

R Camp 2019 HW3

Alex Pugh

8/6/2019

Problem 1

Unilaterally taking a dispute to court is not generally viewed as a friendly act between states. I am interested in the impact of a unilateral appeal to the ICJ on a population's perception of another state. For a case to go to the court, there must already exist some sort of dispute. Presumably, the presence of a dispute sours the perception of the other state in the mind of the population. Does the unilateral appeal by State A against State B have an additional negative impact on the population of State B's perception of State A? Or, does a unilateral appeal to the court make the population of State B perceive State A as being more willing to settle the dispute and find a peaceful solution?

To explore this question, I will simulate an experiment. I will make State B the United States, so the population of my experiment would be the adult population of the United States. Relevant characteristics of the population that I will simulate are age, gender, and political affiliation. I will also consider the state-level percentage of bachelor's degree holder. I expect that younger people, women, Democrats, and people in states with higher educational attainment will generally have more positive perceptions of the other state than others because they may be more open to the international community and have fewer nationalist tendencies. They may be more likely to recognize that the US having a dispute with another state does not make the other state a "bad" state or an enemy.

In my experiment, I would inform people that it is 2025 and the US and Canada are in a dispute, potentially over the maritime boundary in the north east given changes in fish migration patterns. I would provide a few examples of the dispute in action and the tensions between the countries. Those in the control would then be asked to rank on a scale ranging from 0 to 10 how favorable they view Canada. Those in the treatment group would be told that Canada had filed a unilateral application to the International Court of Justice and that the issue would go before the court. Then they would be asked a few question and will get a perception score on a scale ranging from 0 to 10 how favorable they view Canada. I expect that the treatment will have a negative effect on their perception and resulting score.

First I will create the population data for the covariates. For computational ease, I have set the adult population of the US to 100 million, with approximately 50% of women. Once we have the individual attributes, we can create a perception score (DV) for if the individual is in the treatment group and if the individual is not in the treatment group. Once we sample and assign individuals to the treatment group, we can then use the appropriate perception score as the observed score.

```
#Adult population of the US, set as 100 million
simN <- 100000000

#set seed
set.seed(111222)
#generate gender variable
female <- rbinom(n=simN, size=1, prob=0.50) #50 percent men and women

head(female)
```

```
## [1] 0 1 1 0 1 1
```

```
#set seed
set.seed(111222)
```

```
#create age variable ranging from 18+
#use normal distribution and add
age <- rpois(n=simN, lambda=38)
summary(age)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##         8      34      38      38      42      79
```

```
#add 10 to all values under 18
age <- ifelse(age < 18, age+10, age)
summary(age)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##        18      34      38      38      42      79
```

```
#set seed
set.seed(111222)
#create political affiliation
p1 <- rep(NA, simN)
p2 <- ifelse(age <= 35 & runif(simN)<0.65, 1, p1) #make 65% of those under 35 democrat
p3 <- ifelse(age >30 & female ==1 & runif(simN)<0.60, 1, p2) # make 60% of women over 30 Democrat
p4 <- ifelse(age >30 & female ==0 & runif(simN)<0.45, 1, p3) # make 45% of men over 30 Democrat
democrat <- ifelse(is.na(p4)==T, 0, 1)
head(democrat)
```

```
## [1] 1 1 0 0 0 1
```

```
#set seed
set.seed(111222)
# create state level % bachelor's degree holders
edu <- rnorm(n=50, mean=27, sd=1.5)
edu <- round(edu, digits = 3)
summary(edu)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    23.20  26.28  26.88  26.87  27.75  29.99
```

```
#set seed
set.seed(111222)
state.edu <- sample(edu, size = simN, replace = TRUE)
#hist(state.edu)
#summary(state.edu)
#make state.edu a decimal
state.edu <- state.edu/100

# create subject number
respondent <- seq(from=1, simN, by=1)

# simulate treated and untreated outcome variable
# set coefficients
a <- 6
```

```

b.age      <- -0.08
b.female   <- 1.1
b.democrat <- 0.7
b.state.edu <- 1.0
tau <- -1.2 #set treatment effect

# set seed and create error term
set.seed(111)
error <- rnorm(n=simN, mean=0, sd=0.17)
summary(error)

```

```

##      Min.    1st Qu.    Median      Mean   3rd Qu.      Max.
## -0.9343730 -0.1146759 -0.0000332 -0.0000355  0.1146269  0.9944444

```

```

#create dataframe
sim.dat <- data.frame(respondent, age, female, democrat, state.edu)

```

```

#create untreated and treated outcome
sim.dat$Y_0 <- a + tau*0 + b.age*age + b.female*female + b.democrat*democrat + b.state.edu*state.edu + error
sim.dat$Y_1 <- a + tau*1 + b.age*age + b.female*female + b.democrat*democrat + b.state.edu*state.edu + error

```

```

#check range and distribution of treated and untreated outcomes
summary(sim.dat$Y_0)

```

```

##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  1.221  4.590   5.298   5.212  5.851   8.290

```

```

summary(sim.dat$Y_1)

```

```

##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## 0.02056 3.39014 4.09826 4.01224 4.65146 7.09024

```

```

#check data
head(sim.dat, n=4)

```

```

##  respondent age female democrat state.edu      Y_0      Y_1
## 1         1  35      0         1  0.24521 5.185198 3.985198
## 2         2  32      1         1  0.28917 6.472945 5.272945
## 3         3  29      1         0  0.28394 6.010964 4.810964
## 4         4  41      0         0  0.23204 3.560641 2.360641

```

```

unique(sim.dat$state.edu) #check to ensure 50 values

```

```

## [1] 0.24521 0.28917 0.28394 0.23204 0.24893 0.26318 0.26925 0.27347
## [9] 0.24652 0.25569 0.27507 0.27942 0.27266 0.26625 0.27774 0.27636
## [17] 0.26660 0.26579 0.25096 0.28007 0.26552 0.28187 0.27477 0.28570
## [25] 0.26730 0.27398 0.26502 0.27174 0.28390 0.27932 0.27372 0.26841
## [33] 0.28779 0.27157 0.25391 0.26211 0.26523 0.26638 0.27280 0.26431
## [41] 0.24645 0.27667 0.24090 0.26134 0.28692 0.26397 0.28141 0.29990
## [49] 0.26268 0.26236

```

```
# DONE WITH POPULATION DATA
```

Next, we will randomly sample 50 individuals from the population. Then we will randomly assign them to the treatment and control group and explore the difference in means between the two groups.

```
#set seed
set.seed(111222)
#sample size
N_samp <- 50
#get sample of 50 names
index <- sample(1:nrow(sim.dat), size = N_samp, replace =FALSE)
index
```

```
## [1] 52031496 41234198 75526188 34979192 60015520 59154793 751508
## [8] 34135490 24389872 31005628 49647273 18471597 71760201 66784057
## [15] 47726505 37090006 18546968 76988891 96467445 32431977 2356190
## [22] 78012274 34717409 364945 36323607 46886223 64093747 44626016
## [29] 31687642 1181287 87159350 48157009 62485559 35889750 70151734
## [36] 17699689 44844860 26097391 48829162 23032364 58755481 39431276
## [43] 49294032 16986708 80803546 9374942 92041205 31143139 2542383
## [50] 73579119
```

```
samp.exp <- sim.dat[index,]
```

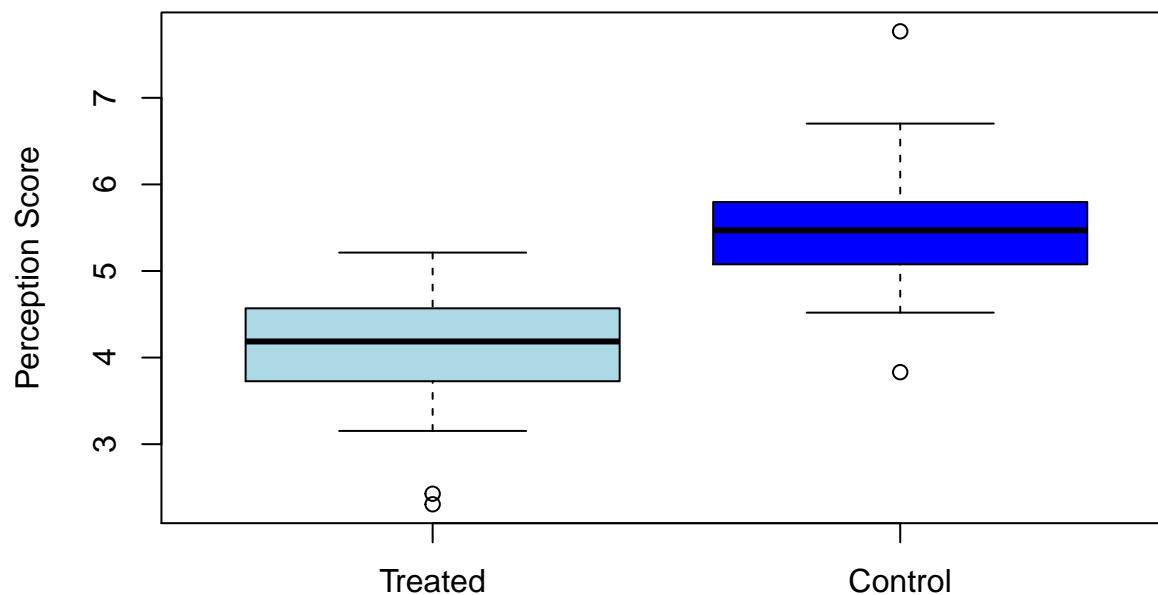
```
#create binary treatment variable
#set seed
set.seed(111)
samp.exp$d <- rbinom(n=nrow(samp.exp), size=1, prob=0.5)
#look at distribution
table(samp.exp$d)
```

```
##
## 0 1
## 26 24
```

```
#create observed outcome
samp.exp$Y_obs <- samp.exp$d*samp.exp$Y_1 + (1-samp.exp$d)*samp.exp$Y_0 # Observed outcomes
# Y_obs = d*Y_1 + (1-d)*Y_0
samp.exp$Status <- ifelse(samp.exp$d==1, "Treated", "Control") # Copying the treatment status into

#Visualize with boxplot
boxplot(samp.exp$Y_obs[samp.exp$Status=="Treated"],
        samp.exp$Y_obs[samp.exp$Status=="Control"],
        names=c("Treated", "Control"), ylab="Perception Score",
        main="Distribution of Perception Scores\nN = 50",
        col=c("light blue", "blue"))
```

Distribution of Perception Scores N = 50



```
#Create plot of mean perception for each group
t <- c(0,1)
y_mean <- c(mean(samp.exp$Y_obs[samp.exp$Status=="Control"]),
             mean(samp.exp$Y_obs[samp.exp$Status=="Treated"]))
y_sdv <- c(sd(samp.exp$Y_obs[samp.exp$Status=="Control"]),
            sd(samp.exp$Y_obs[samp.exp$Status=="Treated"]))

plot(y_mean ~ t, pch=19, col="red", ylim=range(c(y_mean-y_sdv, y_mean+y_sdv)),
     xlab="Treatment Status", ylab="Y_obs ± sd", xaxt="n") # Without x-axis label
axis(1, at = seq(0, 1, by = 1), las=1) # If las=2, numbers will be flipped by 90 degree
arrows(t, y_mean-y_sdv, t, y_mean+y_sdv, length=0, angle=90, lwd=3)
title("Simulated Result")
```

Simulated Result



```
#perform a T-test (difference in means test)
```

```
t.test(Y_obs ~ Status, data=samp.exp)
```

```
##
```

```
## Welch Two Sample t-test
```

```
##
```

```
## data: Y_obs by Status
```

```
## t = 6.929, df = 47.985, p-value = 9.462e-09
```

```
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
```

```
## 1.062859 1.931863
```

```
## sample estimates:
```

```
## mean in group Control mean in group Treated
```

```
## 5.555876 4.058515
```

```
# Estimating the average treatment effect
```

```
tau_hat <- mean(samp.exp$Y_obs[samp.exp$Status=="Treated"]) -  
            mean(samp.exp$Y_obs[samp.exp$Status=="Control"])
```

```
tau_hat
```

```
## [1] -1.497361
```

```
# Population level treatment (fixed=constant) effect
```

```
tau_true <- mean(sim.dat$Y_1 - sim.dat$Y_0) # Mean Difference of Individual Potential Outcomes
```

```
tau_true
```

```
## [1] -1.2
```

```
tau_true == tau
```

```
## [1] FALSE
```

With the sample of 50 individuals, I was not able to perfectly recover the treatment effects. Instead, I over-estimated the treatment effect. Below are graphs of the distribution of the covariates across the treatment and control groups. The control group had more women and more democrats, which likely inflated the estimate of the treatment effect as the coefficients for female and democrat were positive.

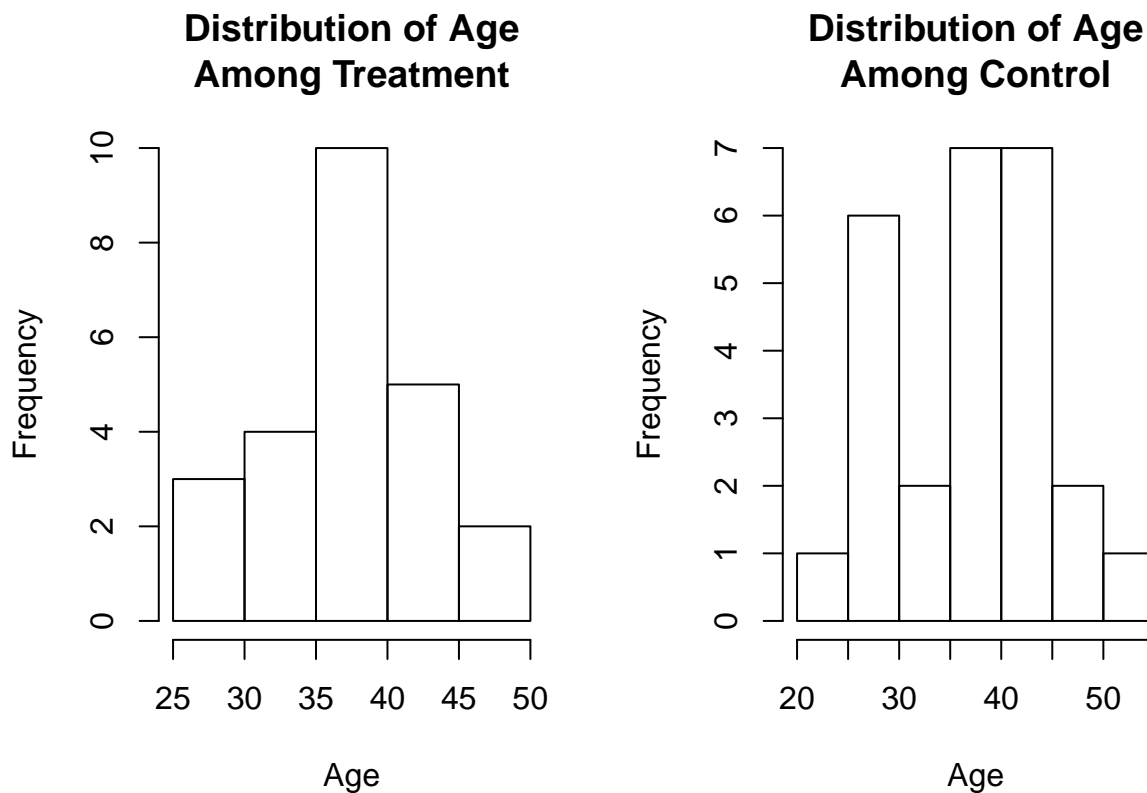
```
#Compare covariates between treatment and control
```

```
#age
```

```
par(mfrow=c(1,2))
```

```
hist(samp.exp$age[samp.exp$Status=="Treated"], breaks = 6, main = "Distribution of Age\nAmong Treatment",
```

```
hist(samp.exp$age[samp.exp$Status=="Control"], breaks = 6, main = "Distribution of Age\nAmong Control",
```



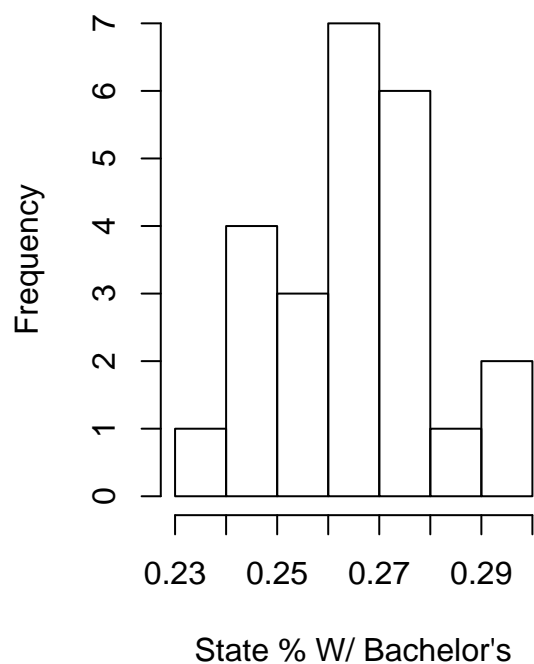
```
#state.edu
```

```
par(mfrow=c(1,2))
```

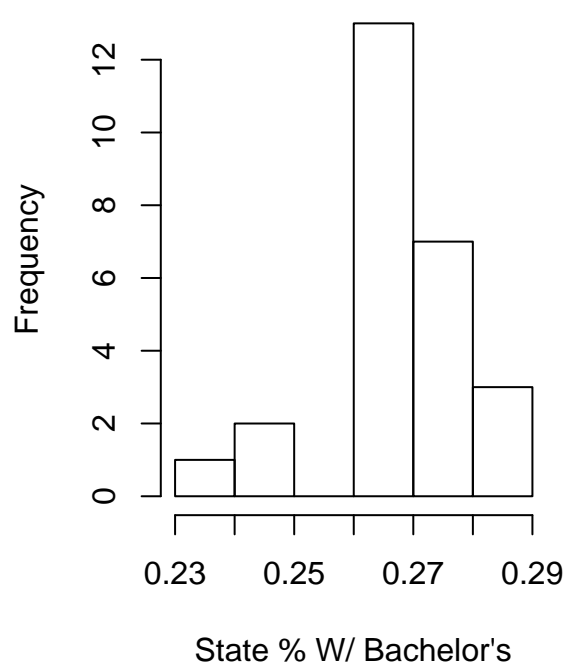
```
hist(samp.exp$state.edu[samp.exp$Status=="Treated"], breaks = 6, main = "Distribution of State Education
```

```
hist(samp.exp$state.edu[samp.exp$Status=="Control"], breaks = 6, main = "Distribution of State Education
```

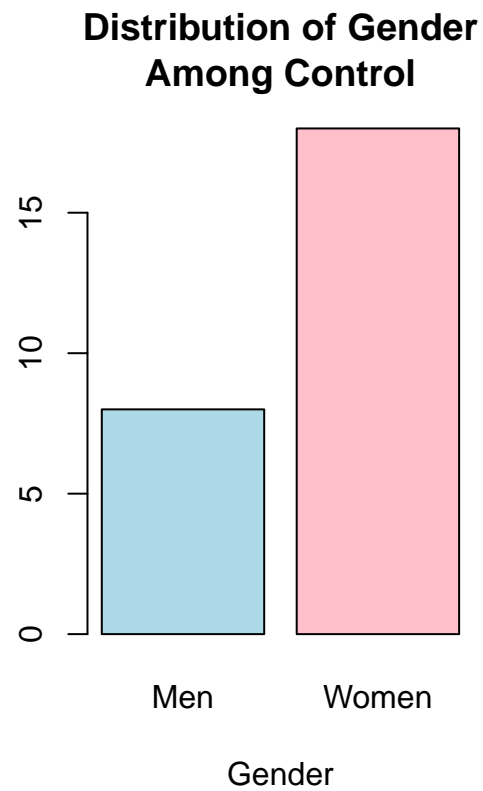
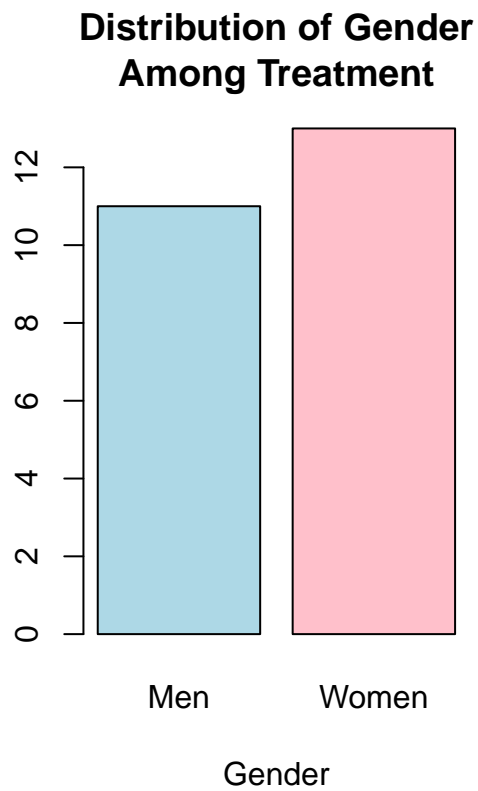
**Distribution of State Education
Among Treatment**



**Distribution of State Education
Among Control**

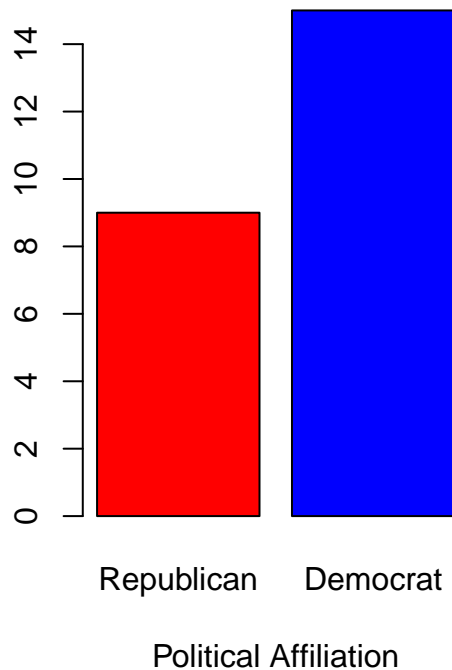


```
#female
par(mfrow=c(1,2))
barplot(table(samp.exp$female[samp.exp$Status=="Treated"]), main = "Distribution of Gender\nAmong Treated")
barplot(table(samp.exp$female[samp.exp$Status=="Control"]), main = "Distribution of Gender\nAmong Control")
```

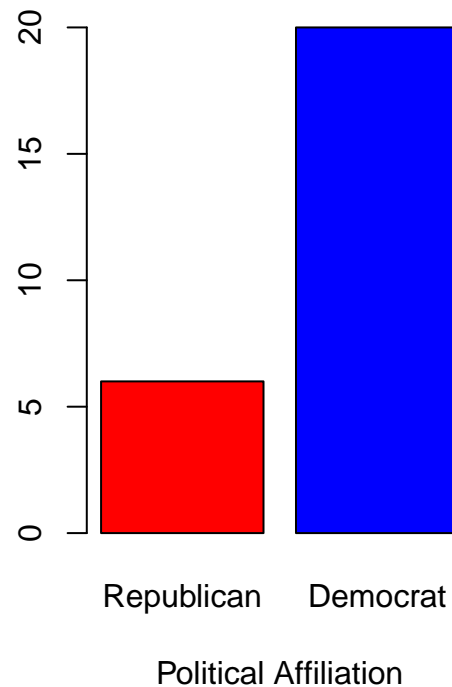



```
#democrat
par(mfrow=c(1,2))
barplot(table(samp.exp$democrat[samp.exp$Status=="Treated"]), main = "Distribution of Political Affilia
        col = c("red", "blue"))
barplot(table(samp.exp$democrat[samp.exp$Status=="Control"]), main = "Distribution of Political Affilia
        col = c("red", "blue"))
```

**Distribution of Political Affiliation
Among Treatment**



**Distribution of Political Affiliation
Among Control**



Problem 2

Next, we repeat the same simulation, changing the number of samples from 50, 100, 150, 200 to 250, and create a graph that combines five plots for the difference in distributions. From these plots, we can see that there is consistently a negative treatment effect.

```
#create vector of sample sizes
sam <- c(50,100,150,200,250)
#create parameters for graphs
par(mfrow=c(3,2))

#create empty vector for treatment effects
t1 <- rep(NA, 5)
#create vector for boolean TF if tau est is == tau true
t2 <- rep(NA, 5)

#create object for i in t1
loop <- 1

#Loop through sample sizes and combine boxplots in one graph
for(i in sam){
  #set seed
  set.seed(111)
  #sample size
  N_samp <- i
  #get sample of names
  index <- sample(1:nrow(sim.dat), size = N_samp, replace =FALSE)
```

```

samp.exp <- sim.dat[index,]

#create binary treatment variable
#set seed
set.seed(111)
samp.exp$d <- rbinom(n=nrow(samp.exp), size=1, prob=0.5)
#look at distribution
table(samp.exp$d)

#create observed outcome
samp.exp$Y_obs <- samp.exp$d*samp.exp$Y_1 + (1-samp.exp$d)*samp.exp$Y_0 # Observed outcomes
# Y_obs = d*Y_1 + (1-d)*Y_0
samp.exp$Status <- ifelse(samp.exp$d==1, "Treated", "Control") # Copying the treatment status into

#Visualize with boxplot
boxplot(samp.exp$Y_obs[samp.exp$Status=="Treated"],
        samp.exp$Y_obs[samp.exp$Status=="Control"],
        names=c("Treated", "Control"), ylab="Perception Score",
        main=paste0("Distribution of Perception Scores\nN =", sep=" ", i),
        col=c("light blue", "blue"))

#perform a T-test (difference in means test)
t.test(Y_obs ~ Status, data=samp.exp)

# Estimating the average treatment effect
tau_hat <- mean(samp.exp$Y_obs[samp.exp$Status=="Treated"]) -
           mean(samp.exp$Y_obs[samp.exp$Status=="Control"])
t1[loop]<- tau_hat

# Population level treatment (fixed=constant) effect
tau_true <- mean(sim.dat$Y_1 - sim.dat$Y_0) # Mean Difference of Individual Potential Outcomes
tau_true
t2[loop] <- tau_true == tau

#add one to loop
loop <- loop + 1
}
#View Boolean
t2

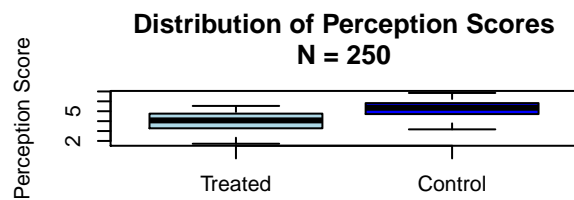
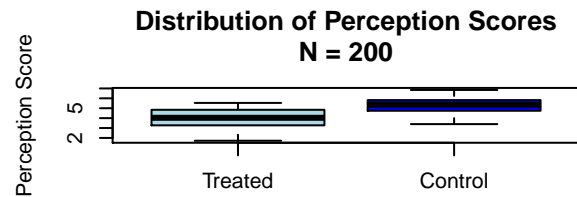
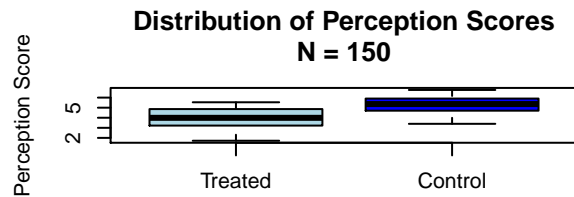
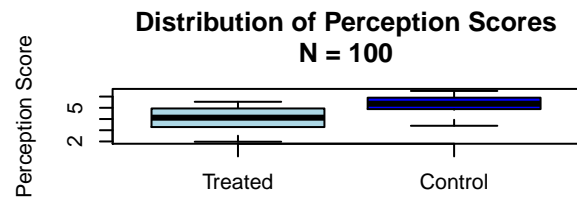
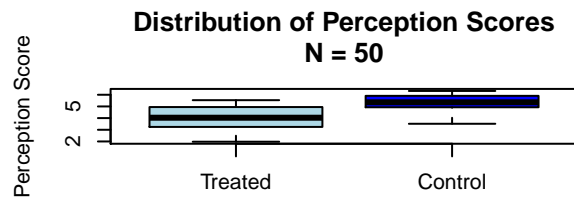
```

```
## [1] FALSE FALSE FALSE FALSE FALSE
```

```

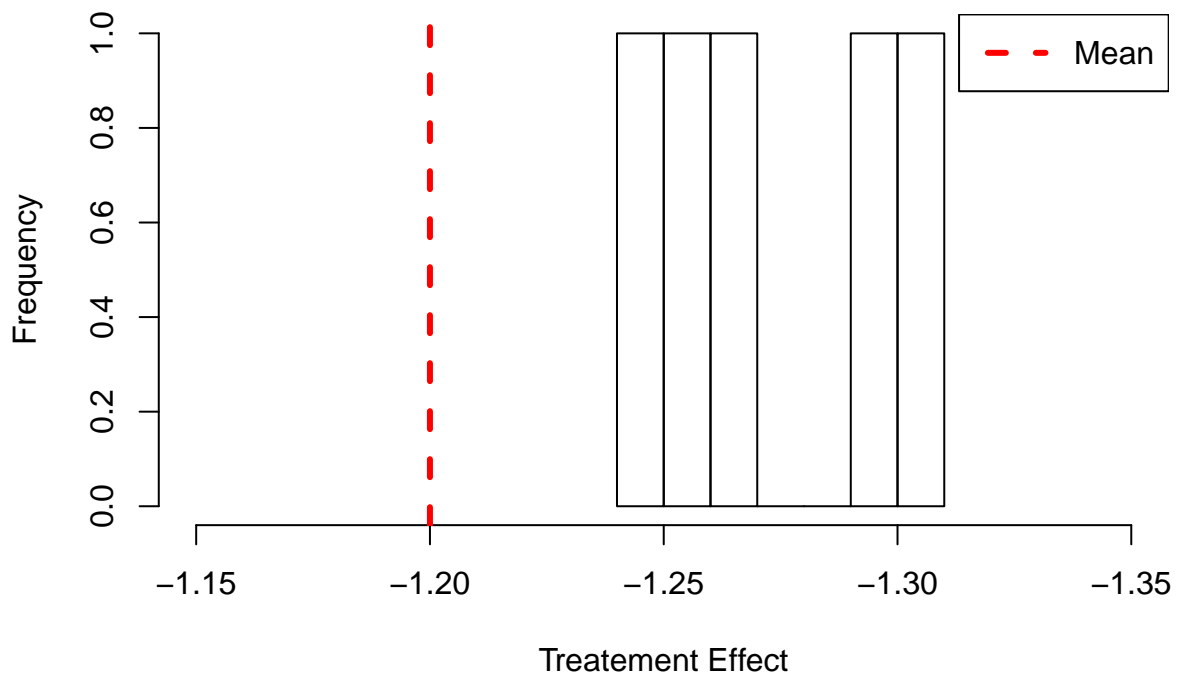
#plot
par(mfrow=c(1,1))

```



```
hist(t1, xlim = c(-1.15,-1.35), main = "Distribution of Estimated Treatment Effect",
     xlab = "Treatement Effect")
abline(v = tau, col="red", lwd=3, lty = 2)
legend("topright", legend="Mean" , lty = 2, lwd = 3, col = "red")
```

Distribution of Estimated Treatment Effect



Even with increasing the sample size, I was not able to accurately recover the treatment effect. My estimated treatment effects were consistently larger in magnitude than the -1.2 true population treatment effect. As will previously, this was likely due to differences in distribution of individuals between the treatment and control, probably with more women/democrats in the control than in the treatment.