matching\_lab\_final.R

81983

2019-11-09

##############################  
#Matching Methods   
##############################  
  
# Install packages  
#install.packages("MatchIt")  
#install.packages("optmatch")  
#install.packages("RItools")  
#install.packages("texreg")  
#install.packages("Matching")  
#install.packages("rgenoud")  
#install.packages("Zelig")  
  
# Load packages  
library(MatchIt)  
library(optmatch)

## Loading required package: survival

## The optmatch package has an academic license. Enter relaxinfo() for more information.

library(RItools)

## Loading required package: SparseM

##   
## Attaching package: 'SparseM'

## The following object is masked from 'package:base':  
##   
## backsolve

library(Matching)

## Loading required package: MASS

## ##   
## ## Matching (Version 4.9-6, Build Date: 2019-04-07)  
## ## See http://sekhon.berkeley.edu/matching for additional documentation.  
## ## Please cite software as:  
## ## Jasjeet S. Sekhon. 2011. ``Multivariate and Propensity Score Matching  
## ## Software with Automated Balance Optimization: The Matching package for R.''  
## ## Journal of Statistical Software, 42(7): 1-52.   
## ##

library(rgenoud)

## ## rgenoud (Version 5.8-3.0, Build Date: 2019-01-22)  
## ## See http://sekhon.berkeley.edu/rgenoud for additional documentation.  
## ## Please cite software as:  
## ## Walter Mebane, Jr. and Jasjeet S. Sekhon. 2011.  
## ## ``Genetic Optimization Using Derivatives: The rgenoud package for R.''  
## ## Journal of Statistical Software, 42(11): 1-26.   
## ##

library(texreg)

## Version: 1.36.23  
## Date: 2017-03-03  
## Author: Philip Leifeld (University of Glasgow)  
##   
## Please cite the JSS article in your publications -- see citation("texreg").

library(tableone)  
library(personalized)

## Loading required package: glmnet

## Loading required package: Matrix

## Loading required package: foreach

## Loaded glmnet 2.0-18

## Loading required package: mgcv

## Loading required package: nlme

## This is mgcv 1.8-28. For overview type 'help("mgcv-package")'.

## Loading required package: gbm

## Loaded gbm 2.1.5

## Loading required package: ggplot2

## Loading required package: plotly

##   
## Attaching package: 'plotly'

## The following object is masked from 'package:ggplot2':  
##   
## last\_plot

## The following object is masked from 'package:MASS':  
##   
## select

## The following object is masked from 'package:stats':  
##   
## filter

## The following object is masked from 'package:graphics':  
##   
## layout

library(Zelig)

##   
## Attaching package: 'Zelig'

## The following object is masked from 'package:ggplot2':  
##   
## stat

## Data  
# Load data: a subset of the LaLonde data is loaded with the MatchIt library  
data(lalonde)  
?lalonde

## starting httpd help server ...

## done

#To read data into R from outside source, use: read.table, read.csv, etc.  
#Or from foreign library, use: read.dta, read.sav, etc.  
#Ex: library(foreign)  
#Ex: dataname <- read.dta('path/filename.dta')  
  
# Examine data briefly  
dim(lalonde)

## [1] 445 12

names(lalonde)

## [1] "age" "educ" "black" "hisp" "married" "nodegr" "re74"   
## [8] "re75" "re78" "u74" "u75" "treat"

str(lalonde)

## 'data.frame': 445 obs. of 12 variables:  
## $ age : int 37 22 30 27 33 22 23 32 22 33 ...  
## $ educ : int 11 9 12 11 8 9 12 11 16 12 ...  
## $ black : int 1 0 1 1 1 1 1 1 1 0 ...  
## $ hisp : int 0 1 0 0 0 0 0 0 0 0 ...  
## $ married: int 1 0 0 0 0 0 0 0 0 1 ...  
## $ nodegr : int 1 1 0 1 1 1 0 1 0 0 ...  
## $ re74 : num 0 0 0 0 0 0 0 0 0 0 ...  
## $ re75 : num 0 0 0 0 0 0 0 0 0 0 ...  
## $ re78 : num 9930 3596 24910 7506 290 ...  
## $ u74 : int 1 1 1 1 1 1 1 1 1 1 ...  
## $ u75 : int 1 1 1 1 1 1 1 1 1 1 ...  
## $ treat : int 1 1 1 1 1 1 1 1 1 1 ...

head(lalonde)

## age educ black hisp married nodegr re74 re75 re78 u74 u75 treat  
## 1 37 11 1 0 1 1 0 0 9930.05 1 1 1  
## 2 22 9 0 1 0 1 0 0 3595.89 1 1 1  
## 3 30 12 1 0 0 0 0 0 24909.50 1 1 1  
## 4 27 11 1 0 0 1 0 0 7506.15 1 1 1  
## 5 33 8 1 0 0 1 0 0 289.79 1 1 1  
## 6 22 9 1 0 0 1 0 0 4056.49 1 1 1

summary(lalonde)

## age educ black hisp   
## Min. :17.00 Min. : 3.0 Min. :0.0000 Min. :0.00000   
## 1st Qu.:20.00 1st Qu.: 9.0 1st Qu.:1.0000 1st Qu.:0.00000   
## Median :24.00 Median :10.0 Median :1.0000 Median :0.00000   
## Mean :25.37 Mean :10.2 Mean :0.8337 Mean :0.08764   
## 3rd Qu.:28.00 3rd Qu.:11.0 3rd Qu.:1.0000 3rd Qu.:0.00000   
## Max. :55.00 Max. :16.0 Max. :1.0000 Max. :1.00000   
## married nodegr re74 re75   
## Min. :0.0000 Min. :0.000 Min. : 0.0 Min. : 0   
## 1st Qu.:0.0000 1st Qu.:1.000 1st Qu.: 0.0 1st Qu.: 0   
## Median :0.0000 Median :1.000 Median : 0.0 Median : 0   
## Mean :0.1685 Mean :0.782 Mean : 2102.3 Mean : 1377   
## 3rd Qu.:0.0000 3rd Qu.:1.000 3rd Qu.: 824.4 3rd Qu.: 1221   
## Max. :1.0000 Max. :1.000 Max. :39570.7 Max. :25142   
## re78 u74 u75 treat   
## Min. : 0 Min. :0.0000 Min. :0.0000 Min. :0.0000   
## 1st Qu.: 0 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000   
## Median : 3702 Median :1.0000 Median :1.0000 Median :0.0000   
## Mean : 5301 Mean :0.7326 Mean :0.6494 Mean :0.4157   
## 3rd Qu.: 8125 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000   
## Max. :60308 Max. :1.0000 Max. :1.0000 Max. :1.0000

table(lalonde$treat)

##   
## 0 1   
## 260 185

#Check Covariate Balance  
vars <- c("age" , "educ" , "black" , "hisp" , "married" ,"nodegr" ,"re74" , "re75" )  
## Construct a table  
covbal0 <- CreateTableOne(vars = vars, strata = "treat", data = lalonde, test = FALSE, smd=TRUE)  
print(covbal0, smd = TRUE)

## Stratified by treat  
## 0 1 SMD   
## n 260 185   
## age (mean (SD)) 25.05 (7.06) 25.82 (7.16) 0.107  
## educ (mean (SD)) 10.09 (1.61) 10.35 (2.01) 0.141  
## black (mean (SD)) 0.83 (0.38) 0.84 (0.36) 0.044  
## hisp (mean (SD)) 0.11 (0.31) 0.06 (0.24) 0.175  
## married (mean (SD)) 0.15 (0.36) 0.19 (0.39) 0.094  
## nodegr (mean (SD)) 0.83 (0.37) 0.71 (0.46) 0.304  
## re74 (mean (SD)) 2107.03 (5687.91) 2095.57 (4886.62) 0.002  
## re75 (mean (SD)) 1266.91 (3102.98) 1532.06 (3219.25) 0.084

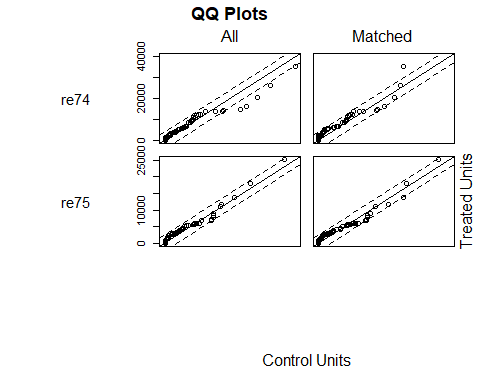
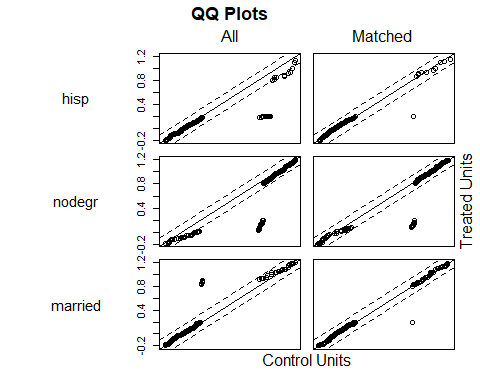
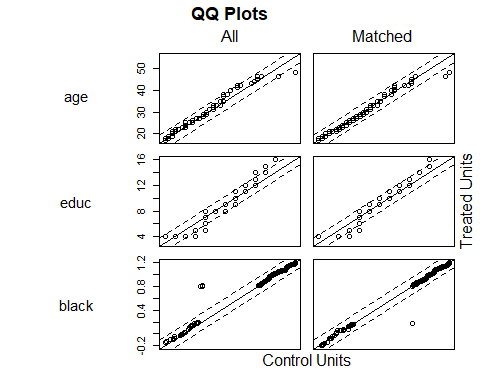
#OTHER FUNCTIONS FOR COVARIATE BALANCE CHECKING  
# Check initial balance (from RItools package):   
# How similar are treatment/control groups on X?  
#?xBalance  
#xBalance(treat ~ . - (re78 + treat), data =lalonde, report=c("adj.means","std.diffs"))   
#xBalance(treat ~ . - (re78 + treat), data =lalonde, report=c("all"))   
#?balanceUV  
#RE75.BAL <- balanceUV(lalonde$re75[lalonde$treat==1],lalonde$re75[lalonde$treat!=1])  
#summary(RE75.BAL)  
  
  
  
## Example USING MATCHIT PACKAGE   
  
# 1. Implement nearest neighbor matching on the propensity score (NN match on PS score: 1:1, greedy, without replacement or calipers)  
psmatch1 <- matchit(treat ~ age + educ + black + hisp + nodegr + married + re74 + re75,   
 distance="logit", method = "nearest", discard = "control", data = lalonde)  
  
# 2. Check balance  
summary(psmatch1, standardize=TRUE)

##   
## Call:  
## matchit(formula = treat ~ age + educ + black + hisp + nodegr +   
## married + re74 + re75, data = lalonde, method = "nearest",   
## distance = "logit", discard = "control")  
##   
## Summary of balance for all data:  
## Means Treated Means Control SD Control Std. Mean Diff. eCDF Med  
## distance 0.4377 0.4001 0.0933 0.3935 0.1143  
## age 25.8162 25.0538 7.0577 0.1066 0.0222  
## educ 10.3459 10.0885 1.6143 0.1281 0.0127  
## black 0.8432 0.8269 0.3790 0.0448 0.0082  
## hisp 0.0595 0.1077 0.3106 -0.2034 0.0241  
## nodegr 0.7081 0.8346 0.3722 -0.2775 0.0633  
## married 0.1892 0.1538 0.3615 0.0900 0.0177  
## re74 2095.5740 2107.0268 5687.9067 -0.0023 0.0158  
## re75 1532.0556 1266.9092 3102.9830 0.0824 0.0620  
## eCDF Mean eCDF Max  
## distance 0.1117 0.2140  
## age 0.0254 0.0652  
## educ 0.0287 0.1265  
## black 0.0082 0.0163  
## hisp 0.0241 0.0482  
## nodegr 0.0633 0.1265  
## married 0.0177 0.0353  
## re74 0.0192 0.0471  
## re75 0.0508 0.1075  
##   
##   
## Summary of balance for matched data:  
## Means Treated Means Control SD Control Std. Mean Diff. eCDF Med  
## distance 0.4377 0.4267 0.0910 0.1148 0.0216  
## age 25.8162 25.8865 7.3094 -0.0098 0.0108  
## educ 10.3459 10.1405 1.7168 0.1022 0.0108  
## black 0.8432 0.8486 0.3594 -0.0148 0.0027  
## hisp 0.0595 0.0649 0.2470 -0.0228 0.0027  
## nodegr 0.7081 0.7730 0.4200 -0.1423 0.0324  
## married 0.1892 0.2000 0.4011 -0.0275 0.0054  
## re74 2095.5740 1653.0266 4509.2061 0.0906 0.0270  
## re75 1532.0556 1352.9973 3345.1146 0.0556 0.0595  
## eCDF Mean eCDF Max  
## distance 0.0353 0.1027  
## age 0.0118 0.0432  
## educ 0.0224 0.0757  
## black 0.0027 0.0054  
## hisp 0.0027 0.0054  
## nodegr 0.0324 0.0649  
## married 0.0054 0.0108  
## re74 0.0360 0.0865  
## re75 0.0509 0.1135  
##   
## Percent Balance Improvement:  
## Std. Mean Diff. eCDF Med eCDF Mean eCDF Max  
## distance 70.8237 81.0909 68.3957 52.0155  
## age 90.7827 51.2881 53.5022 33.6523  
## educ 20.2261 14.7541 21.9664 40.1808  
## black 66.8790 66.8790 66.8790 66.8790  
## hisp 88.7931 88.7931 88.7931 88.7931  
## nodegr 48.7264 48.7264 48.7264 48.7264  
## married 69.4118 69.4118 69.4118 69.4118  
## re74 -3764.0926 -71.0526 -87.4591 -83.6645  
## re75 32.4681 4.0268 -0.1187 -5.6093  
##   
## Sample sizes:  
## Control Treated  
## All 260 185  
## Matched 185 185  
## Unmatched 64 0  
## Discarded 11 0

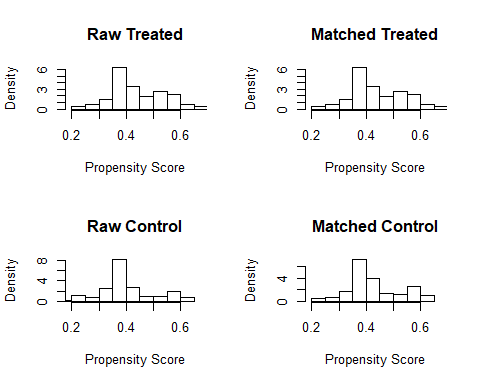
#compare with initial sample  
print(covbal0, smd = TRUE)

## Stratified by treat  
## 0 1 SMD   
## n 260 185   
## age (mean (SD)) 25.05 (7.06) 25.82 (7.16) 0.107  
## educ (mean (SD)) 10.09 (1.61) 10.35 (2.01) 0.141  
## black (mean (SD)) 0.83 (0.38) 0.84 (0.36) 0.044  
## hisp (mean (SD)) 0.11 (0.31) 0.06 (0.24) 0.175  
## married (mean (SD)) 0.15 (0.36) 0.19 (0.39) 0.094  
## nodegr (mean (SD)) 0.83 (0.37) 0.71 (0.46) 0.304  
## re74 (mean (SD)) 2107.03 (5687.91) 2095.57 (4886.62) 0.002  
## re75 (mean (SD)) 1266.91 (3102.98) 1532.06 (3219.25) 0.084

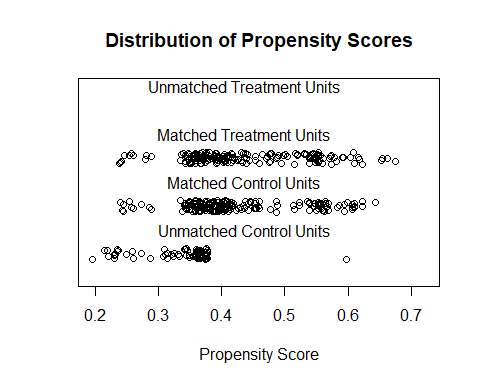
plot(psmatch1)



plot(psmatch1, type="hist")



par(mfrow = c(1, 1))  
plot(psmatch1, type="jitter", interactive=FALSE)



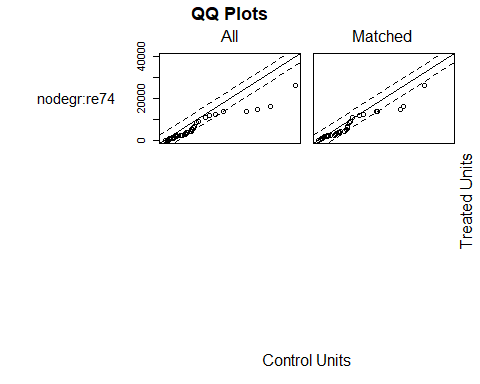
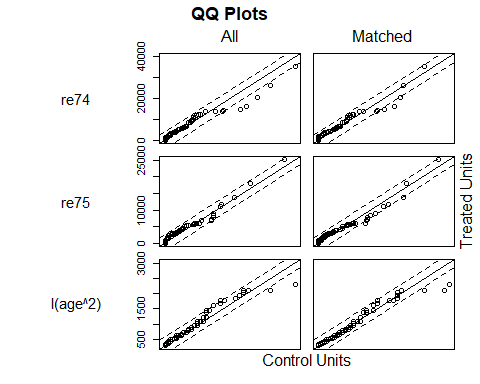
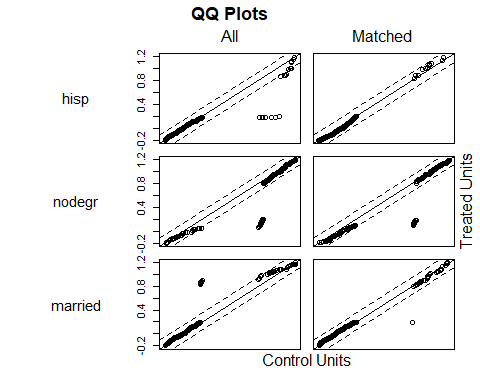
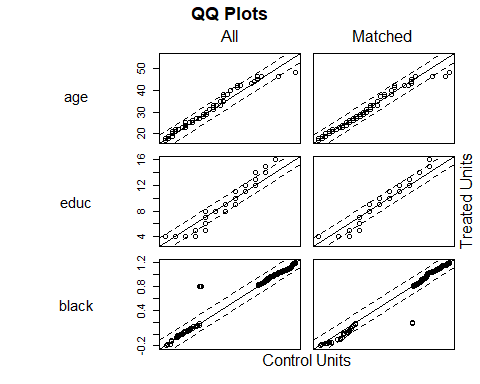
## More on nearest neighbor matching and the propensity score  
  
# Estimating a propensity score, implementing a matching procedure, checking balance  
  
# Optimal (instead of greedy) matching on PS score: 1:1, without replacement or calipers  
# Now Add polynomials and interactions to propensity score model  
psmatch2 <- matchit(treat ~ age + educ + black + hisp + nodegr + married + re74 +re75  
 + I(age^2) + nodegr:re74, distance = "logit", method = "optimal", data = lalonde)

## Warning in optmatch::fullmatch(d, min.controls = ratio, max.controls = ratio, : Without 'data' argument the order of the match is not guaranteed  
## to be the same as your original data.

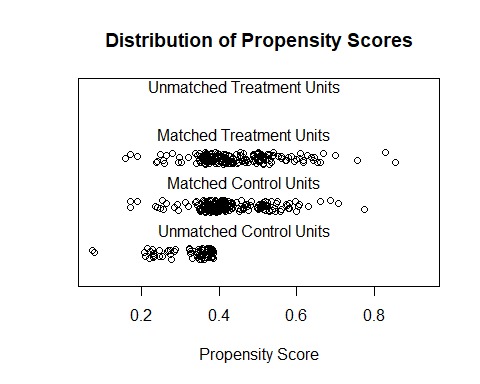
#authors of the R package say to ignore this warning..  
  
# Check balance  
summary(psmatch2, standardize=TRUE)

##   
## Call:  
## matchit(formula = treat ~ age + educ + black + hisp + nodegr +   
## married + re74 + re75 + I(age^2) + nodegr:re74, data = lalonde,   
## method = "optimal", distance = "logit")  
##   
## Summary of balance for all data:  
## Means Treated Means Control SD Control Std. Mean Diff.  
## distance 0.4425 0.3967 0.0990 0.4241  
## age 25.8162 25.0538 7.0577 0.1066  
## educ 10.3459 10.0885 1.6143 0.1281  
## black 0.8432 0.8269 0.3790 0.0448  
## hisp 0.0595 0.1077 0.3106 -0.2034  
## nodegr 0.7081 0.8346 0.3722 -0.2775  
## married 0.1892 0.1538 0.3615 0.0900  
## re74 2095.5740 2107.0268 5687.9067 -0.0023  
## re75 1532.0556 1266.9092 3102.9830 0.0824  
## I(age^2) 717.3946 677.3154 428.7844 0.0929  
## nodegr:re74 1094.1482 1778.4835 5375.0171 -0.2019  
## eCDF Med eCDF Mean eCDF Max  
## distance 0.1173 0.1232 0.2467  
## age 0.0222 0.0254 0.0652  
## educ 0.0127 0.0287 0.1265  
## black 0.0082 0.0082 0.0163  
## hisp 0.0241 0.0241 0.0482  
## nodegr 0.0633 0.0633 0.1265  
## married 0.0177 0.0177 0.0353  
## re74 0.0158 0.0192 0.0471  
## re75 0.0620 0.0508 0.1075  
## I(age^2) 0.0222 0.0254 0.0652  
## nodegr:re74 0.0290 0.0276 0.0534  
##   
##   
## Summary of balance for matched data:  
## Means Treated Means Control SD Control Std. Mean Diff.  
## distance 0.4425 0.4260 0.0940 0.1526  
## age 25.8162 25.8595 7.3980 -0.0060  
## educ 10.3459 10.1081 1.7160 0.1183  
## black 0.8432 0.8541 0.3540 -0.0297  
## hisp 0.0595 0.0595 0.2371 0.0000  
## nodegr 0.7081 0.7676 0.4235 -0.1304  
## married 0.1892 0.1946 0.3970 -0.0138  
## re74 2095.5740 1922.3512 5034.2494 0.0354  
## re75 1532.0556 1456.7589 3412.2044 0.0234  
## I(age^2) 717.3946 723.1459 459.7272 -0.0133  
## nodegr:re74 1094.1482 1460.6146 4498.6622 -0.1081  
## eCDF Med eCDF Mean eCDF Max  
## distance 0.0324 0.0443 0.1297  
## age 0.0108 0.0138 0.0541  
## educ 0.0108 0.0239 0.1189  
## black 0.0054 0.0054 0.0108  
## hisp 0.0000 0.0000 0.0000  
## nodegr 0.0297 0.0297 0.0595  
## married 0.0027 0.0027 0.0054  
## re74 0.0162 0.0235 0.0649  
## re75 0.0459 0.0376 0.0811  
## I(age^2) 0.0108 0.0138 0.0541  
## nodegr:re74 0.0108 0.0133 0.0378  
##   
## Percent Balance Improvement:  
## Std. Mean Diff. eCDF Med eCDF Mean eCDF Max  
## distance 64.0080 72.3404 64.0638 47.4083  
## age 94.3278 51.2881 45.4677 17.0654  
## educ 7.6302 14.7541 16.5847 5.9984  
## black 33.7580 33.7580 33.7580 33.7580  
## hisp 100.0000 100.0000 100.0000 100.0000  
## nodegr 52.9992 52.9992 52.9992 52.9992  
## married 84.7059 84.7059 84.7059 84.7059  
## re74 -1412.4914 -2.6316 -22.1508 -37.7483  
## re75 71.6018 25.8389 26.0345 24.5648  
## I(age^2) 85.6500 51.2881 45.4677 17.0654  
## nodegr:re74 46.4493 62.6571 51.9828 29.1829  
##   
## Sample sizes:  
## Control Treated  
## All 260 185  
## Matched 185 185  
## Unmatched 75 0  
## Discarded 0 0

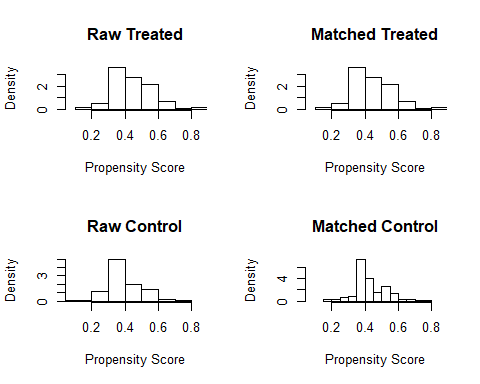
plot(psmatch2)



plot(psmatch2, type="jitter", interactive = FALSE)



plot(psmatch2, type="hist")



# 3. Create matched data for analysis  
psmatch1.data <- match.data(psmatch1)  
psmatch2.data <- match.data(psmatch2)  
  
  
# 4. Analyize: Difference of means  
# Outcome Regression Analysis: AVERAGE TREATMENT EFFECT ON THE TREATED  
psmatch1.mod <- lm(re78 ~ treat + age + educ + black + hisp + nodegr + married + re74 + re75,   
 data = psmatch1.data)  
summary(psmatch1.mod) #beta of treat：ATT, among the treated

##   
## Call:  
## lm(formula = re78 ~ treat + age + educ + black + hisp + nodegr +   
## married + re74 + re75, data = psmatch1.data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -10347 -4421 -1535 3034 54098   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -3.182e+02 3.592e+03 -0.089 0.92946   
## treat 1.881e+03 6.710e+02 2.804 0.00533 \*\*  
## age 7.930e+01 4.818e+01 1.646 0.10064   
## educ 4.374e+02 2.449e+02 1.786 0.07491 .   
## black -2.341e+03 1.177e+03 -1.990 0.04739 \*   
## hisp -8.511e+01 1.763e+03 -0.048 0.96153   
## nodegr -8.340e+01 1.061e+03 -0.079 0.93741   
## married 1.045e+02 9.063e+02 0.115 0.90825   
## re74 8.805e-02 1.026e-01 0.858 0.39127   
## re75 3.116e-02 1.497e-01 0.208 0.83523   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6429 on 360 degrees of freedom  
## Multiple R-squared: 0.07003, Adjusted R-squared: 0.04679   
## F-statistic: 3.012 on 9 and 360 DF, p-value: 0.001763

?zelig  
psmatch1.mod.zelig <- zelig(re78 ~ treat + age + educ + black + hisp + nodegr + married + re74 + re75, data = psmatch1.data, model = "ls")

## How to cite this model in Zelig:  
## R Core Team. 2007.  
## ls: Least Squares Regression for Continuous Dependent Variables  
## in Christine Choirat, Christopher Gandrud, James Honaker, Kosuke Imai, Gary King, and Olivia Lau,  
## "Zelig: Everyone's Statistical Software," http://zeligproject.org/

summary(psmatch1.mod.zelig)

## Model:   
##   
## Call:  
## z5$zelig(formula = re78 ~ treat + age + educ + black + hisp +   
## nodegr + married + re74 + re75, data = psmatch1.data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -10347 -4421 -1535 3034 54098   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -3.182e+02 3.592e+03 -0.089 0.92946  
## treat 1.881e+03 6.710e+02 2.804 0.00533  
## age 7.930e+01 4.818e+01 1.646 0.10064  
## educ 4.374e+02 2.449e+02 1.786 0.07491  
## black -2.341e+03 1.177e+03 -1.990 0.04739  
## hisp -8.511e+01 1.763e+03 -0.048 0.96153  
## nodegr -8.340e+01 1.061e+03 -0.079 0.93741  
## married 1.045e+02 9.063e+02 0.115 0.90825  
## re74 8.805e-02 1.026e-01 0.858 0.39127  
## re75 3.116e-02 1.497e-01 0.208 0.83523  
##   
## Residual standard error: 6429 on 360 degrees of freedom  
## Multiple R-squared: 0.07003, Adjusted R-squared: 0.04679   
## F-statistic: 3.012 on 9 and 360 DF, p-value: 0.001763  
##   
## Statistical Warning: The GIM test suggests this model is misspecified  
## (based on comparisons between classical and robust SE's; see http://j.mp/GIMtest).  
## We suggest you run diagnostics to ascertain the cause, respecify the model  
## and run it again.  
##   
## Next step: Use 'setx' method

psmatch1.mod <- lm(re78 ~ treat + age + educ + black + hisp + nodegr + married + re74 + re75 +   
 I(age^2) + I(educ^2) + re74 + nodegr:re74,   
 data = psmatch1.data)  
summary(psmatch1.mod)

##   
## Call:  
## lm(formula = re78 ~ treat + age + educ + black + hisp + nodegr +   
## married + re74 + re75 + I(age^2) + I(educ^2) + re74 + nodegr:re74,   
## data = psmatch1.data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -11669 -4083 -1614 3008 53591   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6203.4831 6809.6343 0.911 0.3629   
## treat 1677.7142 671.6945 2.498 0.0129 \*  
## age 145.8751 294.9614 0.495 0.6212   
## educ -1625.7305 1290.6379 -1.260 0.2086   
## black -2499.7040 1169.9855 -2.137 0.0333 \*  
## hisp -224.2470 1755.3434 -0.128 0.8984   
## nodegr 1835.6996 1320.4480 1.390 0.1653   
## married -117.3395 914.6177 -0.128 0.8980   
## re74 0.2326 0.1206 1.929 0.0545 .  
## re75 0.1734 0.1588 1.092 0.2757   
## I(age^2) -1.3882 4.8159 -0.288 0.7733   
## I(educ^2) 117.0572 71.8735 1.629 0.1043   
## nodegr:re74 -0.3585 0.1543 -2.323 0.0207 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6385 on 357 degrees of freedom  
## Multiple R-squared: 0.09035, Adjusted R-squared: 0.05977   
## F-statistic: 2.955 on 12 and 357 DF, p-value: 0.0006105

psmatch2.mod <- lm(re78 ~ treat + age + educ + black + hisp + nodegr + married + re74 + re75 +   
 I(age^2) + I(educ^2) + re74 + nodegr:re74,   
 data = psmatch2.data) # optimal matching  
summary(psmatch2.mod)

##   
## Call:  
## lm(formula = re78 ~ treat + age + educ + black + hisp + nodegr +   
## married + re74 + re75 + I(age^2) + I(educ^2) + re74 + nodegr:re74,   
## data = psmatch2.data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -11632 -4248 -1538 2975 53383   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4422.5941 7164.7184 0.617 0.5374   
## treat 1395.9141 706.6411 1.975 0.0490 \*  
## age 243.4706 301.8886 0.806 0.4205   
## educ -1597.6938 1340.7119 -1.192 0.2342   
## black -2382.7918 1226.6819 -1.942 0.0529 .  
## hisp 107.3512 1859.8312 0.058 0.9540   
## nodegr 2293.1741 1400.2362 1.638 0.1024   
## married -199.4948 964.5742 -0.207 0.8363   
## re74 0.2566 0.1259 2.038 0.0423 \*  
## re75 0.1093 0.1635 0.668 0.5044   
## I(age^2) -3.3032 4.9116 -0.673 0.5017   
## I(educ^2) 119.0317 75.0092 1.587 0.1134   
## nodegr:re74 -0.3321 0.1539 -2.158 0.0316 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6700 on 357 degrees of freedom  
## Multiple R-squared: 0.07609, Adjusted R-squared: 0.04504   
## F-statistic: 2.45 on 12 and 357 DF, p-value: 0.004445

### AVERAGE TREATMENT EFFECT   
z.out1 <- zelig(re78 ~ treat + age + educ + black + hisp + nodegr + married + re74 + re75 +   
 I(age^2) + I(educ^2) + re74 + nodegr:re74, data = match.data(psmatch1), model = "ls")

## How to cite this model in Zelig:  
## R Core Team. 2007.  
## ls: Least Squares Regression for Continuous Dependent Variables  
## in Christine Choirat, Christopher Gandrud, James Honaker, Kosuke Imai, Gary King, and Olivia Lau,  
## "Zelig: Everyone's Statistical Software," http://zeligproject.org/

x.out1 <- setx(z.out1, treat=0)  
x1.out1 <- setx(z.out1, treat=1)  
s.out1 <- Zelig:::sim(z.out1, x = x.out1, x1 = x1.out1)  
summary(s.out1)

##   
## sim x :  
## -----  
## ev  
## mean sd 50% 2.5% 97.5%  
## 1 4075.016 548.5975 4072.384 3013.199 5149.151  
## pv  
## mean sd 50% 2.5% 97.5%  
## [1,] 3947.454 6372.591 3913.297 -8720.277 15802.69  
##   
## sim x1 :  
## -----  
## ev  
## mean sd 50% 2.5% 97.5%  
## 1 5737.357 610.7944 5709.385 4513.563 6928.498  
## pv  
## mean sd 50% 2.5% 97.5%  
## [1,] 5961.031 6421.044 6137.802 -6204.766 18975.12  
## fd  
## mean sd 50% 2.5% 97.5%  
## 1 1662.34 681.8868 1664.577 337.1618 2951.94

z.out2 <- zelig(re78 ~ treat + age + educ + black + hisp + nodegr + married + re74 + re75 +   
 I(age^2) + I(educ^2) + re74 + nodegr:re74, data = match.data(psmatch2), model = "ls")

## How to cite this model in Zelig:  
## R Core Team. 2007.  
## ls: Least Squares Regression for Continuous Dependent Variables  
## in Christine Choirat, Christopher Gandrud, James Honaker, Kosuke Imai, Gary King, and Olivia Lau,  
## "Zelig: Everyone's Statistical Software," http://zeligproject.org/

x.out2 <- setx(z.out2, treat=0)  
x1.out2 <- setx(z.out2, treat=1)  
s.out2 <- Zelig:::sim(z.out2, x = x.out2, x1 = x1.out2)  
summary(s.out2)

##   
## sim x :  
## -----  
## ev  
## mean sd 50% 2.5% 97.5%  
## 1 4430.302 594.3034 4406.54 3326.112 5599.165  
## pv  
## mean sd 50% 2.5% 97.5%  
## [1,] 4466.441 6701.395 4319.52 -9138.501 17832.64  
##   
## sim x1 :  
## -----  
## ev  
## mean sd 50% 2.5% 97.5%  
## 1 5833.22 627.0897 5812.316 4658.353 7058.466  
## pv  
## mean sd 50% 2.5% 97.5%  
## [1,] 5628.744 6644.857 5854.259 -6879.091 19342.5  
## fd  
## mean sd 50% 2.5% 97.5%  
## 1 1402.918 711.023 1398.383 55.1986 2806.566

#COMPARE WITH ASSOCIATION  
# Treatment effect without matching: difference of means  
association <- lm(re78 ~ treat, data=lalonde)  
summary(association)

##   
## Call:  
## lm(formula = re78 ~ treat, data = lalonde)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6349 -4555 -1829 2917 53959   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4554.8 408.0 11.162 < 2e-16 \*\*\*  
## treat 1794.3 632.9 2.835 0.00479 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6580 on 443 degrees of freedom  
## Multiple R-squared: 0.01782, Adjusted R-squared: 0.01561   
## F-statistic: 8.039 on 1 and 443 DF, p-value: 0.004788

#COMPARE WITH DIRECT REGRESSION ADJUSTMENT  
# Treatment effect without matching: regression  
unmatched <- lm(re78 ~ treat + age +I(age^2)+ educ + I(educ^2) + black + hisp + nodegr + married +nodegr\*re74+ re75,   
 data = lalonde)  
summary(unmatched)

##   
## Call:  
## lm(formula = re78 ~ treat + age + I(age^2) + educ + I(educ^2) +   
## black + hisp + nodegr + married + nodegr \* re74 + re75, data = lalonde)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -11327 -4306 -1631 2875 53758   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5137.9066 6539.6043 0.786 0.4325   
## treat 1486.1422 641.6566 2.316 0.0210 \*  
## age 169.1049 273.5058 0.618 0.5367   
## I(age^2) -2.1082 4.4909 -0.469 0.6390   
## educ -1232.1829 1235.0520 -0.998 0.3190   
## I(educ^2) 91.7247 69.1569 1.326 0.1854   
## black -2349.1062 1167.9616 -2.011 0.0449 \*  
## hisp -81.6848 1550.3483 -0.053 0.9580   
## nodegr 1520.1419 1263.5903 1.203 0.2296   
## married -355.0499 891.7335 -0.398 0.6907   
## re74 0.2623 0.1196 2.193 0.0288 \*  
## re75 0.1000 0.1367 0.732 0.4649   
## nodegr:re74 -0.2681 0.1330 -2.017 0.0444 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6492 on 432 degrees of freedom  
## Multiple R-squared: 0.0676, Adjusted R-squared: 0.0417   
## F-statistic: 2.61 on 12 and 432 DF, p-value: 0.002279

## Final comparison  
association$coeff["treat"]

## treat   
## 1794.343

unmatched$coeff["treat"]

## treat   
## 1486.142

summary(s.out1)

##   
## sim x :  
## -----  
## ev  
## mean sd 50% 2.5% 97.5%  
## 1 4075.016 548.5975 4072.384 3013.199 5149.151  
## pv  
## mean sd 50% 2.5% 97.5%  
## [1,] 3947.454 6372.591 3913.297 -8720.277 15802.69  
##   
## sim x1 :  
## -----  
## ev  
## mean sd 50% 2.5% 97.5%  
## 1 5737.357 610.7944 5709.385 4513.563 6928.498  
## pv  
## mean sd 50% 2.5% 97.5%  
## [1,] 5961.031 6421.044 6137.802 -6204.766 18975.12  
## fd  
## mean sd 50% 2.5% 97.5%  
## 1 1662.34 681.8868 1664.577 337.1618 2951.94

summary(s.out2)

##   
## sim x :  
## -----  
## ev  
## mean sd 50% 2.5% 97.5%  
## 1 4430.302 594.3034 4406.54 3326.112 5599.165  
## pv  
## mean sd 50% 2.5% 97.5%  
## [1,] 4466.441 6701.395 4319.52 -9138.501 17832.64  
##   
## sim x1 :  
## -----  
## ev  
## mean sd 50% 2.5% 97.5%  
## 1 5833.22 627.0897 5812.316 4658.353 7058.466  
## pv  
## mean sd 50% 2.5% 97.5%  
## [1,] 5628.744 6644.857 5854.259 -6879.091 19342.5  
## fd  
## mean sd 50% 2.5% 97.5%  
## 1 1402.918 711.023 1398.383 55.1986 2806.566

####END OF LAB####  
  
  
  
  
  
  
  
  
  
####OPTIMAL MATCHING WITH GENMATCH  
  
  
#This function finds optimal balance using multivariate matching where a search   
#algorithm determines the weight each covariate is given. Balance is determined by   
#examining cumulative probability distribution functions of a variety of standardized   
#statistics. By default, these statistics include t-tests and Kolmogorov-Smirnov tests.  
#A variety of descriptive statistics based on empirical- QQ (eQQ) plots can also be used   
#or any user provided measure of balance. The statistics are not used to conduct formal  
#hypothesis tests, because no measure of balance is a monotonic function of bias and   
#because balance should be maximized without limit.   
#The object returned by GenMatch can be supplied to the Match function   
#(via the Weight.matrix option) to obtain causal estimates.  
  
?GenMatch  
attach(lalonde)  
#The covariates we want to match on  
X = cbind(age, educ, black, hisp, married, nodegr, re75, re74)  
#The covariates we want to obtain balance on   
BalanceMat <- cbind(age,I(age^2), educ,I(educ^2), black, hisp, married, nodegr, re75, re74, I(nodegr\*re74))  
#  
#Let's call GenMatch() to find the optimal weight to give each  
#covariate in 'X' so as we have achieved balance on the covariates in

BalanceMat’. This is only an example so we want GenMatch to be quick

#so the population size has been set to be only 16 via the 'pop.size'  
#option. This is \*WAY\* too small for actual problems.  
#For details see http://sekhon.berkeley.edu/papers/MatchingJSS.pdf.  
#  
out <- GenMatch(Tr=treat, X=X, BalanceMatrix=BalanceMat, estimand="ATE", M=1,  
 pop.size=200, max.generations=10, wait.generations=1)

##   
##   
## Sat Nov 09 21:01:53 2019  
## Domains:  
## 0.000000e+00 <= X1 <= 1.000000e+03   
## 0.000000e+00 <= X2 <= 1.000000e+03   
## 0.000000e+00 <= X3 <= 1.000000e+03   
## 0.000000e+00 <= X4 <= 1.000000e+03   
## 0.000000e+00 <= X5 <= 1.000000e+03   
## 0.000000e+00 <= X6 <= 1.000000e+03   
## 0.000000e+00 <= X7 <= 1.000000e+03   
## 0.000000e+00 <= X8 <= 1.000000e+03   
##   
## Data Type: Floating Point  
## Operators (code number, name, population)   
## (1) Cloning........................... 22  
## (2) Uniform Mutation.................. 25  
## (3) Boundary Mutation................. 25  
## (4) Non-Uniform Mutation.............. 25  
## (5) Polytope Crossover................ 25  
## (6) Simple Crossover.................. 26  
## (7) Whole Non-Uniform Mutation........ 25  
## (8) Heuristic Crossover............... 26  
## (9) Local-Minimum Crossover........... 0  
##   
## SOFT Maximum Number of Generations: 10  
## Maximum Nonchanging Generations: 1  
## Population size : 200  
## Convergence Tolerance: 1.000000e-03  
##   
## Not Using the BFGS Derivative Based Optimizer on the Best Individual Each Generation.  
## Not Checking Gradients before Stopping.  
## Using Out of Bounds Individuals.  
##   
## Maximization Problem.  
## GENERATION: 0 (initializing the population)  
## Lexical Fit..... 1.710672e-01 4.648276e-01 4.796243e-01 4.796243e-01 6.086968e-01 6.548664e-01 6.548664e-01 7.253821e-01 7.826131e-01 7.923521e-01 8.509656e-01 8.961588e-01 8.961588e-01 9.709321e-01 9.709321e-01 9.969691e-01 9.969691e-01 9.990352e-01 1.000000e+00 1.000000e+00 1.000000e+00 1.000000e+00   
## #unique......... 200, #Total UniqueCount: 200  
## var 1:  
## best............ 2.681455e+02  
## mean............ 5.156537e+02  
## variance........ 8.753088e+04  
## var 2:  
## best............ 4.310157e+02  
## mean............ 5.053488e+02  
## variance........ 8.062645e+04  
## var 3:  
## best............ 4.444041e+02  
## mean............ 5.053601e+02  
## variance........ 8.338302e+04  
## var 4:  
## best............ 3.383783e+02  
## mean............ 5.130571e+02  
## variance........ 8.067212e+04  
## var 5:  
## best............ 7.406809e+02  
## mean............ 5.073551e+02  
## variance........ 8.770186e+04  
## var 6:  
## best............ 1.047467e+02  
## mean............ 5.341342e+02  
## variance........ 7.763256e+04  
## var 7:  
## best............ 2.832501e+02  
## mean............ 5.075298e+02  
## variance........ 8.297934e+04  
## var 8:  
## best............ 9.659624e+02  
## mean............ 5.223700e+02  
## variance........ 8.298078e+04  
##   
## GENERATION: 1  
## Lexical Fit..... 2.848167e-01 4.930091e-01 6.351056e-01 6.548664e-01 6.548664e-01 7.110517e-01 7.131602e-01 8.633067e-01 8.633067e-01 8.969553e-01 9.535947e-01 9.535947e-01 9.673921e-01 9.928121e-01 9.997927e-01 9.997927e-01 1.000000e+00 1.000000e+00 1.000000e+00 1.000000e+00 1.000000e+00 1.000000e+00   
## #unique......... 161, #Total UniqueCount: 361  
## var 1:  
## best............ 1.302870e+02  
## mean............ 3.455743e+02  
## variance........ 6.798047e+04  
## var 2:  
## best............ 2.825108e+01  
## mean............ 3.617729e+02  
## variance........ 3.836339e+04  
## var 3:  
## best............ 7.761736e+02  
## mean............ 5.451329e+02  
## variance........ 4.508298e+04  
## var 4:  
## best............ 8.174664e+02  
## mean............ 5.228697e+02  
## variance........ 5.666422e+04  
## var 5:  
## best............ 9.320669e+02  
## mean............ 6.685690e+02  
## variance........ 5.757884e+04  
## var 6:  
## best............ 1.933878e+02  
## mean............ 3.938389e+02  
## variance........ 1.116560e+05  
## var 7:  
## best............ 3.272325e+02  
## mean............ 3.510125e+02  
## variance........ 4.625129e+04  
## var 8:  
## best............ 9.123560e+02  
## mean............ 7.478796e+02  
## variance........ 6.001829e+04  
##   
## GENERATION: 2  
## Lexical Fit..... 3.139654e-01 5.766567e-01 6.030243e-01 6.257337e-01 6.548664e-01 6.548664e-01 7.055926e-01 7.055926e-01 7.263274e-01 7.606488e-01 8.461369e-01 8.967123e-01 8.967123e-01 9.523162e-01 9.523162e-01 9.970046e-01 9.997787e-01 9.997787e-01 1.000000e+00 1.000000e+00 1.000000e+00 1.000000e+00   
## #unique......... 137, #Total UniqueCount: 498  
## var 1:  
## best............ 1.539011e+02  
## mean............ 1.908591e+02  
## variance........ 9.057005e+03  
## var 2:  
## best............ 7.799765e+01  
## mean............ 1.685279e+02  
## variance........ 3.439067e+04  
## var 3:  
## best............ 6.592820e+02  
## mean............ 5.879455e+02  
## variance........ 2.818173e+04  
## var 4:  
## best............ 6.486703e+02  
## mean............ 5.615154e+02  
## variance........ 4.868692e+04  
## var 5:  
## best............ 8.574934e+02  
## mean............ 7.994744e+02  
## variance........ 1.825043e+04  
## var 6:  
## best............ 1.810109e+02  
## mean............ 1.899047e+02  
## variance........ 2.248963e+04  
## var 7:  
## best............ 3.219226e+02  
## mean............ 3.274160e+02  
## variance........ 7.639676e+03  
## var 8:  
## best............ 9.198410e+02  
## mean............ 8.943824e+02  
## variance........ 1.072135e+04  
##   
## GENERATION: 3  
## Lexical Fit..... 3.139654e-01 5.766567e-01 6.030243e-01 6.257337e-01 6.548664e-01 6.548664e-01 7.055926e-01 7.055926e-01 7.263274e-01 7.606488e-01 8.461369e-01 8.967123e-01 8.967123e-01 9.523162e-01 9.523162e-01 9.970046e-01 9.997787e-01 9.997787e-01 1.000000e+00 1.000000e+00 1.000000e+00 1.000000e+00   
## #unique......... 150, #Total UniqueCount: 648  
## var 1:  
## best............ 1.539011e+02  
## mean............ 1.595523e+02  
## variance........ 3.079793e+03  
## var 2:  
## best............ 7.799765e+01  
## mean............ 9.075371e+01  
## variance........ 1.079933e+04  
## var 3:  
## best............ 6.592820e+02  
## mean............ 6.544363e+02  
## variance........ 6.457891e+03  
## var 4:  
## best............ 6.486703e+02  
## mean............ 6.633008e+02  
## variance........ 8.178972e+03  
## var 5:  
## best............ 8.574934e+02  
## mean............ 8.445677e+02  
## variance........ 8.310706e+03  
## var 6:  
## best............ 1.810109e+02  
## mean............ 1.991866e+02  
## variance........ 5.951557e+03  
## var 7:  
## best............ 3.219226e+02  
## mean............ 3.332574e+02  
## variance........ 5.832663e+03  
## var 8:  
## best............ 9.198410e+02  
## mean............ 9.115828e+02  
## variance........ 1.245990e+03  
##   
## GENERATION: 4  
## Lexical Fit..... 3.139654e-01 5.766567e-01 6.030243e-01 6.257337e-01 6.548664e-01 6.548664e-01 7.055926e-01 7.055926e-01 7.263274e-01 7.606488e-01 8.461369e-01 8.967123e-01 8.967123e-01 9.523162e-01 9.523162e-01 9.970046e-01 9.997787e-01 9.997787e-01 1.000000e+00 1.000000e+00 1.000000e+00 1.000000e+00   
## #unique......... 140, #Total UniqueCount: 788  
## var 1:  
## best............ 1.539011e+02  
## mean............ 1.609962e+02  
## variance........ 3.088773e+03  
## var 2:  
## best............ 7.799765e+01  
## mean............ 9.105943e+01  
## variance........ 6.170662e+03  
## var 3:  
## best............ 6.592820e+02  
## mean............ 6.523201e+02  
## variance........ 2.944434e+03  
## var 4:  
## best............ 6.486703e+02  
## mean............ 6.485935e+02  
## variance........ 4.144641e+03  
## var 5:  
## best............ 8.574934e+02  
## mean............ 8.504764e+02  
## variance........ 3.775469e+03  
## var 6:  
## best............ 1.810109e+02  
## mean............ 1.857736e+02  
## variance........ 1.691815e+03  
## var 7:  
## best............ 3.219226e+02  
## mean............ 3.283919e+02  
## variance........ 2.428365e+03  
## var 8:  
## best............ 9.198410e+02  
## mean............ 9.088475e+02  
## variance........ 3.321831e+03  
##   
## 'wait.generations' limit reached.  
## No significant improvement in 1 generations.  
##   
## Solution Lexical Fitness Value:  
## 3.139654e-01 5.766567e-01 6.030243e-01 6.257337e-01 6.548664e-01 6.548664e-01 7.055926e-01 7.055926e-01 7.263274e-01 7.606488e-01 8.461369e-01 8.967123e-01 8.967123e-01 9.523162e-01 9.523162e-01 9.970046e-01 9.997787e-01 9.997787e-01 1.000000e+00 1.000000e+00 1.000000e+00 1.000000e+00   
##   
## Parameters at the Solution:  
##   
## X[ 1] : 1.539011e+02  
## X[ 2] : 7.799765e+01  
## X[ 3] : 6.592820e+02  
## X[ 4] : 6.486703e+02  
## X[ 5] : 8.574934e+02  
## X[ 6] : 1.810109e+02  
## X[ 7] : 3.219226e+02  
## X[ 8] : 9.198410e+02  
##   
## Solution Found Generation 2  
## Number of Generations Run 4  
##   
## Sat Nov 09 21:02:00 2019  
## Total run time : 0 hours 0 minutes and 7 seconds

#The outcome variable  
Y=re78/1000  
#  
  
  
#  
#Let's determine if balance has actually been obtained on the variables of interest  
#  
mb <- MatchBalance(treat~age +educ+black+ hisp+ married+ nodegr+ re75+ re74+ I(re74\*re75) + I(re74\*nodegr),  
 match.out=out, nboots=500)

## Warning in MatchBalance(treat ~ age + educ + black + hisp + married +  
## nodegr + : Object not of class 'Match'

##   
## \*\*\*\*\* (V1) age \*\*\*\*\*  
## before matching:  
## mean treatment........ 25.816   
## mean control.......... 25.054   
## std mean diff......... 10.655   
##   
## mean raw eQQ diff..... 0.94054   
## med raw eQQ diff..... 1   
## max raw eQQ diff..... 7   
##   
## mean eCDF diff........ 0.025364   
## med eCDF diff........ 0.022193   
## max eCDF diff........ 0.065177   
##   
## var ratio (Tr/Co)..... 1.0278   
## T-test p-value........ 0.26594   
## KS Bootstrap p-value.. 0.5   
## KS Naive p-value...... 0.7481   
## KS Statistic.......... 0.065177   
##   
##   
## \*\*\*\*\* (V2) educ \*\*\*\*\*  
## before matching:  
## mean treatment........ 10.346   
## mean control.......... 10.088   
## std mean diff......... 12.806   
##   
## mean raw eQQ diff..... 0.40541   
## med raw eQQ diff..... 0   
## max raw eQQ diff..... 2   
##   
## mean eCDF diff........ 0.028698   
## med eCDF diff........ 0.012682   
## max eCDF diff........ 0.12651   
##   
## var ratio (Tr/Co)..... 1.5513   
## T-test p-value........ 0.15017   
## KS Bootstrap p-value.. 0.016   
## KS Naive p-value...... 0.062873   
## KS Statistic.......... 0.12651   
##   
##   
## \*\*\*\*\* (V3) black \*\*\*\*\*  
## before matching:  
## mean treatment........ 0.84324   
## mean control.......... 0.82692   
## std mean diff......... 4.4767   
##   
## mean raw eQQ diff..... 0.016216   
## med raw eQQ diff..... 0   
## max raw eQQ diff..... 1   
##   
## mean eCDF diff........ 0.0081601   
## med eCDF diff........ 0.0081601   
## max eCDF diff........ 0.01632   
##   
## var ratio (Tr/Co)..... 0.92503   
## T-test p-value........ 0.64736   
##   
##   
## \*\*\*\*\* (V4) hisp \*\*\*\*\*  
## before matching:  
## mean treatment........ 0.059459   
## mean control.......... 0.10769   
## std mean diff......... -20.341   
##   
## mean raw eQQ diff..... 0.048649   
## med raw eQQ diff..... 0   
## max raw eQQ diff..... 1   
##   
## mean eCDF diff........ 0.024116   
## med eCDF diff........ 0.024116   
## max eCDF diff........ 0.048233   
##   
## var ratio (Tr/Co)..... 0.58288   
## T-test p-value........ 0.064043   
##   
##   
## \*\*\*\*\* (V5) married \*\*\*\*\*  
## before matching:  
## mean treatment........ 0.18919   
## mean control.......... 0.15385   
## std mean diff......... 8.9995   
##   
## mean raw eQQ diff..... 0.037838   
## med raw eQQ diff..... 0   
## max raw eQQ diff..... 1   
##   
## mean eCDF diff........ 0.017672   
## med eCDF diff........ 0.017672   
## max eCDF diff........ 0.035343   
##   
## var ratio (Tr/Co)..... 1.1802   
## T-test p-value........ 0.33425   
##   
##   
## \*\*\*\*\* (V6) nodegr \*\*\*\*\*  
## before matching:  
## mean treatment........ 0.70811   
## mean control.......... 0.83462   
## std mean diff......... -27.751   
##   
## mean raw eQQ diff..... 0.12432   
## med raw eQQ diff..... 0   
## max raw eQQ diff..... 1   
##   
## mean eCDF diff........ 0.063254   
## med eCDF diff........ 0.063254   
## max eCDF diff........ 0.12651   
##   
## var ratio (Tr/Co)..... 1.4998   
## T-test p-value........ 0.0020368   
##   
##   
## \*\*\*\*\* (V7) re75 \*\*\*\*\*  
## before matching:  
## mean treatment........ 1532.1   
## mean control.......... 1266.9   
## std mean diff......... 8.2363   
##   
## mean raw eQQ diff..... 367.61   
## med raw eQQ diff..... 0   
## max raw eQQ diff..... 2110.2   
##   
## mean eCDF diff........ 0.050834   
## med eCDF diff........ 0.061954   
## max eCDF diff........ 0.10748   
##   
## var ratio (Tr/Co)..... 1.0763   
## T-test p-value........ 0.38527   
## KS Bootstrap p-value.. 0.062   
## KS Naive p-value...... 0.16449   
## KS Statistic.......... 0.10748   
##   
##   
## \*\*\*\*\* (V8) re74 \*\*\*\*\*  
## before matching:  
## mean treatment........ 2095.6   
## mean control.......... 2107   
## std mean diff......... -0.23437   
##   
## mean raw eQQ diff..... 487.98   
## med raw eQQ diff..... 0   
## max raw eQQ diff..... 8413   
##   
## mean eCDF diff........ 0.019223   
## med eCDF diff........ 0.0158   
## max eCDF diff........ 0.047089   
##   
## var ratio (Tr/Co)..... 0.7381   
## T-test p-value........ 0.98186   
## KS Bootstrap p-value.. 0.58   
## KS Naive p-value...... 0.97023   
## KS Statistic.......... 0.047089   
##   
##   
## \*\*\*\*\* (V9) I(re74 \* re75) \*\*\*\*\*  
## before matching:  
## mean treatment........ 13118591   
## mean control.......... 14530303   
## std mean diff......... -2.7799   
##   
## mean raw eQQ diff..... 3278733   
## med raw eQQ diff..... 0   
## max raw eQQ diff..... 188160151   
##   
## mean eCDF diff........ 0.022723   
## med eCDF diff........ 0.014449   
## max eCDF diff........ 0.061019   
##   
## var ratio (Tr/Co)..... 0.69439   
## T-test p-value........ 0.79058   
## KS Bootstrap p-value.. 0.31   
## KS Naive p-value...... 0.81575   
## KS Statistic.......... 0.061019   
##   
##   
## \*\*\*\*\* (V10) I(re74 \* nodegr) \*\*\*\*\*  
## before matching:  
## mean treatment........ 1094.1   
## mean control.......... 1778.5   
## std mean diff......... -20.192   
##   
## mean raw eQQ diff..... 650.66   
## med raw eQQ diff..... 0   
## max raw eQQ diff..... 15568   
##   
## mean eCDF diff........ 0.027631   
## med eCDF diff........ 0.02895   
## max eCDF diff........ 0.05343   
##   
## var ratio (Tr/Co)..... 0.39757   
## T-test p-value........ 0.10083   
## KS Bootstrap p-value.. 0.344   
## KS Naive p-value...... 0.91719   
## KS Statistic.......... 0.05343   
##   
##   
## Before Matching Minimum p.value: 0.0020368   
## Variable Name(s): nodegr Number(s): 6

# Now that GenMatch() has found the optimal weights, let's estimate  
# our causal effect of interest using those weights  
#  
mout <- Match(Y=Y, Tr=treat, X=X, estimand="ATE", Weight.matrix=out)  
  
psmatch3.mod <-mout  
  
summary(psmatch3.mod)

##   
## Estimate... 1.9961   
## AI SE...... 0.76864   
## T-stat..... 2.5969   
## p.val...... 0.0094067   
##   
## Original number of observations.............. 445   
## Original number of treated obs............... 185   
## Matched number of observations............... 445   
## Matched number of observations (unweighted). 607

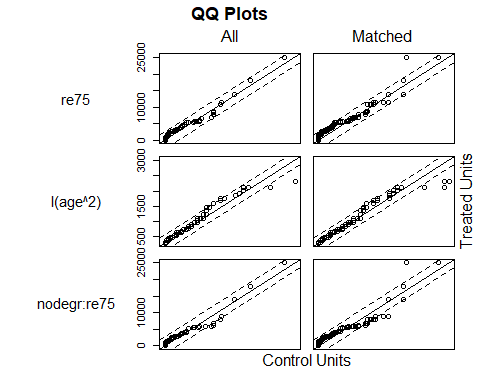
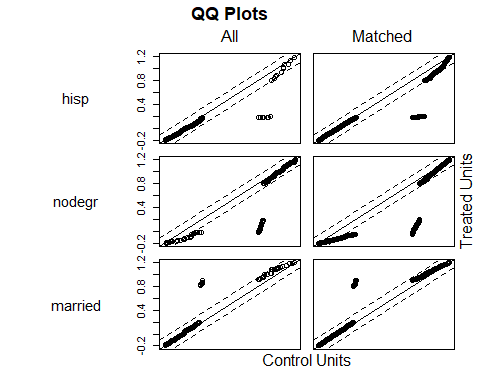
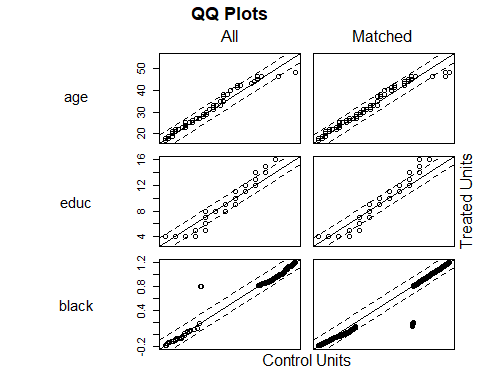
#########OTHER MATCHING TECHNIQUES WITH MATCHIT  
  
# NN match on PS score: k:1, greedy, without replacement or calipers  
psmatch4 <- matchit(treat ~ age + educ + black + hisp + nodegr + married + re75  
 + I(age^2) + nodegr:re75, distance = "logit", method = "nearest", discard= "control",   
 ratio = 3, data = lalonde)

## Warning in matchit2nearest(c(`1` = 1L, `2` = 1L, `3` = 1L, `4` = 1L, `5`  
## = 1L, : Not enough control units for 3 matches for each treated unit when  
## matching without replacement. Not all treated units will receive 3 matches

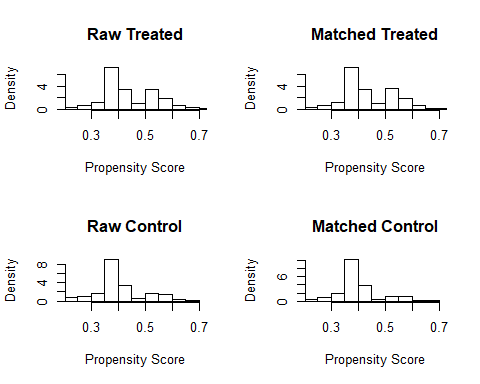
# Check balance  
summary(psmatch4, standardize=TRUE)

##   
## Call:  
## matchit(formula = treat ~ age + educ + black + hisp + nodegr +   
## married + re75 + I(age^2) + nodegr:re75, data = lalonde,   
## method = "nearest", distance = "logit", discard = "control",   
## ratio = 3)  
##   
## Summary of balance for all data:  
## Means Treated Means Control SD Control Std. Mean Diff.  
## distance 0.4358 0.4014 0.0873 0.3648  
## age 25.8162 25.0538 7.0577 0.1066  
## educ 10.3459 10.0885 1.6143 0.1281  
## black 0.8432 0.8269 0.3790 0.0448  
## hisp 0.0595 0.1077 0.3106 -0.2034  
## nodegr 0.7081 0.8346 0.3722 -0.2775  
## married 0.1892 0.1538 0.3615 0.0900  
## re75 1532.0556 1266.9092 3102.9830 0.0824  
## I(age^2) 717.3946 677.3154 428.7844 0.0929  
## nodegr:re75 1134.9556 1081.9148 3016.7335 0.0179  
## eCDF Med eCDF Mean eCDF Max  
## distance 0.1133 0.1043 0.2164  
## age 0.0222 0.0254 0.0652  
## educ 0.0127 0.0287 0.1265  
## black 0.0082 0.0082 0.0163  
## hisp 0.0241 0.0241 0.0482  
## nodegr 0.0633 0.0633 0.1265  
## married 0.0177 0.0177 0.0353  
## re75 0.0620 0.0508 0.1075  
## I(age^2) 0.0222 0.0254 0.0652  
## nodegr:re75 0.0275 0.0283 0.0672  
##   
##   
## Summary of balance for matched data:  
## Means Treated Means Control SD Control Std. Mean Diff.  
## distance 0.4358 0.3957 0.0735 0.4260  
## age 25.8162 25.4432 7.2852 0.0521  
## educ 10.3459 10.0081 1.5160 0.1680  
## black 0.8432 0.8703 0.3367 -0.0741  
## hisp 0.0595 0.0757 0.2650 -0.0684  
## nodegr 0.7081 0.8838 0.3211 -0.3854  
## married 0.1892 0.1568 0.3643 0.0826  
## re75 1532.0556 1109.8678 2853.0850 0.1311  
## I(age^2) 717.3946 700.2216 447.9698 0.0398  
## nodegr:re75 1134.9556 979.8717 2781.2453 0.0524  
## eCDF Med eCDF Mean eCDF Max  
## distance 0.1015 0.0925 0.2002  
## age 0.0176 0.0188 0.0423  
## educ 0.0111 0.0266 0.1206  
## black 0.0067 0.0067 0.0133  
## hisp 0.0081 0.0081 0.0162  
## nodegr 0.0603 0.0603 0.1206  
## married 0.0149 0.0149 0.0298  
## re75 0.0526 0.0491 0.1109  
## I(age^2) 0.0176 0.0188 0.0423  
## nodegr:re75 0.0282 0.0279 0.0722  
##   
## Percent Balance Improvement:  
## Std. Mean Diff. eCDF Med eCDF Mean eCDF Max  
## distance -16.7760 10.4412 11.3633 7.5192  
## age 51.0772 20.5762 25.8294 35.1722  
## educ -31.2071 12.5465 7.4563 4.6876  
## black -65.6051 18.3191 18.3191 18.3191  
## hisp 66.3793 66.3347 66.3347 66.3347  
## nodegr -38.8661 4.6876 4.6876 4.6876  
## married 8.2353 15.6082 15.6082 15.6082  
## re75 -59.2282 15.1154 3.4350 -3.1649  
## I(age^2) 57.1524 20.5762 25.8294 35.1722  
## nodegr:re75 -192.3864 -2.2957 1.6559 -7.5303  
##   
## Sample sizes:  
## Control Treated  
## All 260 185  
## Matched 251 185  
## Unmatched 0 0  
## Discarded 9 0

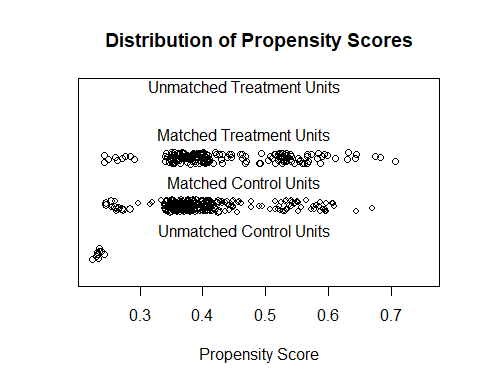
plot(psmatch4)



plot(psmatch4, type="hist")



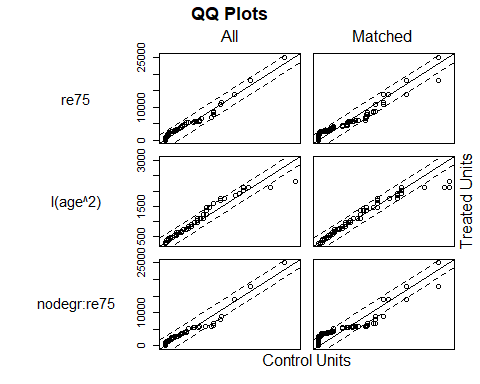
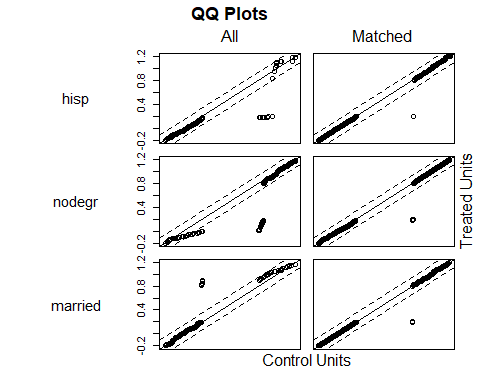
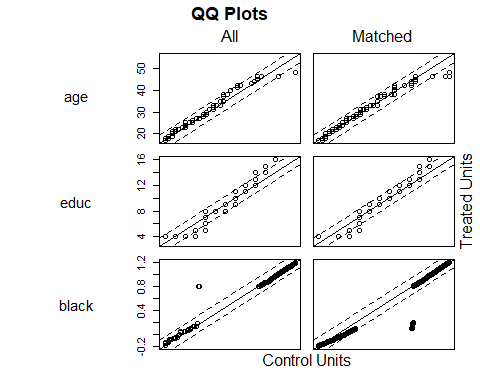
par(mfrow = c(1, 1))  
plot(psmatch4, type="jitter", interactive = FALSE)



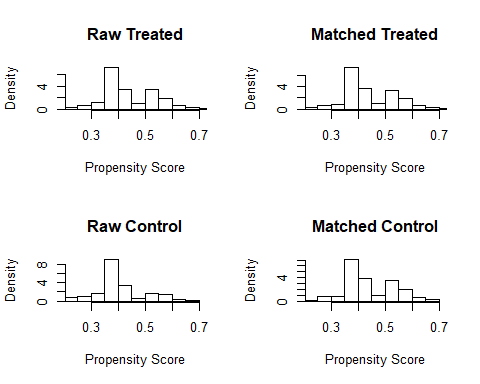
# Create matched data for analysis  
psmatch4.data <- match.data(psmatch4)  
  
# NN match on PS score: k:1, with replacement (greedy doesn't matter), no calipers  
psmatch5 <- matchit(treat ~ age + educ + black + hisp + nodegr + married + re75+ I(age^2) + nodegr:re75,   
 distance = "logit", method = "nearest", discard= "control",   
 ratio = 3, replace= TRUE, data = lalonde)  
# Check balance  
summary(psmatch5, standardize=TRUE)

##   
## Call:  
## matchit(formula = treat ~ age + educ + black + hisp + nodegr +   
## married + re75 + I(age^2) + nodegr:re75, data = lalonde,   
## method = "nearest", distance = "logit", discard = "control",   
## ratio = 3, replace = TRUE)  
##   
## Summary of balance for all data:  
## Means Treated Means Control SD Control Std. Mean Diff.  
## distance 0.4358 0.4014 0.0873 0.3648  
## age 25.8162 25.0538 7.0577 0.1066  
## educ 10.3459 10.0885 1.6143 0.1281  
## black 0.8432 0.8269 0.3790 0.0448  
## hisp 0.0595 0.1077 0.3106 -0.2034  
## nodegr 0.7081 0.8346 0.3722 -0.2775  
## married 0.1892 0.1538 0.3615 0.0900  
## re75 1532.0556 1266.9092 3102.9830 0.0824  
## I(age^2) 717.3946 677.3154 428.7844 0.0929  
## nodegr:re75 1134.9556 1081.9148 3016.7335 0.0179  
## eCDF Med eCDF Mean eCDF Max  
## distance 0.1133 0.1043 0.2164  
## age 0.0222 0.0254 0.0652  
## educ 0.0127 0.0287 0.1265  
## black 0.0082 0.0082 0.0163  
## hisp 0.0241 0.0241 0.0482  
## nodegr 0.0633 0.0633 0.1265  
## married 0.0177 0.0177 0.0353  
## re75 0.0620 0.0508 0.1075  
## I(age^2) 0.0222 0.0254 0.0652  
## nodegr:re75 0.0275 0.0283 0.0672  
##   
##   
## Summary of balance for matched data:  
## Means Treated Means Control SD Control Std. Mean Diff.  
## distance 0.4358 0.4349 0.0922 0.0094  
## age 25.8162 25.8955 7.6765 -0.0111  
## educ 10.3459 10.1802 1.8381 0.0824  
## black 0.8432 0.8775 0.3287 -0.0939  
## hisp 0.0595 0.0559 0.2302 0.0152  
## nodegr 0.7081 0.7153 0.4524 -0.0158  
## married 0.1892 0.1982 0.3996 -0.0229  
## re75 1532.0556 1315.0271 3480.1219 0.0674  
## I(age^2) 717.3946 729.2144 493.0412 -0.0274  
## nodegr:re75 1134.9556 937.8923 3318.4357 0.0666  
## eCDF Med eCDF Mean eCDF Max  
## distance 0.0804 0.0703 0.1530  
## age 0.0156 0.0169 0.0385  
## educ 0.0168 0.0251 0.0850  
## black 0.0094 0.0094 0.0188  
## hisp 0.0048 0.0048 0.0095  
## nodegr 0.0425 0.0425 0.0850  
## married 0.0059 0.0059 0.0118  
## re75 0.0668 0.0664 0.1494  
## I(age^2) 0.0156 0.0169 0.0385  
## nodegr:re75 0.0330 0.0468 0.1166  
##   
## Percent Balance Improvement:  
## Std. Mean Diff. eCDF Med eCDF Mean eCDF Max  
## distance 97.4245 29.0514 32.5882 29.2921  
## age 89.6009 29.6916 33.4750 40.9653  
## educ 35.6210 -32.8030 12.6148 32.8139  
## black -109.7665 -15.3525 -15.3525 -15.3525  
## hisp 92.5287 80.2913 80.2913 80.2913  
## nodegr 94.3029 32.8139 32.8139 32.8139  
## married 74.5098 66.4735 66.4735 66.4735  
## re75 18.1477 -7.7495 -30.5976 -38.9791  
## I(age^2) 70.5089 29.6916 33.4750 40.9653  
## nodegr:re75 -271.5319 -19.9591 -65.2250 -73.6400  
##   
## Sample sizes:  
## Control Treated  
## All 260 185  
## Matched 203 185  
## Unmatched 48 0  
## Discarded 9 0

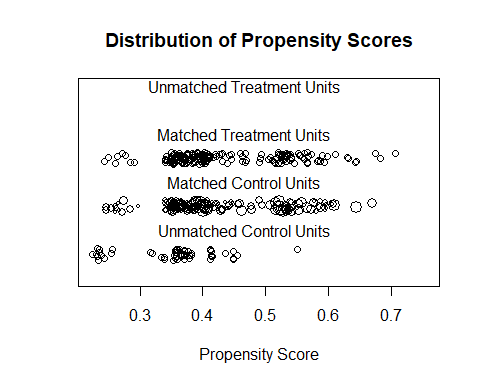
plot(psmatch5)



plot(psmatch5, type="hist")



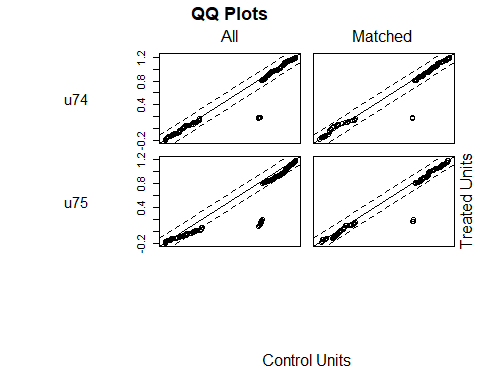
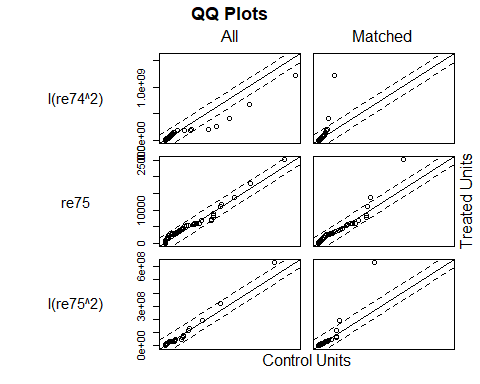
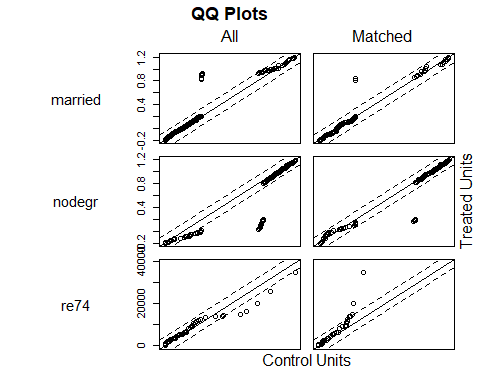
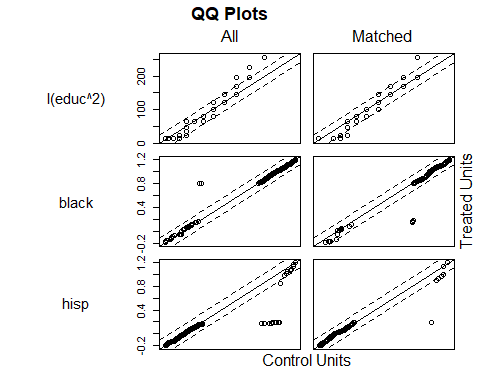
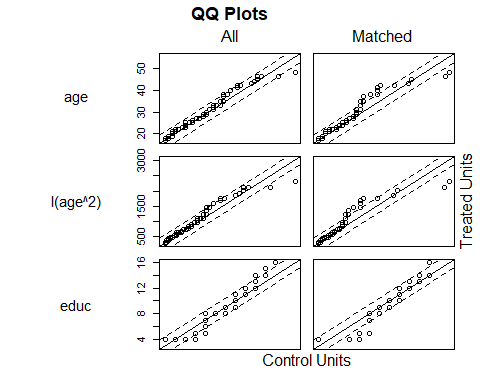
par(mfrow = c(1, 1))  
plot(psmatch5, type="jitter", interactive = FALSE)



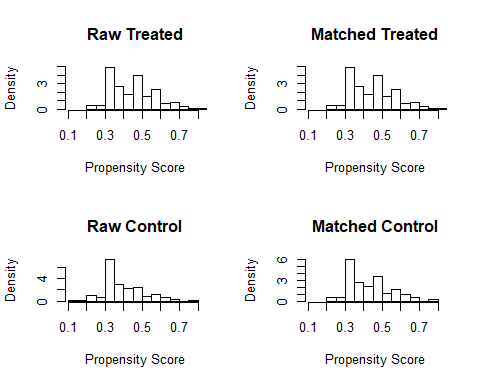
#now ratio 1  
psmatch6 <- matchit(treat ~ age + I(age^2) + educ + I(educ^2) + black +  
 hisp + married + nodegr + re74 + I(re74^2) + re75 + I(re75^2) +  
 u74 + u75, ratio = 1, replace= TRUE, data = lalonde)  
  
# Check balance  
summary(psmatch6, standardize=TRUE)

##   
## Call:  
## matchit(formula = treat ~ age + I(age^2) + educ + I(educ^2) +   
## black + hisp + married + nodegr + re74 + I(re74^2) + re75 +   
## I(re75^2) + u74 + u75, data = lalonde, ratio = 1, replace = TRUE)  
##   
## Summary of balance for all data:  
## Means Treated Means Control SD Control Std. Mean Diff.  
## distance 4.468000e-01 3.936000e-01 1.065000e-01 0.4533  
## age 2.581620e+01 2.505380e+01 7.057700e+00 0.1066  
## I(age^2) 7.173946e+02 6.773154e+02 4.287844e+02 0.0929  
## educ 1.034590e+01 1.008850e+01 1.614300e+00 0.1281  
## I(educ^2) 1.110595e+02 1.043731e+02 3.048290e+01 0.1701  
## black 8.432000e-01 8.269000e-01 3.790000e-01 0.0448  
## hisp 5.950000e-02 1.077000e-01 3.106000e-01 -0.2034  
## married 1.892000e-01 1.538000e-01 3.615000e-01 0.0900  
## nodegr 7.081000e-01 8.346000e-01 3.722000e-01 -0.2775  
## re74 2.095574e+03 2.107027e+03 5.687907e+03 -0.0023  
## I(re74^2) 2.814143e+07 3.666741e+07 1.607542e+08 -0.0747  
## re75 1.532056e+03 1.266909e+03 3.102983e+03 0.0824  
## I(re75^2) 1.265475e+07 1.119653e+07 4.635935e+07 0.0260  
## u74 7.081000e-01 7.500000e-01 4.338000e-01 -0.0919  
## u75 6.000000e-01 6.846000e-01 4.656000e-01 -0.1723  
## eCDF Med eCDF Mean eCDF Max  
## distance 0.1574 0.1340 0.2244  
## age 0.0222 0.0254 0.0652  
## I(age^2) 0.0222 0.0254 0.0652  
## educ 0.0127 0.0287 0.1265  
## I(educ^2) 0.0127 0.0287 0.1265  
## black 0.0082 0.0082 0.0163  
## hisp 0.0241 0.0241 0.0482  
## married 0.0177 0.0177 0.0353  
## nodegr 0.0633 0.0633 0.1265  
## re74 0.0158 0.0192 0.0471  
## I(re74^2) 0.0158 0.0192 0.0471  
## re75 0.0620 0.0508 0.1075  
## I(re75^2) 0.0620 0.0508 0.1075  
## u74 0.0209 0.0209 0.0419  
## u75 0.0423 0.0423 0.0846  
##   
##   
## Summary of balance for matched data:  
## Means Treated Means Control SD Control Std. Mean Diff.  
## distance 4.468000e-01 4.461000e-01 1.155000e-01 0.0055  
## age 2.581620e+01 2.477300e+01 6.786800e+00 0.1458  
## I(age^2) 7.173946e+02 6.593459e+02 4.415750e+02 0.1346  
## educ 1.034590e+01 1.049190e+01 1.836400e+00 -0.0726  
## I(educ^2) 1.110595e+02 1.134216e+02 3.476370e+01 -0.0601  
## black 8.432000e-01 8.595000e-01 3.491000e-01 -0.0445  
## hisp 5.950000e-02 4.860000e-02 2.161000e-01 0.0456  
## married 1.892000e-01 1.676000e-01 3.752000e-01 0.0551  
## nodegr 7.081000e-01 6.919000e-01 4.638000e-01 0.0356  
## re74 2.095574e+03 1.497326e+03 3.070620e+03 0.1224  
## I(re74^2) 2.814143e+07 1.158575e+07 3.152906e+07 0.1451  
## re75 1.532056e+03 1.263170e+03 2.612701e+03 0.0835  
## I(re75^2) 1.265475e+07 8.360307e+06 2.728585e+07 0.0766  
## u74 7.081000e-01 7.081000e-01 4.567000e-01 0.0000  
## u75 6.000000e-01 6.216000e-01 4.872000e-01 -0.0440  
## eCDF Med eCDF Mean eCDF Max  
## distance 0.0613 0.0531 0.1027  
## age 0.0468 0.0453 0.0955  
## I(age^2) 0.0468 0.0453 0.0955  
## educ 0.0225 0.0241 0.0595  
## I(educ^2) 0.0225 0.0241 0.0595  
## black 0.0153 0.0153 0.0306  
## hisp 0.0018 0.0018 0.0036  
## married 0.0090 0.0090 0.0180  
## nodegr 0.0288 0.0288 0.0577  
## re74 0.0270 0.0289 0.0667  
## I(re74^2) 0.0270 0.0289 0.0667  
## re75 0.0252 0.0315 0.0775  
## I(re75^2) 0.0252 0.0315 0.0775  
## u74 0.0063 0.0063 0.0126  
## u75 0.0153 0.0153 0.0306  
##   
## Percent Balance Improvement:  
## Std. Mean Diff. eCDF Med eCDF Mean eCDF Max  
## distance 98.7932 61.0873 60.3567 54.2381  
## age -36.8421 -111.0851 -78.4836 -46.5178  
## I(age^2) -44.8348 -111.0851 -78.4836 -46.5178  
## educ 43.3185 -77.5956 16.1363 52.9992  
## I(educ^2) 64.6720 -77.5956 16.1363 52.9992  
## black 0.6369 -87.6858 -87.6858 -87.6858  
## hisp 77.5862 92.5287 92.5287 92.5287  
## married 38.8235 49.0196 49.0196 49.0196  
## nodegr 87.1816 54.4234 54.4234 54.4234  
## re74 -5123.5908 -71.0526 -50.5459 -41.5747  
## I(re74^2) -94.1793 -71.0526 -50.5459 -41.5747  
## re75 -1.4102 59.2841 38.0495 27.9175  
## I(re75^2) -194.4986 59.2841 38.0495 27.9175  
## u74 100.0000 69.8925 69.8925 69.8925  
## u75 74.4472 63.8002 63.8002 63.8002  
##   
## Sample sizes:  
## Control Treated  
## All 260 185  
## Matched 111 185  
## Unmatched 149 0  
## Discarded 0 0

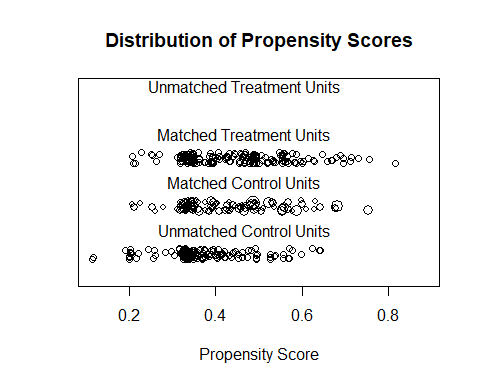
plot(psmatch6)



plot(psmatch6, type="hist")



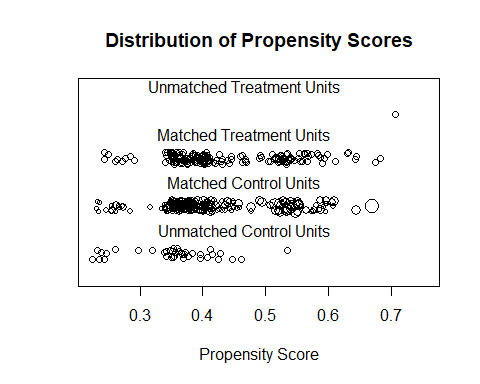
par(mfrow = c(1, 1))  
plot(psmatch6, type="jitter", interactive = FALSE)



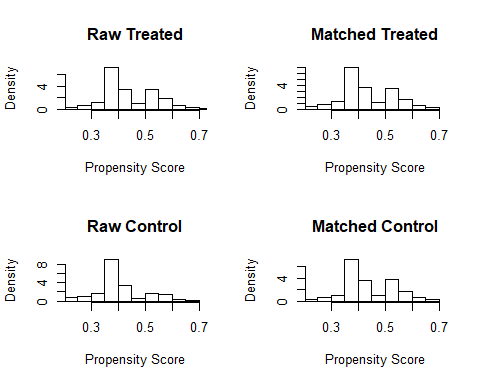
# NN match on PS score: k:1, with replacement, add calipers (instead of discarding outside common support)  
psmatch7 <- matchit(treat ~ age + educ + black + hisp + nodegr + married + re75 +   
 I(age^2) + nodegr:re75,   
 distance = "logit", method = "nearest", caliper = 0.3,   
 ratio = 3, replace= TRUE, data = lalonde)  
  
# Check balance  
summary(psmatch7, standardize=TRUE)

##   
## Call:  
## matchit(formula = treat ~ age + educ + black + hisp + nodegr +   
## married + re75 + I(age^2) + nodegr:re75, data = lalonde,   
## method = "nearest", distance = "logit", caliper = 0.3, ratio = 3,   
## replace = TRUE)  
##   
## Summary of balance for all data:  
## Means Treated Means Control SD Control Std. Mean Diff.  
## distance 0.4358 0.4014 0.0873 0.3648  
## age 25.8162 25.0538 7.0577 0.1066  
## educ 10.3459 10.0885 1.6143 0.1281  
## black 0.8432 0.8269 0.3790 0.0448  
## hisp 0.0595 0.1077 0.3106 -0.2034  
## nodegr 0.7081 0.8346 0.3722 -0.2775  
## married 0.1892 0.1538 0.3615 0.0900  
## re75 1532.0556 1266.9092 3102.9830 0.0824  
## I(age^2) 717.3946 677.3154 428.7844 0.0929  
## nodegr:re75 1134.9556 1081.9148 3016.7335 0.0179  
## eCDF Med eCDF Mean eCDF Max  
## distance 0.1133 0.1043 0.2164  
## age 0.0222 0.0254 0.0652  
## educ 0.0127 0.0287 0.1265  
## black 0.0082 0.0082 0.0163  
## hisp 0.0241 0.0241 0.0482  
## nodegr 0.0633 0.0633 0.1265  
## married 0.0177 0.0177 0.0353  
## re75 0.0620 0.0508 0.1075  
## I(age^2) 0.0222 0.0254 0.0652  
## nodegr:re75 0.0275 0.0283 0.0672  
##   
##   
## Summary of balance for matched data:  
## Means Treated Means Control SD Control Std. Mean Diff.  
## distance 0.4344 0.4338 0.0935 0.0062  
## age 25.7772 25.8225 7.2230 -0.0063  
## educ 10.3370 10.3025 1.7519 0.0171  
## black 0.8424 0.8804 0.3252 -0.1044  
## hisp 0.0598 0.0670 0.2506 -0.0306  
## nodegr 0.7120 0.7065 0.4564 0.0119  
## married 0.1848 0.1920 0.3948 -0.0185  
## re75 1480.9179 1597.4991 3682.0656 -0.0362  
## I(age^2) 715.3750 718.7355 447.5820 -0.0078  
## nodegr:re75 1141.1238 1187.8224 3530.8998 -0.0158  
## eCDF Med eCDF Mean eCDF Max  
## distance 0.0849 0.0785 0.1843  
## age 0.0150 0.0177 0.0431  
## educ 0.0168 0.0268 0.0980  
## black 0.0004 0.0004 0.0008  
## hisp 0.0131 0.0131 0.0262  
## nodegr 0.0490 0.0490 0.0980  
## married 0.0155 0.0155 0.0309  
## re75 0.0472 0.0449 0.0980  
## I(age^2) 0.0150 0.0177 0.0431  
## nodegr:re75 0.0317 0.0297 0.0708  
##   
## Percent Balance Improvement:  
## Std. Mean Diff. eCDF Med eCDF Mean eCDF Max  
## distance 98.3114 25.0997 24.8151 14.8360  
## age 94.0593 32.3523 30.0604 33.8577  
## educ 86.6321 -32.6359 6.6195 22.5355  
## black -133.1072 95.3288 95.3288 95.3288  
## hisp 84.9763 45.7004 45.7004 45.7004  
## nodegr 95.7040 22.5355 22.5355 22.5355  
## married 79.4970 12.4680 12.4680 12.4680  
## re75 56.0314 23.7489 11.6108 8.7799  
## I(age^2) 91.6153 32.3523 30.0604 33.8577  
## nodegr:re75 11.9570 -15.2065 -4.7399 -5.4690  
##   
## Sample sizes:  
## Control Treated  
## All 260 185  
## Matched 221 184  
## Unmatched 39 1  
## Discarded 0 0

plot(psmatch7, type="jitter", interactive = FALSE)



plot(psmatch7, type="hist")



# Create matched data for analysis  
psmatch4.data <- match.data(psmatch4)  
psmatch5.data <- match.data(psmatch5)  
psmatch6.data <- match.data(psmatch6)  
psmatch7.data <- match.data(psmatch7)  
  
#OUTCOME ANALYSIS  
psmatch4.mod <- lm(re78 ~ treat + age + educ + black + hisp + nodegr + married + re74 + re75 +   
 I(age^2) + I(educ^2) + educ:re74,   
 data = psmatch4.data)  
summary(psmatch4.mod)

##   
## Call:  
## lm(formula = re78 ~ treat + age + educ + black + hisp + nodegr +   
## married + re74 + re75 + I(age^2) + I(educ^2) + educ:re74,   
## data = psmatch4.data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -10960 -4335 -1826 3054 53929   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5.540e+03 6.611e+03 0.838 0.4025   
## treat 1.548e+03 6.483e+02 2.388 0.0174 \*  
## age 1.771e+02 2.799e+02 0.633 0.5273   
## educ -1.180e+03 1.249e+03 -0.945 0.3453   
## black -2.221e+03 1.178e+03 -1.885 0.0601 .  
## hisp 8.434e+01 1.666e+03 0.051 0.9597   
## nodegr 9.301e+02 1.238e+03 0.751 0.4529   
## married -2.939e+02 8.995e+02 -0.327 0.7441   
## re74 -3.724e-01 3.033e-01 -1.228 0.2202   
## re75 5.696e-02 1.372e-01 0.415 0.6782   
## I(age^2) -2.226e+00 4.585e+00 -0.486 0.6276   
## I(educ^2) 8.492e+01 7.004e+01 1.212 0.2260   
## educ:re74 4.462e-02 2.918e-02 1.529 0.1271   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6558 on 423 degrees of freedom  
## Multiple R-squared: 0.06404, Adjusted R-squared: 0.03749   
## F-statistic: 2.412 on 12 and 423 DF, p-value: 0.00496

psmatch5.mod <- lm(re78 ~ treat + age + educ + black + hisp + nodegr + married + re74 + re75 +   
 I(age^2) + I(educ^2) + educ:re74,   
 data = psmatch5.data, weights = weights)  
summary(psmatch5.mod)

##   
## Call:  
## lm(formula = re78 ~ treat + age + educ + black + hisp + nodegr +   
## married + re74 + re75 + I(age^2) + I(educ^2) + educ:re74,   
## data = psmatch5.data, weights = weights)  
##   
## Weighted Residuals:  
## Min 1Q Median 3Q Max   
## -11362 -4090 -1503 2956 53601   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.825e+03 6.207e+03 0.777 0.43740   
## treat 1.906e+03 6.525e+02 2.921 0.00371 \*\*  
## age 2.824e+02 2.780e+02 1.016 0.31033   
## educ -1.826e+03 1.227e+03 -1.488 0.13750   
## black -1.856e+03 1.189e+03 -1.561 0.11928   
## hisp 6.139e+02 1.772e+03 0.346 0.72926   
## nodegr 1.946e+03 1.223e+03 1.591 0.11241   
## married -3.277e+02 9.024e+02 -0.363 0.71669   
## re74 -5.269e-01 3.258e-01 -1.617 0.10666   
## re75 3.158e-02 1.423e-01 0.222 0.82448   
## I(age^2) -3.246e+00 4.470e+00 -0.726 0.46811   
## I(educ^2) 1.204e+02 6.869e+01 1.752 0.08053 .   
## educ:re74 6.306e-02 2.976e-02 2.119 0.03474 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6314 on 375 degrees of freedom  
## Multiple R-squared: 0.09344, Adjusted R-squared: 0.06443   
## F-statistic: 3.221 on 12 and 375 DF, p-value: 0.0002018

# Added weights option, weights = weights  
# matchit automatically created frequency weights (named `weights') in the matched data frame  
psmatch6.mod <- lm(re78 ~ treat + age + educ + black + hisp + nodegr + married + re74 + re75 +   
 I(age^2) + I(educ^2) + educ:re74,   
 data = psmatch5.data, weights = weights)  
summary(psmatch6.mod)

##   
## Call:  
## lm(formula = re78 ~ treat + age + educ + black + hisp + nodegr +   
## married + re74 + re75 + I(age^2) + I(educ^2) + educ:re74,   
## data = psmatch5.data, weights = weights)  
##   
## Weighted Residuals:  
## Min 1Q Median 3Q Max   
## -11362 -4090 -1503 2956 53601   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.825e+03 6.207e+03 0.777 0.43740   
## treat 1.906e+03 6.525e+02 2.921 0.00371 \*\*  
## age 2.824e+02 2.780e+02 1.016 0.31033   
## educ -1.826e+03 1.227e+03 -1.488 0.13750   
## black -1.856e+03 1.189e+03 -1.561 0.11928   
## hisp 6.139e+02 1.772e+03 0.346 0.72926   
## nodegr 1.946e+03 1.223e+03 1.591 0.11241   
## married -3.277e+02 9.024e+02 -0.363 0.71669   
## re74 -5.269e-01 3.258e-01 -1.617 0.10666   
## re75 3.158e-02 1.423e-01 0.222 0.82448   
## I(age^2) -3.246e+00 4.470e+00 -0.726 0.46811   
## I(educ^2) 1.204e+02 6.869e+01 1.752 0.08053 .   
## educ:re74 6.306e-02 2.976e-02 2.119 0.03474 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6314 on 375 degrees of freedom  
## Multiple R-squared: 0.09344, Adjusted R-squared: 0.06443   
## F-statistic: 3.221 on 12 and 375 DF, p-value: 0.0002018

psmatch6.mod <- lm(re78 ~ treat + age + educ + black + hisp + nodegr + married + re74 + re75 +   
 I(age^2) + I(educ^2) + educ:re74,   
 data = psmatch6.data, weights = weights)  
summary(psmatch6.mod)

##   
## Call:  
## lm(formula = re78 ~ treat + age + educ + black + hisp + nodegr +   
## married + re74 + re75 + I(age^2) + I(educ^2) + educ:re74,   
## data = psmatch6.data, weights = weights)  
##   
## Weighted Residuals:  
## Min 1Q Median 3Q Max   
## -11423 -4282 -934 3183 53079   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.502e+03 7.238e+03 0.622 0.53443   
## treat 1.963e+03 7.957e+02 2.467 0.01422 \*   
## age 5.736e+02 3.345e+02 1.715 0.08751 .   
## educ -2.350e+03 1.405e+03 -1.673 0.09545 .   
## black -2.570e+03 1.337e+03 -1.922 0.05562 .   
## hisp 6.330e+02 2.099e+03 0.302 0.76320   
## nodegr 1.586e+03 1.416e+03 1.120 0.26384   
## married -5.239e+02 1.095e+03 -0.478 0.63273   
## re74 -1.558e+00 5.631e-01 -2.767 0.00604 \*\*  
## re75 2.552e-01 1.867e-01 1.367 0.17276   
## I(age^2) -8.133e+00 5.454e+00 -1.491 0.13704   
## I(educ^2) 1.453e+02 7.780e+01 1.868 0.06279 .   
## educ:re74 1.337e-01 4.683e-02 2.855 0.00462 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6576 on 283 degrees of freedom  
## Multiple R-squared: 0.1185, Adjusted R-squared: 0.0811   
## F-statistic: 3.17 on 12 and 283 DF, p-value: 0.0002889

psmatch7.mod <- lm(re78 ~ treat + age + educ + black + hisp + nodegr + married + re74 + re75 +   
 I(age^2) + I(educ^2) + educ:re74,   
 data = psmatch7.data, weights = weights)  
summary(psmatch7.mod)

##   
## Call:  
## lm(formula = re78 ~ treat + age + educ + black + hisp + nodegr +   
## married + re74 + re75 + I(age^2) + I(educ^2) + educ:re74,   
## data = psmatch7.data, weights = weights)  
##   
## Weighted Residuals:  
## Min 1Q Median 3Q Max   
## -12527 -4393 -1502 2894 53717   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6.870e+03 6.876e+03 0.999 0.3184   
## treat 1.467e+03 6.937e+02 2.114 0.0351 \*  
## age 1.655e+02 3.119e+02 0.530 0.5961   
## educ -1.658e+03 1.328e+03 -1.248 0.2127   
## black -2.195e+03 1.325e+03 -1.656 0.0985 .  
## hisp 3.581e+02 1.869e+03 0.192 0.8481   
## nodegr 1.764e+03 1.286e+03 1.371 0.1711   
## married -3.968e+02 9.587e+02 -0.414 0.6791   
## re74 -3.116e-01 3.260e-01 -0.956 0.3397   
## re75 3.162e-04 1.459e-01 0.002 0.9983   
## I(age^2) -2.213e+00 5.106e+00 -0.433 0.6650   
## I(educ^2) 1.159e+02 7.362e+01 1.574 0.1163   
## educ:re74 3.834e-02 3.057e-02 1.254 0.2105   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6833 on 392 degrees of freedom  
## Multiple R-squared: 0.05354, Adjusted R-squared: 0.02457   
## F-statistic: 1.848 on 12 and 392 DF, p-value: 0.0393