

TrainingStudy_DM.R

danny

2020-02-26

```
#install.packages("MatchIt")
#install.packages("cem")
#install.packages("cobalt")
#install.packages("Matching")
#install.packages("rbinds")
```

```
library(MatchIt)
```

```
## Warning: package 'MatchIt' was built under R version 3.6.2
```

```
library(cem)
```

```
## Warning: package 'cem' was built under R version 3.6.2
```

```
## Loading required package: tcltk
```

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

```
##
```

```
## How to use CEM? Type vignette("cem")
```

```
library(cobalt)
```

```
## Warning: package 'cobalt' was built under R version 3.6.2
```

```
## cobalt (Version 4.0.0, Build Date: 2020-01-08 14:20:10 UTC)
```

```
## Please read the documentation at ?bal.tab to understand the default outputs.
```

```
## Submit bug reports and feature requests to https://github.com/ngreifer/cobalt/issues
```

```
## Install the development version (not guaranteed to be stable) with:
```

```
## devtools::install_github("ngreifer/cobalt")
```

```
## Thank you for using cobalt!
```

```
##
```

```
## Attaching package: 'cobalt'
```

```
## The following object is masked from 'package:MatchIt':
```

```
##
```

```
## lalonde
```

```
library(Matching)
```

```
## Warning: package 'Matching' was built under R version 3.6.2
```

```
## Loading required package: MASS
```

```
## ##  
## ## Matching (Version 4.9-7, Build Date: 2020-02-05)  
## ## 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(rbounds)
```

```
## Warning: package 'rbounds' was built under R version 3.6.2
```

```
library(data.table)  
library(stargazer)
```

```
##  
## Please cite as:
```

```
## Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.
```

```
## R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.6.2
```

```
**** MSBA 6440 ****  
**** Gordon Burtch and Gautam Ray****  
**** Updated Feb 2020 ****  
**** Code for Lecture 4 ****
```

```
# Analyzing Training Program Data
```

```
# Load Data
```

```
MyData<-read.csv("TrainingProgram.csv")
```

```
#Estimate Average Treatment Effect
```

```
mean(MyData$re78[MyData$treat == 1]) - mean(MyData$re78[MyData$treat == 0])
```

```
## [1] 886.3035
```

```
ols.model <- lm(re78 ~ treat + age + education + black + hispanic + married + nodegree + re75 , data = MyData)
```

```
summary(lm(re78 ~ treat + age + education + black + hispanic + married + nodegree + re75 , data = MyData))
```

```
##
## Call:
## lm(formula = re78 ~ treat + age + education + black + hispanic +
##     married + nodegree + re75, data = MyData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9949  -4401  -1558   3056   54903
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3879.6048   2604.4327    1.490  0.136767
## treat         806.5111    467.8870    1.724  0.085190 .
## age          17.3925     36.1876    0.481  0.630934
## education    175.3231    179.5810    0.976  0.329252
## black       -1445.5412    801.4514   -1.804  0.071708 .
## hispanic      98.4215   1046.1070    0.094  0.925069
## married       71.8648    652.2659    0.110  0.912300
## nodegree     -470.3970    742.9283   -0.633  0.526828
## re75           0.1705     0.0467    3.651  0.000281 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6153 on 713 degrees of freedom
## Multiple R-squared:  0.0426, Adjusted R-squared:  0.03186
## F-statistic: 3.966 on 8 and 713 DF, p-value: 0.0001309
```

#Exact Matching

```
matchit(formula = treat ~ age + education + black + married + nodegree + re75 + hispanic, data = MyData)
```

```
##
## Call:
## matchit(formula = treat ~ age + education + black + married +
##     nodegree + re75 + hispanic, data = MyData, method = "exact")
##
## Exact Subclasses: 38
##
## Sample sizes:
##           Control Treated
## All           425     297
## Matched        80      57
## Unmatched     345     240
```

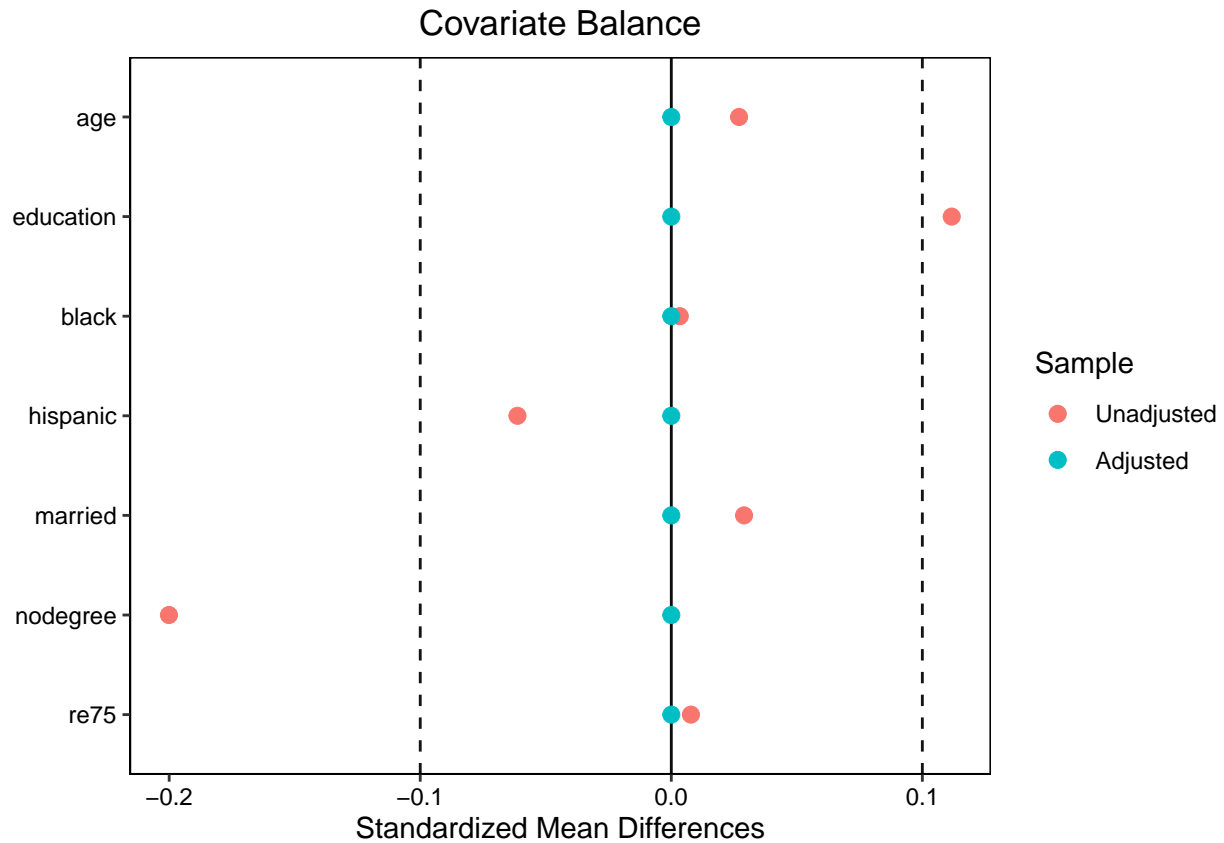
```
exact.match <- matchit(formula= treat ~ age + education + black + married + nodegree + re75 + hispanic,
exact.data <- match.data(exact.match)
```

Lets Check Covariate Balance

```
covs <- subset(MyData, select = -c(treat, re78))
m.out <- matchit(f.build("treat", covs), data = MyData, method = "exact")
```

```
love.plot(m.out, binary = "std", threshold = .1)
```

```
## Note: s.d.denom not specified; assuming pooled.
```



```
#Estimate Average Treatment Effect
```

```
exact.model <- lm(re78 ~ treat+ age + education + black + married + nodegree + re75 + hispanic, data = d)
```

```
stargazer(ols.model, exact.model, title="Model with Different Types of Matches", type="text", column.labels = c("OLS", "Exact Matches"))
```

```
##
## Model with Different Types of Matches
## =====
##                               Dependent variable:
##                               -----
##                               re78
##                               OLS Model      Exact Matches
##                               (1)            (2)
## -----
## treat                        806.511*      1,111.872
##                               (467.887)      (1,001.031)
##
## age                          17.392        39.186
##                               (36.188)        (106.032)
```

```
##
## education          175.323          183.483
##                   (179.581)        (547.027)
##
## black              -1,445.541*
##                   (801.451)
##
## hispanic           98.422
##                   (1,046.107)
##
## married            71.865          -704.116
##                   (652.266)        (2,476.021)
##
## nodegree           -470.397        -3,501.047*
##                   (742.928)        (1,910.287)
##
## re75               0.170***
##                   (0.047)
##
## Constant           3,879.605        4,154.232
##                   (2,604.433)      (6,646.500)
##
## -----
## Observations          722          137
## R2                   0.043          0.075
## Adjusted R2          0.032          0.040
## Residual Std. Error  6,152.525 (df = 713)  5,679.853 (df = 131)
## F Statistic          3.966*** (df = 8; 713)  2.134* (df = 5; 131)
## =====
## Note:                  *p<0.1; **p<0.05; ***p<0.01
```

#Propensity Score Matching

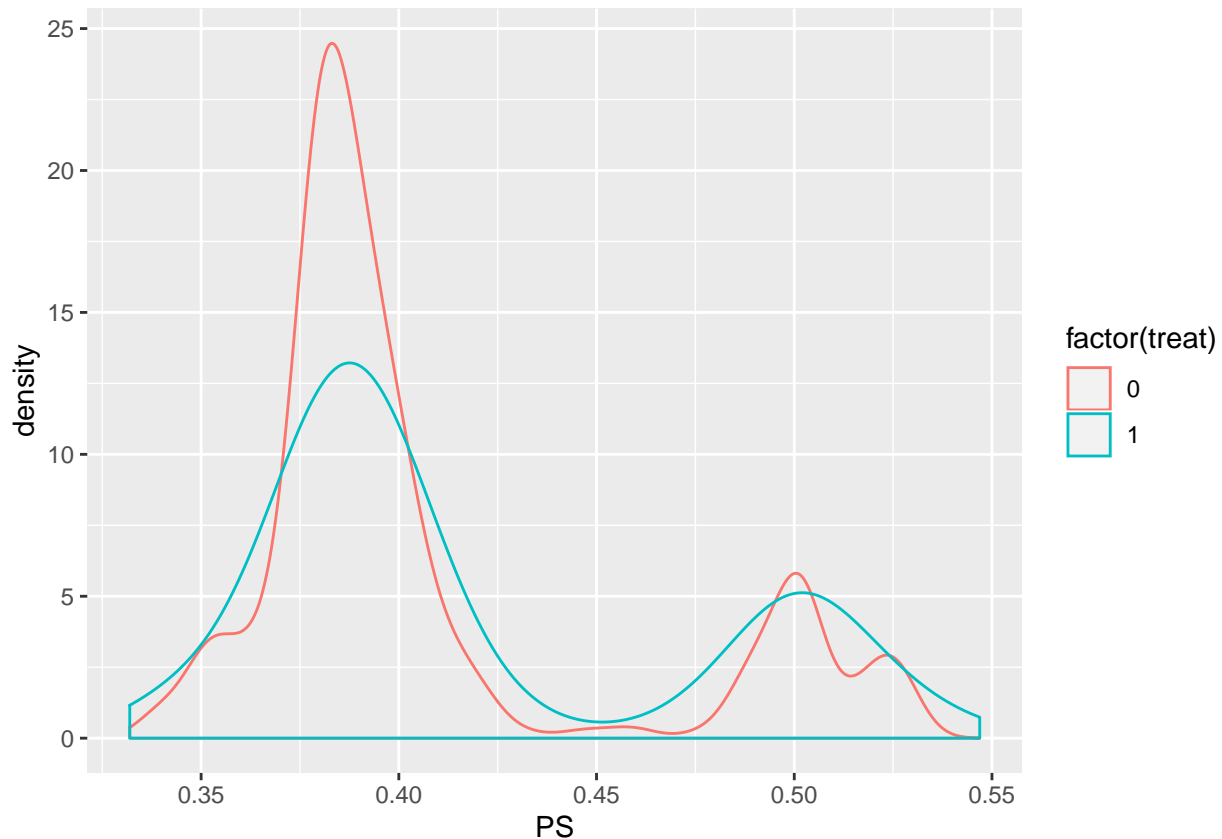
```
PS<- glm(treat ~ age + education + black + married + nodegree + re75 + hispanic, family = "binomial", data = MyData)
summary(PS)
```

```
##
## Call:
## glm(formula = treat ~ age + education + black + married + nodegree +
##      re75 + hispanic, family = "binomial", data = MyData)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2318  -0.9981  -0.9696   1.3521   1.4851
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  4.343e-01  8.599e-01   0.505   0.6135
## age         -1.624e-03  1.204e-02  -0.135   0.8927
## education   -2.378e-02  5.966e-02  -0.399   0.6901
## black       -9.743e-02  2.635e-01  -0.370   0.7116
## married      9.292e-02  2.158e-01   0.431   0.6668
## nodegree    -5.292e-01  2.443e-01  -2.167   0.0303 *
```

```
## re75          -3.121e-06  1.550e-05  -0.201   0.8404
## hispanic      -2.525e-01  3.481e-01  -0.725   0.4682
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 978.09  on 721  degrees of freedom
## Residual deviance: 970.25  on 714  degrees of freedom
## AIC: 986.25
##
## Number of Fisher Scoring iterations: 4
```

No Matching

```
MyData$PS<-glm(treat ~ age + education + black + married + nodegree + re75 + hispanic, data=MyData, fam
ggplot(MyData, aes(x = PS, color = factor(treat))) +geom_density()
```



#Propensity Score Matching - Nearest Match

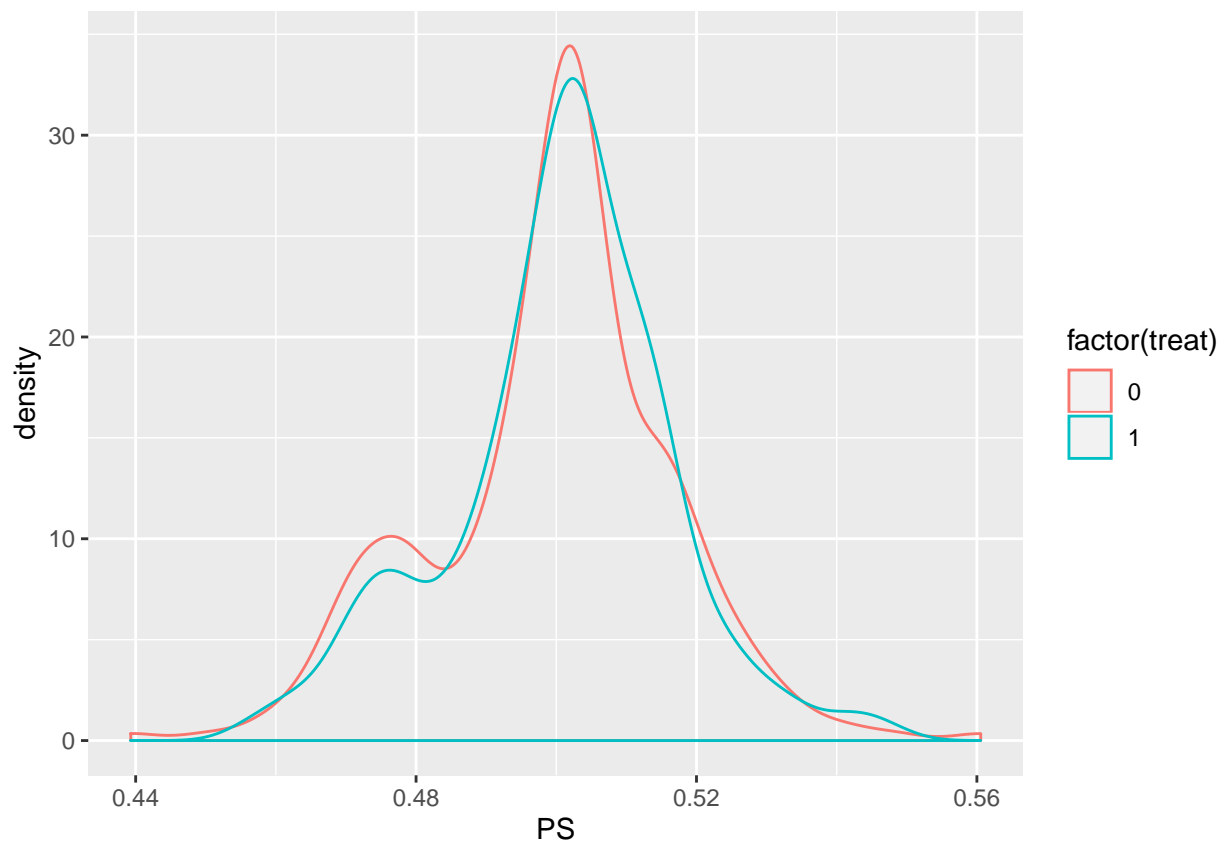
```
nearest.match <- matchit(formula = treat ~ age + education + black + married + nodegree + re75 + hispan
matchit(formula = treat ~ age + education + black + married + nodegree + re75 + hispanic, data = MyData
##
```

```
## Call:
## matchit(formula = treat ~ age + education + black + married +
##         nodegree + re75 + hispanic, data = MyData, method = "nearest",
##         distance = "logit")
##
## Sample sizes:
##           Control Treated
## All           425      297
## Matched        297      297
## Unmatched      128        0
## Discarded        0        0
```

```
nearest.data <- match.data(nearest.match)

nearest.data$PS<-glm(treat ~ age + education + black + married + nodegree + re75 + hispanic, data=nearest.data)$fitted.values

ggplot(nearest.data, aes(x = PS, color = factor(treat))) +geom_density()
```

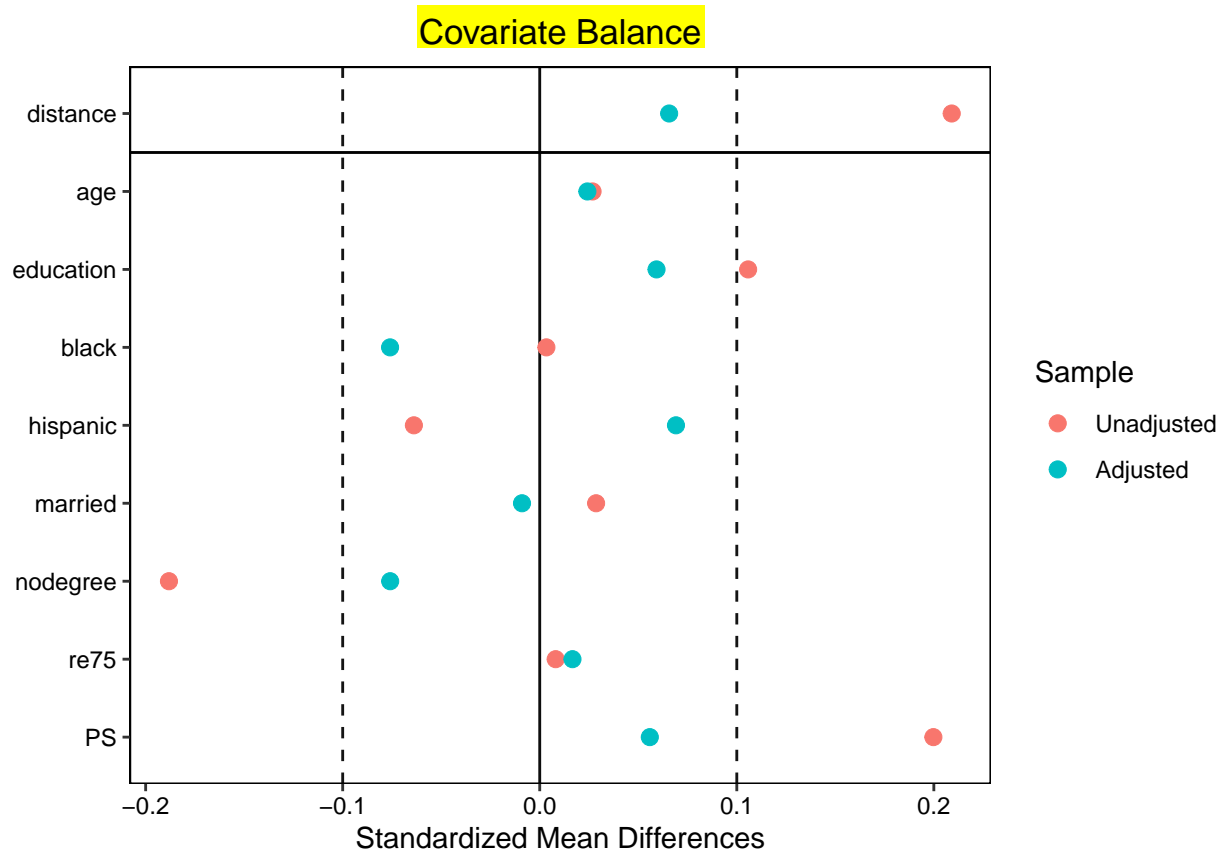


```
# Lets Check Covariate Balance

covs <- subset(MyData, select = -c(treat, re78))

m.out <- matchit(f.build("treat", covs), data = MyData, method = "nearest", distance = "logit")

love.plot(m.out, binary = "std", threshold = .1)
```



#Estimate Average Treatment Effect

```
nearest.model <- lm(re78 ~ treat + age + education + black + married + nodegree + re75 + hispanic, data = data)
```

```
stargazer(exact.model, nearest.model, title = "Model with Different Types of Matches", type = "text", column.labels = c("Exact", "PSM - Nearest"))
```

```
##
## Model with Different Types of Matches
## =====
##                               Dependent variable:
##                               -----
##                               re78
##                               Exact Matches      PSM - Nearest
##                               (1)                (2)
## -----
```

treat	1,111.872 (1,001.031)	768.950 (520.887)
age	39.186 (106.032)	22.080 (40.633)
education	183.483 (547.027)	231.817 (200.599)
black		-1,524.208* (856.030)

```
##
```



```
##
## married                -704.116          466.634
##                        (2,476.021)        (737.256)
##
## nodegree               -3,501.047*        -204.301
##                        (1,910.287)        (814.603)
##
## re75                   0.156***
##                        (0.053)
##
## hispanic               124.220
##                        (1,156.996)
##
## Constant               4,154.232          3,060.760
##                        (6,646.500)        (2,918.129)
##
## -----
## Observations           137                594
## R2                     0.075              0.041
## Adjusted R2            0.040              0.028
## Residual Std. Error 5,679.853 (df = 131) 6,344.118 (df = 585)
## F Statistic           2.134* (df = 5; 131) 3.107*** (df = 8; 585)
## =====
## Note:                  *p<0.1; **p<0.05; ***p<0.01
```

#Propensity Score Matching - Caliper

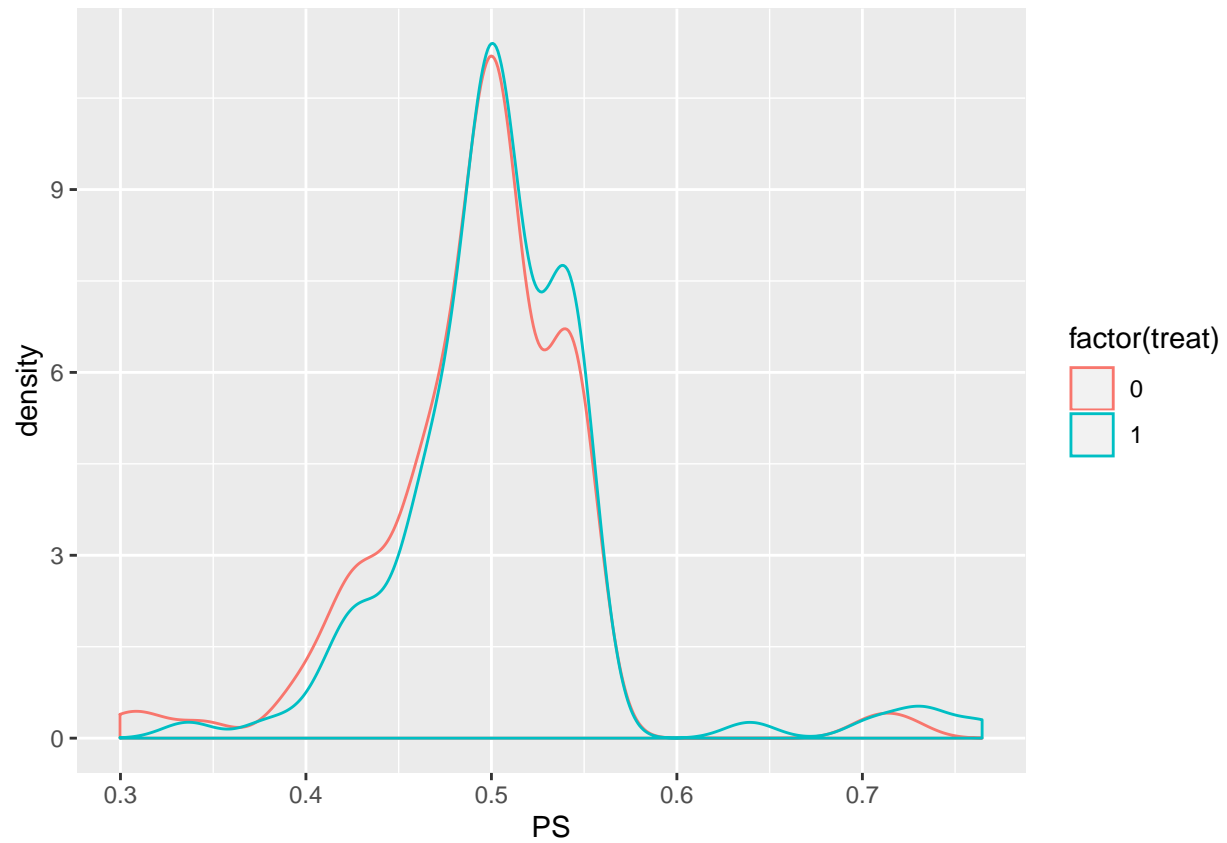
```
caliper.match <- matchit(formula = treat ~ age + education + black + married + nodegree + re75 + hispanic,
data = MyData)
```

```
##
## Call:
## matchit(formula = treat ~ age + education + black + married +
## nodegree + re75 + hispanic, data = MyData, method = "nearest",
## distance = "logit", caliper = 0.001)
##
## Sample sizes:
##           Control Treated
## All           425      297
## Matched       113      113
## Unmatched     312      184
## Discarded      0        0
```

```
caliper.data <- match.data(caliper.match)

caliper.data$PS<-glm(treat ~ age + education + black + married + nodegree + re75 + hispanic, data=caliper.data)

ggplot(caliper.data, aes(x = PS, color = factor(treat))) +geom_density()
```



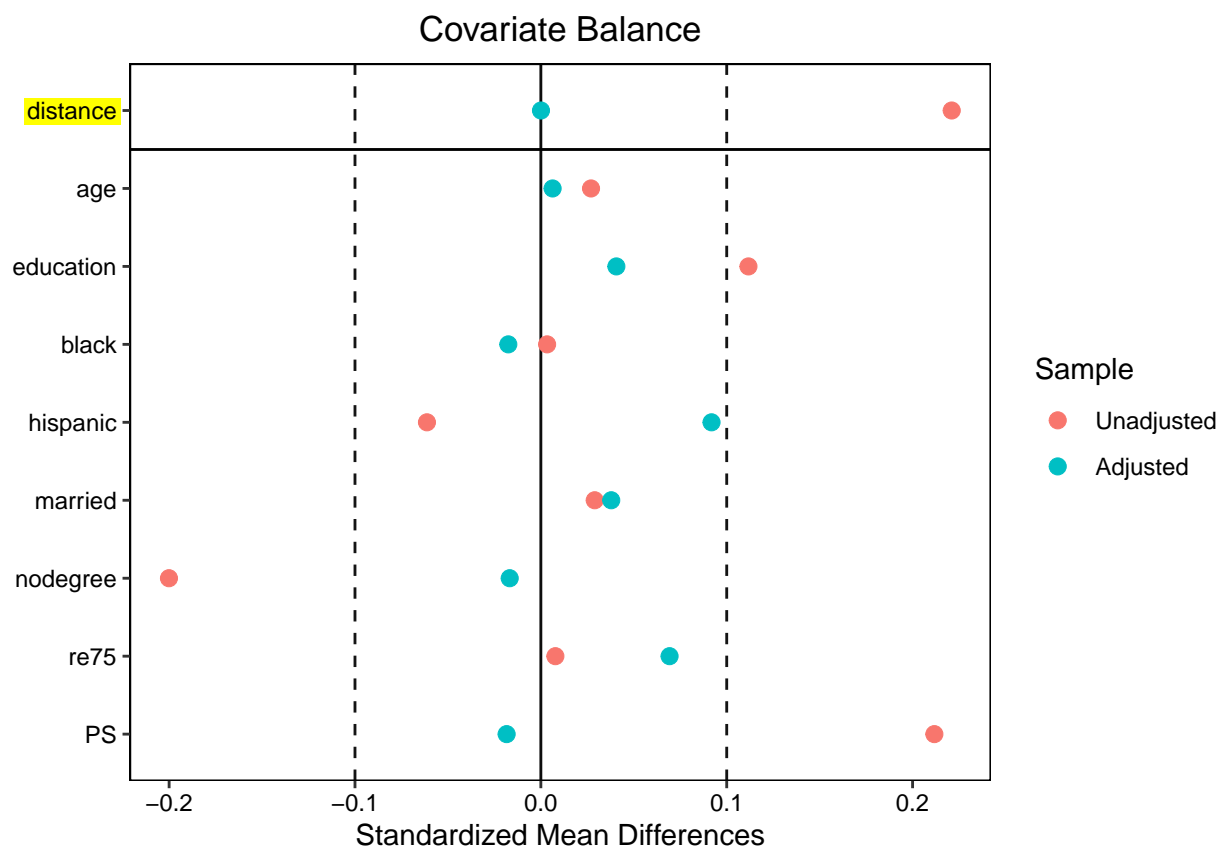
```
# Lets Check Covariate Balance
```

```
covs <- subset(MyData, select = -c(treat, re78))
```

```
m.out <- matchit(f.build("treat", covs), data = MyData, distance = "logit", method = "nearest", caliper = 0.1)
```

```
love.plot(m.out, binary = "std", threshold = .1)
```

```
## Note: s.d.denom not specified; assuming pooled.
```



#Estimate Average Treatment Effect

```
caliper.model <- lm(re78 ~ treat + age + education + black + married + nodegree + re75 + hispanic, data = data)
stargazer(nearest.model, caliper.model, title="Model with Different Types of Matches", type="text", column.labels=c("Nearest", "Caliper"))
```

```
##
## Model with Different Types of Matches
## =====
##                               Dependent variable:
##                               -----
##                               re78
##                               Nearest Matching      Caliper Matching
##                               (1)                   (2)
## -----
## treat                        768.950              651.665
##                               (520.887)            (801.438)
##
## age                          22.080              -38.951
##                               (40.633)            (71.619)
##
## education                    231.817              362.379
##                               (200.599)            (411.837)
##
## black                       -1,524.208*           -1,230.201
##                               (856.030)            (2,470.760)
```

```
##
## married                466.634          2,826.426*
##                        (737.256)        (1,492.779)
##
## nodegree               -204.301         -1,381.853
##                        (814.603)        (1,461.980)
##
## re75                   0.156***          0.171
##                        (0.053)        (0.114)
##
## hispanic               124.220          120.484
##                        (1,156.996)      (3,158.114)
##
## Constant               3,060.760         3,448.040
##                        (2,918.129)      (5,816.400)
##
## -----
## Observations            594             226
## R2                      0.041             0.057
## Adjusted R2             0.028             0.023
## Residual Std. Error  6,344.118 (df = 585)  5,983.065 (df = 217)
## F Statistic          3.107*** (df = 8; 585) 1.649 (df = 8; 217)
## =====
## Note:                    *p<0.1; **p<0.05; ***p<0.01
```

Runs the sensitivity test based on the matched sample using Wilcoxon's rank sign

```
Match <- Match(Y=MyData$re78, Tr=MyData$treat, X=PS$fitted, replace=FALSE)
psens(Match, Gamma = 2, GammaInc = 0.1)
```

```
##
## Rosenbaum Sensitivity Test for Wilcoxon Signed Rank P-Value
##
## Unconfounded estimate .... 0.0919
##
## Gamma Lower bound Upper bound
## 1.0      0.0919      0.0919
## 1.1      0.0218      0.2602
## 1.2      0.0040      0.4930
## 1.3      0.0006      0.7114
## 1.4      0.0001      0.8623
## 1.5      0.0000      0.9439
## 1.6      0.0000      0.9801
## 1.7      0.0000      0.9937
## 1.8      0.0000      0.9982
## 1.9      0.0000      0.9995
## 2.0      0.0000      0.9999
##
## Note: Gamma is Odds of Differential Assignment To
## Treatment Due to Unobserved Factors
##
```

```

# Coarsend Exact Matching

# Automatic Coarsend Exact Matching

autocem.match <- matchit(formula = treat ~ age + education + black + married + nodegree + re75 + hispani

##
## Using 'treat'='1' as baseline group

matchit(formula = treat ~ age + education + black + married + nodegree + re75 + hispanic, data = MyData

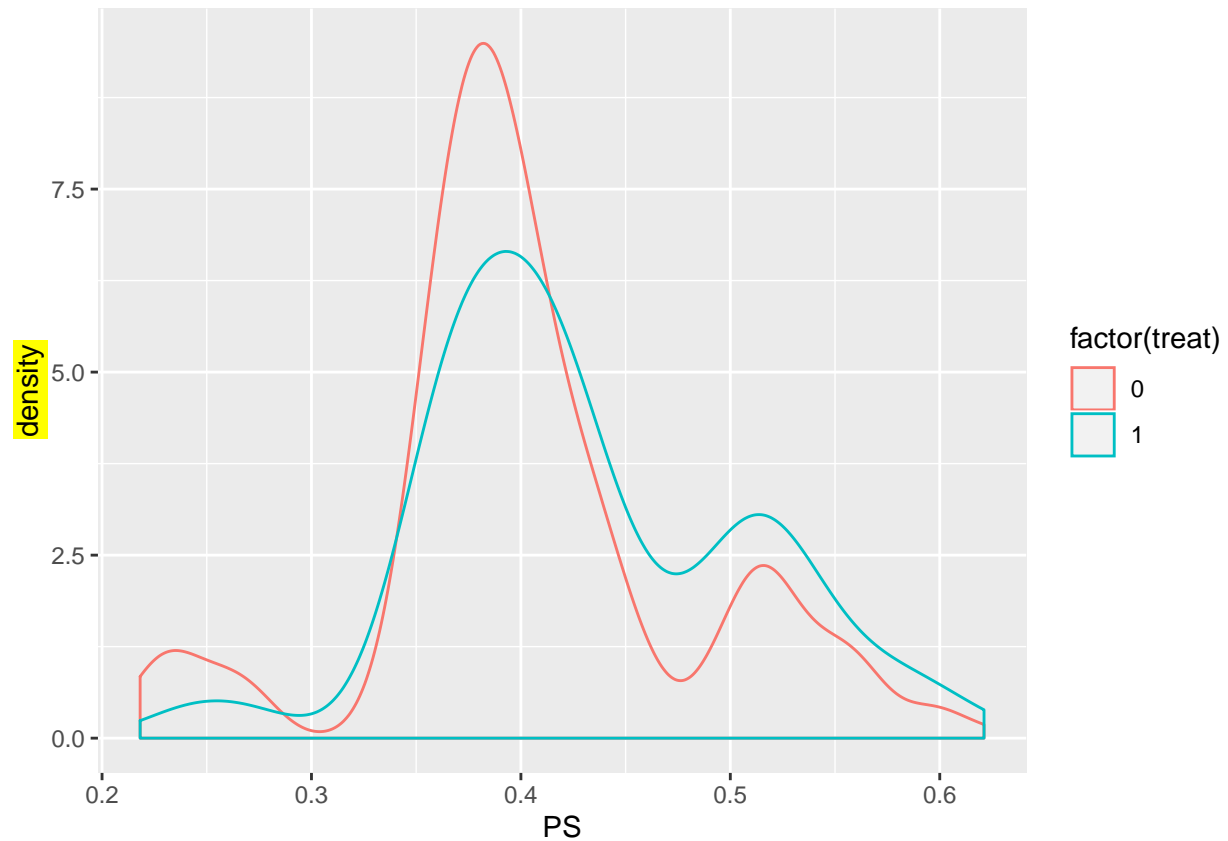
##
## Using 'treat'='1' as baseline group

##
## Call:
## matchit(formula = treat ~ age + education + black + married +
##         nodegree + re75 + hispanic, data = MyData, method = "cem")
##
## Sample sizes:
##           Control Treated
## All           425      297
## Matched       299      212
## Unmatched     126       85
## Discarded      0        0

autocem.data <- match.data(autocem.match)

autocem.data$PS<-glm(treat ~ age + education + black + married + nodegree + re75 + hispanic, data=autocem
ggplot(autocem.data, aes(x = PS, color = factor(treat))) +geom_density()

```



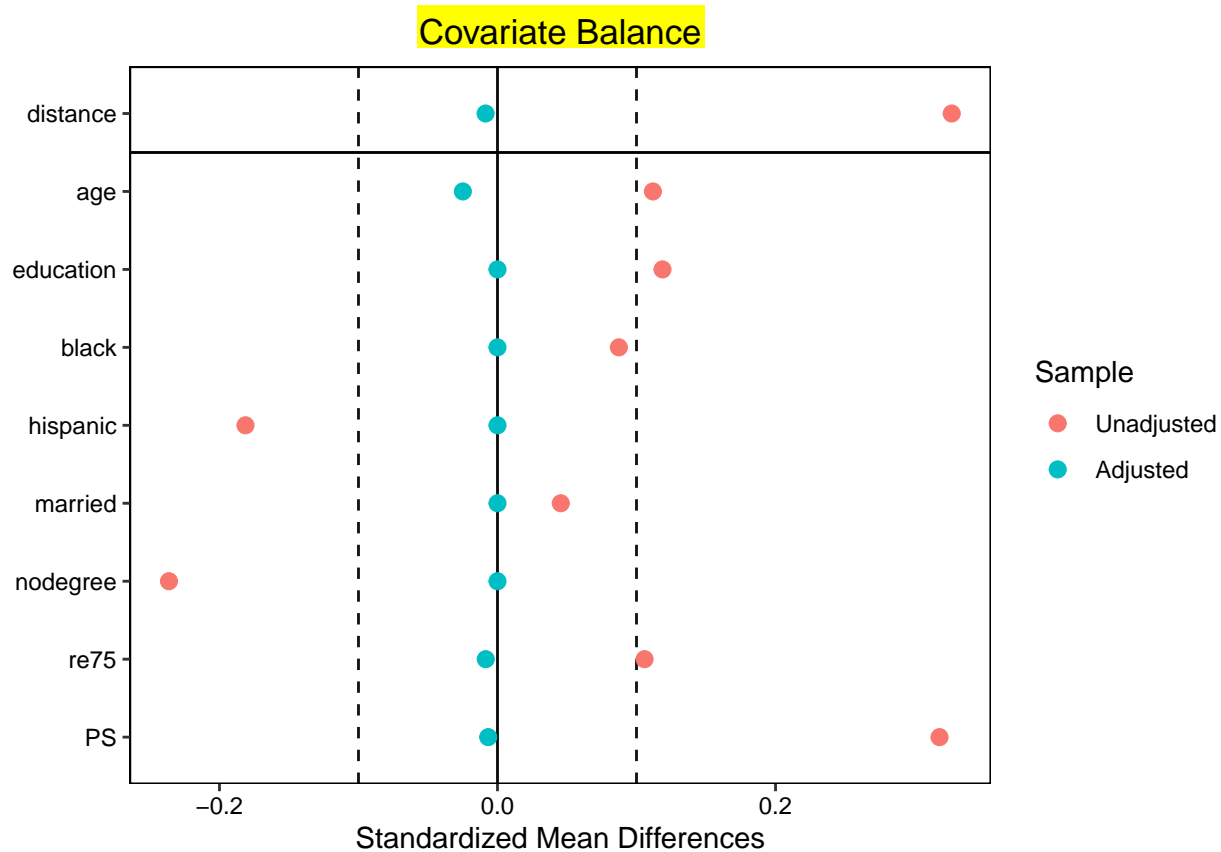
```
#Lets Check Covariate balance
```

```
m.out <- matchit(f.build("treat", covs), data = autocem.data, method = "cem")
```

```
##  
## Using 'treat'='1' as baseline group
```

```
love.plot(m.out, binary = "std", threshold = .1)
```

```
## Note: s.d.denom not specified; assuming pooled.
```



```
autocem.model <- lm(re78 ~ treat + age + education + black + married + nodegree + re75 + hispanic, data = data)
stargazer(exact.model, nearest.model, autocem.model, title="Model with Different Types of Matches", type="text")
```

```
##
## Model with Different Types of Matches
## =====
##                               Dependent variable:
##                               -----
##                               re78
##                               PSM - Nearest
##                               Auto CEM
##                               (1)      (2)      (3)
## -----
## treat                1,111.872        768.950        1,028.247*
##                      (1,001.031)      (520.887)      (553.855)
##
## age                   39.186          22.080          29.878
##                      (106.032)        (40.633)        (58.560)
##
## education            183.483          231.817          122.674
##                      (547.027)        (200.599)        (255.022)
##
## black                -1,524.208*       -2,461.259**
##                      (856.030)        (1,129.102)
```

```
## married                -704.116          466.634          -427.753
##                        (2,476.021)        (737.256)        (942.867)
##
## nodegree                -3,501.047*        -204.301          -290.313
##                        (1,910.287)        (814.603)        (914.372)
##
## re75                    0.156***          0.209*
##                        (0.053)          (0.117)
##
## hispanic                124.220          -319.967
##                        (1,156.996)        (1,578.470)
##
## Constant                4,154.232          3,060.760          4,841.363
##                        (6,646.500)        (2,918.129)        (3,472.525)
##
## -----
## Observations            137              594              511
## R2                      0.075              0.041              0.031
## Adjusted R2             0.040              0.028              0.015
## Residual Std. Error 5,679.853 (df = 131) 6,344.118 (df = 585) 6,093.433 (df = 502)
## F Statistic             2.134* (df = 5; 131) 3.107*** (df = 8; 585) 1.984** (df = 8; 502)
## =====
## Note:                                     *p<0.1; **p<0.05; ***p<0.01
```

```
# User Specified Coarsend Exact Matching
```

```
re75cut <- seq(0, max(MyData$re75), by=1000)
agecut <- c(20.5, 25.5, 30.5, 35.5, 40.5)
educut <- c(0, 6.5, 8.5, 12.5, 17)
```

```
my.cutpoints <- list(re75=re75cut, age=agecut, education = educut)
```

```
usercem.match <- matchit(treat ~ age + education + black + married + nodegree + re75 + hispanic, data =
```

```
##
## Using 'treat'='1' as baseline group
```

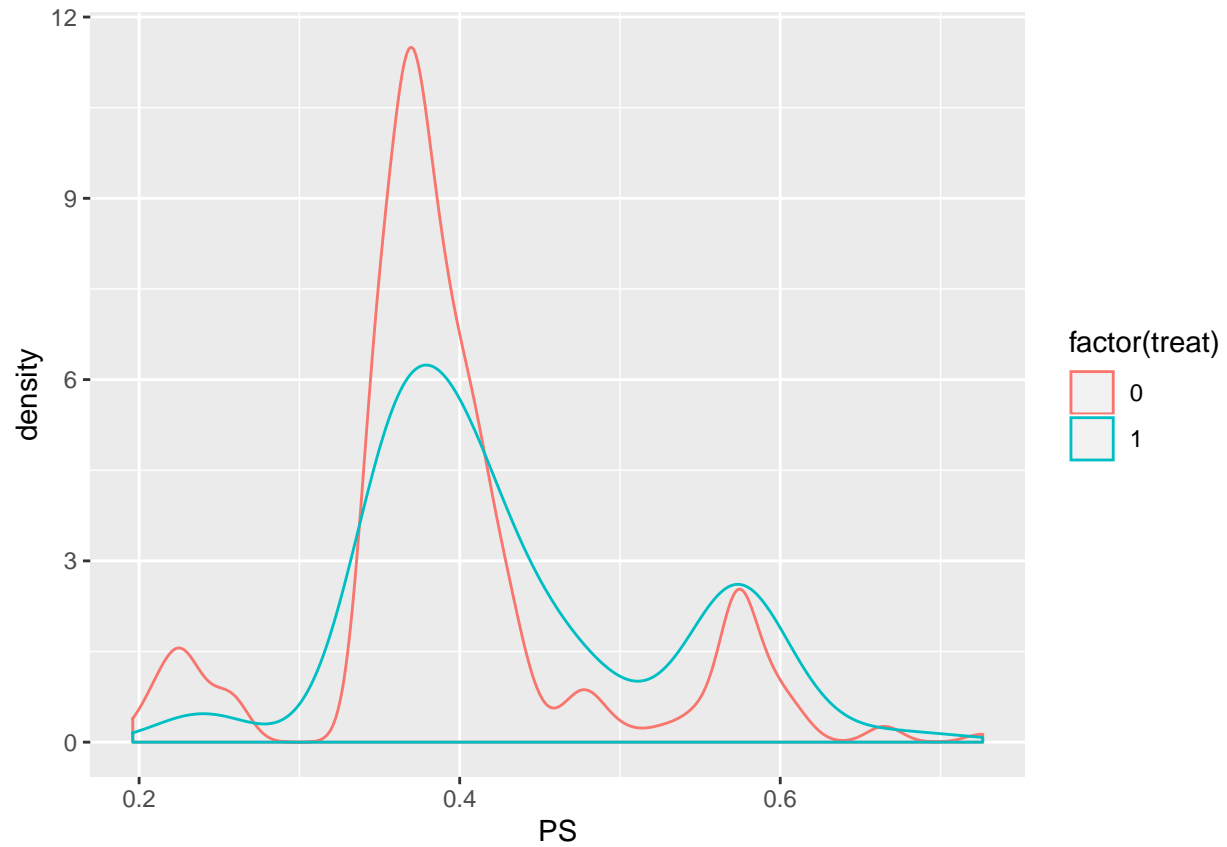
```
summary(usercem.match)$nn
```

```
##           Control Treated
## All           425      297
## Matched       284      199
## Unmatched     141       98
## Discarded       0        0
```

```
usercem.data <- match.data(usercem.match)
```

```
usercem.data$PS<-glm(treat ~ age + education + black + married + nodegree + re75 + hispanic, data=usercem
```

```
ggplot(usercem.data, aes(x = PS, color = factor(treat))) +geom_density()
```

```
#Let's check Covariate Balance
```

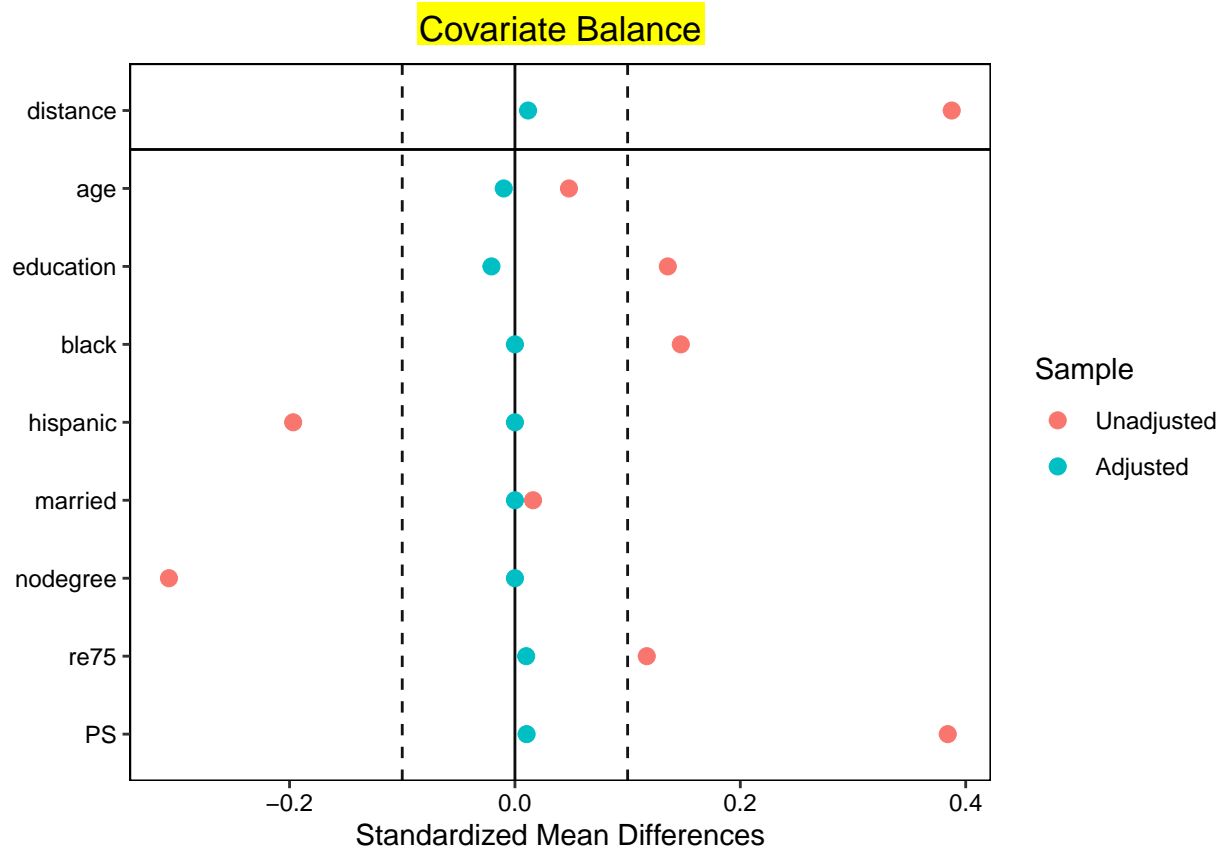
```
m.out <- matchit(f.build("treat", covs), data = usercem.data, method = "cem")
```

```
##
```

```
## Using 'treat'='1' as baseline group
```

```
love.plot(m.out, binary = "std", threshold = .1)
```

```
## Note: s.d.denom not specified; assuming pooled.
```



```
#Estimate Average Treatment Effect
```

```
usercem.model <- lm(re78 ~ treat + age + education + black + married + nodegree + re75 + hispanic, data =
```

```
stargazer(nearest.model, autocem.model, usercem.model, title="Model with Different Types of Matches", type="
```

```
##
```

```
## Model with Different Types of Matches
```

```
## =====
```

```
## Dependent variable:
```

```
## -----
```

```
## re78
```

```
## PSM - Nearest Auto CEM User CEM
```

```
## (1) (2) (3)
```

```
## -----
```

```
## treat 768.950 1,028.247* 918.704
```

```
## (520.887) (553.855) (572.979)
```

```
##
```

```
## age 22.080 29.878 23.047
```

```
## (40.633) (58.560) (45.349)
```

```
##
```

```
## education 231.817 122.674 242.918
```

```
## (200.599) (255.022) (244.044)
```

```
##
```

```
## black -1,524.208* -2,461.259** -1,219.881
```

```
## (856.030) (1,129.102) (1,565.652)
```

```

##
## married          466.634          -427.753          -978.750
##                  (737.256)          (942.867)          (1,028.592)
##
## nodegree         -204.301          -290.313           30.472
##                  (814.603)          (914.372)          (967.360)
##
## re75              0.156***          0.209*           0.019
##                  (0.053)          (0.117)          (0.145)
##
## hispanic          124.220          -319.967          571.143
##                  (1,156.996)          (1,578.470)          (1,902.790)
##
## Constant          3,060.760          4,841.363          2,422.117
##                  (2,918.129)          (3,472.525)          (3,574.739)
##
## -----
## Observations          594           511           483
## R2                    0.041           0.031           0.015
## Adjusted R2           0.028           0.015           -0.002
## Residual Std. Error  6,344.118 (df = 585)  6,093.433 (df = 502)  6,088.005 (df = 474)
## F Statistic          3.107*** (df = 8; 585)  1.984** (df = 8; 502)  0.908 (df = 8; 474)
## =====
## Note:                                     *p<0.1; **p<0.05; ***p<0.01

```