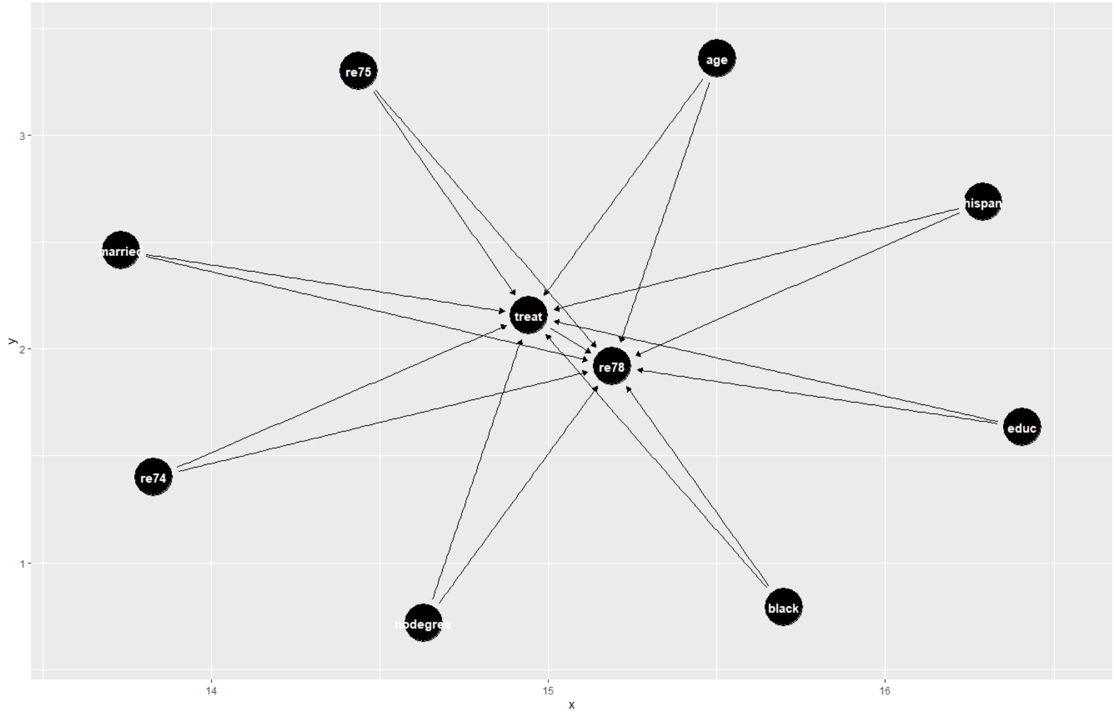
Homework 5

Xinyi Lin

10/13/2019

Question 1

The treatment is `treat`, the outcome is `re78` and cofounders are `age`, `educ`, `black`, `hispan`, `married`, `nodegree`, `re74`, `re75`.The treatment has casual effect on the outcome and confounders have casual effect on both the treatment and the outcome.



Question 2

```
##                     Stratified by treat
##                      0                 1                 SMD
##   n                      429               185
##   age (mean (SD))      28.03 (10.79)     25.82 (7.16)     0.242
##   educ (mean (SD))     10.24 (2.86)      10.35 (2.01)     0.045
##   black = 1 (%)           87 (20.3)        156 (84.3)     1.671
##   hispan = 1 (%)          61 (14.2)         11 ( 5.9)     0.277
##   married = 1 (%)        220 (51.3)         35 (18.9)     0.721
##   nodegree = 1 (%)       256 (59.7)        131 (70.8)     0.235
##   re74 (mean (SD)) 5619.24 (6788.75) 2095.57 (4886.62)  0.596
##   re75 (mean (SD)) 2466.48 (3292.00) 1532.06 (3219.25)  0.287
```

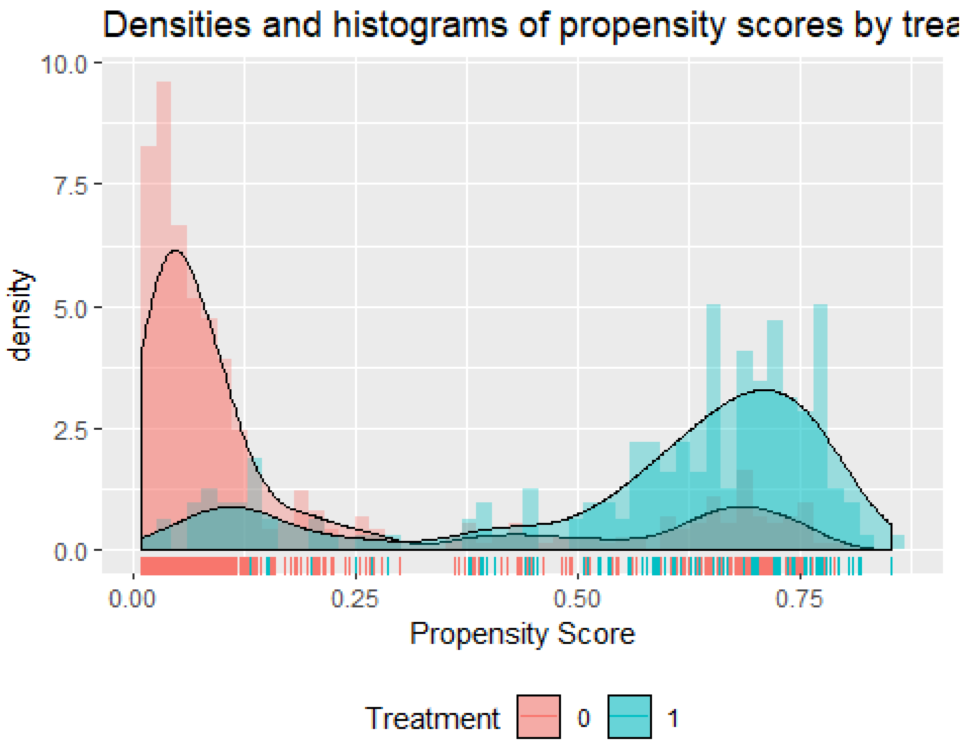
The covariate balance table is shown as above. SMD means standardized mean difference. We want SMDs are smaller than 0.2. However, according to the table above, we can find out that except for educ, SMDs are larger than 0.2, which means covariates are not balanced.

Question 3

```
##
## Call:
## glm(formula = treat ~ age + educ + black + hispan + married +
##     nodegree + re74 + re75, family = binomial, data = q2_data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7645  -0.4736  -0.2862   0.7508   2.7169
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.729e+00  1.017e+00  -4.649 3.33e-06 ***
## age          1.578e-02  1.358e-02   1.162  0.24521
## educ         1.613e-01  6.513e-02   2.477  0.01325 *
## black1       3.065e+00  2.865e-01  10.699  < 2e-16 ***
## hispan1      9.836e-01  4.257e-01   2.311  0.02084 *
## married1    -8.321e-01  2.903e-01  -2.866  0.00415 **
## nodegree1    7.073e-01  3.377e-01   2.095  0.03620 *
## re74        -7.178e-05  2.875e-05  -2.497  0.01253 *
## re75         5.345e-05  4.635e-05   1.153  0.24884
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 751.49  on 613  degrees of freedom
## Residual deviance: 487.84  on 605  degrees of freedom
## AIC: 505.84
##
## Number of Fisher Scoring iterations: 5
```

The propensity score is shown above.

Question 4



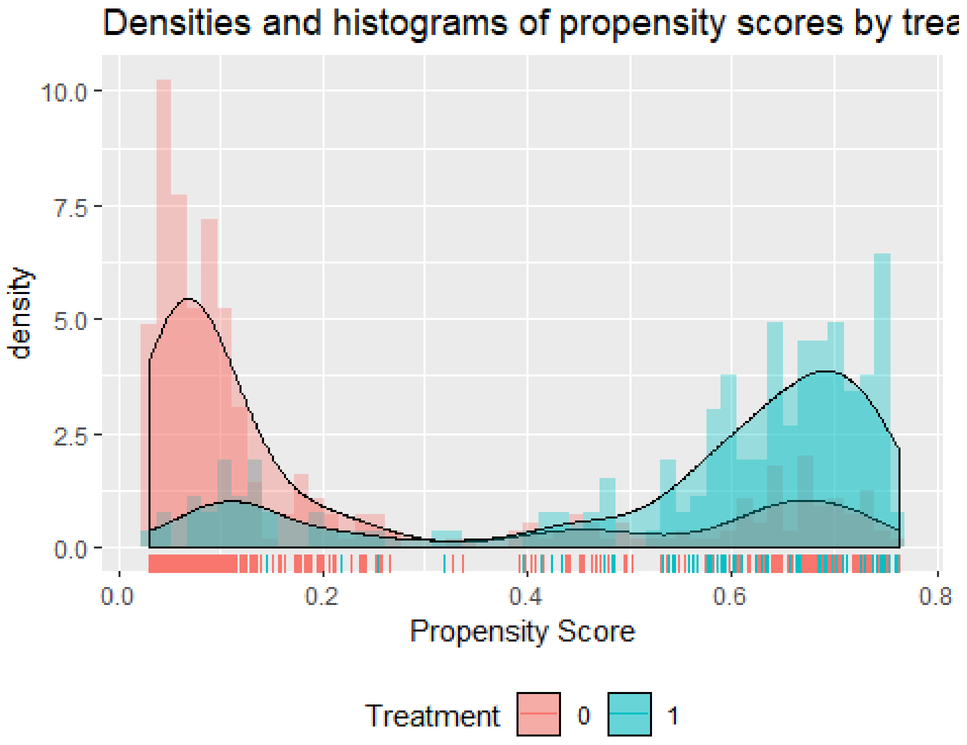
Distributions of propensity score in two treatment groups are shown above. We can find that for those propensity scores that are close to 1, two distributions not overlap. We need to trim data.

Eliminate controls for whom the P(A=1|C) is less than the min(P(A=1|C)) found in the treated group and eliminate treated for whom the P(A=1|C) is greater than the max(P(A=1|C)) found in the control group.

There are 65 observations have been eliminated and there are 549 observations left.

Trimming can improve covariate balance, improving internal validity, so efficiency is improved. But trimming will hurts external validity(generalizability).

```
##
## Call:
## glm(formula = treat ~ age + educ + black + hispan + married +
##     nodegree + re74 + re75, family = binomial, data = trim_data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6953  -0.5085  -0.3413   0.8349   2.6217
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.285e+00  1.038e+00  -4.130 3.63e-05 ***
## age          1.174e-02  1.383e-02   0.849   0.3961
## educ         1.386e-01  6.596e-02   2.102   0.0355 *
## black1       2.964e+00  2.865e-01  10.345  < 2e-16 ***
## hispan1      9.087e-01  4.237e-01   2.145   0.0320 *
## married1    -7.324e-01  2.906e-01  -2.520   0.0117 *
## nodegree1    5.978e-01  3.406e-01   1.755   0.0793 .
## re74        -5.734e-05  2.978e-05  -1.925   0.0542 .
## re75         3.700e-05  4.747e-05   0.779   0.4358
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 690.28  on 548  degrees of freedom
## Residual deviance: 481.82  on 540  degrees of freedom
## AIC: 499.82
##
## Number of Fisher Scoring iterations: 5
```



Distributions of propensity score of trimmed data is shown above, we can find that distributions overlap better which means balance of covariates improve.

Question 5

The table for covariate balance is as following. We can find that balance among covariate is improved as SMD decrease and SMD of age, educ, nodegree are less than 0.2.

```
##                    Stratified by treat
##                     0                 1                 SMD
##   n                     372               177
##   age (mean (SD))    26.92 (10.38)     25.45 (6.99)     0.166
##   educ (mean (SD))   10.24 (2.79)      10.32 (2.05)     0.035
##   black = 1 (%)         87 (23.4)        148 (83.6)     1.515
##   hispan = 1 (%)        61 (16.4)         11 ( 6.2)     0.326
##   married = 1 (%)      163 (43.8)         35 (19.8)     0.534
##   nodegree = 1 (%)     234 (62.9)        123 (69.5)     0.140
##   re74 (mean (SD)) 4051.32 (5341.39) 2179.39 (4978.21)  0.363
##   re75 (mean (SD)) 2329.24 (3246.09) 1485.92 (3215.90)  0.261
```

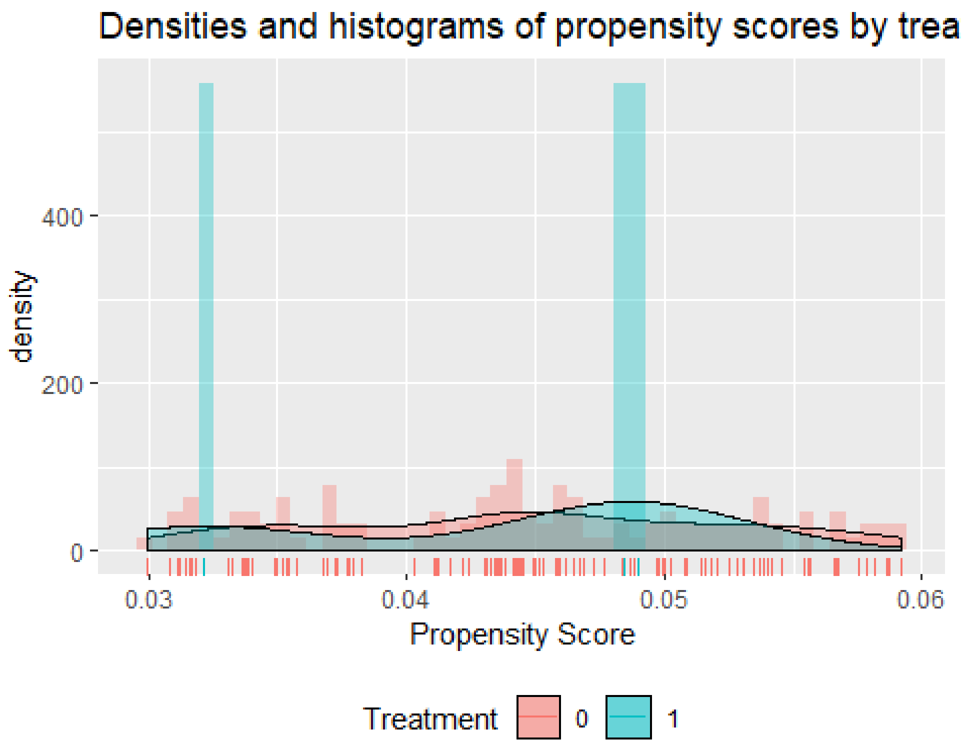
The propensity score of trimmed data is as following.

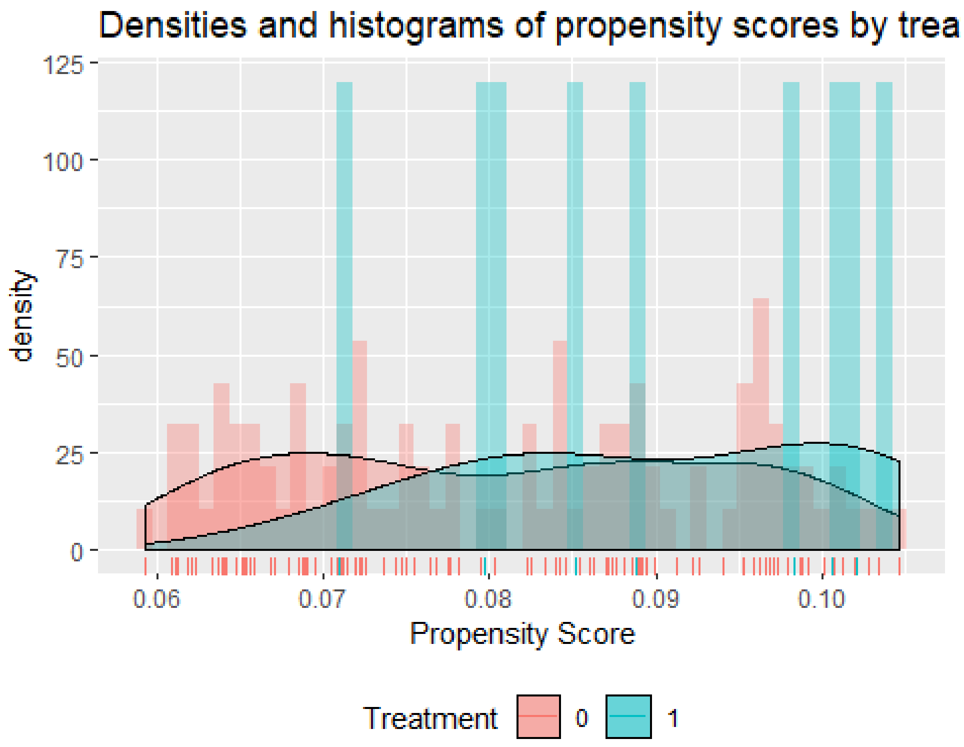
```
##
## Call:
## glm(formula = treat ~ age + educ + black + hispan + married +
##     nodegree + re74 + re75, family = binomial, data = trim_data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6953  -0.5085  -0.3413   0.8349   2.6217
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.285e+00  1.038e+00  -4.130 3.63e-05 ***
## age          1.174e-02  1.383e-02   0.849   0.3961
## educ         1.386e-01  6.596e-02   2.102   0.0355 *
## black1       2.964e+00  2.865e-01  10.345  < 2e-16 ***
## hispan1      9.087e-01  4.237e-01   2.145   0.0320 *
## married1    -7.324e-01  2.906e-01  -2.520   0.0117 *
## nodegree1    5.978e-01  3.406e-01   1.755   0.0793 .
## re74        -5.734e-05  2.978e-05  -1.925   0.0542 .
## re75         3.700e-05  4.747e-05   0.779   0.4358
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 690.28  on 548  degrees of freedom
## Residual deviance: 481.82  on 540  degrees of freedom
## AIC: 499.82
##
## Number of Fisher Scoring iterations: 5
```

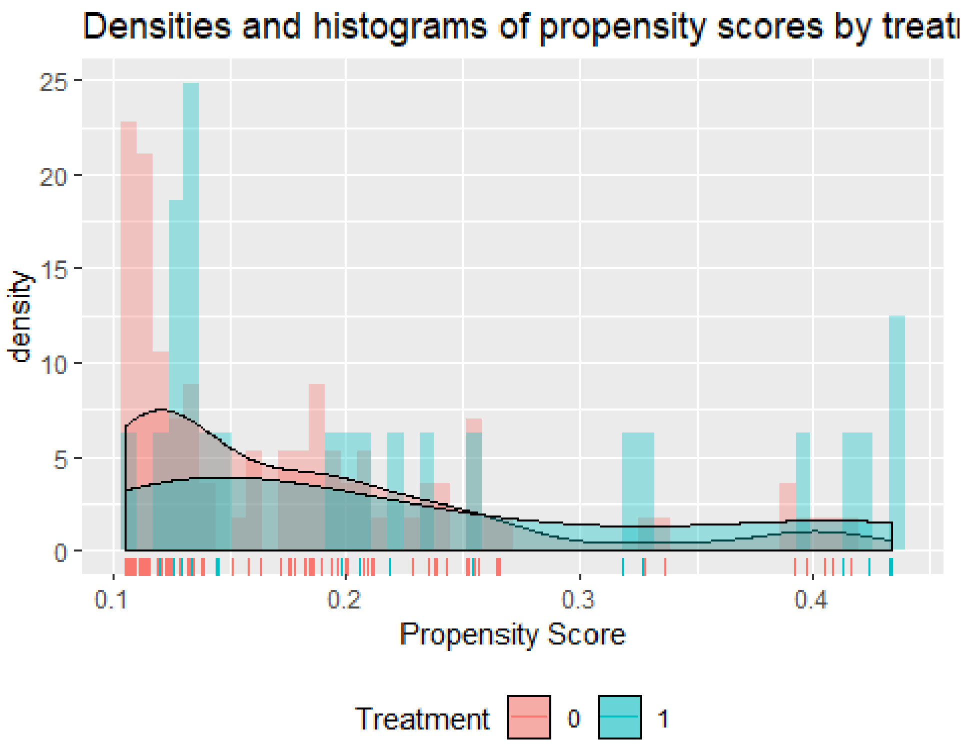
Question 6

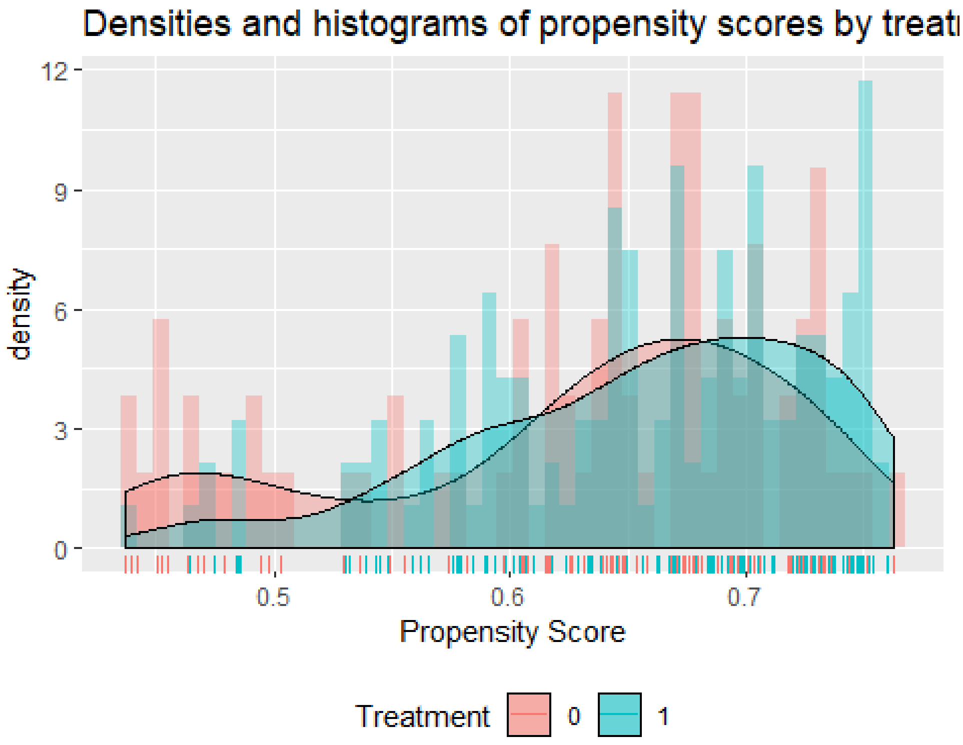
```
##        20%        40%        60%        80%
## 0.05923431 0.10473370 0.43605402 0.66770086
##    subclass
##       0   1   2   3   4
##   0 107 101  85  43  36
##   1   3   9  24  67  74
```

I choose 20%, 40%, 60% and 80% quantiles as breaks. Breaks are 0.05923431, 0.10473370, 0.43605402, 0.66770086.


Densities and histograms of propensity scores by trea
density
400
200
0
0.03
0.04
0.05
0.06
Propensity Score
Treatment
0
1


Densities and histograms of propensity scores by trea
density
125
100
75
50
25
0
0.06
0.07
0.08
0.09
0.10
Propensity Score
Treatment
0
1





Above are plots of propensity scores in each subclass. According to them, we can find that in each subclass, distributions of propensity score overlap better.

```
##                    Stratified by treat
##                     0                 1                 SMD
##   n                    107                 3
##   age (mean (SD))    30.47 (9.52)      27.67 (5.03)     0.368
##   educ (mean (SD))    9.65 (3.01)      11.00 (2.65)     0.475
##   black = 1 (%)          0 ( 0.0)          0 (  0.0)   <0.001
##   hispan = 1 (%)         9 ( 8.4)          0 (  0.0)    0.429
##   married = 1 (%)       99 (92.5)          3 (100.0)    0.402
##   nodegree = 1 (%)      69 (64.5)          1 ( 33.3)    0.656
##   re74 (mean (SD)) 7287.10 (6294.61) 3127.19 (5416.45)  0.708
##   re75 (mean (SD)) 3870.39 (4012.39)  855.62 (856.58)   1.039
```

```
##                    Stratified by treat
##                     0                 1                 SMD
##   n                    101                 9
##   age (mean (SD))    26.38 (10.33)     24.22 (6.42)     0.250
##   educ (mean (SD))   10.38 (2.66)      10.67 (2.06)     0.122
##   black = 1 (%)          0 ( 0.0)          0 ( 0.0)    <0.001
##   hispan = 1 (%)        14 (13.9)          1 (11.1)     0.083
##   married = 1 (%)       33 (32.7)          1 (11.1)     0.540
##   nodegree = 1 (%)      57 (56.4)          3 (33.3)     0.478
##   re74 (mean (SD)) 3436.38 (4074.18) 1158.99 (2425.99)  0.679
##   re75 (mean (SD)) 1884.93 (2565.98) 1266.75 (2828.07)  0.229
```

```
##                    Stratified by treat
##                     0                 1                 SMD
##   n                     85                24
##   age (mean (SD))    24.00 (9.60)      26.00 (6.69)     0.242
##   educ (mean (SD))   10.87 (2.63)      10.42 (2.06)     0.192
##   black = 1 (%)          8 ( 9.4)          7 (29.2)     0.517
##   hispan = 1 (%)        38 (44.7)         10 (41.7)     0.061
##   married = 1 (%)       14 (16.5)          8 (33.3)     0.398
##   nodegree = 1 (%)      55 (64.7)         17 (70.8)     0.131
##   re74 (mean (SD)) 1990.66 (3582.09) 5696.63 (9407.57)  0.521
##   re75 (mean (SD)) 1501.60 (2186.58) 3182.66 (3811.06)  0.541
```

```
##                    Stratified by treat
##                     0                 1                 SMD
##   n                     43                67
##   age (mean (SD))    28.91 (12.26)     25.99 (7.08)     0.292
##   educ (mean (SD))    9.88 (3.27)      10.07 (2.41)     0.066
##   black = 1 (%)         43 (100.0)        67 (100.0)   <0.001
##   hispan = 1 (%)         0 (  0.0)         0 (  0.0)   <0.001
##   married = 1 (%)       17 ( 39.5)        23 ( 34.3)    0.108
##   nodegree = 1 (%)      24 ( 55.8)        40 ( 59.7)    0.079
##   re74 (mean (SD)) 4411.86 (6256.01) 2926.13 (4861.49)  0.265
##   re75 (mean (SD)) 2573.53 (4002.31) 1795.30 (4183.12)  0.190
```

Above are tables showing covariate balance in each subclass. We can find that SMD are smaller in each subclass, so balance of covariate improve for each subclass.

Question 7

The point estimate of the marginal average causal effect is 621.156. The confidence interval is (-1334.344, 2576.656). Assuming $z = \frac{v_{ACE}-0}{sd_{ACE}}$ follows normal distribution, then the p-value is 0.263.

As the p-value is larger than 0.05 and the confidence interval contains 0, there is no marginal causal effect between job training and the income in 1978. With 95% confidence, we can conculde that the true marginal causal effect lies between -1334.344 and 2576.656.

Question 8

```
## 
## Call:
## lm(formula = re78 ~ treat + age + educ + black + hispan + married + 
##     nodegree + re74 + re75, data = q2_data)
## 
## Residuals:
##    Min     1Q Median     3Q    Max 
## -13595  -4894  -1662   3929  54570 
## 
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  6.651e+01  2.437e+03   0.027   0.9782    
## treat1       1.548e+03  7.813e+02   1.982   0.0480 *  
## age          1.298e+01  3.249e+01   0.399   0.6897    
## educ         4.039e+02  1.589e+02   2.542   0.0113 *  
## black1      -1.241e+03  7.688e+02  -1.614   0.1071    
## hispan1      4.989e+02  9.419e+02   0.530   0.5966    
## married1     4.066e+02  6.955e+02   0.585   0.5590    
## nodegree1    2.598e+02  8.474e+02   0.307   0.7593    
## re74         2.964e-01  5.827e-02   5.086 4.89e-07 ***
## re75         2.315e-01  1.046e-01   2.213   0.0273 *  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 6948 on 604 degrees of freedom
## Multiple R-squared:  0.1478, Adjusted R-squared:  0.1351 
## F-statistic: 11.64 on 9 and 604 DF,  p-value: < 2.2e-16
```

```
##                      2.5 %       97.5 %
## (Intercept) -4.719009e+03 4852.0379796
## treat1       1.388991e+01 3082.5976942
## age         -5.082731e+01   76.7825747
## educ         9.186538e+01  716.0170862
## black1      -2.750420e+03  269.1318254
## hispan1     -1.350983e+03 2348.7772098
## married1    -9.592168e+02 1772.4584772
## nodegree1   -1.404474e+03 1924.1082505
## re74         1.819359e-01    0.4108190
## re75         2.606296e-02    0.4369888
```

According to results above, we can find that p-value of variable `treat` is 0.048 which is smalller than 0.05, thus we can reject the null hypothesis and conclude that there exists marginal causal effect.

Compared to people who did not receive job training, the average income of people who receive job training in 1978 is 1548 higher.

With 95% confidence, we can conculde that the true marginal causal effect lies between 13.88991 and 3082.5976942.

Compred to results in question 7, we can find that the p-value in question 7 is larger. It is larger than 0.05, which means there is no significant marginal causal effect of treat. The marginal causal effect in question 7 is smaller. Two methods give different conclusions.

Question 9

For the regression based approach to confounding adjustment, the advantage is that it is easy to get point estimate and p-value of marginal average causal effect and the disadvantage is that it might not adjust confounders properly, because it ignores whether treatment group and control group is comparable.

For the subclassification approach, the advantage is that it allows treatment group and control group comparable in each subgroup and the disadvantage is that we need to do multiple steps to get results.

Appendix

```
knitr::opts_chunk$set(echo = FALSE)
library(ggdag)
library(mlogit)
library(dplyr)
library(tableone)
library(personalized)
redata = read.csv("./hw5_data.csv")
# Question 1
tidy_ggdag <- dagify(re78 ~ treat + age + educ + black + hispan + married +
nodegree + re74 + re75,
             treat ~ age + educ + black + hispan + married + nodegree + re74
+ re75,
             exposure = "treat",
             outcome = "y") %>% tidy_dagitty()
ggdag(tidy_ggdag)
# Question 2
q2_data = redata %>%
  mutate(treat = as.factor(treat),
         black = as.factor(black),
         hispan = as.factor(hispan),
         married = as.factor(married),
         nodegree = as.factor(nodegree))
vars = c("age", "educ", "black", "hispan", "married", "nodegree", "re74",
"re75")

## Construct a table
cov_bal <- CreateTableOne(vars = vars, strata = "treat", data = q2_data, test
= FALSE)

## Show table with SMD
print(cov_bal, smd = TRUE)
# Question 3
ps.model<-glm(treat~age + educ + black + hispan + married + nodegree + re74 +
re75,data=q2_data, family = binomial)
summary(ps.model)
# Question 4
# propensity score of each unit
ps <- predict(ps.model, type="response")

x = q2_data

prop.func <- function(x, trt){
  # fit propensity score model
  propens.model <- glm(trt~age + educ + black + hispan + married + nodegree +
re74 + re75, data=x, family = binomial)
  pi.x <- predict(propens.model, type = "response")
  pi.x
}

# now add density plot with histogram
check.overlap(x = x,
              trt = q2_data$treat,
              type = "both",
              propensity.func = prop.func)
trim_data = x[ps>=min(ps[q2_data$treat==1]) & ps <=
max(ps[q2_data$treat==0]),]
ps.model<-glm(treat~age + educ + black + hispan + married + nodegree + re74 +
re75, data=trim_data, family = binomial)
summary(ps.model)

ps <- predict(ps.model, type="response") #gets the propensity scores for each
unit, based on the model

x = trim_data

# now add density plot with histogram
check.overlap(x = x,
              trt = trim_data$treat,
              type = "both",
              propensity.func = prop.func)
# Question 5
## Construct a table
cov_bal <- CreateTableOne(vars = vars, strata = "treat", data = trim_data,
test = FALSE)

## Show table with SMD
print(cov_bal, smd = TRUE)
ps.model<-glm(treat~age + educ + black + hispan + married + nodegree + re74 +
re75, data=trim_data, family = binomial)
summary(ps.model)

ps <- predict(ps.model, type="response") #gets the propensity scores for each
unit, based on the model
# Question 6
#creating subclasses
subclass.breaks = quantile(ps, c(.20, .40, .60, .80)) # bins (initial try -
modify as needed)
subclass.breaks
subclass = ps
subclass = as.numeric(ps>subclass.breaks[1])
subclass[which(ps>subclass.breaks[1]& ps<=subclass.breaks[2])]<- 1
subclass[which(ps>subclass.breaks[2]& ps<=subclass.breaks[3])]<- 2
subclass[which(ps>subclass.breaks[3]& ps<=subclass.breaks[4])]<- 3
subclass[which(ps>subclass.breaks[4])]<- 4
#looking at sample sizes within each subclass

table(trim_data$treat, subclass)
#looking at propensity scores within subclasses
prop.func <- function(x, trt) {
  ps[which(ps <= subclass.breaks[1])]
}
trim_data$ps <-ps
check.overlap(x = trim_data[which(trim_data$ps <=subclass.breaks[1]),],
              trt = trim_data$treat[which(trim_data$ps <=
subclass.breaks[1])],
              type = "both",
              propensity.func = prop.func)


prop.func <- function(x, trt)
{

  ps[which(ps>subclass.breaks[1]&ps<=subclass.breaks[2])]
}
trim_data$ps <-ps
check.overlap(x =
trim_data[which(ps>subclass.breaks[1]&ps<=subclass.breaks[2]),],
              trt =
trim_data$treat[which(ps>subclass.breaks[1]&ps<=subclass.breaks[2])],
              type = "both",
              propensity.func = prop.func)

prop.func <- function(x, trt)
{

  ps[which(ps>subclass.breaks[2]&ps<=subclass.breaks[3])]
}
trim_data$ps <-ps
check.overlap(x =
trim_data[which(ps>subclass.breaks[2]&ps<=subclass.breaks[3]),],
              trt =
trim_data$treat[which(ps>subclass.breaks[2]&ps<=subclass.breaks[3])],
              type = "both",
              propensity.func = prop.func)



 prop.func <- function(x, trt)
 {

   ps[which(ps>subclass.breaks[3])]
 }
 trim_data$ps <-ps
 check.overlap(x = trim_data[which(ps>subclass.breaks[3]),],
               trt = trim_data$treat[which(ps>subclass.breaks[3])],
               type = "both",
               propensity.func = prop.func)
tabUnmatched_s0 <- CreateTableOne(vars = vars, strata = "treat", data =
trim_data[which(subclass==0),], test = FALSE)
tabUnmatched_s1 <- CreateTableOne(vars = vars, strata = "treat", data =
trim_data[which(subclass==1),], test = FALSE)
tabUnmatched_s2 <- CreateTableOne(vars = vars, strata = "treat", data =
trim_data[which(subclass==2),], test = FALSE)
tabUnmatched_s3 <- CreateTableOne(vars = vars, strata = "treat", data =
trim_data[which(subclass==3),], test = FALSE)

## Show table with SMD

print(tabUnmatched_s0, smd = TRUE)
print(tabUnmatched_s1, smd = TRUE)
print(tabUnmatched_s2, smd = TRUE)
print(tabUnmatched_s3, smd = TRUE)
# Question 7
#AVERAGE CAUSAL EFFECT WITHIN STRATA
ACE0 <- mean(trim_data$re78[which(subclass==0 & trim_data$treat==1)])-
mean(trim_data$re78[which(subclass==0 & trim_data$treat==0)])
ACE1 <- mean(trim_data$re78[which(subclass==1 & trim_data$treat==1)])-
```

```
mean(trim_data$re78[which(subclass==1 & trim_data$treat==0)])
ACE2 <- mean(trim_data$re78[which(subclass==2 & trim_data$treat==1)])-
mean(trim_data$re78[which(subclass==2 & trim_data$treat==0)])
ACE3 <- mean(trim_data$re78[which(subclass==3 & trim_data$treat==1)])-
mean(trim_data$re78[which(subclass==3 & trim_data$treat==0)])
ACE4 <- mean(trim_data$re78[which(subclass==4 & trim_data$treat==1)])-
mean(trim_data$re78[which(subclass==4 & trim_data$treat==0)])

ace <- (nrow(trim_data[which(subclass==0),])/nrow(trim_data))*ACE0+
  (nrow(trim_data[which(subclass==1),])/nrow(trim_data))*ACE1+
  (nrow(trim_data[which(subclass==2),])/nrow(trim_data))*ACE2+
  (nrow(trim_data[which(subclass==3),])/nrow(trim_data))*ACE3+
  (nrow(trim_data[which(subclass==4),])/nrow(trim_data))*ACE4


v01 <- var(trim_data$re78[which(subclass==0 & trim_data$treat==1)])
v00 <- var(trim_data$re78[which(subclass==0 & trim_data$treat==0)])
v11 <- var(trim_data$re78[which(subclass==1 & trim_data$treat==1)])
v10 <- var(trim_data$re78[which(subclass==1 & trim_data$treat==0)])
v21 <- var(trim_data$re78[which(subclass==2 & trim_data$treat==1)])
v20 <- var(trim_data$re78[which(subclass==2 & trim_data$treat==0)])
v31 <- var(trim_data$re78[which(subclass==3 & trim_data$treat==1)])
v30 <- var(trim_data$re78[which(subclass==3 & trim_data$treat==0)])
v41 <- var(trim_data$re78[which(subclass==4 & trim_data$treat==1)])
v40 <- var(trim_data$re78[which(subclass==4 & trim_data$treat==0)])

n0 <- nrow(trim_data[which(subclass==0),])
n1 <- nrow(trim_data[which(subclass==1),])
n2 <- nrow(trim_data[which(subclass==2),])
n3 <- nrow(trim_data[which(subclass==3),])
n4 <- nrow(trim_data[which(subclass==4),])

n01 <- nrow(trim_data[which(subclass==0& trim_data$treat==1),])
n11 <- nrow(trim_data[which(subclass==1& trim_data$treat==1),])
n21 <- nrow(trim_data[which(subclass==2& trim_data$treat==1),])
n31 <- nrow(trim_data[which(subclass==3& trim_data$treat==1),])
n41 <- nrow(trim_data[which(subclass==4& trim_data$treat==1),])
n00 <- nrow(trim_data[which(subclass==0& trim_data$treat==0),])

n10 <- nrow(trim_data[which(subclass==1& trim_data$treat==0),])
n20 <- nrow(trim_data[which(subclass==2& trim_data$treat==0),])
n30 <- nrow(trim_data[which(subclass==3& trim_data$treat==0),])
n40 <- nrow(trim_data[which(subclass==4& trim_data$treat==0),])

varace <-
(n1)^2/nrow(trim_data)^2*((v11/n11)+(v10/n10))+(n2)^2/nrow(trim_data)^2*((v21
/n21)+(v20/n20))+(n3)^2/nrow(trim_data)^2*((v31/n31)+(v30/n30))+(n4)^2/nrow(t
rim_data)^2*((v41/n41)+(v40/n40))+(n0)^2/nrow(trim_data)^2*((v01/n01)+(v00/n0
0))
```

```
sdace<-sqrt(varace)

CIL=ace-sdace*2
CIU=ace+sdace*2

z = (ace-0)/sdace
pvale = pnorm(z, lower.tail = FALSE)
# Question 8
lm_model = lm(re78~treat + age + educ + black + hispan + married + nodegree +
re74 + re75, data=q2_data)
summary(lm_model)
confint(lm_model)
```