SISCER Module 15 Practice 2: Matching/Weighting and Sensitivity Analysis

Ting Ye and Qingyuan Zhao

July 2022

1

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.7     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggplot2)
library(tableone) # for creating table 1
source("optmatch.R")
```

```
## Loading required package: lattice
```

```
##
## Attaching package: 'caret'
```

```
## The following object is masked from 'package:purrr':
##
##     lift
```

```
df<-read.csv("nhanesi_class_dataset.csv",stringsAsFactors = TRUE)
df<-df[,-1]
summary(df)
```

```
##  physically.inactive years.lived.since.1971.up.to.1992     sex
##  Mode :logical       Min.   : 1.00                     Female:962
##  FALSE:1481          1st Qu.:14.00                     Male  :989
##  TRUE :470           Median :21.00
##                      Mean   :17.32
##                      3rd Qu.:21.00
##                      Max.   :21.00
##
##  smoking.status income.poverty.ratio age.at.interview       race
##  Current:604    Min.   :0.130        Min.   :45.00    Nonwhite: 274
##  Former :416    1st Qu.:1.215        1st Qu.:52.00    White   :1677
```

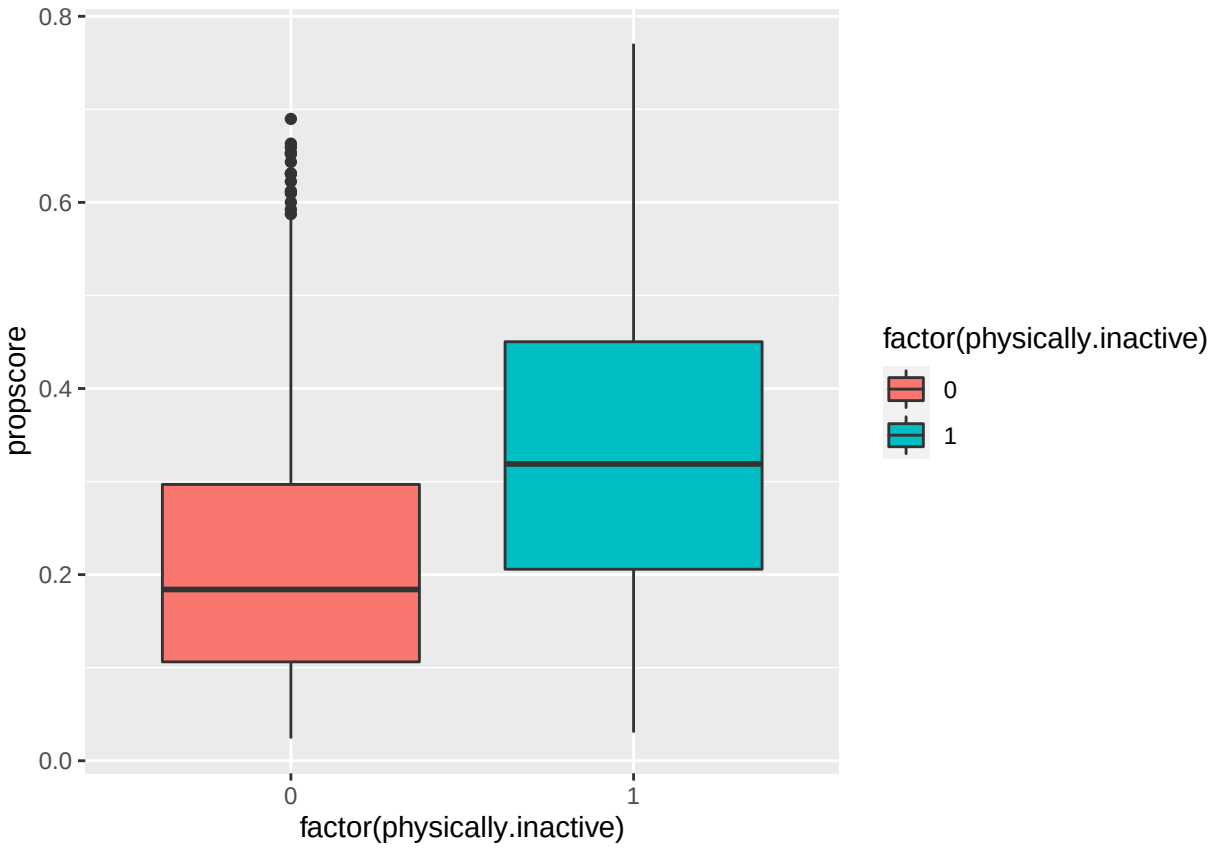
```
##  Never  :931    Median :2.280        Median :60.00
##                 Mean   :2.706        Mean   :59.84
##                 3rd Qu.:3.510        3rd Qu.:67.00
##                 Max.   :9.980        Max.   :74.00
##                 NA's   :544
##         education   working.last.three.months  married
##  0-8          :739   Mode :logical             Mode :logical
##  12           :439   FALSE:1063                FALSE:505
##  9-11         :415   TRUE :888                 TRUE :1446
##  College Grad:151
##  Missing      : 12
##  Some College:195
##
##                    alcohol.consumption dietary.adequacy
##  <1 time per month           :377      Min.   :0.000
##  1-4 times per month         :410      1st Qu.:3.000
##  2+ times per week           :205      Median :4.000
##  Just about everyday/everyday:300      Mean   :3.565
##  Never                       :659      3rd Qu.:5.000
##                                        Max.   :5.000
##                                        NA's   :479
```

```
# (a) create missing indicators for income.poverty.ratio and dietary.adequacy and imputate by observed
df<-df %>% mutate(income.poverty.ratio.missingind=1*(is.na(income.poverty.ratio)),
                  dietary.adequacy.missingind=1*(is.na(dietary.adequacy)))
df<-df %>% mutate(income.poverty.ratio=ifelse(is.na(income.poverty.ratio),
                                              mean(income.poverty.ratio,na.rm=TRUE),income.poverty.ratio
  mutate(dietary.adequacy=ifelse(is.na(dietary.adequacy),
                                 mean(dietary.adequacy,na.rm=TRUE),dietary.adequacy))
df<- df %>% mutate(physically.inactive=ifelse(physically.inactive,1,0))
```

2

```
# PS model should include all confounders and missing indicators
ps.formula=physically.inactive~sex+smoking.status+income.poverty.ratio+
  age.at.interview+race+education+working.last.three.months+married+
  alcohol.consumption+dietary.adequacy+income.poverty.ratio.missingind+
  dietary.adequacy.missingind
propscore.model=glm(ps.formula,family=binomial,data=df)
propscore<-propscore.model$fitted.values

ggplot(data=df, aes(x = factor(physically.inactive), y = propscore,
                    fill = factor(physically.inactive))) + geom_boxplot()
```




```
# excluding subjects lack of overlap
exclud1<-which(propscore>max(propscore[df$physically.inactive==0]))
exclud2<-which(propscore<min(propscore[df$physically.inactive==1]))
cbind(propscore,df)[c(exclud1,exclud2),] # check the subjects to be excluded
```

```
##       propscore physically.inactive years.lived.since.1971.up.to.1992    sex
## 272  0.74720177                   1                                6 Female
## 328  0.77056752                   1                               14 Female
## 371  0.70173323                   1                                1   Male
## 815  0.74575568                   1                               21 Female
## 123  0.03025210                   0                               14 Female
## 429  0.02386385                   0                               21   Male
## 602  0.02913807                   0                               21   Male
## 754  0.02532975                   0                               19   Male
## 1282 0.02928994                   0                               21 Female
##      smoking.status income.poverty.ratio age.at.interview     race    education
## 272         Current              2.70634               70 Nonwhite          0-8
## 328         Current              2.70634               63    White College Grad
## 371         Current              0.73000               55 Nonwhite          0-8
## 815         Current              0.40000               68 Nonwhite          0-8
## 123           Never              9.05000               65    White Some College
## 429          Former              8.89000               66    White           12
## 602          Former              8.86000               45    White Some College
## 754           Never              9.98000               62    White Some College
## 1282          Never              8.34000               57    White         9-11
##      working.last.three.months married          alcohol.consumption
## 272                       FALSE   FALSE                        Never
```

```
## 328                      FALSE   FALSE Just about everyday/everyday
## 371                      FALSE   FALSE                      Never
## 815                      FALSE   FALSE                      Never
## 123                       TRUE    TRUE Just about everyday/everyday
## 429                       TRUE    TRUE          1-4 times per month
## 602                       TRUE    TRUE            2+ times per week
## 754                       TRUE    TRUE                      Never
## 1282                      TRUE    TRUE            <1 time per month
##      dietary.adequacy income.poverty.ratio.missingind
## 272                 4                               1
## 328                 4                               1
## 371                 1                               0
## 815                 1                               0
## 123                 4                               0
## 429                 5                               0
## 602                 5                               0
## 754                 3                               0
## 1282                5                               0
##      dietary.adequacy.missingind
## 272                            0
## 328                            0
## 371                            0
## 815                            0
## 123                            0
## 429                            0
## 602                            0
## 754                            0
## 1282                           0
```

```
length(c(exclud1,exclud2)) # the number of units excluded
```

```
## [1] 9
```

```
df<-df[-c(exclud1,exclud2),]
```

3

```
#(d) Perform matching using optmatch
# Specify the model for propensity score (fitted by logistic regression)
ps.formula
```

```
## physically.inactive ~ sex + smoking.status + income.poverty.ratio +
##     age.at.interview + race + education + working.last.three.months +
##     married + alcohol.consumption + dietary.adequacy + income.poverty.ratio.missingind +
##     dietary.adequacy.missingind
```

```
# Specify all the variables you want to use in the Mahalanobis distance
# on the right handside of ~
mahal.formula=physically.inactive~sex+smoking.status+age.at.interview
# Perform paired matching
# Print standardized difference (both before and after matching) and the Love plot
match_res1<-optmatch_caliper(df,nocontrols.per.match = 1, calipersd=0.5,
ps.formula=ps.formula,mahal.formula=mahal.formula)
```

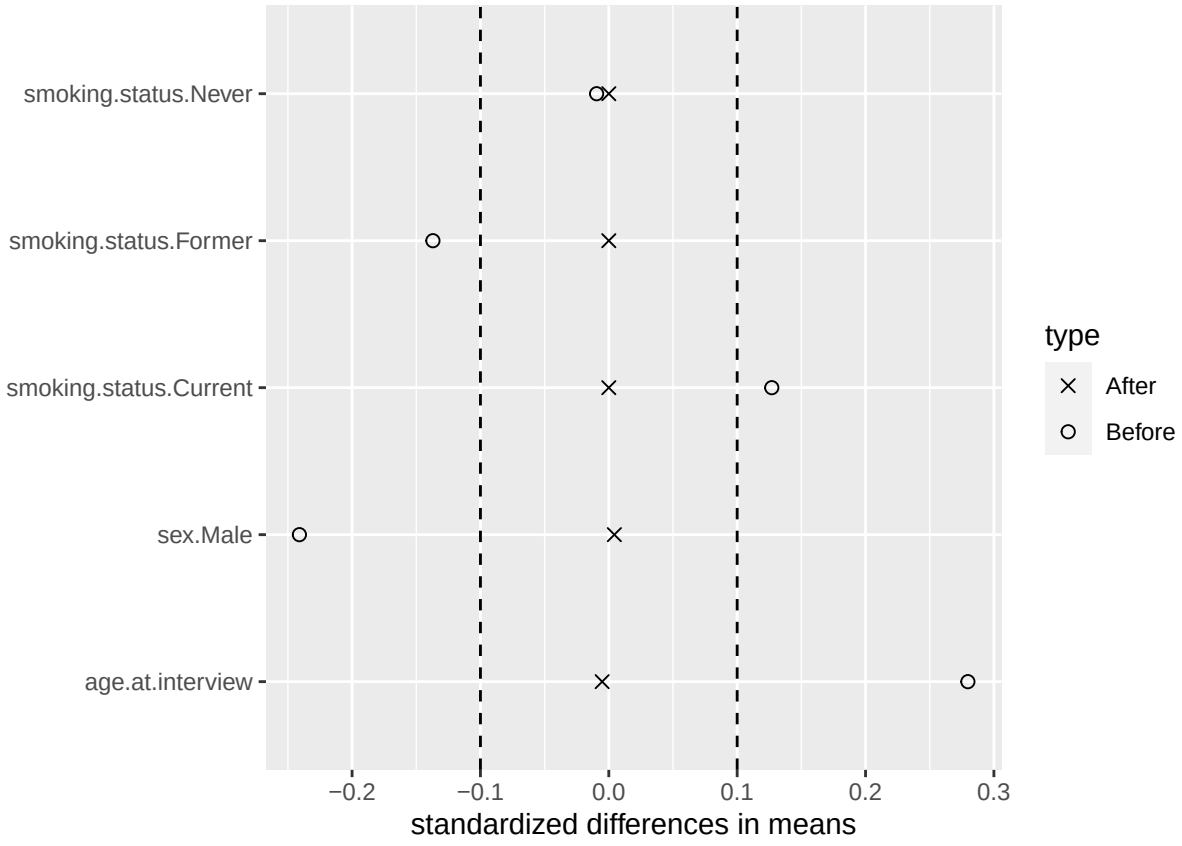
```
##
```

```
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##     select
```

```
##                         stand.diff.before stand.diff.after
## sex.Female                           0.24             0.00
## sex.Male                            -0.24             0.00
## smoking.status.Current               0.13             0.00
## smoking.status.Former               -0.14             0.00
## smoking.status.Never                -0.01             0.00
## age.at.interview                     0.28            -0.01
##                         treatmean.after controlmean.before controlmean.after
## sex.Female                         0.58               0.46              0.59
## sex.Male                           0.42               0.54              0.41
## smoking.status.Current             0.35               0.29              0.35
## smoking.status.Former              0.17               0.23              0.17
## smoking.status.Never               0.47               0.48              0.47
## age.at.interview                  61.69              59.25             61.73
```

```
match_res1$p
```




```
# Matched samples
df_treated<-match_res1$datatemp[match_res1$treated.subject.index,] # the treated data
df_control<-match_res1$datatemp[match_res1$matched.control.subject.index,] # the matched controls
df_matched<-rbind(df_treated,df_control)

# create a table 1
```

```
tb1<-CreateTableOne(vars=names(df)[-c(1:2)],strata="physically.inactive",data=df_matched)
tab1Mat<-print(tb1,showAllLevels = TRUE,smd=FALSE,quote=FALSE, noSpaces = TRUE, printToggle = FALSE)
tab1Mat # note this SMD is different than the standardized mean difference discussed in the lecture,
```

```
##                                                Stratified by physically.inactive
##                                                 level
##   n                                             ""
##   sex (%)                                       "Female"
##                                                 "Male"
##   smoking.status (%)                            "Current"
##                                                 "Former"
##                                                 "Never"
##   income.poverty.ratio (mean (SD))              ""
##   age.at.interview (mean (SD))                  ""
##   race (%)                                      "Nonwhite"
##                                                 "White"
##   education (%)                                 "0-8"
##                                                 "12"
##                                                 "9-11"
##                                                 "College Grad"
##                                                 "Missing"
##                                                 "Some College"
##   working.last.three.months (%)                 "FALSE"
##                                                 "TRUE"
##   married (%)                                   "FALSE"
##                                                 "TRUE"
##   alcohol.consumption (%)                       "<1 time per month"
##                                                 "1-4 times per month"
##                                                 "2+ times per week"
##                                                 "Just about everyday/everyday"
##                                                 "Never"
##   dietary.adequacy (mean (SD))                  ""
##   income.poverty.ratio.missingind (mean (SD)) ""
##   dietary.adequacy.missingind (mean (SD))       ""
##                                                Stratified by physically.inactive
##                                                 0              1
##   n                                             "466"          "466"
##   sex (%)                                       "273 (58.6)"   "272 (58.4)"
##                                                 "193 (41.4)"   "194 (41.6)"
##   smoking.status (%)                            "165 (35.4)"   "165 (35.4)"
##                                                 "80 (17.2)"    "80 (17.2)"
##                                                 "221 (47.4)"   "221 (47.4)"
##   income.poverty.ratio (mean (SD))              "2.36 (1.29)"  "2.20 (1.51)"
##   age.at.interview (mean (SD))                  "61.73 (8.75)" "61.69 (8.79)"
##   race (%)                                      "73 (15.7)"    "96 (20.6)"
##                                                 "393 (84.3)"   "370 (79.4)"
##   education (%)                                 "205 (44.0)"   "229 (49.1)"
##                                                 "85 (18.2)"    "78 (16.7)"
##                                                 "81 (17.4)"    "86 (18.5)"
##                                                 "46 (9.9)"     "39 (8.4)"
##                                                 "2 (0.4)"      "3 (0.6)"
##                                                 "47 (10.1)"    "31 (6.7)"
##   working.last.three.months (%)                 "346 (74.2)"   "350 (75.1)"
##                                                 "120 (25.8)"   "116 (24.9)"
```

```
##   married (%)                                    "152 (32.6)"   "174 (37.3)"
##                                                  "314 (67.4)"   "292 (62.7)"
##   alcohol.consumption (%)                        "95 (20.4)"    "93 (20.0)"
##                                                  "84 (18.0)"    "81 (17.4)"
##                                                  "45 (9.7)"     "42 (9.0)"
##                                                  "65 (13.9)"    "62 (13.3)"
##                                                  "177 (38.0)"   "188 (40.3)"
##   dietary.adequacy (mean (SD))                   "3.40 (1.26)"  "3.28 (1.33)"
##   income.poverty.ratio.missingind (mean (SD)) "0.35 (0.48)"  "0.27 (0.45)"
##   dietary.adequacy.missingind (mean (SD))     "0.32 (0.47)"  "0.23 (0.42)"
##                                                 Stratified by physically.inactive
##                                                  p       test
##   n                                              ""      ""
##   sex (%)                                        "1.000" ""
##                                                  ""      ""
##   smoking.status (%)                             "1.000" ""
##                                                  ""      ""
##                                                  ""      ""
##   income.poverty.ratio (mean (SD))               "0.089" ""
##   age.at.interview (mean (SD))                   "0.937" ""
##   race (%)                                       "0.061" ""
##                                                  ""      ""
##   education (%)                                  "0.323" ""
##                                                  ""      ""
##                                                  ""      ""
##                                                  ""      ""
##                                                  ""      ""
##                                                  ""      ""
##   working.last.three.months (%)                  "0.821" ""
##                                                  ""      ""
##   married (%)                                    "0.149" ""
##                                                  ""      ""
##   alcohol.consumption (%)                        "0.965" ""
##                                                  ""      ""
##                                                  ""      ""
##                                                  ""      ""
##                                                  ""      ""
##   dietary.adequacy (mean (SD))                   "0.168" ""
##   income.poverty.ratio.missingind (mean (SD)) "0.013" ""
##   dietary.adequacy.missingind (mean (SD))     "0.002" ""
```

```
# the latter uses the pre-matching data to estimate the pooled standard error
write.csv(tab1Mat, file = "Table1_nhanesi.csv") # saving this to your computer, so it is easy to paste

head(df)
```

```
##   physically.inactive years.lived.since.1971.up.to.1992    sex smoking.status
## 1                   1                                8 Female          Never
## 2                   0                               16   Male        Current
## 3                   0                               21   Male        Current
## 4                   0                               13   Male          Never
## 5                   0                               21 Female        Current
## 6                   0                               17   Male        Current
##   income.poverty.ratio age.at.interview     race    education
## 1              2.70634               74 Nonwhite          0-8
```

```
## 2              1.52000               74    White         9-11
## 3              3.11000               53 Nonwhite           12
## 4              9.18000               65    White College Grad
## 5              2.62000               47 Nonwhite         9-11
## 6              2.62000               53 Nonwhite         9-11
##   working.last.three.months married          alcohol.consumption
## 1                      FALSE   FALSE                        Never
## 2                      FALSE    TRUE Just about everyday/everyday
## 3                       TRUE    TRUE          1-4 times per month
## 4                       TRUE   FALSE            2+ times per week
## 5                       TRUE    TRUE                        Never
## 6                       TRUE    TRUE Just about everyday/everyday
##   dietary.adequacy income.poverty.ratio.missingind dietary.adequacy.missingind
## 1                2                               1                           0
## 2                5                               0                           0
## 3                3                               0                           0
## 4                2                               0                           0
## 5                4                               0                           0
## 6                2                               0                           0
```

```
catVars<-c("cm1ethrace","cm1age_c","cm1edu",
           "cm1relf","cm1hhinc_c","cm1bsex",
           "cf1ethrace","cf1age_c","cf1edu")
tb1<-CreateTableOne(vars=names(df),data=df,factorVars = catVars)
```

```
## Warning in ModuleReturnVarsExist(factorVars, data): The data frame does not
## have: cm1ethrace cm1age_c cm1edu cm1relf cm1hhinc_c cm1bsex cf1ethrace cf1age_c
## cf1edu Dropped
```

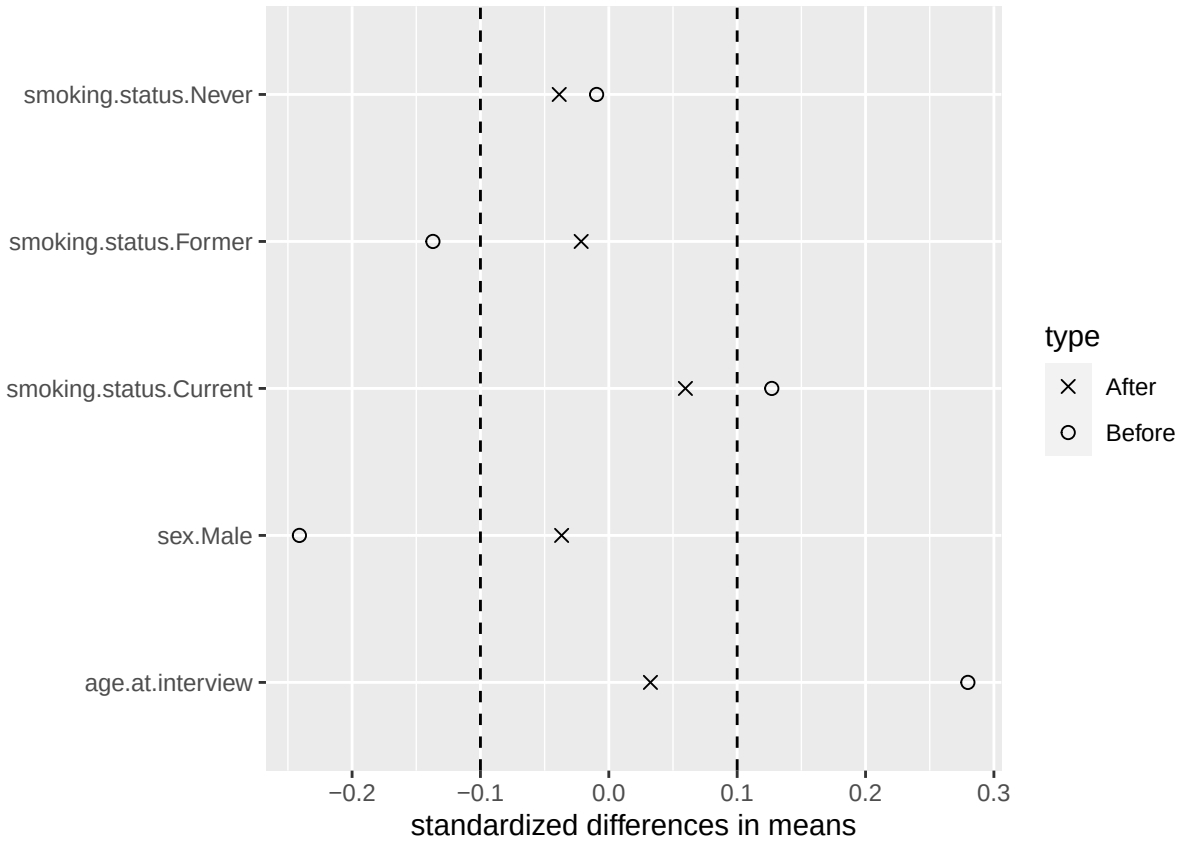
```
tab1Mat<-print(tb1,showAllLevels = TRUE,quote=FALSE, noSpaces = TRUE, printToggle = FALSE)
```

4

```
# Consider matching 2 controls to each treated unit. Still found adequate balance,
# all absolute standardized diff < 0.1
match_res2<-optmatch_caliper(df,nocontrols.per.match = 2, calipersd=0.5,
                             ps.formula=ps.formula,mahal.formula=mahal.formula)
```

```
##                         stand.diff.before stand.diff.after
## sex.Female                           0.24            0.04
## sex.Male                            -0.24           -0.04
## smoking.status.Current               0.13            0.06
## smoking.status.Former               -0.14           -0.02
## smoking.status.Never                -0.01           -0.04
## age.at.interview                     0.28            0.03
##                         treatmean.after controlmean.before controlmean.after
## sex.Female                         0.58              0.46             0.57
## sex.Male                           0.42              0.54             0.43
## smoking.status.Current             0.35              0.29             0.33
## smoking.status.Former              0.17              0.23             0.18
## smoking.status.Never               0.47              0.48             0.49
## age.at.interview                  61.69             59.25            61.40
```

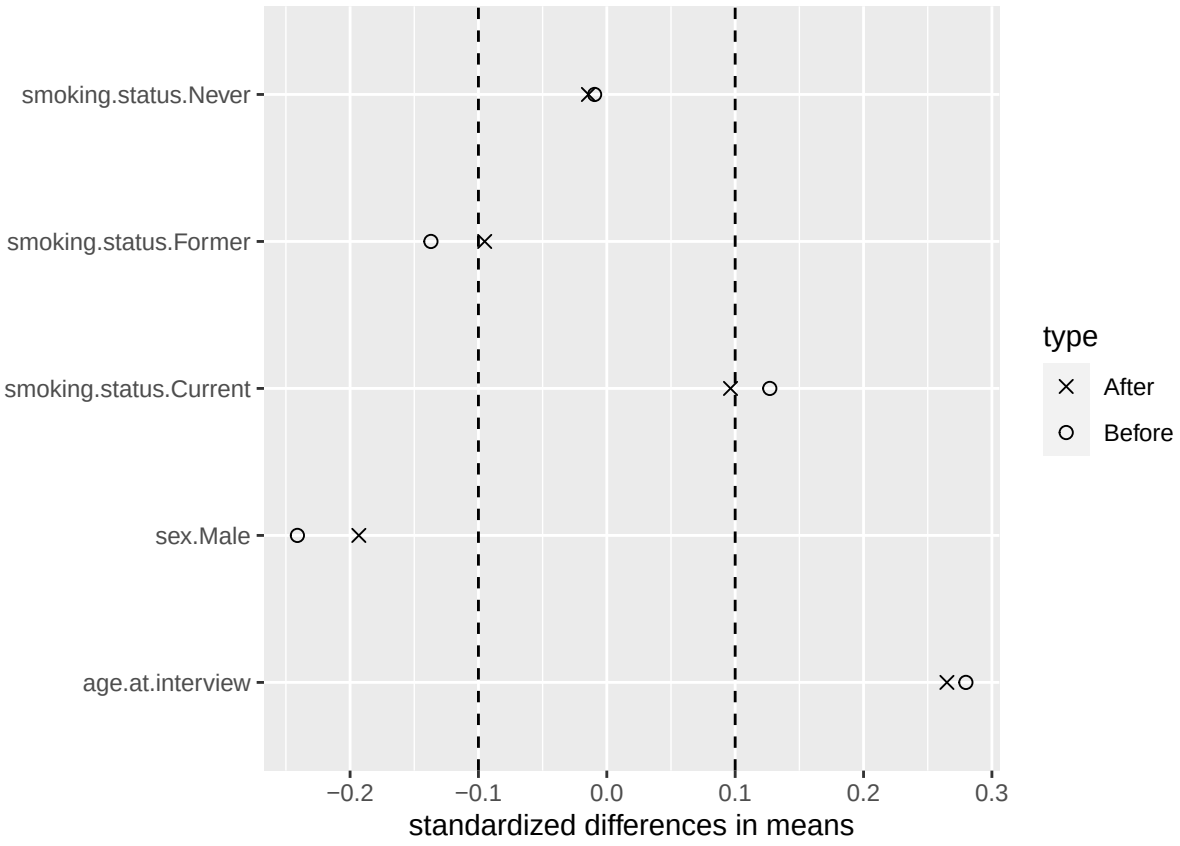
```
match_res2$p
```




```
#consider matching 3 controls to each treated unit. Not adequate balance
match_res3<-optmatch_caliper(df,nocontrols.per.match = 3, calipersd=0.5,
                             ps.formula=ps.formula,mahal.formula=mahal.formula)
```

```
##                          stand.diff.before stand.diff.after
## sex.Female                            0.24             0.19
## sex.Male                             -0.24            -0.19
## smoking.status.Current                0.13             0.10
## smoking.status.Former                -0.14            -0.10
## smoking.status.Never                 -0.01            -0.01
## age.at.interview                      0.28             0.26
##                          treatmean.after controlmean.before controlmean.after
## sex.Female                          0.58                0.46               0.49
## sex.Male                            0.42                0.54               0.51
## smoking.status.Current              0.35                0.29               0.31
## smoking.status.Former               0.17                0.23               0.21
## smoking.status.Never                0.47                0.48               0.48
## age.at.interview                   61.69               59.25              59.38
```

```
match_res3$p
```



5

I chose 1-v-2 match. Because it uses more controls and all standardized diff is below 0.1.

6

```
# For paired matching
df_treated<-match_res1$datatemp[match_res1$treated.subject.index,] # the treated data
df_control<-match_res1$datatemp[match_res1$matched.control.subject.index,] # the matched controls
df_matched<-rbind(df_treated,df_control)
# Wilcoxon signed rank test, a randomization test.
# This is only appliable to matched pairs
# Exact = FALSE implements the large sample approximation (the familiar chi-square test)
# which is fine for most samples of practical size.
library(exactRankTests)
```

```
##  Package 'exactRankTests' is no longer under development.
##  Please consider using package 'coin' instead.
```

```
wilcox.exact(df_treated$years.lived.since.1971.up.to.1992,
    df_control$years.lived.since.1971.up.to.1992,
    paired=TRUE,conf.int=TRUE,exact = TRUE)
```

```
##
##  Exact Wilcoxon signed rank test
##
```

```
## data:  df_treated$years.lived.since.1971.up.to.1992 and df_control$years.lived.since.1971.up.to.1992
## V = 16134, p-value = 3.206e-11
## alternative hypothesis: true mu is not equal to 0
## 95 percent confidence interval:
##  -4.5 -2.5
## sample estimates:
## (pseudo)median
##          -3.25
```

```
# this is the confidence interval about location parameter mu, not difference in means

# Regression method, adjusting for all match set indicators
matched.reg.model=lm(years.lived.since.1971.up.to.1992~physically.inactive+matchvec+
                       sex+smoking.status+income.poverty.ratio+age.at.interview+race+
                       education+working.last.three.months+married+alcohol.consumption+
                       dietary.adequacy+income.poverty.ratio.missingind+
                       dietary.adequacy.missingind,data=df_matched)

coef(matched.reg.model)[2] # Point estimate of treatment effect
```

```
## physically.inactive
##           -1.916665
```

```
confint(matched.reg.model)[2,] # Confidence interval
```

```
##     2.5 %    97.5 %
## -2.646187 -1.187144
```

```
# For 1-v-2 matching
df_treated2<-match_res2$datatemp[match_res2$treated.subject.index,] # the treated data
df_control2<-match_res2$datatemp[match_res2$matched.control.subject.index,] # the matched controls
df_matched2<-rbind(df_treated2,df_control2)
# Regression method, adjusting for all match set indicators
matched.reg.model2=lm(years.lived.since.1971.up.to.1992~physically.inactive+matchvec+
                       sex+smoking.status+income.poverty.ratio+age.at.interview+race+
                       education+working.last.three.months+married+alcohol.consumption+
                       dietary.adequacy+income.poverty.ratio.missingind+
                       dietary.adequacy.missingind,data=df_matched2)
coef(matched.reg.model2)[2] # Point estimate of treatment effect
```

```
## physically.inactive
##           -1.415638
```

```
confint(matched.reg.model2)[2,] # Confidence interval
```

```
##      2.5 %     97.5 %
## -2.2782274 -0.5530493
```

7

Potential unmeasured confounders: stress, ...

8

```
#### Sensitivity Analysis for Signed Rank Statistic
#### Assumes that a one-sided test of whether the treatment has
#### a positive effect is being done.
#### diff is the difference between the treated and control units
#### gamma is the sensitivity parameter

library(DOS)
diff=df_treated$years.lived.since.1971.up.to.1992-
  df_control$years.lived.since.1971.up.to.1992
senWilcox(diff, gamma = 1, conf.int = FALSE, alpha = 0.05, alternative = "less")
```

```
## $pval
## [1] 9.539214e-11
```

```
# gamma=1 is the primary analysis assuming no unmeasured confounders
senWilcox(diff, gamma = 1.7, conf.int = FALSE, alpha = 0.05, alternative = "less")
```

```
## $pval
## [1] 0.02630124
```

```
# unmeasured bias up to gamma=1.7 cannot explain away the treatment effect
senWilcox(diff, gamma = 1.8, conf.int = FALSE, alpha = 0.05, alternative = "less")
```

```
## $pval
## [1] 0.07017079
```

```
# gamma=1.8 would explain away the treatment effect because the p-value is now larger than alpha=0.05
```

9: Alternative estimators: outcome regression, IPW, AIPW

```
# unadjusted. Would be biased due to confounding
reg.model=lm(years.lived.since.1971.up.to.1992~physically.inactive,data=df)
coef(reg.model)[2] # Point estimate of treatment effect
```

```
## physically.inactive
##           -2.889424
```

```
confint(reg.model)[2,] # Confidence interval
```

```
##     2.5 %    97.5 %
## -3.434285 -2.344563
```

```
library(CausalGAM) # this package implements the stablized version
```

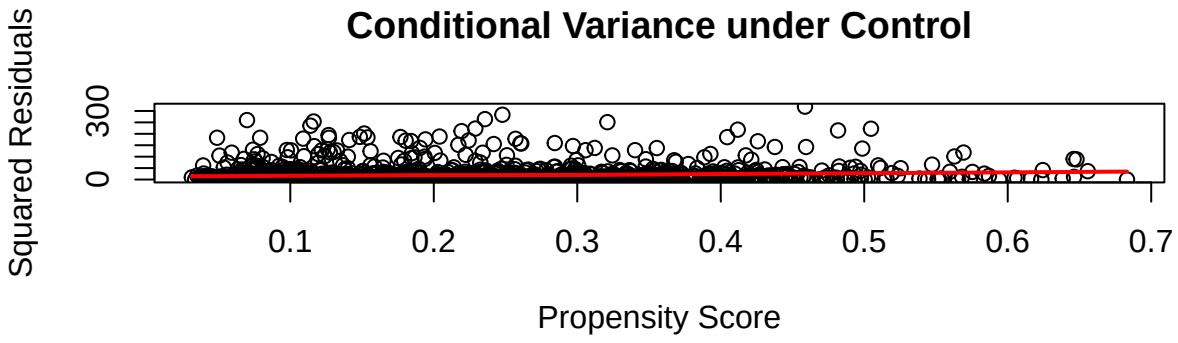
```
## Loading required package: gam
## Loading required package: splines
## Loading required package: foreach
##
## Attaching package: 'foreach'
## The following objects are masked from 'package:purrr':
##
##     accumulate, when
## Loaded gam 1.20.2
```

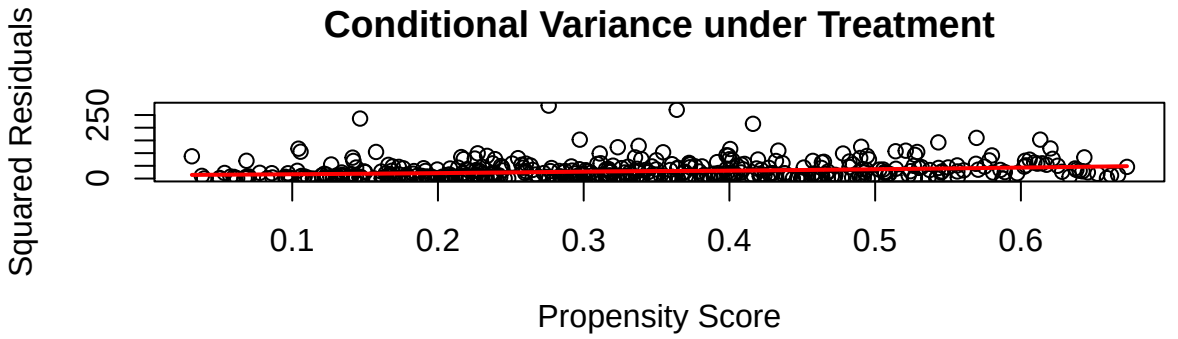
```
## ##
## ## CausalGAM Package

## ## Copyright (C) 2009 Adam Glynn and Kevin Quinn
```

```
out1.formula<-years.lived.since.1971.up.to.1992~sex+smoking.status+income.poverty.ratio+
  age.at.interview+race+education+working.last.three.months+married+
  alcohol.consumption+dietary.adequacy+income.poverty.ratio.missingind+
                       dietary.adequacy.missingind

ATE.out<-estimate.ATE(pscore.formula=ps.formula,pscore.family = binomial,
                      outcome.formula.t=out1.formula,outcome.formula.c=out1.formula,
                      outcome.family=gaussian,
                      treatment.var="physically.inactive",data=df,nboot=0)
```






```
print(ATE.out)
```

```
## ###########################################################################
## AIPW Estimator:
## ---------------------------------------------------------------------------
## Estimated ATE:  -1.8731
##                      SE           z-statistic            Pr(>|z|)
## Emp. Sandwich      0.3144            -5.9571              < 0.0001
## Estim. Asymptotic  0.2828            -6.6231              < 0.0001
## ###########################################################################
##
## ###########################################################################
## IPW Estimator:
## ---------------------------------------------------------------------------
## Estimated ATE:  -1.6107
##                      SE           z-statistic            Pr(>|z|)
```

```
## Estim. Asymptotic  0.2829            -5.6946               < 0.0001
## ################################################################################
##
## ################################################################################
## Regression Estimator:
## --------------------------------------------------------------------------------
## Estimated ATE:  -1.8375
##                       SE            z-statistic            Pr(>|z|)
## Estim. Asymptotic  0.2828              -6.4975               < 0.0001
## ################################################################################
##
## ################################################################################
## General Information
## --------------------------------------------------------------------------------
## Control Value:  physically.inactive = 0
## Treated Value:  physically.inactive = 1
## Number of Discarded Units:  0
## Number of Truncated Propensity Scores:  0
## Number of Treated Units Before Discards/Truncations:  466
## Number of Treated Units not Discarded/Truncated:      466
## Number of Control Units Before Discards/Truncations:  1476
## Number of Control Units not Discarded/Truncated:      1476
## ################################################################################
```

```
# validate the IPW estimator from the CausalGAM package
ps.fitted<-glm(ps.formula,family=binomial,data=df)$fitted.values
my.ipw<-mean(with(df,physically.inactive*years.lived.since.1971.up.to.1992/ps.fitted))-
mean(with(df,(1-physically.inactive)*years.lived.since.1971.up.to.1992/(1-ps.fitted))) # non-stablized
print(paste("my ipw =", round(my.ipw,4)))
```

```
## [1] "my ipw = -1.1186"
```

```
my.sipw<-sum(with(df,physically.inactive*years.lived.since.1971.up.to.1992/ps.fitted))/
  sum(with(df,physically.inactive/ps.fitted))-
sum(with(df,(1-physically.inactive)*years.lived.since.1971.up.to.1992/(1-ps.fitted)))/
  sum(with(df,(1-physically.inactive)/(1-ps.fitted))) # stablized IPW
print(paste("my sipw =", round(my.sipw,4)))
```

```
## [1] "my sipw = -1.6107"
```

```
#entropy balancing
library(ebal)
```

```
## ##
## ## ebal Package: Implements Entropy Balancing.
```

```
## ## See http://www.stanford.edu/~jhain/ for additional information.
```

```
A<-df$physically.inactive
X<-model.matrix(glm(ps.formula,family=binomial,data=df))
X<-X[,-1]
Y<-df$years.lived.since.1971.up.to.1992

eb.out<-ebalance(Treatment = A, X=X)
```

```
## Converged within tolerance
```

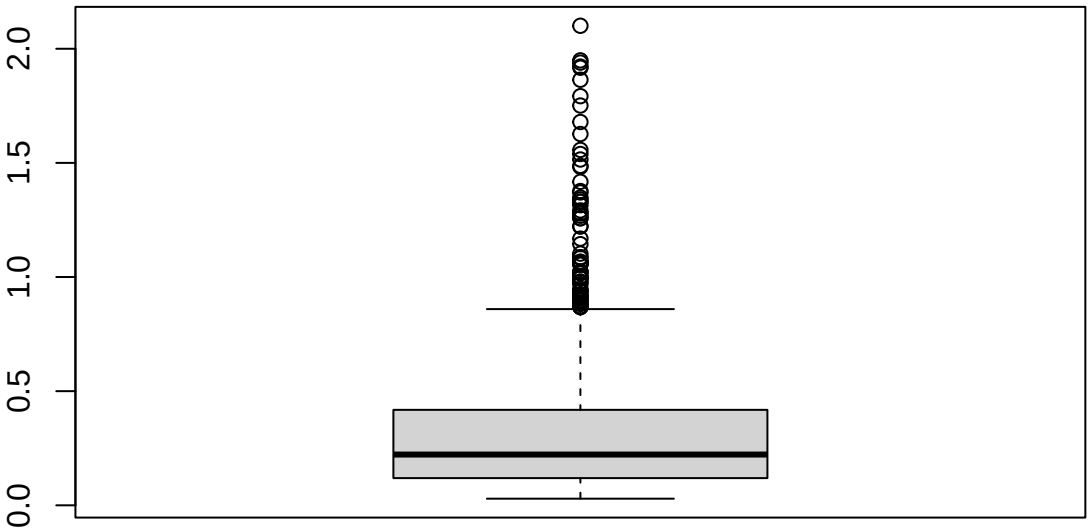
```
res<-data.frame(treated=apply(X[A==1,],2,mean),weighted.control=apply(X[A==0,],2,weighted.mean,w=eb.out$
res
```

```
##                                                    treated weighted.control
## sexMale                                        0.416309013      0.416309736
## smoking.statusFormer                           0.171673820      0.171674149
## smoking.statusNever                            0.474248927      0.474248555
## income.poverty.ratio                           2.202049669      2.202054549
## age.at.interview                              61.686695279     61.686678489
## raceWhite                                      0.793991416      0.793991578
## education12                                    0.167381974      0.167382219
## education9-11                                  0.184549356      0.184549418
## educationCollege Grad                          0.083690987      0.083691073
## educationMissing                               0.006437768      0.006437763
## educationSome College                          0.066523605      0.066523842
## working.last.three.monthsTRUE                  0.248927039      0.248928948
## marriedTRUE                                    0.626609442      0.626609654
## alcohol.consumption1-4 times per month         0.173819742      0.173819909
## alcohol.consumption2+ times per week           0.090128755      0.090128826
## alcohol.consumptionJust about everyday/everyday 0.133047210     0.133047498
## alcohol.consumptionNever                       0.403433476      0.403433044
## dietary.adequacy                               3.277570442      3.277571376
## income.poverty.ratio.missingind                0.272532189      0.272531904
## dietary.adequacy.missingind                    0.225321888      0.225321630
```

```
summary(eb.out$w)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0291  0.1192  0.2223  0.3157  0.4181  2.1006
```

```
boxplot(eb.out$w)
```



10

```
# sensitivity analysis using bootsens
library(bootsens)
# ipw
```

```r
A<-df$physically.inactive
X<-model.matrix(glm(ps.formula,family=binomial,data=df))
X<-X[,-1]
Y<-df$years.lived.since.1971.up.to.1992

## IPW, assuming no unmeasure confounder (i.e. gamma = 0 or Gamma = e^0 = 1)
extrema.os(A, X, Y) # point estimate
```

```
## [1] -1.610729 -1.610729
```

```r
bootsens.os(A, X, Y, parallel = FALSE) # bootstrap confidence interval
```

```
##      2.5%     97.5%
## -2.257772 -1.023293
```

```r
## IPW, Sensitivity analysis (Gamma = exp(gamma))
extrema.os(A, X, Y, gamma = log(1.2)) # point estimate
```

```
## [1] -2.5119367 -0.7512871
```

```r
bootsens.os(A, X, Y, gamma =  log(1.2), parallel = FALSE) # bootstrap CI
```

```
##       2.5%      97.5%
## -3.1532805 -0.2151229
```

```r
# the IPW estimator is robust to Gamma=1.2

## AIPW, assuming no unmeasure confounder (i.e. gamma = 0 or Gamma = e^0 = 1)
extrema.os(A, X, Y, reg.adjust = TRUE) # point estimate
```

```
## [1] -1.872209 -1.872209
```

```r
options(warn=-1) # there is a category with small number,
# which leads to some bootstrap samples without that category and generates warnings.
# this is to temporarily turn off the warning
bootsens.os(A, X, Y, reg.adjust = TRUE, parallel = FALSE) # bootstrap confidence interval
```

```
##      2.5%     97.5%
## -2.556446 -1.274457
```

```r
## AIPW, Sensitivity analysis (Gamma = exp(gamma))
extrema.os(A, X, Y, reg.adjust = TRUE,gamma = log(1.2)) # point estimate
```

```
## [1] -2.555729 -1.206745
```

```r
bootsens.os(A, X, Y, reg.adjust = TRUE, gamma = log(1.2),parallel = FALSE) # bootstrap CI
```

```
##       2.5%      97.5%
## -3.1896053 -0.6580529
```

```r
bootsens.os(A, X, Y, reg.adjust = TRUE, gamma = log(1.4),parallel = FALSE) # bootstrap CI
```

```
##       2.5%      97.5%
## -3.8506294 -0.0941925
```

```r
# the AIPW estimator is robust to Gamma=1.4, because it has higher power compared to the IPW
options(warn=0)
```