

Homework - Day 3: Simulation II (Experimental Data)

Social Analysis and Simulation in R

Contents

Question 1	1
Research Question, Theoretical Interests, and Data	2
Simulation	2
Problem 2	13

Yui Nishimura
23 August 2019

Question 1

Instruction: In this question, you are asked to simulate an experimental data you would want to analyze if you have all the resources and tools to perform a set of experiments. First, briefly describe a research question that you are interested in, but have not empirically examined or collected any data about. In so doing, consider an experiment with a continuous outcome variable and a binary treatment variable in which some number of units is sampled from a clearly defined population. Consider also other relevant characteristics of individuals in the population such as age, gender, race, and others that you want to take into consideration in your experiment. Next, simulate N data points for these covariates and define them as the population data. Here, choose the most realistic value for N according to your theoretical interest. For example, N can be as large as the U.S. population (about 328,915,700), may be as moderate as students at Rice (about 6700), or could be as small as legislators in Argentine Chamber of Deputies (257). Create also at least one aggregate level covariate (e.g., city level median income).

Now perform any type of sampling, including but not limited to simple random sampling and stratified sampling, on the simulated population data, and draw 50 data points as your initial sample. These are your subjects. Then create a binary treatment variable which indicates units' treatment status and randomly assign a value to it (0 for the control group and 1 for the treatment group). And simulate the value of the continuous outcome variable as a realization of a random variable following a normal distribution (with a positive variance) whose expected value is a function of all the covariates and the treatment variable. Here set the parameters according to your theoretical expectation and make the most realistic guess about them.

Given the sample data, visualize the difference in the values of the continuous outcome variables for the control and treatment groups at least two ways. Also visualize the differences in the values of covariates for the two groups in your favorite approach. Finally, add the true "treatment effect" (i.e., the parameter you set in the simulation) to the graphs and comment on how much your experiments recover the "treatment effect(s)."

Research Question, Theoretical Interests, and Data

My research question is whether the recognition of discrimination changes attitudes towards the majority rights. Almost all countries historically experienced to govern with the political or economic classes and gave different political and social privileges. Global human rights movements changed the normative recognition about discrimination, and in today's society, such a treatment is recognized as unequal policy or categorized as discrimination in some contexts. However, there is a controversy about education of discrimination. While some argues that it is important to share and educate the history of discrimination to prevent further unfortunate situations for the discriminated group of people, others consider that such a recognition can create a foundation of discriminatory attitudes, since they could have a chance to acknowledge their difference ("in-group" and "out-group"). However, there is limited empirical work to assess whether and how the recognition creates negative or positive effects on issues addressing the discrimination.

My theoretical expectation is that the information about discrimination against the marginalized population creates positive impact on the general citizens' recognition about the majority rights. The logic behind it is that the discrimination happens usually against the group of people belonging to the minority, based on a specific categorization. If people recognize the categorization can be changed and the new one is applicable to them, they might change their attitudes towards the majority rights.

Variables of interest, population, and relevant characteristics of individuals

Accordingly, the outcome variable is the feeling thermometer, which can take any values from 0 to 1 (basically the percentage of strong support for the majority rights). The treatment is the exposure to the information about discrimination. The population is the citizens of the Japan, the total is 126,807,358.¹ Individuals vary in age, gender, ethnicity, living place (47 prefectures), education level.

Simulation

Data generation

First, cleaning up the environment.

```
rm(list=ls())
```

Next, setting the global parameters.

¹The reason why I choose Japan for this research question is that the Japanese education tend to teach the society as a single ethnic nation. Thus the treatment can work better as new information, compared to other nations where the general population recognize that their country has a variety of population in terms of race, ethnicity, and other identification.

```
# Global parameters
```

```
N = 126807358
```

```
# Population size
```

```
N_samp <- 50
```

```
# Sample size
```

Here, I generate population parameters. The treatment is randomly assigned to the population. Approximately, a half of the population should be assigned the treatment, but the other half is not. For age, the variable should be positive integer or 0, less than 112.

The gender balance in Japan is almost even. 98.5% of the population is estimated as “Japanese people”. Here, it is hard to use the real sizes of ethnic groups, since the Japanese government does not publicly report any information about this.

```
set.seed(7272)
```

```
# Population parameters
```

```
treat <- rbinom(n=N, 1, 0.5)
```

```
# treatment
```

```
#
```

```
library("truncnorm")
```

```
age <- round(rtruncnorm(n=N, a=0, b=117, mean = 45, sd = 35))
```

```
gender <- rbinom(n=N, 1, 0.5)
```

```
ethnicity <- rbinom(n=N, 1, 0.15)
```

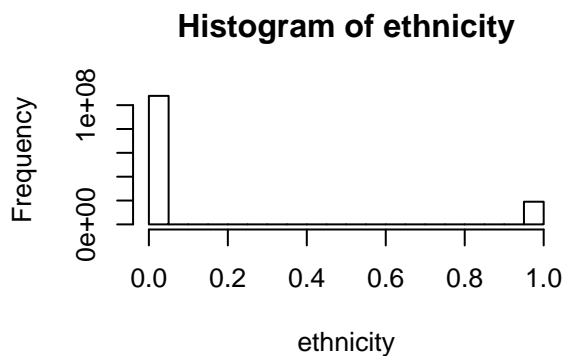
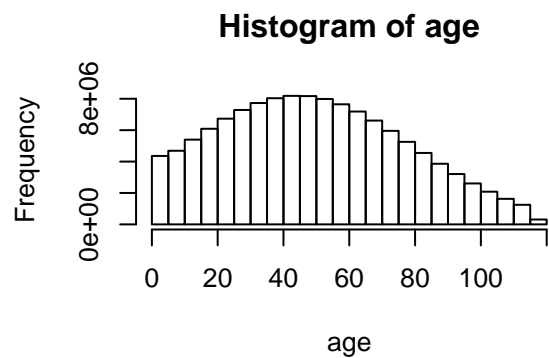
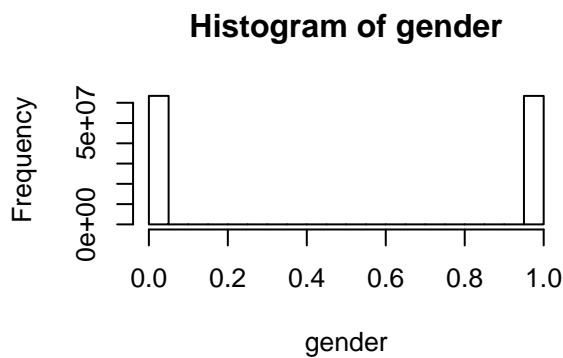
To check whether the simulated data are successfully created or not, here I check the visualization of variables.

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

```
hist(gender)
```

```
hist(age)
```

```
hist(ethnicity)
```



Now, turning into the creation of the outcome variable, the feeling thermometer. First, I fix parameter values, and then create the potential outcomes produced by treated and untreated cases. In general, it is expected that people do not have a strong opinion about the majority or minority rights.

```
a = 50                                # Intercept: Ground mean - neutral
b1 = .02                              # (Fixed=constant) effects of age
b2 = -.1                              # (Fixed=constant) effects of gender
b3 = -5                               # (Fixed=constant) effects of ethnicity
tau = -10                             # (Fixed=constant) positive treatment effect

# Potential outcomes
set.seed(7272)
Y_0 <- rnorm(n = N,
             mean = a + tau*0 + (b1*age) + (b2*gender) + (b3*ethnicity), # not treated
             sd = 1)
Y_1 <- rnorm(n = N,
             mean = a + tau*1 + (b1*age) + (b2*gender) + (b3*ethnicity), # treated
             sd = 1)
```

Finally, storing the all variables in one dataset.

```
# population level data
pop_dat <- data.frame(Y_0, Y_1, age, gender, ethnicity)
head(pop_dat)
```

```
##           Y_0           Y_1 age gender ethnicity
## 1 51.11922 40.87326 41      0          0
## 2 46.31275 36.02440 99      0          1
## 3 50.79764 40.88993 25      0          0
## 4 52.42444 42.14888 51      0          0
## 5 52.35458 40.08578 45      1          0
## 6 51.00963 40.45921 56      1          0
```

Now, sampling from the saved dataset.

```
sample_ind <- sample(1:nrow(pop_dat), size=N_samp) # sampling index
sample_dat <- pop_dat[sample_ind, ]                # only keep obs that match the index
head(sample_dat)
```

```
##           Y_0           Y_1 age gender ethnicity
## 104552389 45.88376 36.44748 62      1          1
## 117138221 44.91884 35.89422 20      0          1
## 18669795  49.75575 41.70117 52      0          0
## 91846675  50.77751 40.89924 80      0          0
## 55156763  46.05135 35.27642 49      1          1
## 97433708  47.51934 35.71602 74      0          1
```

```
sample_dat$d <- ifelse(runif(N_samp)<=0.5, 1, 0) # treatment assignment indicator (1=Treated, 0=Control)
```

```
# Observed outcomes:  $Y_{obs} = d \cdot Y_1 + (1-d) \cdot Y_0$ 
```

```
#  $Y_{obs} = d \cdot Y_1 + (1-d) \cdot Y_0$ 
```

```
sample_dat$Y_obs <- sample_dat$d*sample_dat$Y_1 + (1-sample_dat$d)*sample_dat$Y_0
```

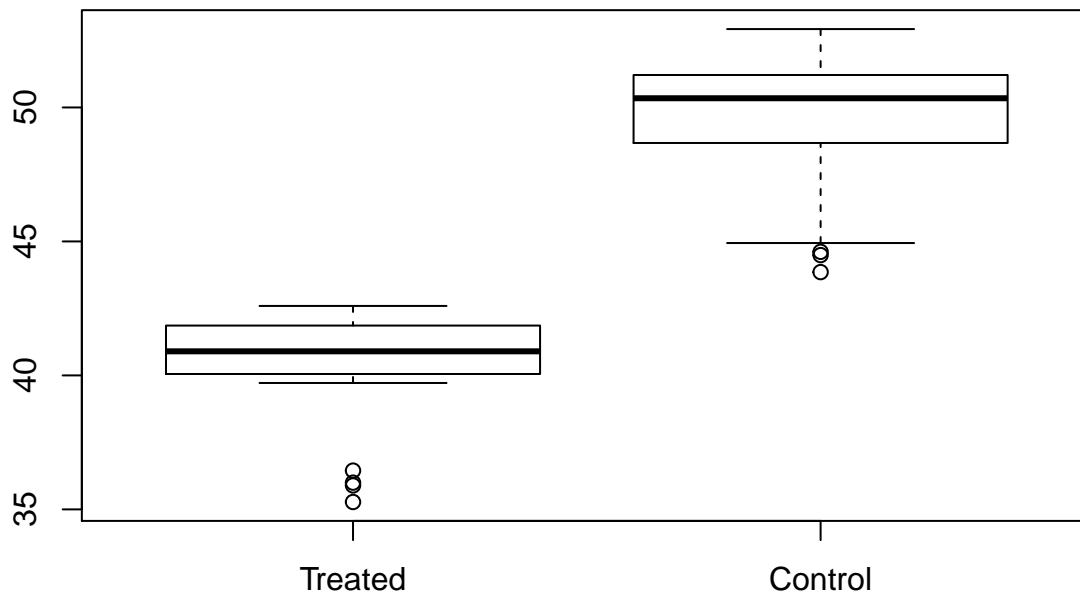
```
sample_dat$Status <- ifelse(sample_dat$d==1, "Treated", "Control") # saving the treatment assignment
```

```
head(sample_dat)
```

```
##           Y_0           Y_1 age gender ethnicity d    Y_obs Status
## 104552389 45.88376 36.44748 62      1          1 1 36.44748 Treated
## 117138221 44.91884 35.89422 20      0          1 1 35.89422 Treated
## 18669795  49.75575 41.70117 52      0          0 1 41.70117 Treated
## 91846675  50.77751 40.89924 80      0          0 1 40.89924 Treated
## 55156763  46.05135 35.27642 49      1          1 1 35.27642 Treated
## 97433708  47.51934 35.71602 74      0          1 0 47.51934 Control
```

Visualization of the difference in the outcome variable

```
boxplot(sample_dat$Y_obs[sample_dat$Status=="Treated"],
        sample_dat$Y_obs[sample_dat$Status=="Control"],
        names=c("Treated", "Control"))
```

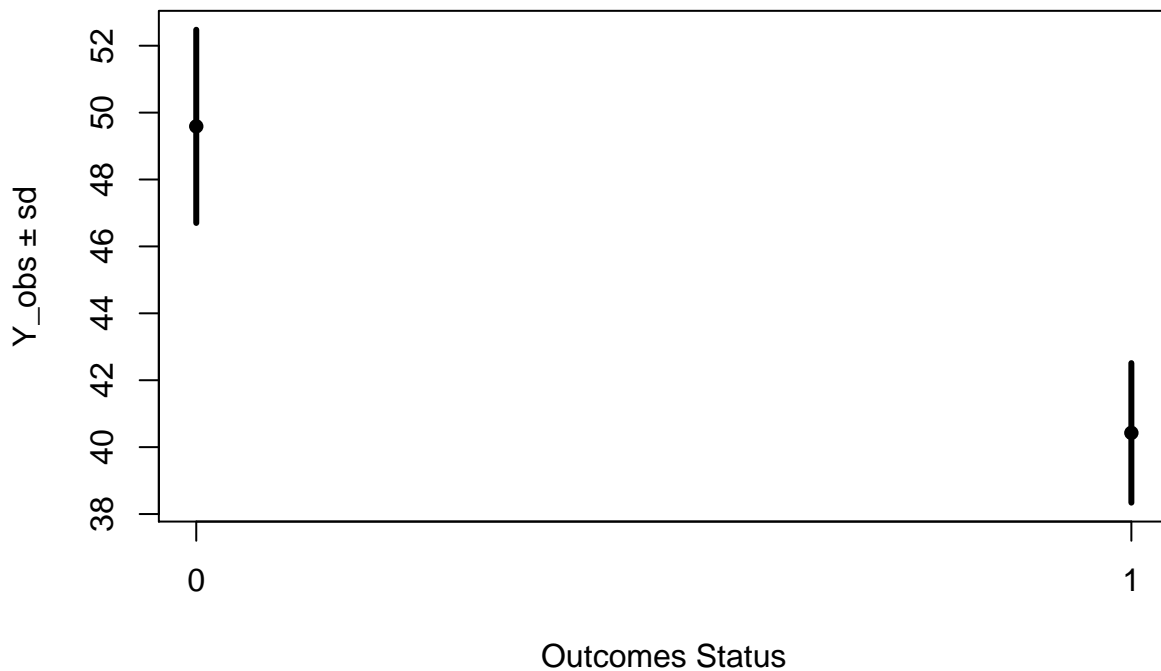


```

y_mean <- c(mean(sample_dat$Y_obs[sample_dat$Status=="Control"]),
            mean(sample_dat$Y_obs[sample_dat$Status=="Treated"]))
y_sdv <- c(sd(sample_dat$Y_obs[sample_dat$Status=="Control"]),
           sd(sample_dat$Y_obs[sample_dat$Status=="Treated"]))
t <- c(0,1)
plot(y_mean ~ t, pch=16, ylim=range(c(y_mean-y_sdv, y_mean+y_sdv)),
     xlab="Outcomes Status", ylab="Y_obs ± sd", xaxt="n") # Without x-axis label
axis(1, at = seq(00, 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

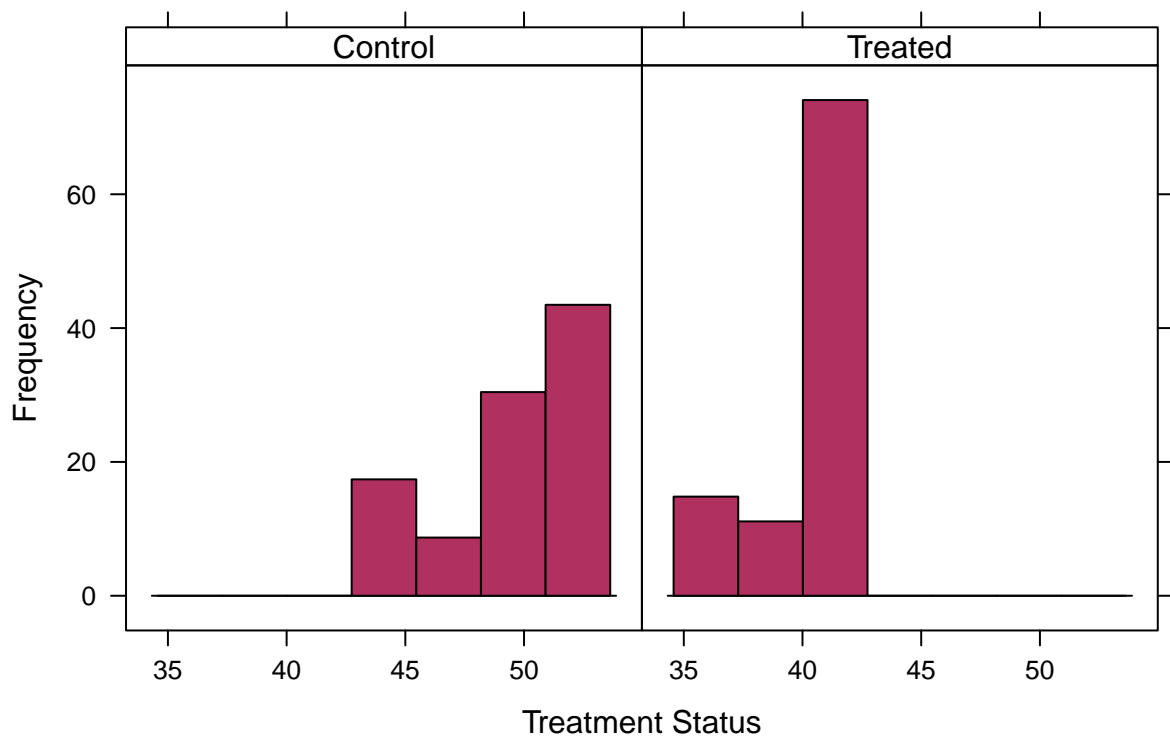


```
library("RColorBrewer")
myColours <- brewer.pal(5,"Blues")

my.settings <- list(
  superpose.polygon=list(col=myColours[2:5], border="transparent"),
  strip.background=list(col=myColours[6]),
  strip.border=list(col="black")
)

library('lattice')
histogram(~ Y_obs | Status, data = sample_dat, scales=list(alternating=1),
  auto.key=list(space="top", columns=4,
  points=FALSE, rectangles=TRUE,
  title="District", cex.title=1),
  col = "maroon",
  main = paste("Histograms of observed outcomes grouping by treatment status"),
  xlab = "Treatment Status", ylab = "Frequency",
  par.settings = my.settings)
```

Histograms of observed outcomes grouping by treatment status



Visualization of covariates

```
par(mfrow=c(1,3))

# age
boxplot(sample_dat$age[sample_dat$Status=="Treated"],
        sample_dat$age[sample_dat$Status=="Control"],
        names=c("Treated", "Control"),
        main=paste("Age"),
        xlab="Treatment Status")

# gender
boxplot(sample_dat$gender[sample_dat$Status=="Treated"],
        sample_dat$gender[sample_dat$Status=="Control"],
        names=c("Treated", "Control"),
        main=paste("Gender"),
        xlab="Treatment Status")

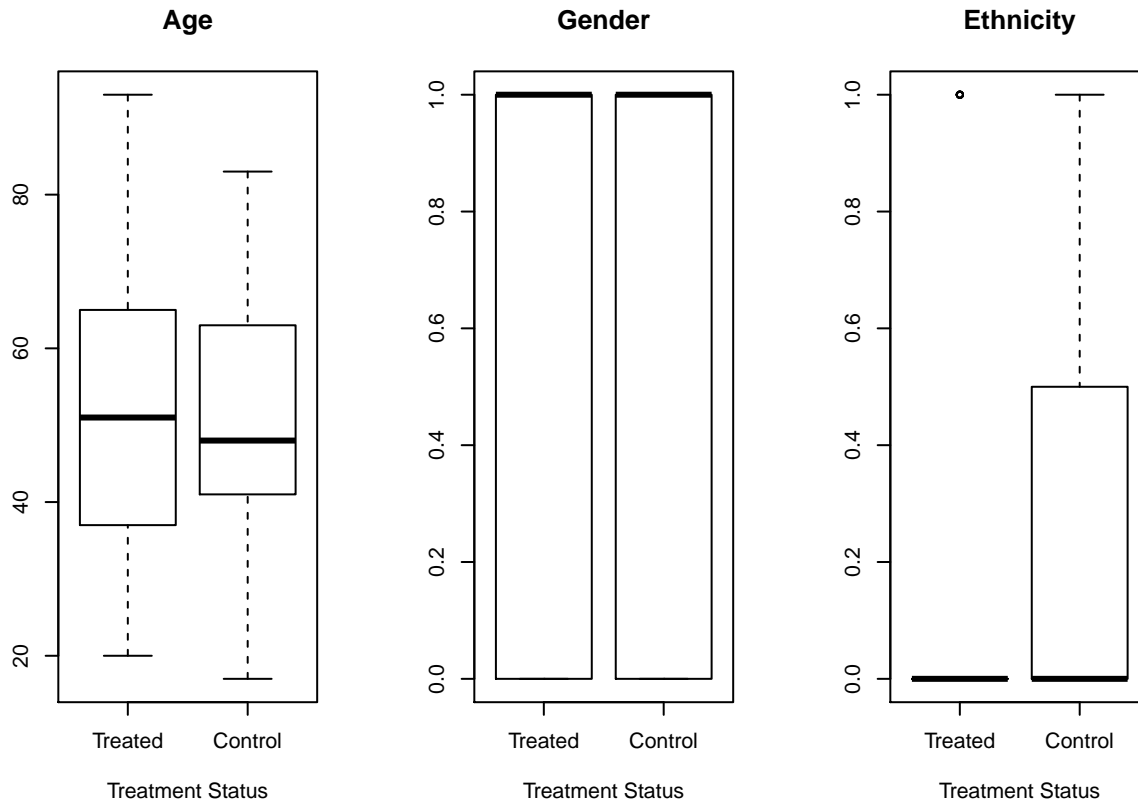
# ethnicity
```



```

boxplot(sample_dat$ethnicity[sample_dat$Status=="Treated"],
        sample_dat$ethnicity[sample_dat$Status=="Control"],
        names=c("Treated", "Control"),
        main=paste("Ethnicity"),
        xlab="Treatment Status")

```



Seemingly, there is no remarkable difference across treatment and control groups for covariates.

```

par(mfrow=c(1,3))

# age
c1_mean <- c(mean(sample_dat$age[sample_dat$Status=="Control"]),
             mean(sample_dat$age[sample_dat$Status=="Treated"]))
c1_sdv <- c(sd(sample_dat$age[sample_dat$Status=="Control"]),
            sd(sample_dat$age[sample_dat$Status=="Treated"]))
plot(c1_mean ~ t, pch=16, ylim=range(c(c1_mean-c1_sdv, c1_mean+c1_sdv)),
     xlab="Treatment Status", ylab="Age ± sd", xaxt="n") # Without x-axis label
axis(1, at = seq(00, 1, by = 1), las=1) # If las=2, numbers will be flipped by 90 degree
arrows(t, c1_mean-c1_sdv, t, c1_mean+c1_sdv, length=0, angle=90, lwd=3)
title("Simulated Result - Age")

# gender

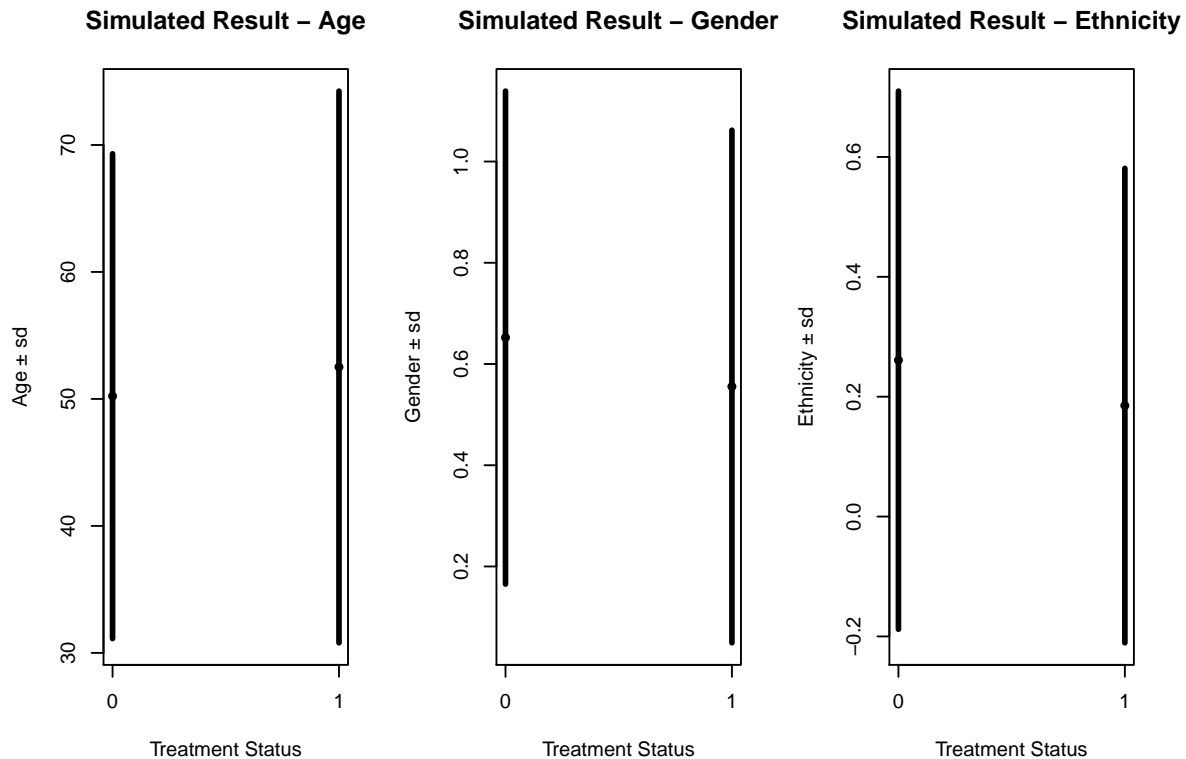
```

```

c2_mean <- c(mean(sample_dat$gender[sample_dat$Status=="Control"]),
             mean(sample_dat$gender[sample_dat$Status=="Treated"]))
c2_sdv <- c(sd(sample_dat$gender[sample_dat$Status=="Control"]),
           sd(sample_dat$gender[sample_dat$Status=="Treated"]))
plot(c2_mean ~ t, pch=16, ylim=range(c(c2_mean-c2_sdv, c2_mean+c2_sdv)),
     xlab="Treatment Status", ylab="Gender ± sd", xaxt="n") # Without x-axis label
axis(1, at = seq(00, 1, by = 1), las=1) # If las=2, numbers will be flipped by 90 degree
arrows(t, c2_mean-c2_sdv, t, c2_mean+c2_sdv, length=0, angle=90, lwd=3)
title("Simulated Result - Gender")

# ethnicity
c3_mean <- c(mean(sample_dat$ethnicity[sample_dat$Status=="Control"]),
             mean(sample_dat$ethnicity[sample_dat$Status=="Treated"]))
c3_sdv <- c(sd(sample_dat$ethnicity[sample_dat$Status=="Control"]),
           sd(sample_dat$ethnicity[sample_dat$Status=="Treated"]))
plot(c3_mean ~ t, pch=16, ylim=range(c(c3_mean-c3_sdv, c3_mean+c3_sdv)),
     xlab="Treatment Status", ylab="Ethnicity ± sd", xaxt="n") # Without x-axis label
axis(1, at = seq(00, 1, by = 1), las=1) # If las=2, numbers will be flipped by 90 degree
arrows(t, c3_mean-c3_sdv, t, c3_mean+c3_sdv, length=0, angle=90, lwd=3)
title("Simulated Result - Ethnicity")

```



This visualization confirms as well.

Visualization of differences between the true ATE and the estimated ATE

Here I would like to show the difference between estimated average effects and the true average effects. To do so, we have to extract the true effect, by taking the mean difference between potential outcomes in the population dataset. On the other hand, the estimated average effect is the mean difference of the sample. Here I compare them.

```
# Population level treatment (fixed=constant) effect
tau_true <- mean(pop_dat$Y_1 - pop_dat$Y_0) # Mean Difference of Individual Potential Outcomes
tau_true
```

```
## [1] -9.999969
```

```
# Estimating the average treatment effect
tau_hat <- mean(sample_dat$Y_obs[sample_dat$Status=="Treated"]) -
            mean(sample_dat$Y_obs[sample_dat$Status=="Control"])
tau_hat
```

```
## [1] -9.162804
```

While the true value of average treatment effect (ATE) is approximately -10, as I setted, the value of the estimated average treatment effect is near -10 but not as near as the true ATE.

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

```
##
## Welch Two Sample t-test
##
## data: Y_obs by Status
## t = 12.661, df = 39.325, p-value = 1.878e-15
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 7.699393 10.626214
## sample estimates:
## mean in group Control mean in group Treated
## 49.58906 40.42625
```

According to the result of the t-test, this difference is statistically significant. Next, I visualize the difference of these two in histogram with the true average treatment effect.

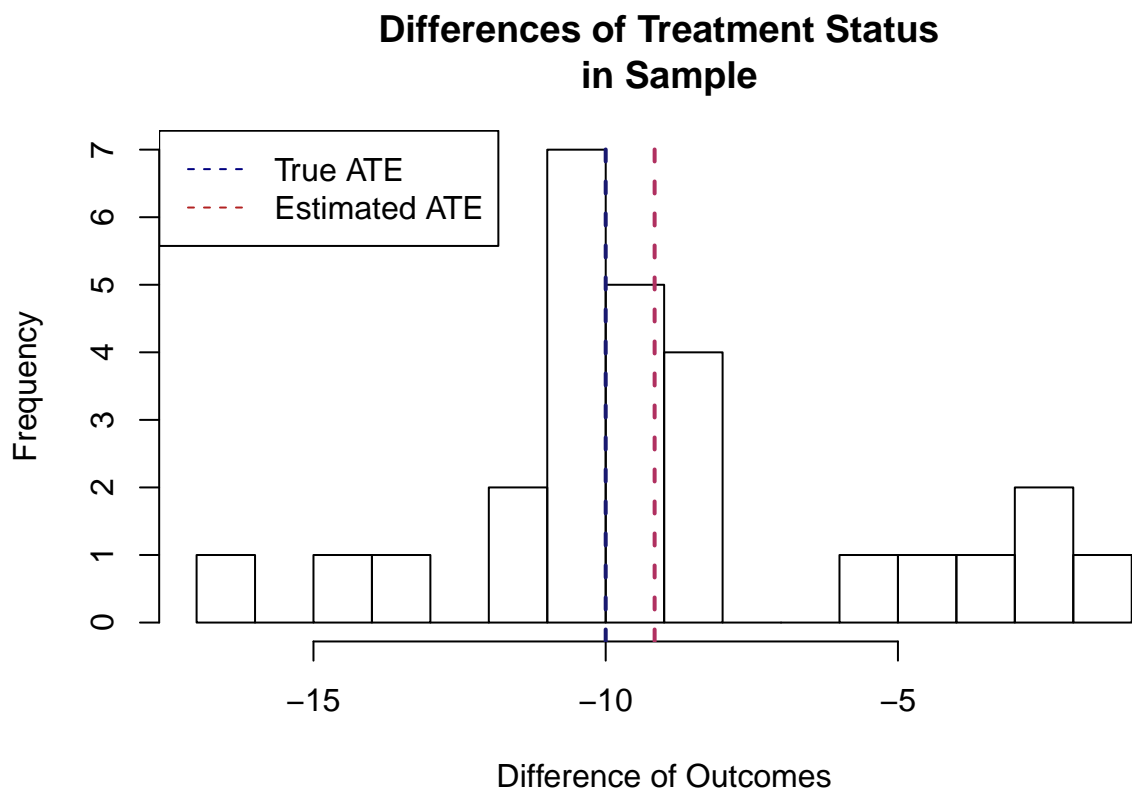
```
diff <- sample_dat$Y_obs[sample_dat$Status=="Treated"] - sample_dat$Y_obs[sample_dat$Status=="Control"]
```

```
## Warning in sample_dat$Y_obs[sample_dat$Status == "Treated"] :
## sample_dat$Y_obs[sample_dat$Status == "Treated"] : longer object length is not a multiple of shorter object length
```

```
hist(diff, main = paste("Differences of Treatment Status \n in Sample"),
      xlab = "Difference of Outcomes", breaks=20)
abline(v=tau_true, col="midnightblue", lwd="2", lty=2) # Add a line: true ATE
abline(v=tau_hat, col="maroon", lwd="2", lty=2) # Add another: estimated ATE
```

```
legend("topleft", legend=c("True ATE", "Estimated ATE"),  
      col=c("navy", "firebrick"), lty=c(2,2))
```

Add legend



Problem 2

Instruction: Repeat the same simulation by changing the number of samples from 50, 100, 150, 200 to 250, and create a graph that combines five plots for the difference in means or distributions.

Simulation of Different Samples

```
N_samp <- 100
sample_ind2 <- sample(1:nrow(pop_dat), size=N_samp) # sampling index
sample_dat2 <- pop_dat[sample_ind2, ] # only keep obs that match the index
head(sample_dat2)

##           Y_0      Y_1 age gender ethnicity
## 76529255  47.76047 35.88306 104      0         1
## 52473555  49.89540 39.42905  13      1         0
## 72009669  50.46367 42.78253  29      0         0
## 105572360 50.81374 39.60539  10      1         0
## 117706448 52.99887 40.86275  48      0         0
## 113465463 50.34141 40.28034  48      0         0

sample_dat2$d <- ifelse(runif(N_samp)<=0.5, 1, 0) # treatment assignment indicator (1=Treated, 0=Control)

# Observed outcomes: Y_obs = d*Y_1 + (1-d)*Y_0
# Y_obs = d*Y_1 + (1-d)*Y_0
sample_dat2$Y_obs <- sample_dat2$d*sample_dat2$Y_1 + (1-sample_dat2$d)*sample_dat2$Y_0

sample_dat2$Status <- ifelse(sample_dat2$d==1, "Treated", "Control") # saving the treatment status

N_samp <- 150
sample_ind3 <- sample(1:nrow(pop_dat), size=N_samp) # sampling index
sample_dat3 <- pop_dat[sample_ind3, ] # only keep obs that match the index
head(sample_dat3)

##           Y_0      Y_1 age gender ethnicity
## 110311776  46.53398 35.41004  36      1         1
## 3108150   45.43093 34.85222  47      1         1
## 65480110  46.02355 37.79521  74      0         1
## 34285856  51.19971 40.29424  41      0         0
## 31403463  51.01000 40.43754  28      0         0
## 79405680  45.16126 35.41984  42      0         1

sample_dat3$d <- ifelse(runif(N_samp)<=0.5, 1, 0) # treatment assignment indicator (1=Treated, 0=Control)

# Observed outcomes: Y_obs = d*Y_1 + (1-d)*Y_0
# Y_obs = d*Y_1 + (1-d)*Y_0
```

```

sample_dat3$Y_obs <- sample_dat3$d*sample_dat3$Y_1 + (1-sample_dat3$d)*sample_dat3$Y_0

sample_dat3$Status <- ifelse(sample_dat3$d==1, "Treated", "Control")      # saving the treat

N_samp <- 200
sample_ind4 <- sample(1:nrow(pop_dat), size=N_samp) # sampling index
sample_dat4 <- pop_dat[sample_ind4, ]                # only keep obs that match the index
head(sample_dat4)

##           Y_0      Y_1 age gender ethnicity
## 116300175 52.13454 42.27271  54      0      0
## 39663623  45.27791 34.69862  12      0      1
## 119490650 50.46177 43.64085  93      1      0
## 39793034  50.36841 41.86554  80      1      0
## 3931563   54.41803 40.75783 115      0      0
## 67582792  51.37218 40.43314  65      0      0

sample_dat4$d <- ifelse(runif(N_samp)<=0.5, 1, 0)      # treatment assignment indicator (1=Tt

# Observed outcomes: Y_obs = d*Y_1 + (1-d)*Y_0
# Y_obs = d*Y_1 + (1-d)*Y_0
sample_dat4$Y_obs <- sample_dat4$d*sample_dat4$Y_1 + (1-sample_dat4$d)*sample_dat4$Y_0
sample_dat4$Status <- ifelse(sample_dat4$d==1, "Treated", "Control")      # saving the treat

N_samp <- 250
sample_ind5 <- sample(1:nrow(pop_dat), size=N_samp) # sampling index
sample_dat5 <- pop_dat[sample_ind5, ]                # only keep obs that match the index
head(sample_dat5)

##           Y_0      Y_1 age gender ethnicity
## 34137337  50.15751 40.59224  31      0      0
## 888836   51.37354 41.02151  72      0      0
## 39991919  51.01932 42.84774  91      0      0
## 79286307  50.70738 40.75841  50      1      0
## 79628248  48.55019 42.26687  17      0      0
## 105430310 49.93684 40.40828  45      1      0

sample_dat5$d <- ifelse(runif(N_samp)<=0.5, 1, 0)      # treatment assignment indicator (1=Tt

# Observed outcomes: Y_obs = d*Y_1 + (1-d)*Y_0
# Y_obs = d*Y_1 + (1-d)*Y_0
sample_dat5$Y_obs <- sample_dat5$d*sample_dat5$Y_1 + (1-sample_dat5$d)*sample_dat5$Y_0
sample_dat5$Status <- ifelse(sample_dat5$d==1, "Treated", "Control")      # saving the treat

```

A Graph of Combined Plots

By using the stored data sets, I create box plots and caterpillar plots for each case. Before that, to show the combined graphs clearly, I enlarged the size of figures by `knitr::opts_chunk()` settings.

```
# this is the setting to show the combined graphs more largely
knitr::opts_chunk$set(warning = FALSE, echo = TRUE, message = TRUE,
  # setting for chunk options
  fig.width = 8, fig.height = 8, fig.align="center")
# setting for figure output
```

Then, as before, I create the box plots.

```
par(mfrow=c(3,2))

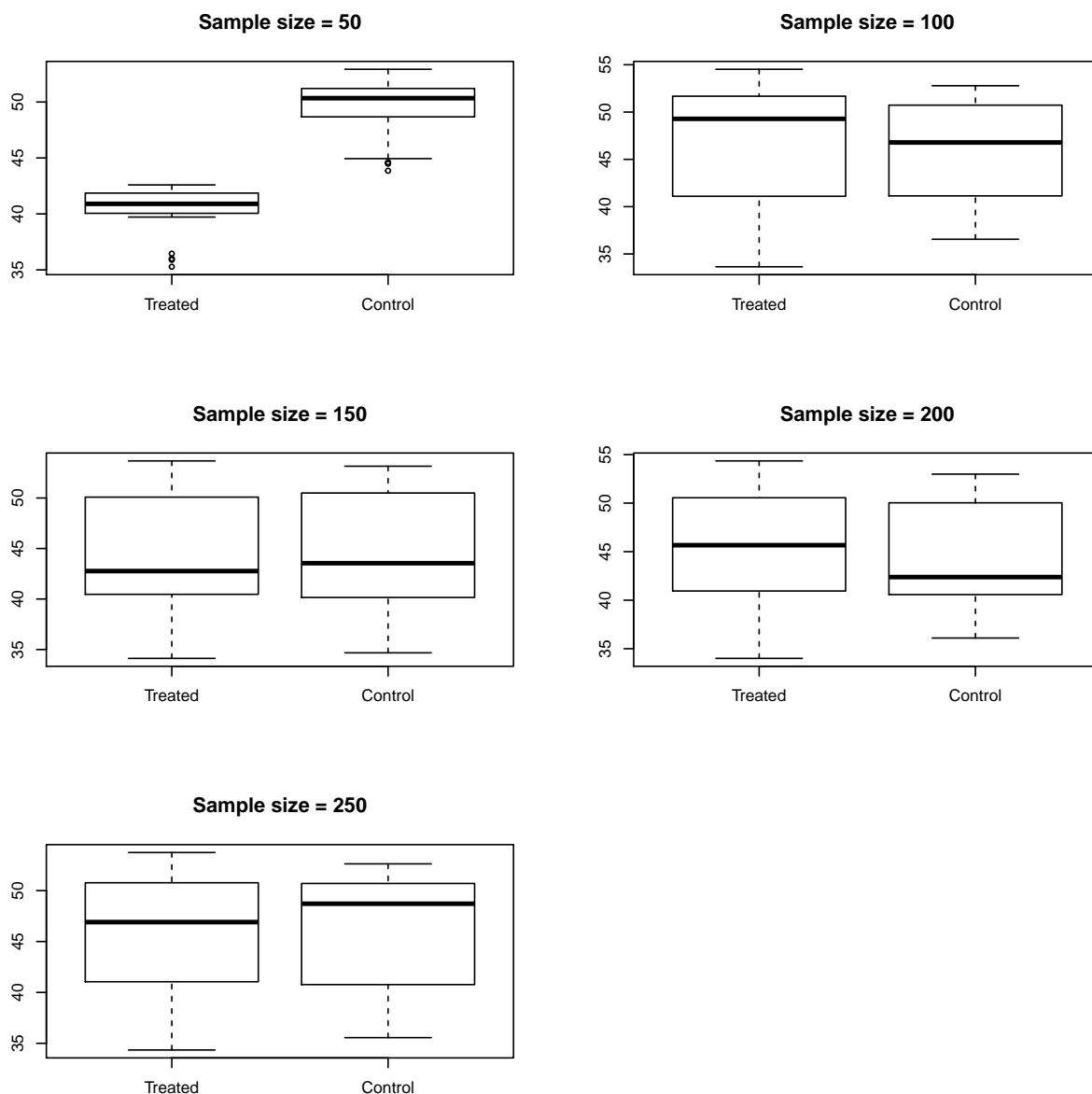
boxplot(sample_dat$Y_obs[sample_dat$Status=="Treated"],
  sample_dat$Y_obs[sample_dat$Status=="Control"],
  names=c("Treated", "Control"),
  main=paste("Sample size = 50"))

boxplot(sample_dat2$Y_obs[sample_dat$Status=="Treated"],
  sample_dat2$Y_obs[sample_dat$Status=="Control"],
  names=c("Treated", "Control"),
  main=paste("Sample size = 100"))

boxplot(sample_dat3$Y_obs[sample_dat$Status=="Treated"],
  sample_dat3$Y_obs[sample_dat$Status=="Control"],
  names=c("Treated", "Control"),
  main=paste("Sample size = 150"))

boxplot(sample_dat4$Y_obs[sample_dat$Status=="Treated"],
  sample_dat4$Y_obs[sample_dat$Status=="Control"],
  names=c("Treated", "Control"),
  main=paste("Sample size = 200"))

boxplot(sample_dat5$Y_obs[sample_dat$Status=="Treated"],
  sample_dat5$Y_obs[sample_dat$Status=="Control"],
  names=c("Treated", "Control"),
  main=paste("Sample size = 250"))
```



Seemingly, as more sample we have, the difference between treated or control groups becomes smaller. To confirm this tendency, I also create the caterpillar plots.

```
par(mfrow=c(3,2))

y_mean <- c(mean(sample_dat$Y_obs[sample_dat$Status=="Control"]),
            mean(sample_dat$Y_obs[sample_dat$Status=="Treated"]))
y_sdv <- c(sd(sample_dat$Y_obs[sample_dat$Status=="Control"]),
           sd(sample_dat$Y_obs[sample_dat$Status=="Treated"]))
plot(y_mean ~ t, pch=16, ylim=range(c(y_mean-y_sdv, y_mean+y_sdv)),
     xlab="Outcomes Status", ylab="Y_obs ± sd", xaxt="n",
     main=paste("Sample size = 50"))
axis(1, at = seq(00, 1, by = 1), las=1)
```



```
arrows(t, y_mean-y_sdv, t, y_mean+y_sdv, length=0, angle=90, lwd=3)
title("Simulated Result")
```

```
y_mean <- c(mean(sample_dat2$Y_obs[sample_dat$Status=="Control"]),
            mean(sample_dat2$Y_obs[sample_dat$Status=="Treated"]))
y_sdv <- c(sd(sample_dat2$Y_obs[sample_dat$Status=="Control"]),
          sd(sample_dat2$Y_obs[sample_dat$Status=="Treated"]))
plot(y_mean ~ t, pch=16, ylim=range(c(y_mean-y_sdv, y_mean+y_sdv)),
     xlab="Outcomes Status", ylab="Y_obs ± sd", xaxt="n",
     main=paste("Sample size = 100"))
axis(1, at = seq(00, 1, by = 1), las=1)
arrows(t, y_mean-y_sdv, t, y_mean+y_sdv, length=0, angle=90, lwd=3)
title("Simulated Result")
```

```
y_mean <- c(mean(sample_dat3$Y_obs[sample_dat$Status=="Control"]),
            mean(sample_dat3$Y_obs[sample_dat$Status=="Treated"]))
y_sdv <- c(sd(sample_dat3$Y_obs[sample_dat$Status=="Control"]),
          sd(sample_dat3$Y_obs[sample_dat$Status=="Treated"]))
plot(y_mean ~ t, pch=16, ylim=range(c(y_mean-y_sdv, y_mean+y_sdv)),
     xlab="Outcomes Status", ylab="Y_obs ± sd", xaxt="n",
     main=paste("Sample size = 150"))
axis(1, at = seq(00, 1, by = 1), las=1)
arrows(t, y_mean-y_sdv, t, y_mean+y_sdv, length=0, angle=90, lwd=3)
title("Simulated Result")
```

```
y_mean <- c(mean(sample_dat4$Y_obs[sample_dat$Status=="Control"]),
            mean(sample_dat4$Y_obs[sample_dat$Status=="Treated"]))
y_sdv <- c(sd(sample_dat4$Y_obs[sample_dat$Status=="Control"]),
          sd(sample_dat4$Y_obs[sample_dat$Status=="Treated"]))
plot(y_mean ~ t, pch=16, ylim=range(c(y_mean-y_sdv, y_mean+y_sdv)),
     xlab="Outcomes Status", ylab="Y_obs ± sd", xaxt="n",
     main=paste("Sample size = 200"))
axis(1, at = seq(00, 1, by = 1), las=1)
arrows(t, y_mean-y_sdv, t, y_mean+y_sdv, length=0, angle=90, lwd=3)
title("Simulated Result")
```

```
y_mean <- c(mean(sample_dat5$Y_obs[sample_dat$Status=="Control"]),
            mean(sample_dat5$Y_obs[sample_dat$Status=="Treated"]))
y_sdv <- c(sd(sample_dat5$Y_obs[sample_dat$Status=="Control"]),
```

```

sd(sample_dat5$Y_obs[sample_dat$Status=="Treated"]))
t <- c(0,1)
plot(y_mean ~ t, pch=16, ylim=range(c(y_mean-y_sdv, y_mean+y_sdv)),
     xlab="Outcomes Status", ylab="Y_obs ± sd", xaxt="n",
     main=paste("Sample size = 250"))
axis(1, at = seq(0, 1, by = 1), las=1)
arrows(t, y_mean-y_sdv, t, y_mean+y_sdv, length=0, angle=90, lwd=3)
title("Simulated Result")

```

