data
Xm<-subset(M0, select=c('age','female','black','hispanic','povertyNA','povertyFill',
'education','height','weight','bmi','cotinine','hd','diabetes','insurance','weighmore','weighless',
'weightchange','physicalact','dietsup','pr'))
btb0<-check(XX,Xm,dat$z,M0$z)
round(btb0,3)

Result0_t=met(Xm,M0$z,method='t',nperm=2000)
Result0_t$pvalue

Result0_smd=met(Xm,M0$z,method='SMD',nperm=2000,xf=XX,zf=dat$z)
Result0_smd$pvalue

Result0_w=met(Xm,M0$z,method='wilcoxon',nperm=2000)
Result0_w$pvalue

Result0_ks=met(Xm,M0$z,method='KS',nperm=2000)
Result0_ks$pvalue

Result0=met(Xm,M0$z,method='GFKS-2',continuous=c(TRUE,FALSE,FALSE,FALSE,FALSE,TRUE,FALSE,
TRUE,TRUE,TRUE,TRUE,TRUE,FALSE,FALSE,FALSE,FALSE,TRUE,FALSE,FALSE,TRUE),nperm=2000,xf=XX,zf=dat$z)
Result0$statistic
Result0$pvalue
Result0$imbalance
Result0$location
Result0$direction

#iteration 1
dist<-maha_dense(dat$z,X)
dist<-addcaliper(dist, dat$z, dat$pr, c(-.2,.2), stdev = TRUE, penalty = 1000)
i1=as.factor(dat$diabetes<=0 & dat$weighmore<=0)
```

```
o<-match(dat$z, dist, dat, fine=as.numeric(factor(dat$education):i1), ncontrol = 4)
M1<-o$data
Xm<-subset(M1, select=c('age','female','black','hispanic','povertyNA','povertyFill',
'education','height','weight','bmi','cotinine','hd','diabetes','insurance','weighmore','weighless',
'weightchange','physicalact','dietsup','pr'))
btb1<-check(XX,Xm,dat$z,M1$z)
round(btb1,3)

Result1_t=met(Xm,M1$z,method='t',nperm=2000)
Result1_t$pvalue

Result1_smd=met(Xm,M1$z,method='SMD',nperm=2000,xf=XX,zf=dat$z)
Result1_smd$pvalue

Result1_w=met(Xm,M1$z,method='wilcoxon',nperm=2000)
Result1_w$pvalue

Result1_ks=met(Xm,M1$z,method='KS',nperm=2000)
Result1_ks$pvalue

Result1=met(Xm,M1$z,method='GFKS-2',continuous=c(TRUE,FALSE,FALSE,FALSE,FALSE,TRUE,FALSE,
TRUE,TRUE,TRUE,TRUE,TRUE,FALSE,FALSE,FALSE,FALSE,TRUE,FALSE,FALSE,TRUE),nperm=2000,xf=XX,zf=dat$z)
Result1$statistic
Result1$pvalue
Result1$imbalance
Result1$location
Result1$direction

#iteration 2
dist<-maha_dense(dat$z,X)
dist<-addcaliper(dist, dat$z, dat$pr, c(-.2,.2), stdev = TRUE, penalty = 1000)
i2=as.factor(dat$weightchange<=quantile(dat$weightchange[dat$z==1],probs=0.65) &
as.numeric(dat$pr)<=quantile(dat$pr[dat$z==1],probs=0.89))
o<-match(dat$z, dist, dat, fine=as.numeric(factor(dat$education):i1:i2), ncontrol = 4)
M2<-o$data
Xm<-subset(M2, select=c('age','female','black','hispanic','povertyNA','povertyFill',
'education','height','weight','bmi','cotinine','hd','diabetes','insurance','weighmore','weighless',
'weightchange','physicalact','dietsup','pr'))
btb2<-check(XX,Xm,dat$z,M2$z)
round(btb2,3)

Result2_t=met(Xm,M2$z,method='t',nperm=2000)
Result2_t$pvalue

Result2_smd=met(Xm,M2$z,method='SMD',nperm=2000,xf=XX,zf=dat$z)
Result2_smd$pvalue

Result2_w=met(Xm,M2$z,method='wilcoxon',nperm=2000)
Result2_w$pvalue

Result2_ks=met(Xm,M2$z,method='KS',nperm=2000)
Result2_ks$pvalue
```

```
Result2=met(Xm,M2$z,method='GFKS-2',continuous=c(TRUE,FALSE,FALSE,FALSE,FALSE,TRUE,FALSE,
TRUE,TRUE,TRUE,TRUE,TRUE,FALSE,FALSE,FALSE,FALSE,TRUE,FALSE,FALSE,TRUE),nperm=2000,xf=XX,zf=dat$z)
Result2$statistic
Result2$pvalue
Result2$imbalance
Result2$location
Result2$direction

#iteration 3
dist<-maha_dense(dat$z,X)
dist<-addcaliper(dist, dat$z, dat$pr, c(-.2,.2), stdev = TRUE, penalty = 1000)
i3=as.factor(dat$weightchange<=quantile(dat$weightchange[dat$z==1],probs=0.35))
o<-match(dat$z, dist, dat, fine=as.numeric(factor(dat$education):i1:i2:i3), ncontrol = 4)
M3<-o$data
Xm<-subset(M3, select=c('age','female','black','hispanic','povertyNA','povertyFill',
'education','height','weight','bmi','cotinine','hd','diabetes','insurance','weighmore','weighless',
'weightchange','physicalact','dietsup','pr'))
btb3<-check(XX,Xm,dat$z,M3$z)
round(btb3,3)

Result3_t=met(Xm,M3$z,method='t',nperm=2000)
Result3_t$pvalue

Result3_smd=met(Xm,M3$z,method='SMD',nperm=2000,xf=XX,zf=dat$z)
Result3_smd$pvalue

Result3_w=met(Xm,M3$z,method='wilcoxon',nperm=2000)
Result3_w$pvalue

Result3_ks=met(Xm,M3$z,method='KS',nperm=2000)
Result3_ks$pvalue

Result3=met(Xm,M3$z,method='GFKS-2',continuous=c(TRUE,FALSE,FALSE,FALSE,FALSE,TRUE,FALSE,
TRUE,TRUE,TRUE,TRUE,TRUE,FALSE,FALSE,FALSE,FALSE,TRUE,FALSE,FALSE,TRUE),nperm=2000,xf=XX,zf=dat$z)
Result3$statistic
Result3$pvalue
Result3$imbalance
Result3$location
Result3$direction
```

`smd_stat` *Standardized mean differences*

Description

Compute the standardized mean differences for each individual covariate in x, i.e., the mean difference of each column of x is standarzied by the pooled standard deviation calculated based on the original data before matching (xf and zf).

Usage

```
smd_stat(x,z,xf,zf)
```

Arguments

x	A matrix with length(z) rows giving the covariates after matching.
z	A vector whose ith coordinate is 1 for a matched treated unit and is 0 for a matched control.
xf	A matrix with length(zf) rows giving the covariates before matching.
zf	A vector whose ith coordinate is 1 for a treated unit and is 0 for a control.

Value

A vector of standardized mean differences for each column of x.

Examples

```
library(optmatch)
data("SSRI")
attach(SSRI)
X<-cbind(female,black,education)
dist<-DiPs::maha_dense(z,X)
o<-DiPs::match(z, dist, SSRI)
M0<-o$data
Xm<-cbind(M0$female,M0$black,M0$education)
smd_stat(Xm,M0$z,X,z)
detach(SSRI)
```

SSRI *Antidepressants and bone density*

Description

NHANES 2009-2010 data on selective serotonin reuptake inhibitors (SSRIs) and bone density for adults.

Usage

```
data("SSRI")
```

Format

A data frame with 2759 observations on the following 11 variables.

`z` Antidepressants status, 1 = SSRI user, 0 = taking neither SSRI nor TCA

`female` 1 = female, 0 = male

`age` Age in years, >=18

`black` 1=black race, 0=other

`hispanic` 1=Hispanic race, 0=other

`povertyNA` 1=missing poverty ratio, 0=observed poverty ratio

`povertyFill` Ratio of family income to the poverty level, capped at 5 times poverty, with missing values imputed as mean

`education` Level of education in 5-point scale, with 1 for less than 9th grade, 3 for high school or equivalent, and 5 for college graduate

`height` Height in cm

`weight` Weight in kg

`bmi` BMI or body-mass-index

`cotinine` serum cotinine level (ng/mL)

`hd` serum 25-hydroxyvitamin D2 + D3 level (nmol/L)

`diabetes` 1=diagnosed with diabetes or borderline diabetes, 0=no diabetes

`insurance` whether covered by insurance

`weighmore` whether would like to weigh more

`weighless` whether would like to weigh less

`weightchange` weight change in the recent 1 year of finishing the questionnaire (kg)

`physicalact` whether has moderate and vigorous activity in the past 30 days

`dietsup` whether has taken any dietary supplements in the past month

`pr` Height

`femurbmd` Femur bone mineral density

`femurbmc` Femur bone mineral content

`fneckbmd` Femoral neck bone mineral density

`fneckbmc` Femoral neck bone mineral content

Source

From the NHANES web page, for NHANES 2009-2010.

References

US National Health and Nutrition Examination Survey, 2009-2010. From the US Center for Health Statistics.

Examples

```
data(SSRI)
summary(SSRI)
```

t_stat *Two sample t statistics*

Description

Compute the two-sample t statistics for each individual covariate in x.

Usage

```
t_stat(x,z)
```

Arguments

x	A matrix with length(z) rows giving the covariates.
z	A vector whose ith coordinate is 1 for a treated unit and is 0 for a control.

Value

A vector of two-sample t statistics for each column of x.

Examples

```
library(optmatch)
data("SSRI")
attach(SSRI)
X<-cbind(female,black,education)
dist<-DiPs::maha_dense(z,X)
o<-DiPs::match(z, dist, SSRI)
M0<-o$data
Xm<-cbind(M0$female,M0$black,M0$education)
t_stat(Xm,M0$z)
detach(SSRI)
```

wilcoxon_stat *Wilcoxon rank statistics*

Description

Compute the Wilcoxon rank statistics for each individual covariate in x.

Usage

```
wilcoxon_stat(x,z)
```

Arguments

x	A matrix with length(z) rows giving the covariates.
z	A vector whose ith coordinate is 1 for a treated unit and is 0 for a control.

Value

A vector of Wilcoxon rank statistics for each column of x.

Examples

```
library(optmatch)
data("SSRI")
attach(SSRI)
X<-cbind(female,black,education)
dist<-DiPs::maha_dense(z,X)
o<-DiPs::match(z, dist, SSRI)
M0<-o$data
Xm<-cbind(M0$female,M0$black,M0$education)
wilcoxon_stat(Xm,M0$z)
detach(SSRI)
```

Index