HW3: The Effect of Court-Ordered Hiring Quotas on the Composition of Police

In McCrary's 2007 AER paper, he argued that the series of court-ordered racial hiring quotas imposed on municipal police departments resulted in an increase of 14-percentage-points in the fraction of African American new hires at these police departments. But, the evidence on police performance after the change in the composition of police is mixed.

In this HW, we investigate the first research question, and analyze the relationship between the court-orders and the percentage of African American police officers. Because the data for this AER paper is confidential, this homework uses a simulated dataset prepared by an RA for this Signature Course (Xiaochen Sun). To simplify the settings from the original paper, we set 1970 as the year for the court-order, and provide the outcome of interest at two time points for each city: 1960 (pre) and 1990 (post).

The names and descriptions of variables in the data file `HW3_Data` are

Name	Description
`city_no`	Index of city
`order`	Factor variable indicating whether a city has the court-ordered hiring quotas (equal to "Quota") or not (equal to "No Quota")
`police_size`	Number of sworn officers (in logarithm format)
`pop`	Population of the city (Unit is in 100K)
`prop_black`	Proportion of African Americans in city population
`pre_prop_pol_blk`	Proportion of African Americans among police officers in 1960
`post_prop_pol_blk`	Proportion of African Americans among police officers in 1990
`pcnt`	Number of police precincts
`south`	Factor variable indicating whether a city is in the south (equal to "In South") or not (equal to "Not in South")
`region`	Census region (factor variable equal to "Region (#)")

Functions you may find helpful

- table()
- proportions()
- group_by()
- summarize()
- summarize_at()
- quantile()
- mean()
- hist()
- boxplot()

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
# ## Load the data and reformat a few variables
quotas <- read.csv("data3/HW3_Data.csv", header = TRUE) %>%
  mutate(order = factor(recode(order,
                               `0` = "No Quota",
                               `1` = "Quota")),
         south = factor(recode(south,
                               `0` = "Not in South",
                               `1` = "In South")),
         region = factor(recode(region,
                               `1` = "Region 1",
                               `2` = "Region 2",
                               `3` = "Region 3",
                               `4` = "Region 4",
                               `5` = "Region 5")))

# restructure the shape of the data
pre_prop_police_blk <-quotas$prop_police_blk[quotas$year== 1960]
post_prop_police_blk <- quotas$prop_police_blk[quotas$year== 1990]

quotas_w <- quotas[quotas$year == 1960,]
quotas_w$pre_prop_pol_blk <- pre_prop_police_blk
quotas_w$post_prop_pol_blk <- post_prop_police_blk

quotas <- quotas_w[, c(1, 3:6, 8:12)]
rm(pre_prop_police_blk)
rm(post_prop_police_blk)
rm(quotas_w)
```

Question 1 [10 pts]

Assume a post-only cross sectional study design for this question - so focus on the level of the outcome in the post period rather than the change over time in the outcome.

(a) With this study design, what is the specific causal question of interest for this study? What are the potential outcomes for a single police department?

(b) For cities experiencing the intervention what is the average factual outcome? For cities experiencing the intervention what is the average missing counterfactual outcome?

(c) Is this study a randomized experiment or an observational study? Explain why in a sentence or two.

(d) How would you estimate the average missing counterfactual for the cities experiencing the intervention with a post-only cross sectional study design? What assumption would need to hold for this to be an unbiased estimate?

Answer 1 Text

a. what the impact of court ordered hiring quotas relative to not having hiring quotas on the average percentage of African American police hires by police departments in american cities is?

potential outcomes #1: What the impact of court ordered hiring quotas would be on the percentage of african american police officers for a police precinct ?

potential outcome #2 : what the impact of no court ordered hiring quotas would be on the percentage of african american police officers for a police precinct?

b. what the average percentage of african american police officers hired for a police precinct is for precincts that have a court ordered racial quotas.

what the average percentage of African American police officer hires would have been if the police precincts that received a court ordered hiring quota instead didn't have an imposed court order and all else remains the same.

c. it is an observational study because the quota and no quota impositions on the precincts are not randomised and it is in the cities jurisdiction to decide whether a city needs intervention or not

d.the average missing counterfactual : the average MCF(what the average proportion of African American police officer hires would have been for police precincts who had racial hiring quotas imposed instead had no racial hiring quotas imposed)is estimated by the average proportion of African Americans police officer hires in precincts that had no court ordered quota imposition.

for this to be unbiased it would have to be the case that no other factor related to the average proportion in African american hires would be different systematically between the cities/ precincts where there was a court order versus no court order. There should be no significant differences int the baseline covariates that can directly bias the mcf.

Question 2 [8 pts]

In our version of the study design, the court-ordered hiring quotas were imposed on the treated cities in 1970.

(a) Calculate the average proportion of African American police officers in the post-period (1990) in cities with and without court-ordered hiring quotas. Also calculate the corresponding standard deviations.

(b) Do cities with the court-ordered hiring quotas have higher, lower, or similar average proportions of African American police officers, compared to the untreated cities in the post period?

(c) Compare the standard deviations of these proportions in each group in the post period - is the spread in these proportions similar? Create a side-by-side boxplot to show the distributions of the proportions of African American police officers in the intervention and non-intervention cities in the post period. Describe what you learn from this figure.

Answer 2 Code

```r
# Your code here :)
quotas_no <- quotas %>% filter(order=="No Quota")
quotas_yes <- quotas %>% filter(order=="Quota")
tapply(quotas$post_prop_pol_blk,quotas$order,summary)
```

```
## $`No Quota`
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 0.0000015 0.0185260 0.1049548 0.1462037 0.2158238 0.8251107
##
## $Quota
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.01578 0.12276 0.19468 0.19949 0.26378 0.56533
```

```r
sd(quotas_no$post_prop_pol_blk,na.rm=FALSE)
```

```
## [1] 0.1590933
```

```r
sd(quotas_yes$post_prop_pol_blk,na.rm=FALSE)
```

```
## [1] 0.1051909
```

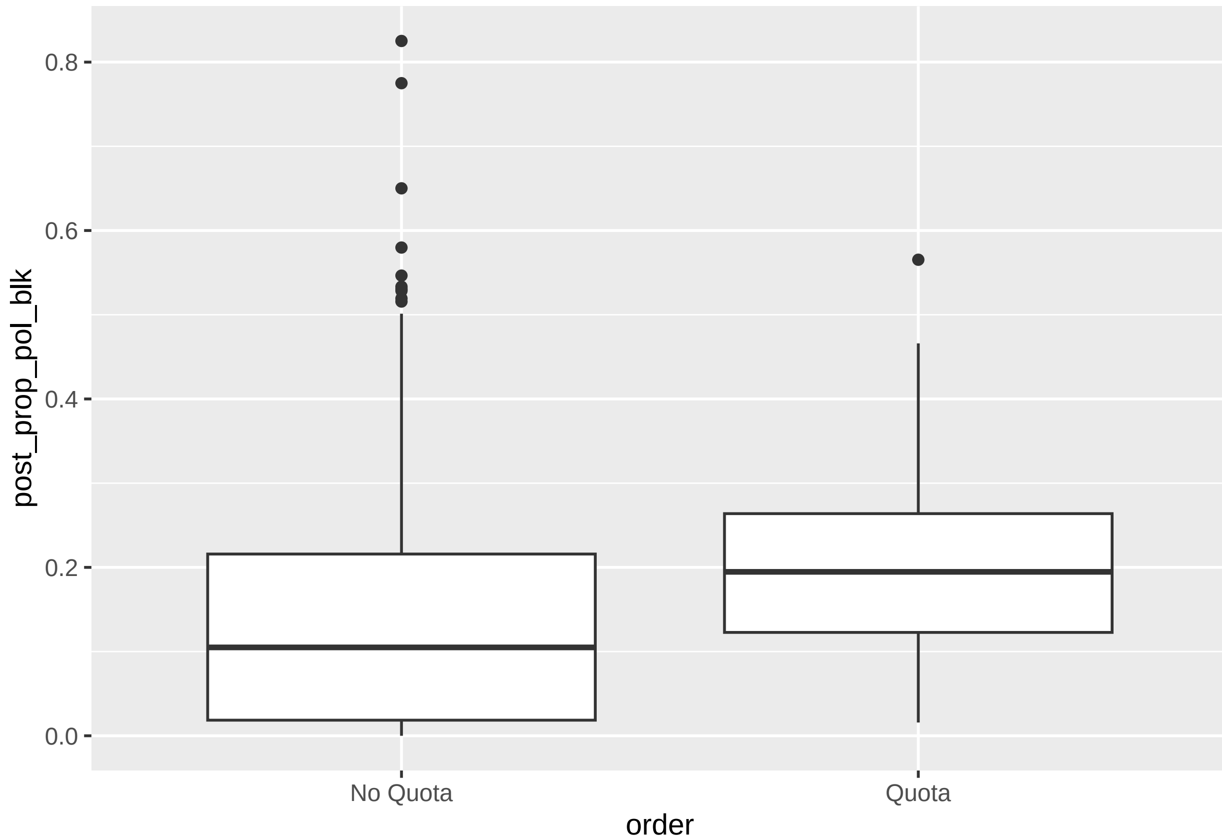
```r
summary(quotas_yes$post_prop_pol_blk)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.01578 0.12276 0.19468 0.19949 0.26378 0.56533
```

```r
summary(quotas_no$post_prop_pol_blk)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 0.0000015 0.0185260 0.1049548 0.1462037 0.2158238 0.8251107
```

```
quotas %>%
  ggplot(aes(y =post_prop_pol_blk, x =order))+
  geom_boxplot()
```



Answer 2 Text

a. the average proportion of african americans in the post period(after 1990) "Quota" is : 19.94 percent. the average proportion of african americans in the post period (after 1990)"no quota" : 14.62 percent the standard deviation : 0.159 or 15.9% for when the no quotas are imposed the standard deviation :0.1051 or 10.51% for when there are quotas imposed.

b. the cities that went through the intervention: QUOTAS-> these precincts have higher proportions of African American police officers compared to the precincts that did not have the QUOTA imposition by 5.33%.

c. The cities that went through the intervention: (Quotas) had a standard deviation of 0.105 and the ones that didn't go through the intervention: (No Quotas) had a standard deviation of 0.159.The difference between the standard deviations between the two is :0.054 or 5.4%. the spread of data for intervention group : 50 percent of the data lies between 12.2% - 26.3% and the spread of data and middle 50 percent of data for alternative group lies : 1.8 % - 21.5%. the spread of data is slightly different between the intervention group and alternative group.

the box plot for No Quotas intervention group has more outliers and the quotas alternative group has less outliers.The proportion for no quotas is significantly higher in the Quotas group as compared to the group that had no quotas.

Question 3 [12 pts]

For this question, consider just cities with court-ordered hiring quotas.

(a) Describe the distributions of the proportion of African American police officers before (1960: pre) and after (1990: post) the imposition of the quotas. Conduct the comparison in terms of the mean, median, and quartiles. Also create a figure to show the difference (or lack of a difference) in those two distributions. Calculate the average treatment effect. Explain your results in words.

(b) What study design is this? How was the average MCF estimated? What assumption would have to hold for this estimate of the MCF to be unbiased? With this study design, name a confounder you think is likely to bias this estimate of the causal effect and state why you think it is a confounder (what two conditions make a feature be a confounder? Why would those potentially occur here)?.

Answer 3 Code

```
# Your code here :)

mean(quotas_yes$pre_prop_pol_blk)
```

```
## [1] 0.04938772
```

```
mean(quotas_yes$post_prop_pol_blk)
```

```
## [1] 0.1994853
```

```
median1 <- median(quotas_yes$pre_prop_pol_blk)
median2 <- median(quotas_yes$post_prop_pol_blk)
median_diff <- median2 - median1

summary(quotas_yes$post_prop_pol_blk)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.01578 0.12276 0.19468 0.19949 0.26378 0.56533
```

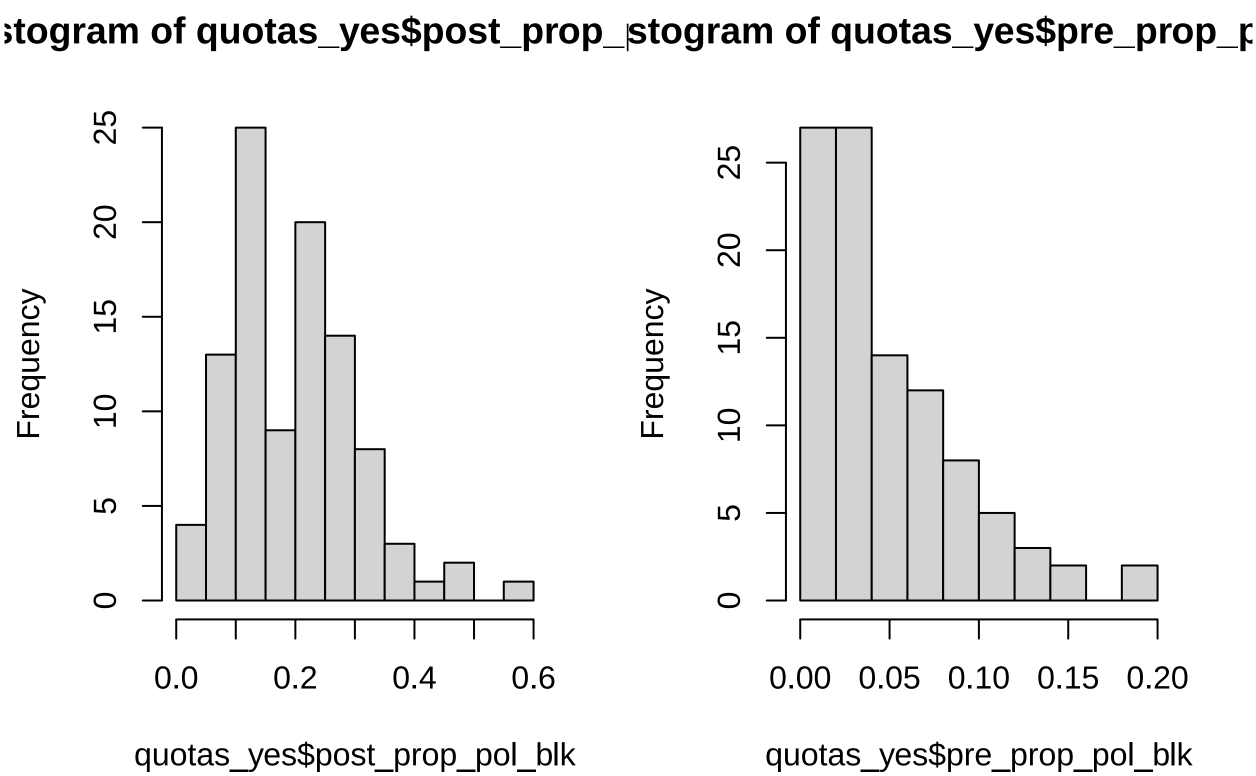
```
summary(quotas_yes$pre_prop_pol_blk)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 0.0006154 0.0186315 0.0349142 0.0493877 0.0697692 0.1899384
```

```
summary_plot <- summary(quotas_yes$post_prop_pol_blk)- summary(quotas_yes$pre_prop_pol_blk)
summary_plot
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.01517 0.10413 0.15976 0.15010 0.19402 0.37539
```

```
par(mfrow=c(1,2))
hist(quotas_yes$post_prop_pol_blk)
hist(quotas_yes$pre_prop_pol_blk)
```



Answer 3 Text

a. the difference in means between the post and the pre treatments is 15 %. The difference in medians is 15.6. the interquartile range differences between pre and post are : 1 st quart : 10.41% 3rd quart: 19.4% the interquartile range for post: 12.2 % - 26.37% the middle 50 percent of the data lies here the interquartile range for pre: 1.86% - 6.9% the middle 50 percent of the data lies here average treatment effect for pre-post: the Average Treatment Effect is the average change in outcomes of the treatment in such studies.The average treatment effect is the difference in the post treatment proportion variable and the pre treatment proportion variable for the individuals in the treatment effect which is calculated to be 15.01% . the treatment effect would mean the increase in the number of African american police officers after the quota mandated.

the proportion of the post variable has two modals and is slightly symmetrical, most of the values lie between 0.00-0.5. the proportion of the pre variable has 1 modal and is not symmetrical and has a bell shaped curve. both distributions are right skewed

b. This is the prepost study design. The average MCF (what the average proportion of African American police officers would have been for precincts who had the court ordered hiring quotas instead had no court order hiring quotas and all else remained the same) is estimated by the proportion of African American police officer hires in these same precincts before the quotas were imposed.(preintervention of the intervention group)

for this estimate to be unbiased, we would assume that nothing would change and that the outcome would stay the same. 1 it would have to be the case that no other factor causes any change of the proportion of African American Police hires in precincts over time. 2 no other changes over time and no changes under the intervention pre and post. 3. the time 1970 (pre quota) which is the year that the pre was measured in can be quite the biasing covariate as the post study(post quota) involves a different year and this could bias the mcf. population can also be a biasing factor, in 1970 the population rates can be different than the post study for the estimate of the mcf.

the 2 confounders could be the year in which the study was conducted. between 1970(pre) and 1990(post).this could be a confounder in that the times could have changed. There could be more diversity in the years after 1970 which would bias the mcf. it could also confound due to the duration of time between the two studies and the nature of the states.for the variables to be confounders: 1. they need to be associated with the outcome and need to systematiclly differ between the intervention and the comparison groups, this can bias the estimation of the mcf. 2 they can also be confounders if they drastically change during the study period (culture, political situation, etc)

Question 4 [7 pts]

(a) Using a difference-in-differences study design, what is the the specific causal question? How is the average MCF estimated in a DiD design? How is the DiD treatment estimate calculated?

(b) Now calculate the difference-in-differences estimate. Give a brief interpretation of this result.

Answer 4 Code

```
# Your code here :)
mean_noquota <- mean(quotas_no$post_prop_pol_blk) - mean(quotas_no$pre_prop_pol_blk)
mean_noquota
```

```
## [1] 0.1150565
```

```
mean_yesquota <-mean(quotas_yes$post_prop_pol_blk)-mean(quotas_yes$pre_prop_pol_blk)
mean_yesquota
```

```
## [1] 0.1500976
```

```
mean_yesquota - mean_noquota
```

```
## [1] 0.03504113
```

```
table(mean_noquota,mean_yesquota)
```

```
##                  mean_yesquota
## mean_noquota      0.150097603
##   0.115056476065            1
```

Answer 4 Text

a. what is the impact of having a court order quota relative to not having a court ordered quota on the change in average proportions of african american police officers(pre- post) for police precincts in cities?

the average change is measured as the: what the average change in proportions of African American police officers (pre to post) would have been if precincts that have the court ordered racial hiring quotas instead had no court ordered racial hiring and all else remains the same.

the DID MCF is calculated by: it is the average change in proportions of African american police officers in precincts that had no court ordered racial hiring quota in cities.

the treatment effect can be calculated by subtracting the change in outcome for the intervention group and the change in outcome of the alternative group. this would be: subtracting the change in the proportions(pre-post) of African american police officers in the QUOTA mandated precincts and the Change in proportions(pre- post) of African American police officers in precincts with no quotas.

b. the mean difference between both the alternative and intervention group is: 3.5%. It is significantly low and the intervention group has a higher mean post intervention(pre post) compared to the mean post alternative treatment(prepost). The difference in the no quota group(pre post - 1970 and 1990) is 11.5% and the different in the intervention QUOTA(prepost) imposition on precincts is 15%.

Question 5 [7 pts]

(a) Compare the distribution of region for the intervention group and the non-intervention group.

(b) What are the possible implications of any differences in these distributions of region for whether the estimate of the average MCF in Q2 may be biased?

(c) What are the possible implications of any differences in these distributions of region for whether the estimate of the average MCF in Q4 may be biased?

Answer 5 Code

```
# Your code here :)
table(quotas_no$region)
```

```
##
## Region 1 Region 2 Region 3 Region 4 Region 5
##       40       43       35       41       41
```

```
table(quotas_yes$region)
```

```
##
## Region 1 Region 2 Region 3 Region 4 Region 5
##       19       18       13       32       18
```

```
summary(quotas_no$region)
```

```
## Region 1 Region 2 Region 3 Region 4 Region 5
##       40       43       35       41       41
```

```
proportions(table(quotas_no$region))
```

```
##
## Region 1 Region 2 Region 3 Region 4 Region 5
##    0.200    0.215    0.175    0.205    0.205
```

```
proportions(table(quotas_yes$region))
```

```
##
## Region 1 Region 2 Region 3 Region 4 Region 5
##     0.19     0.18     0.13     0.32     0.18
```

Answer 5 Text

a. the distributions : NO QUOTA imposition -> region 1 = 20% , region 2 = 21.5 % ,region 3 = 17.5% ,region 4=20.5% , region 5= 20.5% the ditributions : QUOTA imposition -> region 1= 19%, region 2 =18%, region 3= 13%, region 4 = 32%, region 5 =18%

b. the MCF can be biased if the percentage of police precincts are significantly different from the treatment and control groups. We can observe that the police percentages between the groups are not significantly higher or lower than each other except for region 4. in region 4 the there is a 11.5 percentage point difference between the region 4 in quota and no quota precincts. the region is a baseline covariate and it can bias the output as the percentage in one region is not the same as the other and this can cause a potential bias in the mcf as the baseline covariates are not balanced. some of the baseline covariates between the treated and untreated precincts differ and this specific "region" covariate is a baseline covariate.if it is related to the proportion of african american hires in treated cities they would be confounders.

c. the mcf can be biased if the percentage change in the covariates is presnet it can bias the change in proportions of african americans and the mcf in DID studies and that except for the result of the intervention , the changes of the time would be the exact same.

Question 6 [6 pts]

(a) Compare the means and standard deviations of three different baseline covariates: police_size, pop, and prop_black between the intervention and non-intervention police departments.

(b) What are the possible implications of any differences in these means and standard deviations for whether the estimate of the average MCF in Q2 may be biased?

(c) What are the possible implications of any differences in these means and standard deviations for whether the estimate of the average MCF in Q4 may be biased?

Answer 6 Code

```r
# Your code here :)
mean(quotas_yes$prop_black)
```

```
## [1] 0.2053316
```

```r
sd(quotas_yes$prop_black)
```

```
## [1] 0.1200983
```

```r
mean(quotas_no$prop_black)
```

```
## [1] 0.1657188
```

```r
sd(quotas_no$prop_black)
```

```
## [1] 0.1410092
```

```r
mean(quotas_yes$police_size)
```

```
## [1] 5.368433
```

```r
sd(quotas_yes$police_size)
```

```
## [1] 0.3773262
```

```r
mean(quotas_no$police_size)
```

```
## [1] 5.031502
```

```r
sd(quotas_no$police_size)
```

```
## [1] 0.3819848
```

```r
mean(quotas_yes$pop)
```

```
## [1] 4.396901
```

```r
sd(quotas_yes$pop)
```

```
## [1] 8.570782
```

```r
mean(quotas_no$pop)
```

```
## [1] 1.427951
```

```r
sd(quotas_no$pop)
```

```
## [1] 2.532853
```

Answer 6 Text

a. means and standard deviations for intervention: police size: 5.36 and 0.377 pop: 4.39 and 8.57 prop_black: 0.20 and 0.12

means and standard deviations for alternative: police size: 5.03 and 0.38 pop: 1.42 and 2.53 prop_black: 0.16 and 0.14

the difference in means for police size : 0.33 the difference in means for pop : 2.97 the difference in means for prop_black: 0.04

b. any bias in these baseline covariates such as the proportion of african americans between both the groups can cause a bias in the outcome of the mcf. while all the baselines seem balanced, only the population for both intervention and control groups is slightly different with 2.95%. a difference in the baseline covariates can cause a significant bias on the mcf for q2. As the differences in the covariates such as the pop adn prop black have a significant impact on the outcome of interest.

c. any bias in these baseline covariates such as the proportion of african americans between both the groups can cause a bias in the outcome of mcf. a bias in the covariates can also alter the average mcf in the DID study: it can alter the average MCF.if the baseline covariates have an impact on the level of mcf and outcome , they can also bias the change in average mcf.