Propensity Score R Code

Data/Environment Setup

```
#Load Libraries
library(tableone)
library(MatchIt)
library(dplyr)
library(ggplot2)

#Load in data
data <- read.csv("http://web.hku.hk/~bcowling/data/propensity.csv", header=TRUE)
```

Estimate Propensity Scores

```
#Fit logistic regression model regressing on treatment/covariates.
pre.ps <- glm(trt ~ age + risk + severity, family="binomial",
              data=data)
#Solve for probabilities using predict function
data$ps<-predict(pre.ps, type="response")
```

Applying Propensity Score Method

Matching

```
#Match by propensity scores
# Nearest neighbor method without replacement
match_dat<-match.data(matchit(trt ~ ps, data=data, ratio=1, replace=F,
                              method = "nearest"))

# Optimal matching without replacement
match_dat<-match.data(matchit(trt ~ ps, data=data, ratio=1, replace=F,
                              method = "optimal"))

# Radius matching
match_dat<-match.data(matchit(trt ~ ps, data=data, caliper=.2,
                              ratio=3, method = "nearest"))

#Using nearest neighbor method (see documentation for details)
#caliper indicates how close scores need to be to be matched
match_dat<-match.data(matchit(trt ~ ps, data=data, caliper=.2,
                              ratio=1, replace=F, method = "nearest"))
```

Stratification

```
#Define groups based on propensity score quintiles
data$ps_grp <- cut(data$ps, breaks=quantile(data$ps, prob=0:5*0.2),
 labels=c("Q1","Q2","Q3","Q4","Q5"), include.lowest = TRUE)
```

$Z_i =$ Indicator variable denoting treatment group$e_i =$ Propensity scoreWeight $w_i = \frac{Z_i}{e_i} + \frac{1 - Z_i}{1 - e_i} w_{i,ATT} = Z_i + \frac{(1 - Z_i)e_i}{1 - e_i} w_{i,A}$

Inverse Probability of Treatment Weighting

```
#Create weights for each row of data
weight<-data$trt/data$ps+(1-data$trt)/(1-data$ps)
```

Covariate Adjustment Using Propensity Score

```
#Run logistic regression adjusting for propensity score
mod1<-glm(death ~ trt+ps, family="binomial",
              data=data)
```

Check Balance Diagnosistics

Example uses matched data but process is similar for IPTW and stratification. With IPTW use weighted sample to compare covariates. With stratification, check balance of covariates for each strata.

```
#Make table comparing smd non matched vs matched samples
tab_before<-CreateTableOne(vars=c("age", "risk", "severity", "male"),
                           strata = "trt", data=data, test=FALSE, smd = TRUE)

tab_match<-CreateTableOne(vars=c("age", "risk", "severity", "male"),
                          strata = "trt", data=match_dat, test=FALSE, smd = TRUE)

print(tab_before, smd=TRUE)
```

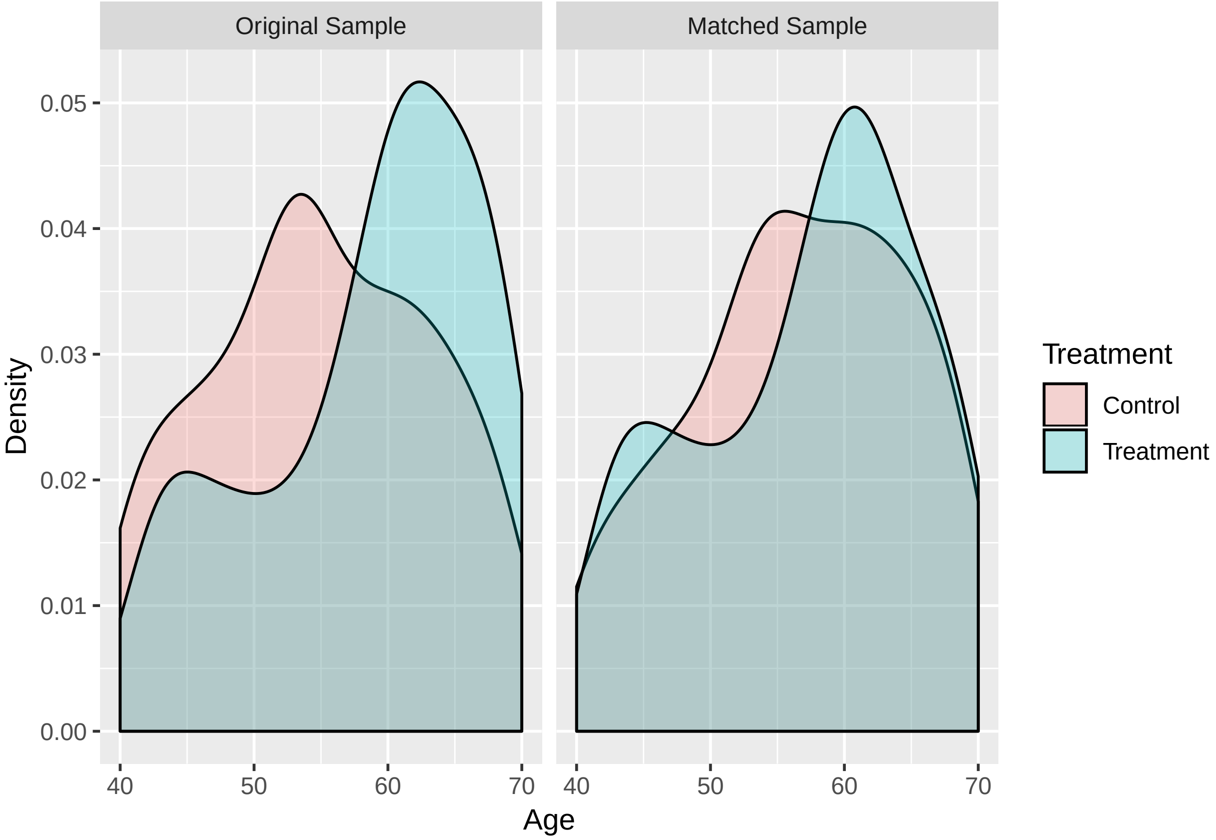
```
##                         Stratified by trt
##                          0            1            SMD
##   n                        208          192
##   age (mean (SD))        54.96 (8.27) 58.64 (8.26)  0.444
##   risk (mean (SD))        1.31 (1.31)  1.99 (1.54)  0.473
##   severity (mean (SD))    4.47 (2.03)  5.28 (2.16)  0.384
##   male (mean (SD))        1.00 (0.00)  1.00 (0.00) <0.001
```

```
print(tab_match, smd = TRUE)
```

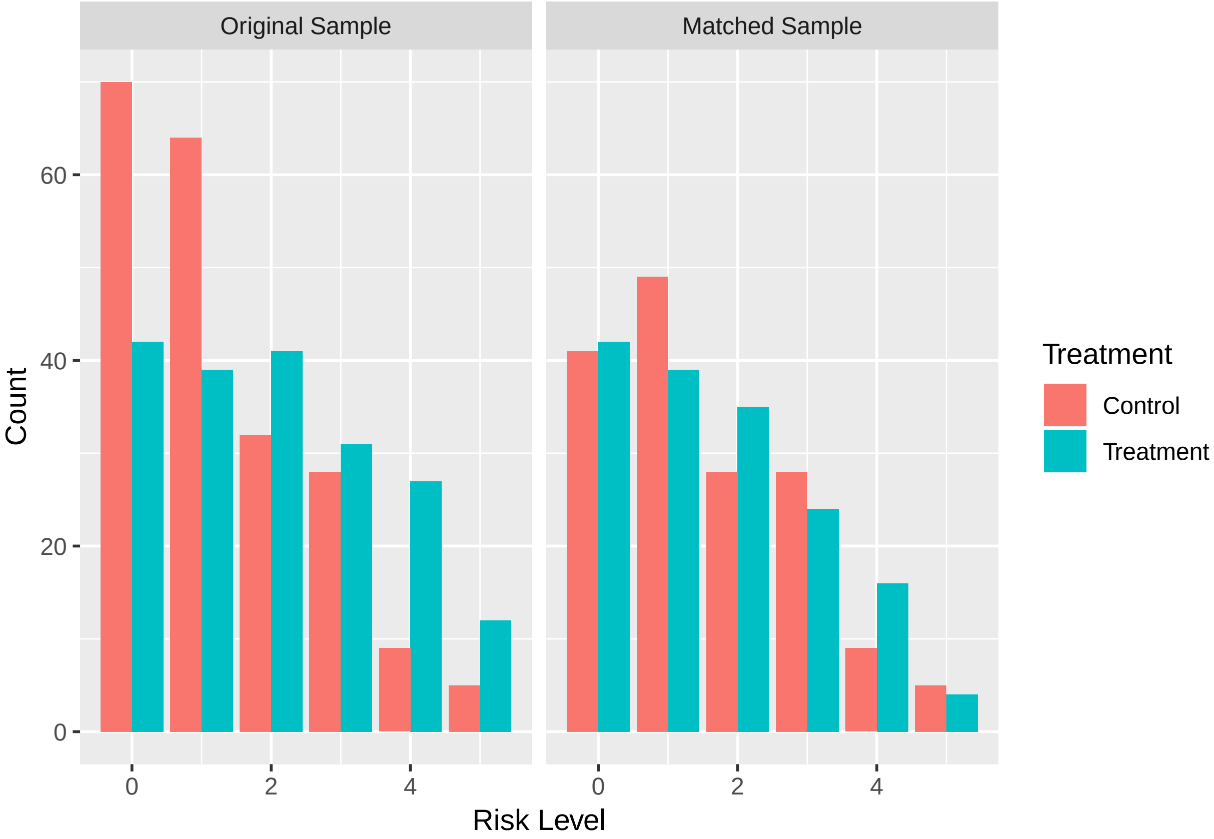
```
##                         Stratified by trt
##                          0            1            SMD
##   n                        160          160
##   age (mean (SD))        56.60 (8.11) 57.23 (8.27)  0.077
##   risk (mean (SD))        1.56 (1.35)  1.66 (1.40)  0.068
##   severity (mean (SD))    4.74 (2.07)  4.94 (2.05)  0.100
##   male (mean (SD))        1.00 (0.00)  1.00 (0.00) <0.001
```

```r
#Set up data for plots
match_dat$matched<-"Matched Sample"
data$matched<-"Original Sample"
comb_data<-bind_rows(match_dat, data)
comb_data$trt<-as.factor(comb_data$trt)
levels(comb_data$trt)<-c("Control", "Treatment")
comb_data$matched<-factor(comb_data$matched, levels=c("Original Sample", "Matched Sample"))

#Age density plots
ggplot(data = comb_data, aes(x=age, fill=trt, group=trt))+
  geom_density(alpha=0.25)+facet_wrap(~matched)+ylab("Density")+xlab("Age")+
  guides(fill=guide_legend(title="Treatment"))
```



```r
#Risk barplots
ggplot(data=comb_data, aes(x=risk, fill=trt, group=trt))+
  geom_bar(position = "dodge")+facet_wrap(~matched)+ylab("Count")+xlab("Risk Level")+
  guides(fill=guide_legend(title="Treatment"))
```


Original Sample
Matched Sample
Count
60
40
20
0
0
2
4
0
2
4
Risk Level
Treatment
Control
Treatment