Homework 6

Xinyi Lin

10/19/2019

Question 1

```
##       Age          AgeGroup Race    Shots   InsuranceType MedAssist Location
##  Min.   :11.00   0:482    0:232   1:153   0:204         0:311     1:216
##  1st Qu.:13.00   1: 33    1:194   2:164   1:171         1:204     2:  0
##  Median :15.00            2: 29   3:198   2: 25                   3:  0
##  Mean   :14.92            3: 60           3:115                   4:299
##  3rd Qu.:17.00
##  Max.   :21.00
##  LocationType
##  0:216
##  1:299
##
##
##
##
```

```
##       Age          AgeGroup Race    Shots   InsuranceType MedAssist Location
##  Min.   :11.00   0:123    0:169   1:117   0: 12         0:353     1:365
##  1st Qu.:17.00   1:242    1:102   2:128   1:188         1: 12     2:  0
##  Median :19.00            2: 10   3:120   2:  9                   3:  0
##  Mean   :19.46            3: 84           3:156                   4:  0
##  3rd Qu.:22.00
##  Max.   :26.00
##  LocationType
##  0:365
##  1:  0
##
##
##
##
```

```
##       Age          AgeGroup Race    Shots   InsuranceType MedAssist Location
##  Min.   :11.00   0: 96    0:331   1:170   0: 59         0:474     1:217
##  1st Qu.:19.00   1:437    1:147   2:144   1:364         1: 59     2:165
##  Median :22.00            2: 13   3:219   2: 50                   3: 89
##  Mean   :21.43            3: 42           3: 60                   4: 62
##  3rd Qu.:24.00
##  Max.   :26.00
##  LocationType
##  0:382
##  1:151
##
##
##
##
```

Question 2

The treatment is received the vaccine at OB-GYN facilities and the control is received the vaccine at facilities other than OB-GYN. Among these variables, `Age`, `Race`, `InsuranceType`, `Location` are covariates. `Shots` contains some information about outcome, information in `AgeGroup` and `LocationType` overlap with `Age` and `Location`. There are NAs when using `MedAssist` to calculate propensity scores. So `Shots`, `MedAssist`, `AgeGroup` and `LocationType` are not regarded as covariates.

```
##       Age        AgeGroup Race    Shots   InsuranceType MedAssist Location
##  Min.   :11.00   0: 96    0:331   1:170   0: 59         0:474     1:217
##  1st Qu.:19.00   1:437    1:147   2:144   1:364         1: 59     2:165
##  Median :22.00            2: 13   3:219   2: 50                   3: 89
##  Mean   :21.43            3: 42           3: 60                   4: 62
##  3rd Qu.:24.00
##  Max.   :26.00
##  LocationType
##  0:382
##  1:151
##
##
##
##
```

```
##       Age       AgeGroup Race    Shots   InsuranceType MedAssist Location
##  Min.   :11.0   0:605    0:401   1:270   0:216         0:664     1:581
##  1st Qu.:14.0   1:275    1:296   2:292   1:359         1:216     2:  0
##  Median :16.0            2: 39   3:318   2: 34                   3:  0
##  Mean   :16.8            3:144           3:271                   4:299
##  3rd Qu.:19.0
##  Max.   :26.0
##  LocationType
##  0:581
##  1:299
##
##
##
##
```

Based on levels of baseline characteristics in treatment group and control group, we can get the eligibility criteria. The eligibility criteria are age between 11-26, Race equals to 0, 1, 2 or 3, InsuranceType equals to 0, 1, 2 or 3, Location equals to 1 or 4 and MedAssist equals to 0 or 1.

Question 3

According to the eligibility criteria, subjects with `LocationType` equals to 2 or 3 in treatment group 1, should be excluded. Descriptive statistics of analytic sample are shown below:

```
##       Age       AgeGroup Race    Shots   InsuranceType MedAssist Location
##  Min.   :11.0   0:605    0:401   1:270   0:216         0:664     1:581
##  1st Qu.:14.0   1:275    1:296   2:292   1:359         1:216     2:  0
##  Median :16.0            2: 39   3:318   2: 34                   3:  0
##  Mean   :16.8            3:144           3:271                   4:299
```

```
##   3rd Qu.:19.0
##   Max.   :26.0
##   LocationType
##   0:581
##   1:299
##
##
##
##
```

```
##       Age        AgeGroup Race    Shots   InsuranceType MedAssist Location
##   Min.   :11.00   0: 47    0:173   1: 92   0: 18         0:261     1:217
##   1st Qu.:19.00   1:232    1: 65   2: 74   1:190         1: 18     2:  0
##   Median :22.00            2:  4   3:113   2: 14                   3:  0
##   Mean   :21.59            3: 37           3: 57                   4: 62
##   3rd Qu.:24.00
##   Max.   :26.00
##   LocationType
##   0:217
##   1: 62
##
##
##
##
```

I redefine the PracticeType variable into binary variable PracticeType_bin. I let PracticeType equals to 2 be treatment group 1 and PracticeType equals to 0 or 1 be control group. Besides, I also exclude subjects with `LocationType` equals to 2 or 3 in treatment group 1. Descriptive statistics are different.

Numbers of subjects in different level of AgeGroup Race, Shots, Completed, InsuranceType, MedAssist, Location and LocationType are different. Compare to study sample, analytic sample do not have subjects with `Location` equals to 2 or 3, and mean, 1st and 3rd quantiles of age are different too.

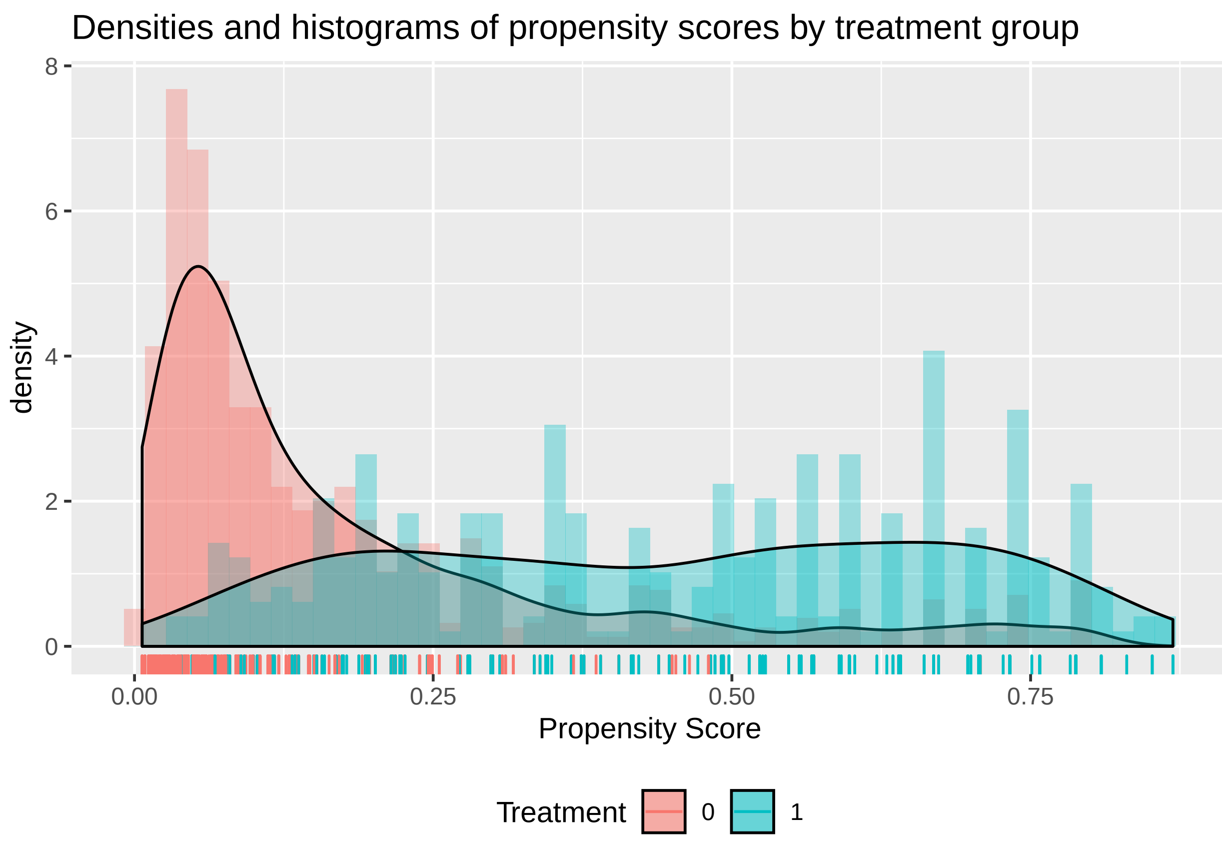
Question 4

```
##
## Call:
## glm(formula = PracticeType_bin ~ Age + Race + InsuranceType +
##     Location, family = binomial, data = q3_data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7610  -0.6181  -0.3636  -0.1536   2.5668
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -7.63403    0.59058 -12.926  < 2e-16 ***
## Age             0.30458    0.02359  12.910  < 2e-16 ***
## Race1          -0.42717    0.19278  -2.216  0.02671 *
## Race2          -1.19596    0.60269  -1.984  0.04721 *
## Race3          -0.64490    0.24149  -2.670  0.00757 **
## InsuranceType1  1.02725    0.32631   3.148  0.00164 **
## InsuranceType2  1.17159    0.45552   2.572  0.01011 *
```

```
## InsuranceType3  0.59322    0.37287   1.591  0.11162
## Location4       0.43685    0.24447   1.787  0.07395 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1279.34  on 1158  degrees of freedom
## Residual deviance:  953.17  on 1150  degrees of freedom
## AIC: 971.17
##
## Number of Fisher Scoring iterations: 5
```

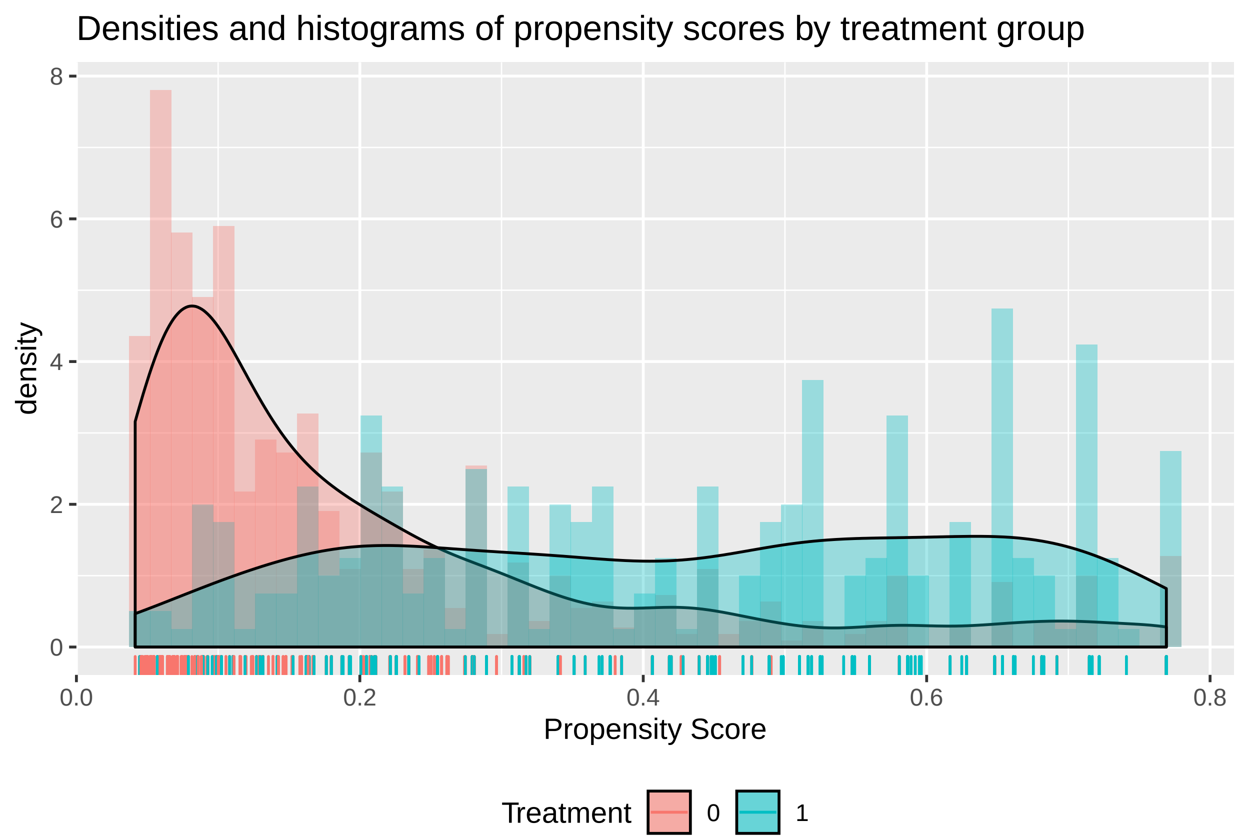
The propensity scores in the analytic sample is shown above.

Question 5



According to the plot, we can find that propensity score of treatment group 1 and treatment group 0 do not overlap when propensity socres are near 1, so we need to trim data.

There are 147 observations have been eliminated and there are 1012 observations left.



According to the plot above, we can find that propensity score of treatment group 1 and treatment group 0 overlap now, which means trimming can improve covariate balance, improving internal validity, so efficiency is improved. But trimming will hurts external validity(generalizability).

Question 6

```
##                       Stratified by PracticeType_bin
##                        0              1              SMD
##   n                      742            270
##   Age (mean (SD))      17.58 (3.53)   21.46 (3.27)    1.141
##   Race (%)                                            0.223
##      0                   371 (50.0)     164 (60.7)
##      1                   234 (31.5)      65 (24.1)
##      2                    18 ( 2.4)       4 ( 1.5)
##      3                   119 (16.0)      37 (13.7)
##   InsuranceType (%)                                   0.492
##      0                   130 (17.5)      18 ( 6.7)
##      1                   341 (46.0)     184 (68.1)
##      2                    34 ( 4.6)      11 ( 4.1)
##      3                   237 (31.9)      57 (21.1)
##   Location (%)                                        0.252
##      1                   516 (69.5)     217 (80.4)
##      2                     0 ( 0.0)       0 ( 0.0)
##      3                     0 ( 0.0)       0 ( 0.0)
##      4                   226 (30.5)      53 (19.6)
```

We want SMD to be small. According to the table above, we can find that SMD of Race and Location is close to 0.2, which mean these two covariates balance relatively well. While SMD of age and Location are relatively large, which means these two covariates do not balance well.

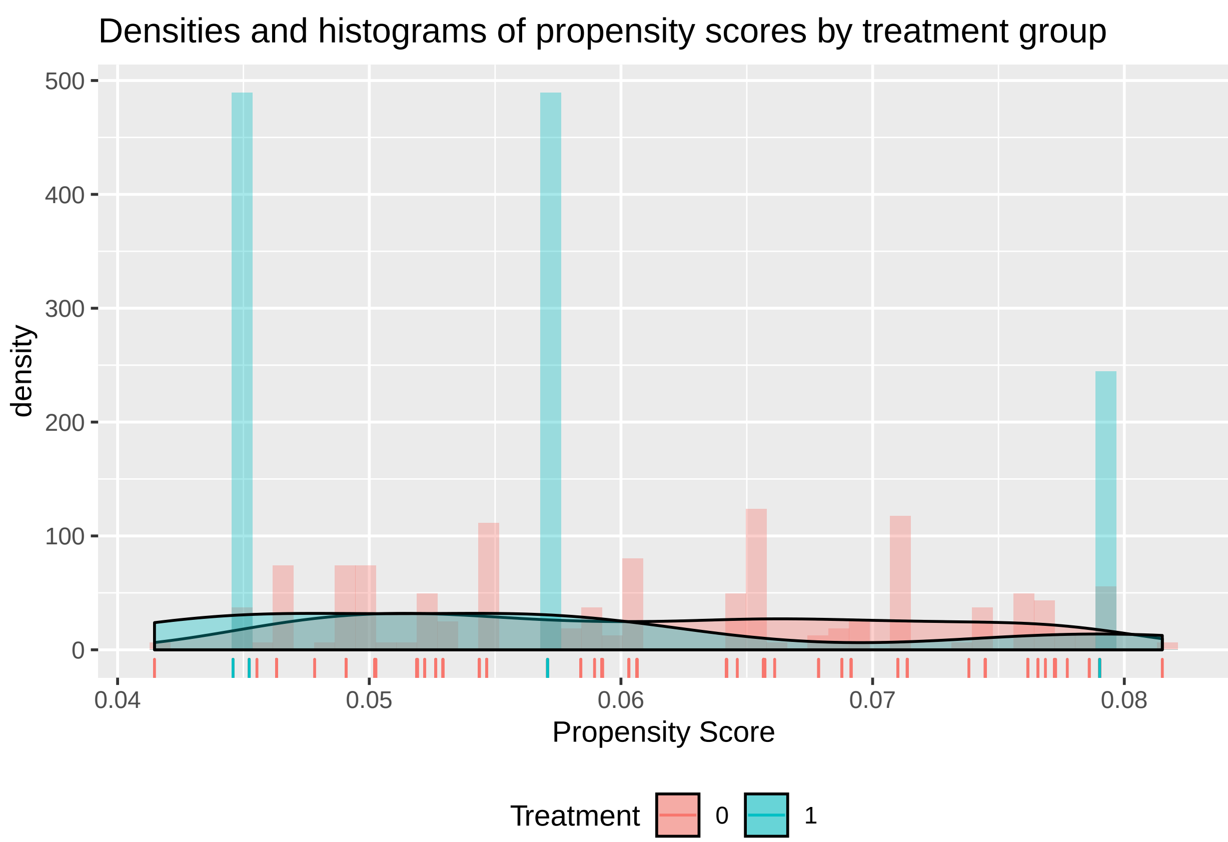
Question 7

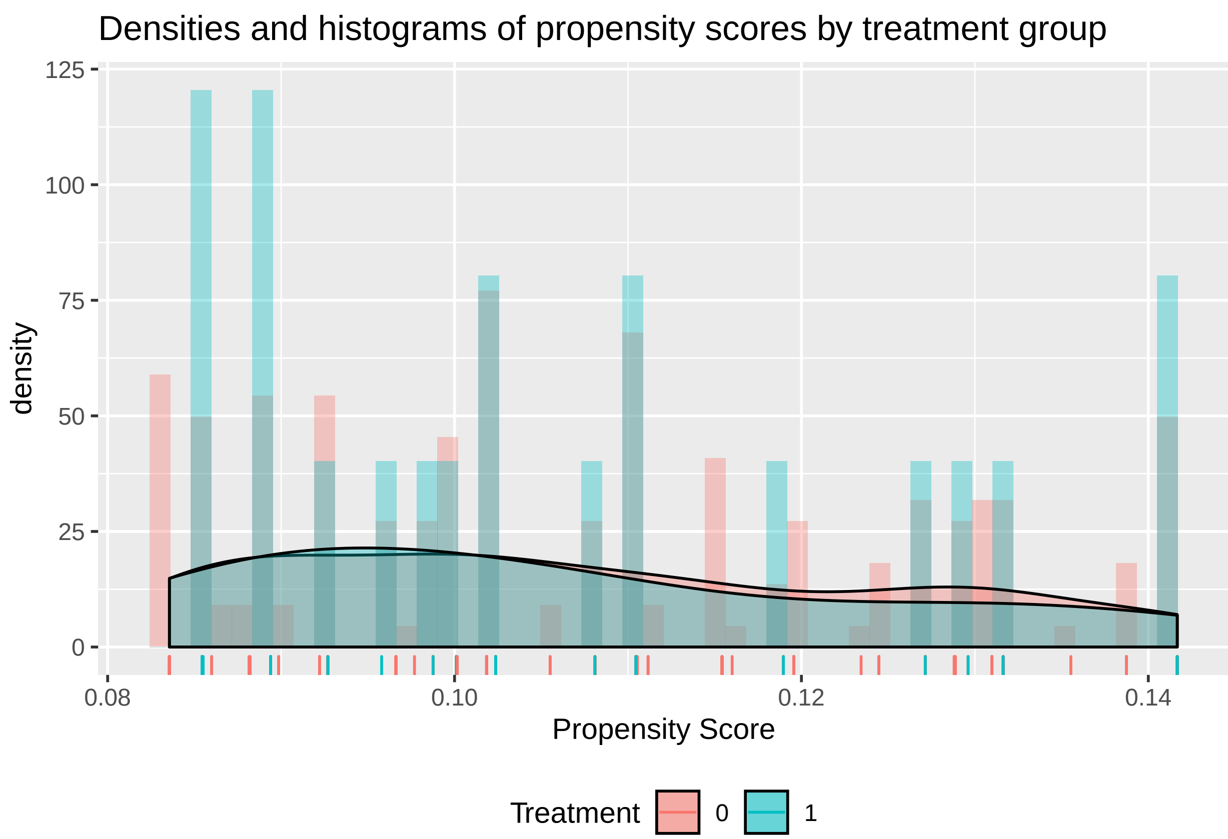
```
##        20%        40%        60%        80%
## 0.08192037 0.14166779 0.24136248 0.47664191
```

```
##    subclass
##       0   1   2   3   4
##   0 198 186 154 128  76
##   1   5  21  43  74 127
```

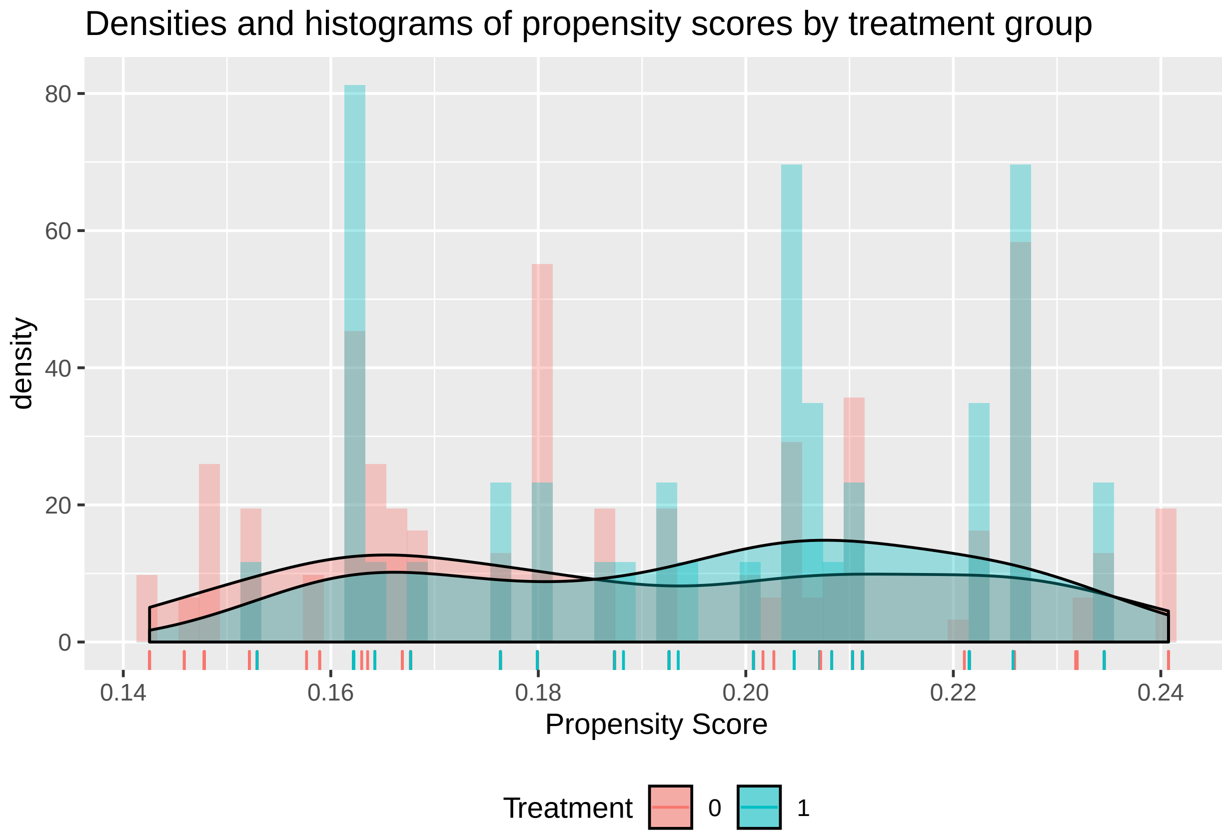
I choose 20%, 40%, 60% and 80% quantiles as breaks. Breaks are 0.082, 0.142, 0.241, 0.477. As these breaks do not violate positivity assumption, they are valid.

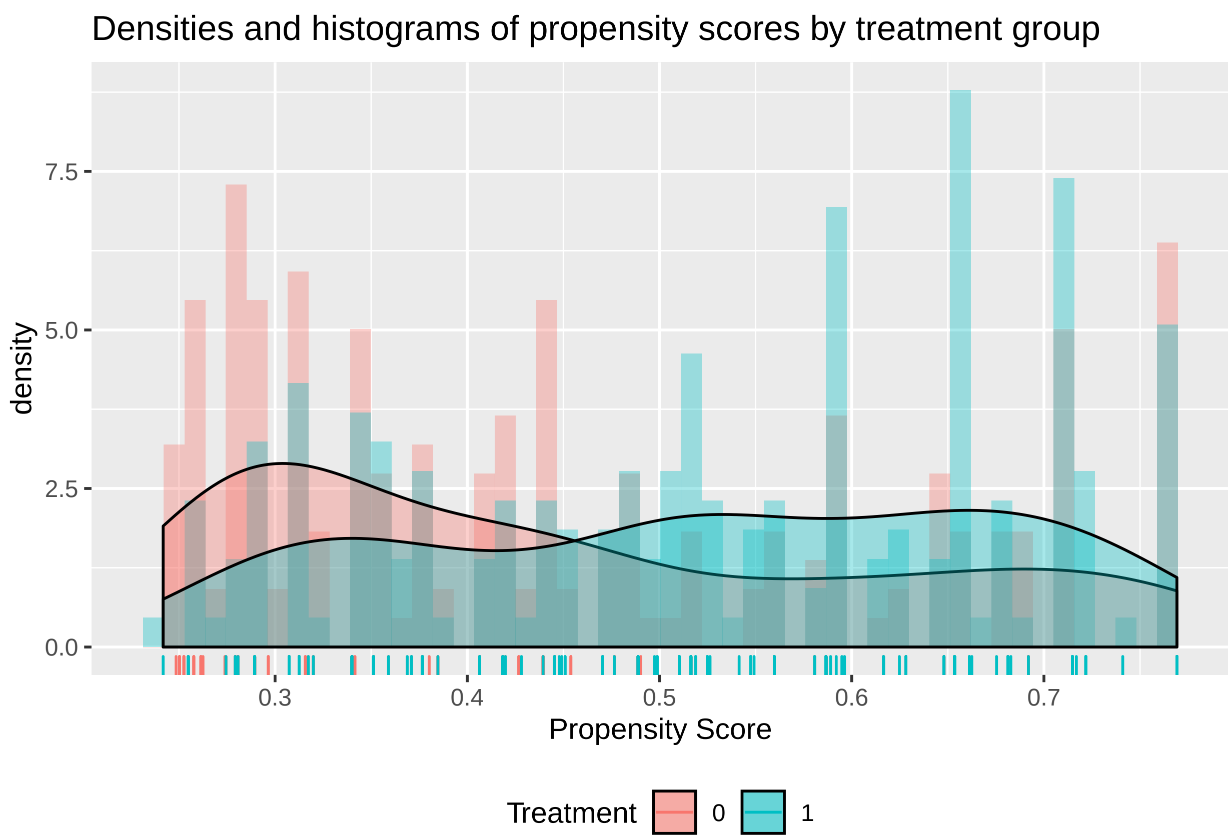
Plots of propensity scores with these breaks are shown below:





Densities and histograms of propensity scores by treatment group
density
125
100
75
50
25
0
0.08
0.10
0.12
0.14
Propensity Score
Treatment
0
1


Densities and histograms of propensity scores by treatment group
80
60
40
20
0
density
0.14
0.16
0.18
0.20
0.22
0.24
Propensity Score
Treatment
0
1



According to above plots, we can find that propensity score of treatment group and control group overlap better, which means covariates are balanced better.

```
##                     Stratified by PracticeType_bin
##                      0              1              SMD
##   n                    198             5
##   Age (mean (SD))    14.27 (1.45)  13.60 (1.52)   0.450
##   Race (%)                                         0.587
##      0                  84 (42.4)      2 (40.0)
##      1                  69 (34.8)      1 (20.0)
##      2                   8 ( 4.0)      0 ( 0.0)
##      3                  37 (18.7)      2 (40.0)
##   InsuranceType (%)                                0.741
##      0                  74 (37.4)      1 (20.0)
##      1                  53 (26.8)      3 (60.0)
##      2                   6 ( 3.0)      0 ( 0.0)
##      3                  65 (32.8)      1 (20.0)
##   Location (%)                                     0.553
##      1                 109 (55.1)      4 (80.0)
##      2                   0 ( 0.0)      0 ( 0.0)
##      3                   0 ( 0.0)      0 ( 0.0)
##      4                  89 (44.9)      1 (20.0)
```

```
##                     Stratified by PracticeType_bin
##                      0              1              SMD
##   n                    186            21
```

```
##   Age (mean (SD))   16.06 (1.52)  16.05 (1.24)   0.012
##   Race (%)                                       0.587
##      0                 85 (45.7)     11 (52.4)
##      1                 60 (32.3)      7 (33.3)
##      2                  3 ( 1.6)      2 ( 9.5)
##      3                 38 (20.4)      1 ( 4.8)
##   InsuranceType (%)                              0.063
##      0                 37 (19.9)      4 (19.0)
##      1                 66 (35.5)      7 (33.3)
##      2                  9 ( 4.8)      1 ( 4.8)
##      3                 74 (39.8)      9 (42.9)
##   Location (%)                                   0.092
##      1                125 (67.2)     15 (71.4)
##      2                  0 ( 0.0)      0 ( 0.0)
##      3                  0 ( 0.0)      0 ( 0.0)
##      4                 61 (32.8)      6 (28.6)

##                     Stratified by PracticeType_bin
##                      0             1             SMD
##   n                   154           43
##   Age (mean (SD))   17.83 (1.56)  18.26 (1.62)   0.267
##   Race (%)                                       0.216
##      0                 86 (55.8)     24 (55.8)
##      1                 48 (31.2)     13 (30.2)
##      2                  3 ( 1.9)      0 ( 0.0)
##      3                 17 (11.0)      6 (14.0)
##   InsuranceType (%)                              0.168
##      0                 16 (10.4)      5 (11.6)
##      1                 75 (48.7)     20 (46.5)
##      2                  8 ( 5.2)      1 ( 2.3)
##      3                 55 (35.7)     17 (39.5)
##   Location (%)                                   0.350
##      1                102 (66.2)     35 (81.4)
##      2                  0 ( 0.0)      0 ( 0.0)
##      3                  0 ( 0.0)      0 ( 0.0)
##      4                 52 (33.8)      8 (18.6)

##                     Stratified by PracticeType_bin
##                      0             1             SMD
##   n                   128           74
##   Age (mean (SD))   20.39 (1.75)  20.88 (1.76)   0.278
##   Race (%)                                       0.131
##      0                 68 (53.1)     36 (48.6)
##      1                 40 (31.2)     25 (33.8)
##      2                  3 ( 2.3)      1 ( 1.4)
##      3                 17 (13.3)     12 (16.2)
##   InsuranceType (%)                              0.195
##      0                  2 ( 1.6)      3 ( 4.1)
##      1                 79 (61.7)     47 (63.5)
##      2                 11 ( 8.6)      4 ( 5.4)
##      3                 36 (28.1)     20 (27.0)
##   Location (%)                                   0.143
##      1                104 (81.2)     64 (86.5)
##      2                  0 ( 0.0)      0 ( 0.0)
```

```
##      3                      0 ( 0.0)      0 ( 0.0)
##      4                     24 (18.8)     10 (13.5)
```

According to tables above, we can find that SMDs decrease a lot in each subclass. Except subclass 1, most of SMDs are smaller than or near 0.2, which means covariates are balanced well in these subclasses. SMDs in subclass 1 are relatively high, which means covariates are not balanced well.

Question 8

The point estimate of the marginal average causal effect is 0.065. The confidence interval is (-0.054, 0.184).

Interpretation:

As the point estimate of the marginal average causal effect is 0.065, the estimated true marginal average causal effect is 0.065.

As the confidence interval is (-0.054, 0.184), it means with 95% confidence, we can conclude that the true marginal average causal effect falls between -0.054 and 0.184.

Question 9

g-formula for observational studies:

$$\begin{aligned}
E[Y_1] - E[Y_0] &= \sum_C E(Y_1|C=c)Pr(C=c) - \sum_C E(Y_0|C=c)Pr(C=c)(IE) \\
&= \sum_C E(Y_1|A=1, C=c)Pr(C=c) - \sum_C E(Y_0|A=1, C=c)Pr(C=c)(RA) \\
&= \sum_{C,U} E(Y_1|A=1, C=c, U=u)Pr(C=c, U=u) - \sum_{C,U} E(Y_0|A=0, C=c, U=u)Pr(C=c, U=u)(IE+C \\
&= E(Y|A=1) - E(Y|A=0)
\end{aligned}$$

```
##
## Call:
## lm(formula = Completed ~ PracticeType_bin + Age + Race + InsuranceType +
##     Location, data = q8_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.5291 -0.3404 -0.2461  0.5851  0.9212
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       0.631520   0.100443   6.287 4.81e-10 ***
## PracticeType_bin  0.100831   0.036918   2.731 0.006421 **
## Age              -0.018211   0.004444  -4.098 4.51e-05 ***
## Race1            -0.114444   0.033860  -3.380 0.000753 ***
## Race2             0.037180   0.100115   0.371 0.710440
## Race3            -0.031418   0.042309  -0.743 0.457901
## InsuranceType1    0.069881   0.051355   1.361 0.173899
## InsuranceType2    0.181203   0.078531   2.307 0.021234 *
## InsuranceType3    0.049890   0.059672   0.836 0.403315
## Location4        -0.083875   0.043978  -1.907 0.056782 .
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.458 on 1002 degrees of freedom
## Multiple R-squared:  0.04086,    Adjusted R-squared:  0.03224
## F-statistic: 4.743 on 9 and 1002 DF,  p-value: 3.362e-06
```

The beta of `PracticeType_bin` is 0.1, which means given coavariates are the same, $E[Y|A=1]-E[Y|A=0]$ is 0.1. If 1)consistency, SUTVA, exchangeability and positivity assumptions are meet, 2)there is no interaction between treatment and covariates and 3)the outcome is continous, then this would be the same as the true marginal average causal effect which I estimate in question 8 as well as the average causa effect get from g-formula.

However, the estimated marginal average causal effect calculated in question 8 is not the same as what we get from linear model. Following are some possible reasons: 1) there is interaction between treatment and covariates, 2) using linear regression to fit model of binary outcome might cause bias, 3) the estimated marginal ACE is slightly different from true marginal ACE.

```r
knitr::opts_chunk$set(echo = FALSE)
library(tidyverse)
library(personalized)
library(tableone)
gard_data = read.table("./gardasil.dat",header = T) %>%
  select(-Completed)
gard_data$id = c(1:nrow(gard_data))
#head(gard_data)
q1_data = gard_data %>%
  mutate(AgeGroup = as.factor(AgeGroup),
         Race = as.factor(Race),
         Shots = as.factor(Shots),
         InsuranceType = as.factor(InsuranceType),
         MedAssist = as.factor(MedAssist),
         Location = as.factor(Location),
         LocationType = as.factor(LocationType))
summary(q1_data[q1_data$PracticeType==0,c(1:8)])
summary(q1_data[q1_data$PracticeType==1,c(1:8)])
summary(q1_data[q1_data$PracticeType==2,c(1:8)])
q2_data = q1_data %>%
  mutate(PracticeType_bin = ifelse(PracticeType==2,1,0)) %>%
  select(-PracticeType)
summary(q2_data[q2_data$PracticeType_bin==1,c(1:8)])
summary(q2_data[q2_data$PracticeType_bin==0,c(1:8)])
q3_data = q2_data %>%
  filter(Location %in% c(1,4))
summary(q3_data[q3_data$PracticeType_bin==0,c(1:8)])
summary(q3_data[q3_data$PracticeType_bin==1,c(1:8)])
ps.model<-glm(PracticeType_bin~Age + Race + InsuranceType + Location,data=q3_data, family = binomial)
summary(ps.model)
ps <- predict(ps.model, type="response")

x = q3_data

prop.func <- function(x, trt){
  # fit propensity score model
  propens.model <- glm(trt ~ Age + Race + InsuranceType + Location, data=x, family = binomial)
```

```r
  pi.x <- predict(propens.model, type = "response")
  pi.x
}

check.overlap(x = x,
              trt = q3_data$PracticeType_bin,
              type = "both",
              propensity.func = prop.func)
trim_data = x[ps>=min(ps[q3_data$PracticeType_bin==1]) & ps <= max(ps[q3_data$PracticeType_bin==0]),]
ps.model<-glm(PracticeType_bin ~ Age + Race + InsuranceType + Location, data=trim_data, family = binomia
ps <- predict(ps.model, type="response")

x = trim_data

prop.func <- function(x, trt){
  # fit propensity score model
  propens.model <- glm(trt~Age + Race + InsuranceType + Location, data=x, family = binomial)
  pi.x <- predict(propens.model, type = "response")
  pi.x
}

check.overlap(x = x,
              trt = trim_data$PracticeType_bin,
              type = "both",
              propensity.func = prop.func)
vars <- c("Age" , "Race", "InsuranceType" ,"Location")
## Construct a table
cov_bal <- CreateTableOne(vars = vars, strata = "PracticeType_bin", data = trim_data, test = FALSE)

## Show table with SMD
print(cov_bal, smd = TRUE)
#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$PracticeType_bin, subclass)
prop.func <- function(x, trt) {
  ps[which(ps <= subclass.breaks[1])]
}
trim_data$ps <-ps
check.overlap(x = trim_data[which(trim_data$ps <=subclass.breaks[1]),],
              trt = trim_data$PracticeType_bin[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$ps <-ps
check.overlap(x = trim_data[which(ps>subclass.breaks[1]&ps<=subclass.breaks[2]),],
              trt = trim_data$PracticeType_bin[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$ps <-ps
check.overlap(x = trim_data[which(ps>subclass.breaks[2]&ps<=subclass.breaks[3]),],
              trt = trim_data$PracticeType_bin[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$ps <-ps
 check.overlap(x = trim_data[which(ps>subclass.breaks[3]),],
               trt = trim_data$PracticeType_bin[which(ps>subclass.breaks[3])],
               type = "both",
               propensity.func = prop.func)
tabUnmatched_s0 <- CreateTableOne(vars = vars, strata = "PracticeType_bin", data = trim_data[which(subcl
tabUnmatched_s1 <- CreateTableOne(vars = vars, strata = "PracticeType_bin", data = trim_data[which(subcl
tabUnmatched_s2 <- CreateTableOne(vars = vars, strata = "PracticeType_bin", data = trim_data[which(subcl
tabUnmatched_s3 <- CreateTableOne(vars = vars, strata = "PracticeType_bin", data = trim_data[which(subcl

## Show table with SMD
print(tabUnmatched_s0, smd = TRUE)
print(tabUnmatched_s1, smd = TRUE)
print(tabUnmatched_s2, smd = TRUE)
print(tabUnmatched_s3, smd = TRUE)
Completed_data <- read.table("./gardasil.dat",header = T) %>%
  mutate(id = c(1:nrow(gard_data))) %>%
  select(Completed, id)
q8_data = merge(Completed_data, trim_data)
ACE0 <- mean(q8_data$Completed[which(subclass==0 & q8_data$PracticeType_bin==1)])-mean(q8_data$Complete
ACE1 <- mean(q8_data$Completed[which(subclass==1 & q8_data$PracticeType_bin==1)])-mean(q8_data$Complete
ACE2 <- mean(q8_data$Completed[which(subclass==2 & q8_data$PracticeType_bin==1)])-mean(q8_data$Complete
ACE3 <- mean(q8_data$Completed[which(subclass==3 & q8_data$PracticeType_bin==1)])-mean(q8_data$Complete
ACE4 <- mean(q8_data$Completed[which(subclass==4 & q8_data$PracticeType_bin==1)])-mean(q8_data$Complete
```

```
ace <- (nrow(q8_data[which(subclass==0),])/nrow(q8_data))*ACE0+
  (nrow(q8_data[which(subclass==1),])/nrow(q8_data))*ACE1+
  (nrow(q8_data[which(subclass==2),])/nrow(q8_data))*ACE2+
  (nrow(q8_data[which(subclass==3),])/nrow(q8_data))*ACE3+
  (nrow(q8_data[which(subclass==4),])/nrow(q8_data))*ACE4


v01 <- var(q8_data$Completed[which(subclass==0 & q8_data$PracticeType_bin==1)])
v00 <- var(q8_data$Completed[which(subclass==0 & q8_data$PracticeType_bin==0)])
v11 <- var(q8_data$Completed[which(subclass==1 & q8_data$PracticeType_bin==1)])
v10 <- var(q8_data$Completed[which(subclass==1 & q8_data$PracticeType_bin==0)])
v21 <- var(q8_data$Completed[which(subclass==2 & q8_data$PracticeType_bin==1)])
v20 <- var(q8_data$Completed[which(subclass==2 & q8_data$PracticeType_bin==0)])
v31 <- var(q8_data$Completed[which(subclass==3 & q8_data$PracticeType_bin==1)])
v30 <- var(q8_data$Completed[which(subclass==3 & q8_data$PracticeType_bin==0)])
v41 <- var(q8_data$Completed[which(subclass==4 & q8_data$PracticeType_bin==1)])
v40 <- var(q8_data$Completed[which(subclass==4 & q8_data$PracticeType_bin==0)])

n0 <- nrow(q8_data[which(subclass==0),])
n1 <- nrow(q8_data[which(subclass==1),])
n2 <- nrow(q8_data[which(subclass==2),])
n3 <- nrow(q8_data[which(subclass==3),])
n4 <- nrow(q8_data[which(subclass==4),])

n01 <- nrow(q8_data[which(subclass==0& q8_data$PracticeType_bin==1),])
n11 <- nrow(q8_data[which(subclass==1& q8_data$PracticeType_bin==1),])
n21 <- nrow(q8_data[which(subclass==2& q8_data$PracticeType_bin==1),])
n31 <- nrow(q8_data[which(subclass==3& q8_data$PracticeType_bin==1),])
n41 <- nrow(q8_data[which(subclass==4& q8_data$PracticeType_bin==1),])
n00 <- nrow(q8_data[which(subclass==0& q8_data$PracticeType_bin==0),])

n10 <- nrow(q8_data[which(subclass==1& q8_data$PracticeType_bin==0),])
n20 <- nrow(q8_data[which(subclass==2& q8_data$PracticeType_bin==0),])
n30 <- nrow(q8_data[which(subclass==3& q8_data$PracticeType_bin==0),])
n40 <- nrow(q8_data[which(subclass==4& q8_data$PracticeType_bin==0),])

varace <-(n1)^2/nrow(q8_data)^2*((v11/n11)+(v10/n10))+(n2)^2/nrow(q8_data)^2*((v21/n21)+(v20/n20))+(n3)

sdace<-sqrt(varace)

CIL=ace-sdace*2
CIU=ace+sdace*2
#glm.model <- glm(Completed~PracticeType_bin + Age + Race + InsuranceType + Location, data=q8_data, fam
lm.model = lm(Completed~PracticeType_bin + Age + Race + InsuranceType + Location, data=q8_data)
#summary(glm.model)
summary(lm.model)
```