Problem Set 3

psid <- read.csv("updated\_psid.csv", header = TRUE)  
attach(psid)  
library(MatchIt)

## Warning: package 'MatchIt' was built under R version 4.0.4

#### 1.

logit\_out <- glm(treat ~ age+education+black+hispanic+married+nodegree+re74+re75+u74+u75,  
 data = psid, family = binomial(logit))  
psid$propensity <- logit\_out$fitted.values

#### 2.

prop\_quan <- as.numeric(quantile(psid$propensity[psid$treat == 1], probs = c(0.2, 0.4, 0.6, 0.8)))  
for (i in (1:dim(psid)[1])){  
 if (psid$propensity[i] < prop\_quan[1]) psid$class[i] = 1  
 if (psid$propensity[i] >= prop\_quan[1] & psid$propensity[i] < prop\_quan[2]) psid$class[i] = 2  
 if (psid$propensity[i] >= prop\_quan[2] & psid$propensity[i] < prop\_quan[3]) psid$class[i] = 3  
 if (psid$propensity[i] >= prop\_quan[3] & psid$propensity[i] < prop\_quan[4]) psid$class[i] = 4  
 if (psid$propensity[i] >= prop\_quan[4]) psid$class[i] = 5  
}  
  
num\_tab <- table(psid$class, psid$treat)  
num\_tab

##   
## 0 1  
## 1 2430 37  
## 2 48 37  
## 3 4 35  
## 4 4 39  
## 5 4 37

mean\_gp <- aggregate(psid$re78, by = list(psid$class, psid$treat), mean)  
sd\_gp <- aggregate(psid$re78, by = list(psid$class, psid$treat), sd)  
  
catt <- mean\_gp[6:10, 3] - mean\_gp[1:5, 3]  
se\_class <- sqrt(sd\_gp[6:10, 3]^2/num\_tab[,2] + sd\_gp[1:5, 3]^2/num\_tab[,1])  
  
catt

## [1] -15.459173 -1.403460 6.233968 3.579682 3.886068

se\_class

## 1 2 3 4 5   
## 1.2113652 1.5828597 2.0071621 2.0392749 0.9691829

#### 3.

att <- sum(catt\*num\_tab[,2]/sum(num\_tab[,2]))  
se <- sqrt(sum(se\_class^2\*(num\_tab[,2]/sum(num\_tab[,2]))^2))

ATT = -0.6612782

SE = 0.7249132

#### 4.

m\_out <- matchit(treat ~ age+education+black+hispanic+married+nodegree+re74+re75+u74+u75,  
 data = psid, method = "subclass", subclass = 5)  
  
psid\_pscore <- match.data(m\_out)  
  
num\_tab <- table(psid\_pscore$subclass, psid\_pscore$treat)  
  
mean\_gp <- aggregate(psid\_pscore$re78, by = list(psid\_pscore$subclass, psid\_pscore$treat), mean)  
  
catt <- mean\_gp[6:10, 3] - mean\_gp[1:5, 3]  
  
att <- sum(catt\*num\_tab[,2]/sum(num\_tab[,2]))

ATT = -0.6612782

#### 5.

Overall average effect on treated =

(2.2 - 15.6)\*(22)/112 + (5.1 - 6.3)\*(23)/112 + (2.2 - 5.3)\*(22)/112 + (7.5 - 16.8)\*(22)/112 + (2.3 - 5.1)\*(23)/112

= -5.88929

Standard error =

sqrt((2.9)^2 \*((22)/112)^2 + (3.9)^2 \*((23)/112)^2 + (3.1)^2 \*((22)/112)^2 + (2.8)^2 \*((22)/112)^2 + (3.1)^2 \*((23)/112)^2)

= 1.429855

#### 6.

DID estimate = (4.55 - 4.75) - (4.85 - 4.80) = -0.25

#### 7.

**i)**

= 4.75 - 4.80 = -0.05

**ii)**

= 4.85 - 4.80 = 0.05

**iii)**

= (4.55 - 4.75) - 0.05 = -0.25