

Problem Set 2

Atit Wongnophadol

1. What happens when pilgrims attend the Hajj pilgrimage to Mecca?

On the one hand, participating in a common task with a diverse group of pilgrims might lead to increased mutual regard through processes identified in *Contact Theories*. On the other hand, media narratives have raised the spectre that this might be accompanied by “antipathy toward non-Muslims”. Clingingsmith, Khwaja and Kremer (2009) investigates the question.

Using the data here, test the sharp null hypothesis that winning the visa lottery for the pilgrimage to Mecca had no effect on the views of Pakistani Muslims toward people from other countries. Assume that the Pakistani authorities assigned visas using complete random assignment. Use, as your primary outcome the `views` variable, and as your treatment feature `success`. If you’re ambitious, write your function generally so that you can also evaluate feelings toward specific nationalities.

```
mecca <- read.csv("/Users/plodium2000/Dropbox/MIDS program/Courses/W241 Causality and Experiments/Assignments/mecca.csv")
```

- a. Using either `dplyr` or `data.table`, group the data by `success` and report whether views toward others are generally more positive among lottery winners or lottery non-winners.

ANS: The views towards others are generally more positive among the lottery winners as shown in the higher average of views for the lottery winners (i.e., `success = 1`) than the lottery non-winners (i.e., `success = 0`).

```
# setup data.table for the dataset
mecca <- data.table(mecca)
```

```
# summarize average views by success (1=lottery winner, 0=lottery non-winners)
mecca[,mean(views), by=success]
```

```
##      success      V1
## 1:         0 1.868304
## 2:         1 2.343137
```

```
# calculate average treatment effect from this sample
ate_actual <- mean(mecca[success==1]$views) - mean(mecca[success==0]$views)
ate_actual
```

```
## [1] 0.4748337
```

- b. But is this a meaningful difference, or could it just be randomization noise? Conduct 10,000 simulated random assignments under the sharp null hypothesis to find out. (Don’t just copy the code from the `async`, think about how to write this yourself.)

```
# count number of winners and non-winners
winners <- sum(mecca$success)
non_winners <- nrow(mecca)-winners

winners
```

```
## [1] 510
```

```

non_winners

## [1] 448

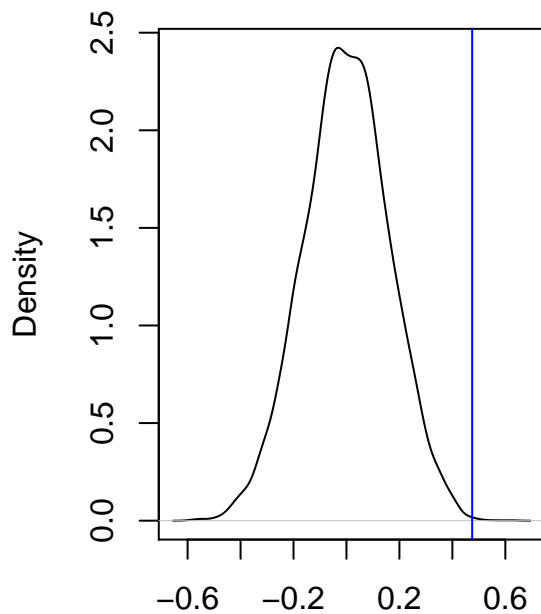
# simulate ATE for 10,000 trials
nsims <- 10000
sim.means <- as.numeric(NULL)
for(i in 1:nsims){
  mecca[, sim.treat := sample(c(rep(0,non_winners), rep(1,winners)))]
  sim.means[i] <- mecca[sim.treat == 1, mean(views)] - mecca[sim.treat == 0, mean(views)]
}

dist_under_sharp_null <- sim.means

# plot the density and histogram under the sharp null hypothesis
par(mfrow = c(1,2))
plot(density(dist_under_sharp_null),
     main = "Density Plot of ATE")
abline(v = ate_actual, col = "blue")
hist(dist_under_sharp_null,
     main = "Histogram of ATE",
     freq = FALSE)
abline(v = ate_actual, col = "blue")

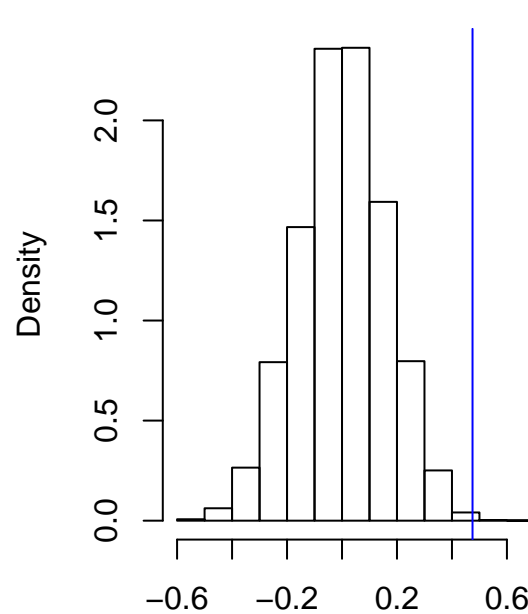
```

Density Plot of ATE



N = 10000 Bandwidth = 0.02298

Histogram of ATE



dist_under_sharp_null

- c. How many of the simulated random assignments generate an estimated ATE that is at least as large as the actual estimate of the ATE?

```

mean(ate_actual < dist_under_sharp_null)*10000

```

```

## [1] 6

```

d. What is the implied *one-tailed* p-value?

```
mean(ate_actual < dist_under_sharp_null)
```

```
## [1] 6e-04
```

e. How many of the simulated random assignments generate an estimated ATE that is at least as large *in absolute value* as the actual estimate of the ATE?

```
mean(ate_actual < abs(dist_under_sharp_null))*10000
```

```
## [1] 18
```

f. What is the implied two-tailed p-value?

```
mean(ate_actual < abs(dist_under_sharp_null))*2
```

```
## [1] 0.0036
```

2. Term Limits Aren't Good.

Naturally occurring experiments sometimes involve what is, in effect, block random assignment. For example, Rocio Titunik , in this paper studies the effect of lotteries that determine whether state senators in TX and AR serve two-year or four-year terms in the aftermath of decennial redistricting. These lotteries are conducted within each state, and so there are effectively two distinct experiments on the effects of term length.

The “thoery” in the news (such as it is), is that legislators who serve 4 year terms have more time to slack off and not produce legislation. If this were true, then it would stand to reason that making terms shorter would increase legislative production.

One way to measure legislative production is to count the number of bills (legislative proposals) that each senator introduces during a legislative session. The table below lists the number of bills introduced by senators in both states during 2003.

```
library(foreign)
```

```
legislator <- read.dta("./data/Titunik.2010.dta")
head(legislator)
```

```
##   term2year bills_introduced texas0_arkansas1
## 1         0             18             0
## 2         0             29             0
## 3         0             41             0
## 4         0             53             0
## 5         0             60             0
## 6         0             67             0
```

a. Using either `dplyr` or `data.table`, group the data by state and report the mean number of bills introduced in each state. Does Texas or Arkansas seem to be more productive? Then, group by two- or four-year terms (ignoring states). Do two- or four-year terms seem to be more productive? **Which of these effects is causal, and which is not?** Finally, using `dplyr` or `data.table` to group by state and term-length. How, if at all, does this change what you learn?

ANS: Texas seems to be more productive given the average bills produced of 68.77 compared to 25.51 that of Arkansas.

```
# setup data.table for the dataset
legislator <- data.table(legislator)
```

```
# summarize average bills produced by state (1=Arkansas, 0=Texas)
legislator[,mean(bills_introduced), by=texas0_arkansas1]
```

```
##      texas0_arkansas1      V1
## 1:                0 68.77419
## 2:                1 25.51429
```

ANS: The 4-year term seem to be less productive given the average bills produced of 53.1 (~26.5 per 2-year) compared to 38.6 that of the 2-year term.

```
# summarize average bills produced by term (1= two years, 0= 4 years)
legislator[,mean(bills_introduced), by=term2year]
```

```
##      term2year      V1
## 1:          0 53.09091
## 2:          1 38.57576
```

ANS: The small marginal difference (77 vs 60), which is only 16 bills difference, in Texas between bills produced given 4-year and 2-year term makes me wonder that the treatment effect would be more pronounced in Texas than in Arkansas.

```
# summarize average bills produced by state and term
legislator[,mean(bills_introduced), by=list(texas0_arkansas1,term2year)]
```

```
##      texas0_arkansas1 term2year      V1
## 1:                0          0 76.87500
## 2:                0          1 60.13333
## 3:                1          0 30.70588
## 4:                1          1 20.61111
```

b. For each state, estimate the standard error of the estimated ATE.

```
# function to calculate standard error (formula 3.6 from FE)
standard_error <- function(var_1, var_2, n1, n2){
  return(sqrt((var_1/n1)+(var_2/n2)))
}
```

```
# est_eta for AK
```

```
est_eta_AK <- legislator[term2year==0 & texas0_arkansas1 == 1, mean(bills_introduced)] - legislator[term2year==1 & texas0_arkansas1 == 1, mean(bills_introduced)]
```

```
# calculate standard error for AK
```

```
var_control_AK <- legislator[term2year==0 & texas0_arkansas1 == 1, var(bills_introduced)]
var_treatment_AK <- legislator[term2year==1 & texas0_arkansas1 == 1, var(bills_introduced)]
n_control_AK <- nrow(legislator[term2year==0 & texas0_arkansas1 == 1])
n_treatment_AK <- nrow(legislator[term2year==1 & texas0_arkansas1 == 1])
```

```
se_AK <- standard_error(var_control_AK,n_control_AK,var_treatment_AK,n_treatment_AK)
paste("Standard Error of the ATE in Arkansas: ", se_AK)
```

```
## [1] "Standard Error of the ATE in Arkansas: 1.97521503241659"
```

```
# est_eta for TX
```

```
est_eta_TX <- legislator[term2year==0 & texas0_arkansas1 == 0, mean(bills_introduced)] - legislator[term2year==1 & texas0_arkansas1 == 0, mean(bills_introduced)]
```

```
# calculate standard error for TX
var_control_TX <- legislator[term2year==0 & texas0_arkansas1 == 0, var(bills_introduced)]
var_treatment_TX <- legislator[term2year==1 & texas0_arkansas1 == 0, var(bills_introduced)]
n_control_TX <- nrow(legislator[term2year==0 & texas0_arkansas1 == 0])
n_treatment_TX <- nrow(legislator[term2year==1 & texas0_arkansas1 == 0])

se_TX <- standard_error(var_control_TX,n_control_TX,var_treatment_TX,n_treatment_TX)
paste("Standard Error of the ATE in Texas: ", se_TX)
```

```
## [1] "Standard Error of the ATE in Texas: 1.83797466005965"
```

c. Use equation (3.10) to estimate the overall ATE for both states combined.

```
total_row <- nrow(legislator)
AK_ratio <- nrow(legislator[texas0_arkansas1 == 1]) / total_row
TX_ratio <- nrow(legislator[texas0_arkansas1 == 0]) / total_row

ATE_overall <- (AK_ratio*est_eta_AK)+(TX_ratio*est_eta_TX)

paste("Estimated overall ATE for both states combined: ", ATE_overall)
```

```
## [1] "Estimated overall ATE for both states combined: 13.2167978807685"
```

d. Explain why, in this study, simply pooling the data for the two states and comparing the average number of bills introduced by two-year senators to the average number of bills introduced by four-year senators leads to biased estimate of the overall ATE.

ANS: By pooling the data and comparing the average number disregards the fact that the weight of each state is not the same. The ATE calculated would assume equal importance / proportion from each state, when in fact it is not.

e. Insert the estimated standard errors into equation (3.12) to estimate the stand error for the overall ATE.

```
se_overall <- sqrt(((se_AK*AK_ratio)**2) + ((se_TX*TX_ratio)**2))

paste("Estimated standard error for the overall ATE: ", se_overall)
```

```
## [1] "Estimated standard error for the overall ATE: 1.35736851921838"
```

f. Use randomization inference to test the sharp null hypothesis that the treatment effect is zero for senators in both states.

ANS: Per the randomization inference described below, the two-tail p-value is over 0.14, suggesting that the sharp null hypothesis cannot be rejected. Therefore the the treatment effect would have no difference on the senator's productivity.

```
# count number of treatment and control
treatment_legislator <- sum(legislator$term2year)
control_legislator <- nrow(legislator)-treatment_legislator

# simulate ATE for 10,000 trials
nsims <- 10000
sim.means_legislator <- as.numeric(NULL)
for(i in 1:nsims){
  legislator[, sim.treat_legislator := sample(c(rep(0,control_legislator), rep(1,treatment_legislator))
  sim.means_legislator[i] <- legislator[sim.treat_legislator == 1, mean(bills_introduced)] - legislator
}
```

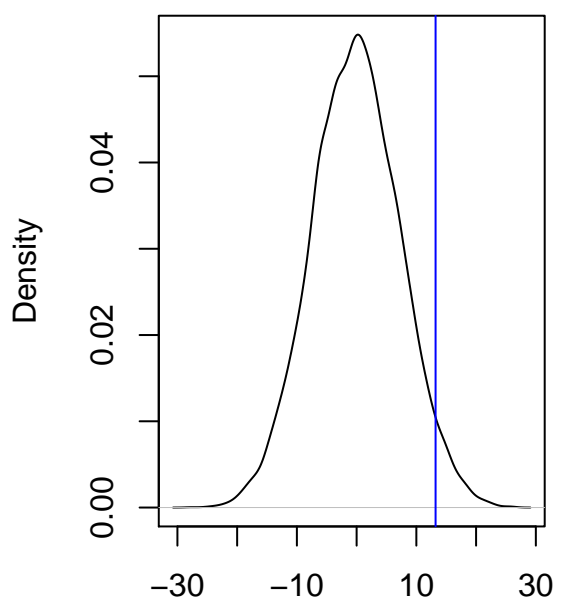
```

dist_under_sharp_null_legislator <- sim.means_legislator

# plot the density and histogram under the sharp null hypothesis
par(mfrow = c(1,2))
plot(density(dist_under_sharp_null_legislator),
     main = "Density Plot of ATE")
abline(v = ATE_overall, col = "blue")
hist(dist_under_sharp_null_legislator,
     main = "Histogram of ATE",
     freq = FALSE)
abline(v = ATE_overall, col = "blue")

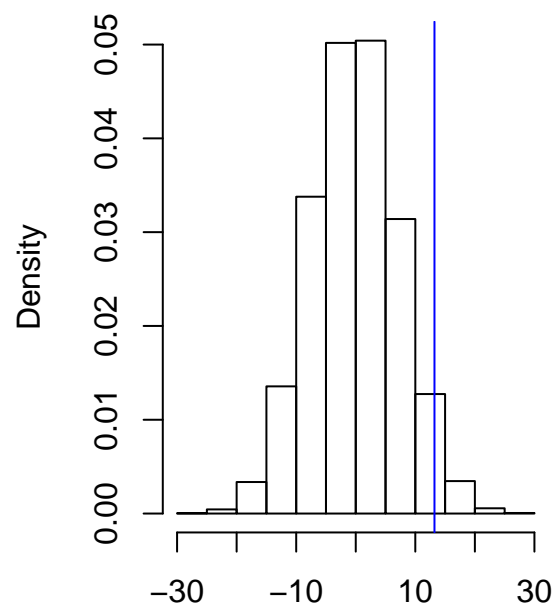
```

Density Plot of ATE



N = 10000 Bandwidth = 1.04

Histogram of ATE



dist_under_sharp_null_legislator

```

# 2-tails p-value
mean(ATE_overall < abs(dist_under_sharp_null_legislator))*2

```

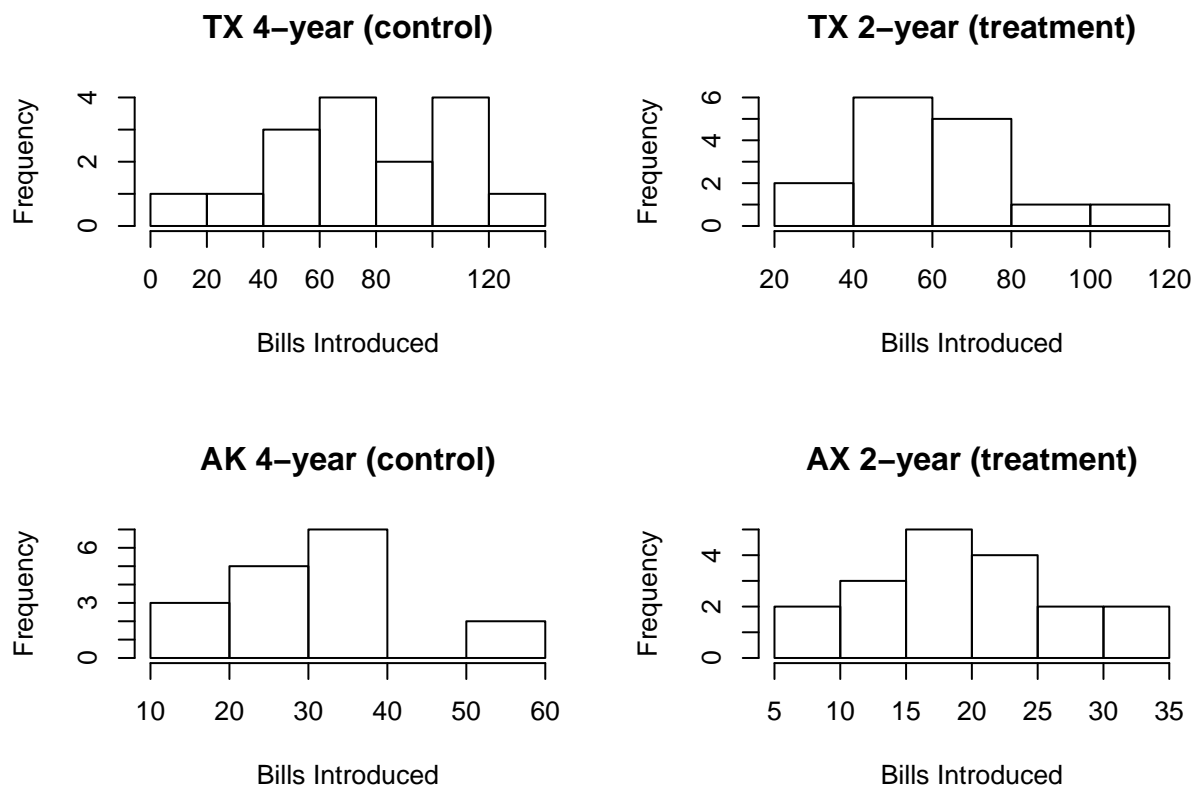
```
## [1] 0.143
```

g. **IN Addition:** Plot histograms for both the treatment and control groups in each state (for 4 histograms in total).

```

par(mfrow = c(2,2))
hist(legislator[term2year==0 & texas0_arkansas1 == 0]$bills_introduced, main = "TX 4-year (control)", xlab = "bills_introduced")
hist(legislator[term2year==1 & texas0_arkansas1 == 0]$bills_introduced, main = "TX 2-year (treatment)", xlab = "bills_introduced")
hist(legislator[term2year==0 & texas0_arkansas1 == 1]$bills_introduced, main = "AK 4-year (control)", xlab = "bills_introduced")
hist(legislator[term2year==1 & texas0_arkansas1 == 1]$bills_introduced, main = "AK 2-year (treatment)", xlab = "bills_introduced")

```



3. Cluster Randomization

Use the data in *Field Experiments* Table 3.3 to simulate cluster randomized assignment. (Notes: (a) Assume 3 clusters in treatment and 4 in control; and (b) When Gerber and Green say *simulate*’, they do not mean run simulations with R code’, but rather, in a casual sense “take a look at what happens if you do this this way.” There is no randomization inference necessary to complete this problem.)

```
## load data
d3 <- read.csv("./data/ggChapter3.csv", stringsAsFactors = FALSE)
d3 <- data.table(d3)
d3
```

```
##      Village Y D Block
## 1:         1 0 0     1
## 2:         2 1 0     1
## 3:         3 2 1     1
## 4:         4 4 2     1
## 5:         5 4 0     1
## 6:         6 6 0     1
## 7:         7 6 2     1
## 8:         8 9 3     1
## 9:         9 14 12    2
## 10:        10 15 9     2
## 11:        11 16 8     2
## 12:        12 16 15    2
## 13:        13 17 5     2
## 14:        14 18 17    2
```

- a. Suppose the clusters are formed by grouping observations {1,2}, {3,4}, {5,6}, ... , {13,14}. Use equation (3.22) to calculate the standard error assuming half of the clusters are randomly assigned to treatment.

ANS: se = 12.12239.

```
var_y0 <- var(c(0.5,3,5,15/2,14.5,16,17.5))
var_y1 <- var(c(0,1.5,0,2.5,10.5,23/2,11))
y0 <- c(0.5,3,5,15/2,14.5,16,17.5)
y1 <- c(0,1.5,0,2.5,10.5,23/2,11)
k <- 2
m <- 7
N <- 14
part1 <- m * var_y0 / (N-m) + (N-m) * var_y1 / m + 2 * cov(y0,y1)
se <- sqrt(1/(k-1) * part1)
se
```

```
## [1] 12.12239
```

- b. Suppose that clusters are instead formed by grouping observations {1,14}, {2,13}, {3,12}, ... , {7,8}. Use equation (3.22) to calculate the standard error assuming half of the clusters are randomly assigned to treatment.

ANS: se = 2.805182.

```
var_y0 <- var(c(9,9,9,10,9.5,10,7.5))
var_y1 <- var(c(8.5,2.5,8,5,4.5,6,2.5))
y0 <- c(9,9,9,10,9.5,10,7.5)
y1 <- c(8.5,2.5,8,5,4.5,6,2.5)
k <- 2
m <- 7
N <- 14
part1 <- m * var_y0 / (N-m) + (N-m) * var_y1 / m + 2 * cov(y0,y1)
se <- sqrt(1/(k-1) * part1)
se
```

```
## [1] 2.805182
```

- c. Why do the two methods of forming clusters lead to different standard errors? What are the implications for the design of cluster randomized experiments?

ANS: It's because the variability of cluster-level in B is much smaller than in A. Therefore, it shows that the penalty associated with clustering depends on the variability of the cluster-level means.

4. Sell Phones?

You are an employee of a newspaper and are planning an experiment to demonstrate to Apple that online advertising on your website causes people to buy iPhones. Each site visitor shown the ad campaign is exposed to \$0.10 worth of advertising for iPhones. (Assume all users could see ads.) There are 1,000,000 users available to be shown ads on your newspaper's website during the one week campaign.

Apple indicates that they make a profit of \$100 every time an iPhone sells and that 0.5% of visitors to your newspaper's website buy an iPhone in a given week in general, in the absence of any advertising.

- a. By how much does the ad campaign need to increase the probability of purchase in order to be "worth it" and a positive ROI (supposing there are no long-run effects and all the effects are measured within that week)?

ANS: The ad campaign need to increase the prob of purchase by another 0.1%.

```
population <- 1000000
ad_cost <- population*0.1
equivalent_iphone_sales <- ad_cost / 100
marginal_increase <- equivalent_iphone_sales / population
```

```
marginal_increase
```

```
## [1] 0.001
```

```
# The cost of placing the ad is 1,000,000 x $0.1 = $100,000
# This is translated into $100,000 / $100 = 1,000 iPhone sales
# Thus the ad campaign need to increase the probability of
# purchase by another 0.1% (1,000 / 1,000,000) to make
# it worth it for Apple to place the ad.
```

- b. Assume the measured effect is 0.2 percentage points. If users are split 50:50 between the treatment group (exposed to iPhone ads) and control group (exposed to unrelated advertising or nothing; something you can assume has no effect), what will be the confidence interval of your estimate on whether people purchase the phone?

```
# Reference: https://newonlinecourses.science.psu.edu/stat500/node/55/

# Function to calculate standard error of two populations being compared
se_apple <-function(p1, p2, n1, n2){
  return(sqrt((p1*(1-p1)/n1)+(p2*(1-p2)/n2)))
}

# Calculate interval
interval_apple <- 1.96*se_apple(0.005, 0.007, 500000, 500000)

# Calculate the length of each tail
confidence_interval <- c(0.002 - interval_apple, 0.002 + interval_apple)

# Print confidence interval
confidence_interval
```

```
## [1] 0.001697296 0.002302704
```

- **Note:** The standard error for a two-sample proportion test is $\sqrt{p(1-p) * (\frac{1}{n_1} + \frac{1}{n_2})}$ where $p = \frac{x_1+x_2}{n_1+n_2}$, where x and n refer to the number of “successes” (here, purchases) over the number of “trials” (here, site visits). The length of each tail of a 95% confidence interval is calculated by multiplying the standard error by 1.96.
- c. Is this confidence interval precise enough that you would recommend running this experiment? Why or why not?

ANS: Yes, the experiment of measured size 0.2% is worth carrying out because the lower bound of the estimated effect of 0.0016 is already above the ROI threshold of 0.001.

- d. Your boss at the newspaper, worried about potential loss of revenue, says he is not willing to hold back a control group any larger than 1% of users. What would be the width of the confidence interval for this experiment if only 1% of users were placed in the control group?

```
# let's re-estimate the interval length by adjusting se
n_control_apple <- population*0.99
n_treatment_apple <- population - n_control_apple
```

```

# Calculate interval
interval_apple2 <- 1.96*se_apple(0.005, 0.007, n_control_apple, n_treatment_apple)

# Calculate the length of each tail
confidence_interval2 <- c(0.002 - interval_apple2, 0.002 + interval_apple2)

# Print confidence interval
confidence_interval2

## [1] 0.0003599996 0.0036400004

```

5. Sports Cards

Here you will find a set of data from an auction experiment by John List and David Lucking-Reiley (2000).

```

d2 <- read.csv("./data/listData.csv", stringsAsFactors = FALSE)
head(d2)

```

```

##   bid uniform_price_auction
## 1   5                      1
## 2   5                      1
## 3  20                      0
## 4   0                      1
## 5  20                      1
## 6   0                      1

```

In this experiment, the experimenters invited consumers at a sports card trading show to bid against one other bidder for a pair trading cards. We abstract from the multi-unit-auction details here, and simply state that the treatment auction format was theoretically predicted to produce lower bids than the control auction format. We provide you a relevant subset of data from the experiment.

- Compute a 95% confidence interval for the difference between the treatment mean and the control mean, using analytic formulas for a two-sample t-test from your earlier statistics course.

ANS: CI = [-20.68596, -3.725801].

```

d2 <- data.table(d2)
t.test(d2[uniform_price_auction == 1, (bid)], d2[uniform_price_auction == 0, (bid)], mu = 0)

##
## Welch Two Sample t-test
##
## data:  d2[uniform_price_auction == 1, (bid)] and d2[uniform_price_auction == 0, (bid)]
## t = -2.8211, df = 61.983, p-value = 0.006421
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -20.854624 -3.557141
## sample estimates:
## mean of x mean of y
## 16.61765 28.82353

```

- In plain language, what does this confidence interval mean?

ANS: If many samples are collected and the confidence interval computed, about 95 % of these intervals would contain the true mean of the population.

- c. Regression on a binary treatment variable turns out to give one the same answer as the standard analytic formula you just used. Demonstrate this by regressing the bid on a binary variable equal to 0 for the control auction and 1 for the treatment auction.

```
summary(lm(d2[,bid] ~ d2[,uniform_price_auction]))
```

```
##
## Call:
## lm(formula = d2[, bid] ~ d2[, uniform_price_auction])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -28.824 -11.618  -3.221   8.382  58.382
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      28.824      3.059   9.421 7.81e-14 ***
## d2[, uniform_price_auction] -12.206      4.327  -2.821  0.00631 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 17.84 on 66 degrees of freedom
## Multiple R-squared:  0.1076, Adjusted R-squared:  0.09409
## F-statistic: 7.959 on 1 and 66 DF,  p-value: 0.006315
```

- d. Calculate the 95% confidence interval you get from the regression. **ANS: CI = [-20.68596, -3.725801].**

```
regression <- summary(lm(d2[,bid] ~ d2[,uniform_price_auction]))
regression
```

```
##
## Call:
## lm(formula = d2[, bid] ~ d2[, uniform_price_auction])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -28.824 -11.618  -3.221   8.382  58.382
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      28.824      3.059   9.421 7.81e-14 ***
## d2[, uniform_price_auction] -12.206      4.327  -2.821  0.00631 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 17.84 on 66 degrees of freedom
## Multiple R-squared:  0.1076, Adjusted R-squared:  0.09409
## F-statistic: 7.959 on 1 and 66 DF,  p-value: 0.006315
```

```
estimate <- regression$coefficients[2, 1]
standard_error <- regression$coefficients[2, 2]
ci_sport <- c(estimate-standard_error*1.997, estimate+standard_error*1.997) # 95% confidence interval of
ci_sport
```

```
## [1] -20.846047 -3.565718
```

- e. On to p-values. What p-value does the regression report? Note: please use two-tailed tests for the entire problem.

ANS: p_value = 0.006314796.

```
p_value <- regression$coefficients[2,4]
p_value
```

```
## [1] 0.006314796
```

- f. Now compute the same p-value using randomization inference.

ANS: p-value = 0.3788.

```
d3 <- d2
total_row <- nrow(d3)
total_row
```

```
## [1] 68
```

```
rownum <- nrow(d3[uniform_price_auction == 0, ])
rownum
```

```
## [1] 34
```

```
randomized_sample <- function (){
  a <- c(rep(0,(total_row -1.5 * rownum )))
  b <- c(rep(1,rownum/2))
  d3[uniform_price_auction == 1, treat := sample(c(a,b))]
  d3[uniform_price_auction == 0, treat := sample(c(rep(0,rownum/2),
                                                    rep(1,rownum/2)))]
  po_control <- d3[(uniform_price_auction == 0) & (treat == 1), bid]
  po_treat <- d3[(uniform_price_auction == 1) & (treat == 1), bid]
  est_ate <- mean(po_treat) - mean(po_control)
  return (est_ate)
}
ate <- randomized_sample()
ate
```

```
## [1] -14.82353
```

```
distribution_under_sharp_null_sport <- replicate(10000, randomized_sample())
head(distribution_under_sharp_null_sport)
```

```
## [1] -13.823529 -8.823529 -9.000000 -10.235294 -13.000000 -14.294118
```

```
p_value <- mean(ate < distribution_under_sharp_null_sport)
p_value
```

```
## [1] 0.7327
```

- g. Compute the same p-value again using analytic formulas for a two-sample t-test from your earlier statistics course. (Also see part (a).)

ANS:p-value = 0.3945283.

```
data <- d3[uniform_price_auction == 1,(bid)]-d3[uniform_price_auction == 0,(bid)]
t_value <- (mean(data) - 0)/(sd(data))/sqrt(length(data))
p_value <- dt(t_value, df=length(data) - 1)
p_value
```

```
## [1] 0.3945283
```

- h. Compare the two p-values in parts (e) and (f). Are they much different? Why or why not? How might your answer to this question change if the sample size were different?

ANS: Compare the two p-values in parts (e) and (f), they are very different. I expect the p-value gap b/w (e) and (f) to be close when I increase the sampling size from 10e5 to 10e6 for randomization inference.