

W241 Problem Set 2

Yang Yang Qian

1. FE exercise 3.6

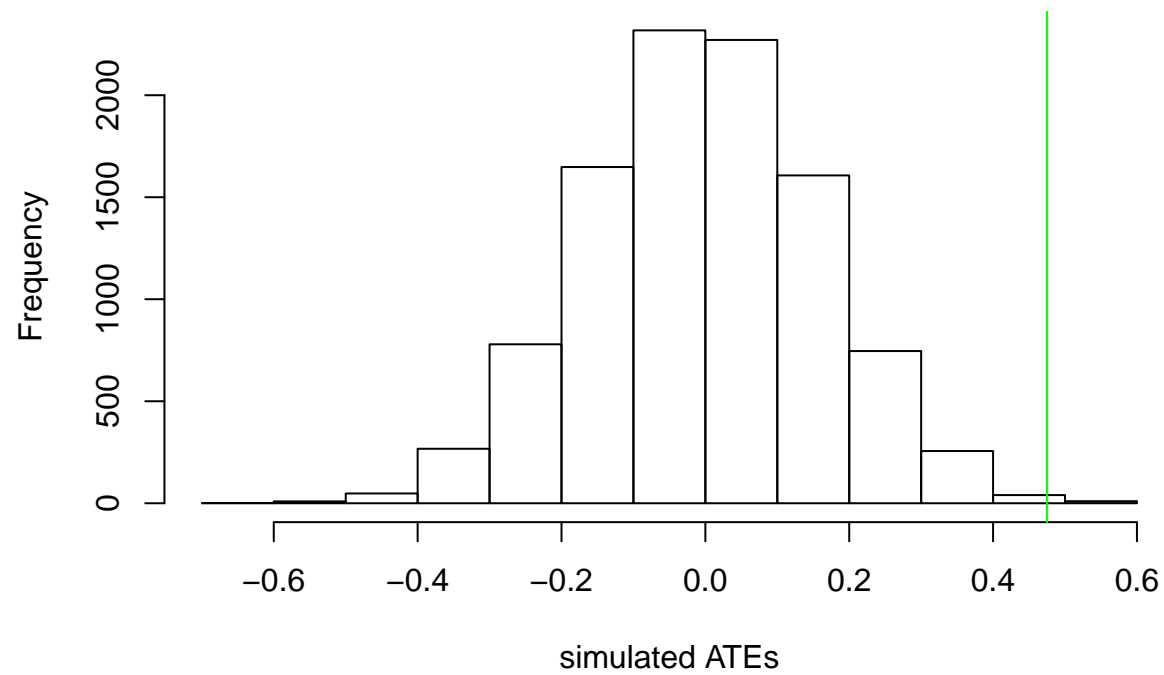
The Clingingsmith, Khwaja, and Kremer study discussed in section 3.5 may be used to test the sharp null hypothesis that winning the visa lottery for the pilgrimage to Mecca had no effect on the views of Pakistani Muslims toward people from other countries. Assume that the Pakistani authorities assigned visas using complete random assignment.

1.a.

Conduct 10,000 simulated random assignments under the sharp null hypothesis. (Don't just copy the code from the async, think about how to write this yourself.)

```
# conducting 1 simulated random assignment under the sharp null hypothesis (e.g. matters not if person won visa, their view would not have  
# so we flip a fair coin for each person in population, 1 = won visa (treatment), 0 = did not win visa (control)  
# ... and we don't change the views, because we are assuming that assignment to win or not win a visa would not have changed it at all  
  
# set.seed(87542)  
# q1$d3.6 <- q1$d3.6[, success.q1.a := replicate(nrow(q1$d3.6), rbinom(1,1,0.5))]  
# q1$m2 <- q1$d3.6[,lm(views ~ success.q1.a)]  
# q1$ate.q1.a <- q1$m2$coefficients[2]  
  
# to conduct 10,000 simulated random assignments under the sharp null hypothesis, just need to create a simulate study function and call r  
# then get a distribution of ATEs  
set.seed(87542)  
q1$sim.study <- function() {  
  q1$d3.6 <- q1$d3.6[, success.q1.a := replicate(nrow(q1$d3.6), rbinom(1,1,0.5))]  
  q1$m2 <- q1$d3.6[,lm(views ~ success.q1.a)]  
  q1$ate.q1.a <- q1$m2$coefficients[2]  
  return(q1$ate.q1.a)  
}  
q1$sim.results <- replicate(10000, q1$sim.study())  
hist(q1$sim.results, xlab="simulated ATEs", main=paste("Frequencies of simulated ATEs, vs actual estimated ATE: ",round(q1$ate,3)))  
abline(v=q1$ate,col="green")
```

Frequencies of simulated ATEs, vs actual estimated ATE: 0.475



1.b.

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

about 15

```
data.table(q1$sim.results)[,sim.ate.as.large.as.actual.ate:=(q1$sim.results >= q1$ate)][,.N,by=sim.ate.as.large.as.actual.ate]
```

```
##      sim.ate.as.large.as.actual.ate      N
## 1:                                FALSE 9985
## 2:                                TRUE   15
```

1.c.

What is the implied one-tailed p-value?

implied one-tailed p-value is about 0.0015

```
mean(q1$sim.results > q1$ate)
```

```
## [1] 0.0015
```

1.d.

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?

about 31, roughly double that of the one-tailed test

```
data.table(q1$sim.results)[,sim.ate.abs.as.large.as.actual.ate:=(abs(q1$sim.results) >= abs(q1$ate))][, .N, by=sim.ate.abs.as.large.as.actual.ate]
```

```
##      sim.ate.abs.as.large.as.actual.ate      N
## 1:                                FALSE 9969
## 2:                                TRUE   31
```

1.e.

What is the implied two-tailed p-value?

implied two-tailed p-value is about 0.0031

```
mean(abs(q1$sim.results) > abs(q1$ate))
```

```
## [1] 0.0031
```

```
# or
```

```
mean(q1$sim.results > q1$ate) * 2
```

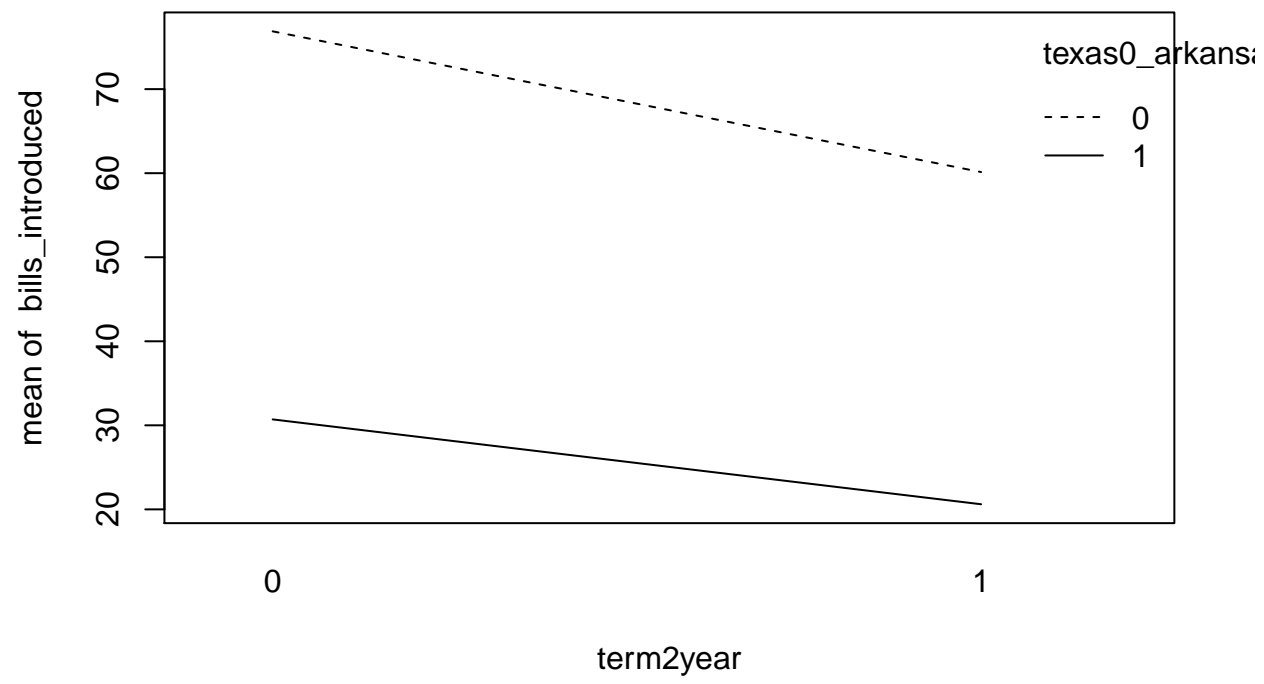
```
## [1] 0.003
```


2.FE exercise 3.8

Naturally occurring experiments sometimes involve what is, in effect, block random assignment. For example, Titunik 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. An interesting outcome variable is 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.

If you're interested, or would like more clarification, the published version of the paper is in the repository.

```
library(foreign)
library(data.table)
q2 <- {}
q2$d3.8 <- data.table( read.dta("./data/Titunik.2010.dta") )
# head(q2$d3.8)
# summary(q2$d3.8)
# str(q2$d3.8)
# hist(q2$d3.8$term2year)
# hist(q2$d3.8$bills_introduced)
# hist(q2$d3.8$texas0_arkansas1)
# q2$d3.8[texas0_arkansas1==0,plot(bills_introduced ~ term2year,main="Texas")]
# q2$d3.8[texas0_arkansas1==1,plot(bills_introduced ~ term2year,main="Arkansas")]
q2$d3.8[,interaction.plot(term2year,texas0_arkansas1,bills_introduced)]
```



NULL

2.a.

For each state, estimate the effect of having a two-year term on the number of bills introduced.

for Texas, the effect of having a two-year term on the number of bills introduced is about -17 bills

for Arkansas, the effect of having a two-year term on the number of bills introduced is about -11 bills

```
# use basic methods to estimate the effect on number of bills introduced, by the 2-year term, and the blocking variable texas-vs-arkansas
dcast(
  q2$d3.8[,list(mu=mean(bills_introduced)),by=c("term2year","texas0_arkansas1")]
  ,texas0_arkansas1 ~ term2year
  ,value.var = "mu"
)
```

```
##      texas0_arkansas1      0      1
## 1:                0 76.87500 60.13333
## 2:                1 30.70588 20.61111
```

```
q2$ate.2y.texas.basic <- q2$d3.8[texas0_arkansas1==0 & term2year==1,mean(bills_introduced)] - q2$d3.8[texas0_arkansas1==0 & term2year==0,mean(bills_introduced)]
# -16.74167
```

```
q2$ate.2y.arkansas.basic <- q2$d3.8[texas0_arkansas1==1 & term2year==1,mean(bills_introduced)] - q2$d3.8[texas0_arkansas1==1 & term2year==0,mean(bills_introduced)]
# -10.09477
```

```
# use OLS regression to estimate the effect on number of bills introduced, by the 2-year term, and the blocking variable texas-vs-arkansas
q2$m1 <- q2$d3.8[texas0_arkansas1==0,lm(bills_introduced ~ term2year)]
(q2$m1)
```

```
##
## Call:
## lm(formula = bills_introduced ~ term2year)
##
## Coefficients:
## (Intercept)      term2year
##          76.88         -16.74
```

```
q2$m2 <- q2$d3.8[texas0_arkansas1==1,lm(bills_introduced ~ term2year)]
(q2$m2)
```

```
##
```

```
## Call:
## lm(formula = bills_introduced ~ term2year)
##
## Coefficients:
## (Intercept)      term2year
##      30.71      -10.09

q2$ate.2y.texas.ols <- q2$m1$coefficients[2] # -16.74167
q2$ate.2y.arkansas.ols <- q2$m2$coefficients[2] # -10.09477

# use OLS in one model, with interaction variable between the term2year and texas0_arkansas1
q2$m3 <- q2$d3.8[,lm(bills_introduced ~ term2year + texas0_arkansas1 + term2year*texas0_arkansas1)]
(q2$m3)

##
## Call:
## lm(formula = bills_introduced ~ term2year + texas0_arkansas1 +
##      term2year * texas0_arkansas1)
##
## Coefficients:
## (Intercept)      term2year
##      76.875      -16.742
##      texas0_arkansas1 term2year:texas0_arkansas1
##      -46.169      6.647

q2$ate.2y.texas.ols2 <- q2$m3$coefficients[2] # -16.74167
q2$ate.2y.arkansas.ols2 <- q2$m3$coefficients[2] + q2$m3$coefficients[4] # -10.09477

# use OLS to predict, incorrectly, the overall ATE
q2$m4 <- q2$d3.8[,lm(bills_introduced ~ term2year)]
(q2$m4)

##
## Call:
## lm(formula = bills_introduced ~ term2year)
##
## Coefficients:
## (Intercept)      term2year
```

```
##          53.09          -14.52
```

```
q2$ate.overall.bad.ols <- q2$m4$coefficients[2] # -14.51515
```

2.b.

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

for texas, se is about 10

for arkansas, se is about 4

note, from GG, equation 3.6 for estimating the standard error of the ATE (no blocking) is:

$$\hat{SE} = \sqrt{\frac{\hat{Var}(Y_i(0))}{N - m} + \frac{\hat{Var}(Y_i(1))}{m}}$$

```
# estimate the se for each state, using basic methods
# var0 = variance of control, n0 = number of units in control
# var1 = variance of treatment, n1 = number of units in treatment
q2$se.func <- function(var0, var1, n0, n1) {
  return( sqrt(var0/n0 + var1/n1) )
}

# calculate the standard error of the estimated ATE, for each state
q2$var0.texas <- var(q2$d3.8[term2year==0 & texas0_arkansas1==0,bills_introduced]) #var0 for texas, 956.25
q2$var1.texas <- var(q2$d3.8[term2year==1 & texas0_arkansas1==0,bills_introduced]) #var1 for texas, 413.6952
q2$var0.arkansas <- var(q2$d3.8[term2year==0 & texas0_arkansas1==1,bills_introduced]) #var0 for arkansas, 148.5956
q2$var1.arkansas <- var(q2$d3.8[term2year==1 & texas0_arkansas1==1,bills_introduced]) #var1 for arkansas, 50.25163
q2$n0.texas <- q2$d3.8[term2year==0 & texas0_arkansas1==0,.N] #n0 for texas, 16
q2$n1.texas <- q2$d3.8[term2year==1 & texas0_arkansas1==0,.N] #n1 for texas, 15
q2$n0.arkansas <- q2$d3.8[term2year==0 & texas0_arkansas1==1,.N] #n0 for arkansas, 17
q2$n1.arkansas <- q2$d3.8[term2year==1 & texas0_arkansas1==1,.N] #n1 for arkansas, 18

q2$se.texas <- q2$se.func(q2$var0.texas, q2$var1.texas, q2$n0.texas, q2$n1.texas) #se for texas block, 9.345871
q2$se.arkansas <- q2$se.func(q2$var0.arkansas, q2$var1.arkansas, q2$n0.arkansas, q2$n1.arkansas) #se for arkansas block, 3.395979

# estimate the se for each state, using libraries
# TBD
```

2.c.

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

overall ATE for both states combined is about -14 bills

note, equation 3.10 is:

$$ATE = \sum_{j=1}^J \frac{N_j}{N} ATE_j$$

```
# total ATE is the sum of the ATEs within each block, scaled by the number of units in each block
q2$ate.total = (q2$n.texas/q2$n.all) * q2$ate.2y.texas.basic + (q2$n.arkansas/q2$n.all) * q2$ate.2y.arkansas.basic
(q2$ate.total) # -13.2168
```

```
## numeric(0)
```

2.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.

When pooling the data for the two states, we get an estimated overall ATE of about -15 bills. Whereas, if we correctly calculated the overall estimated ATE, it should be about -14 bills. Pooling the data leads to a biased estimate of the overall ATE because the variances and ATE for Texas & Arkansas blocks were different from each other.

2.e.

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

overall se is about 18

The equation 3.12 for estimating the standard error of the ATE with 2 blocks is:

$$SE(\hat{ATE}) = \sqrt{(SE_1)^2 \left(\frac{N_1}{N}\right)^2 + (SE_2)^2 \left(\frac{N_2}{N}\right)^2}$$

```
# using results from previous section, calculate the blocked standard error of the estimated ATE
q2$n.texas <- q2$d3.8[texas0_arkansas1==0,.N] #n for texas block, 31
q2$n.arkansas <- q2$d3.8[texas0_arkansas1==1,.N] #n for arkansas block, 35
q2$n.all <- q2$d3.8[,.N] #n for all, 66

q2$se.all <- sqrt( q2$se.texas^2 * (q2$n.texas/q2$n.all)^2 + q2$n.texas^2 * (q2$n.arkansas/q2$n.all)^2 ) #se for overall ATE, 17.01539
(q2$se.all)

## [1] 17.01539
```

2.f.

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

using randomization inference technique, we saw that implied p-values for both states were less than 0.05, so we can reject the sharp null hypothesis.

```
# for a round of randomization inference, we will randomly assign senators to 2-year terms for both texas and arkansas
set.seed(12659)
q2$d3.8 <- q2$d3.8[,term2yearRI:=replicate(nrow(q2$d3.8), rbinom(1,1,0.5))]]

# use basic methods to calculate the ATE for each state
q2$ate.texas.ri <- q2$d3.8[texas0_arkansas1==0 & term2yearRI==1,mean(bills_introduced)] - q2$d3.8[texas0_arkansas1==0 & term2yearRI==0,mean(bills_introduced)]
# -19.48246

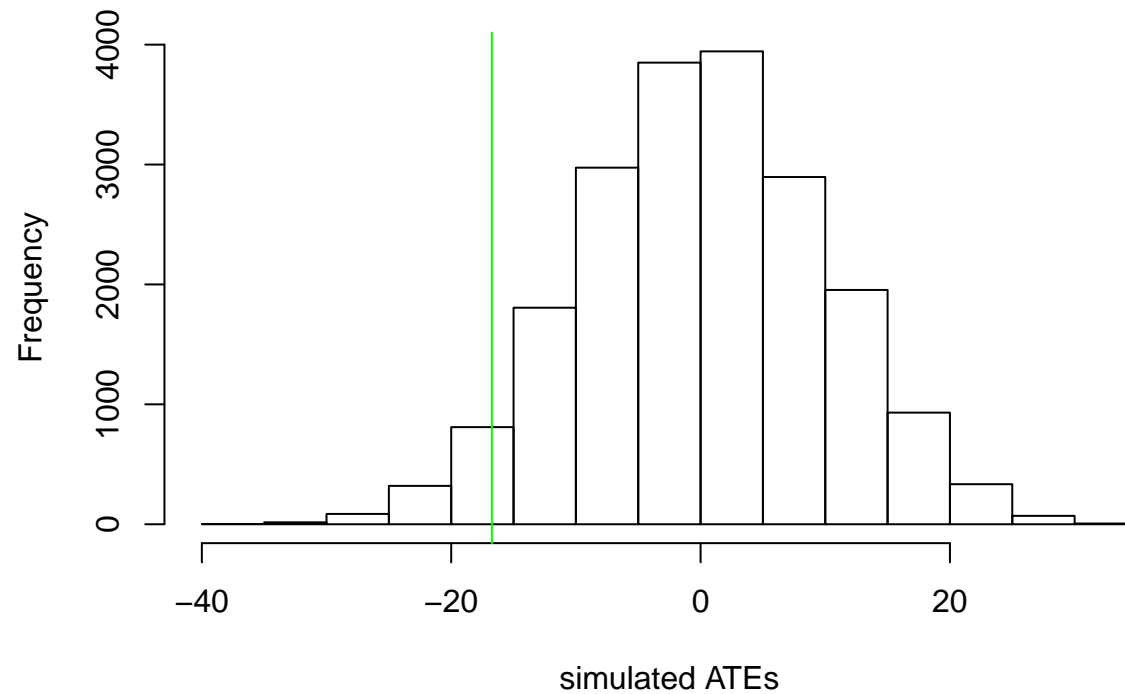
q2$ate.arkansas.ri <- q2$d3.8[texas0_arkansas1==1 & term2yearRI==1,mean(bills_introduced)] - q2$d3.8[texas0_arkansas1==1 & term2yearRI==0,mean(bills_introduced)]
# -0.7156863

# creates a function to simulate the randomization inference
q2$sim.study <- function() {
  q2$d3.8 <- q2$d3.8[,term2yearRI:=replicate(nrow(q2$d3.8), rbinom(1,1,0.5))]]
  q2$m5 <- q2$d3.8[,lm(bills_introduced ~ term2yearRI + texas0_arkansas1 + term2yearRI*texas0_arkansas1)]
  q2$ate.texas.ri <- q2$m5$coefficients[2]
  q2$ate.arkansas.ri <- q2$m5$coefficients[2] + q2$m5$coefficients[4]
  return(list(q2$ate.texas.ri, q2$ate.arkansas.ri))
}

q2$sim.results <- replicate(10000, q2$sim.study())

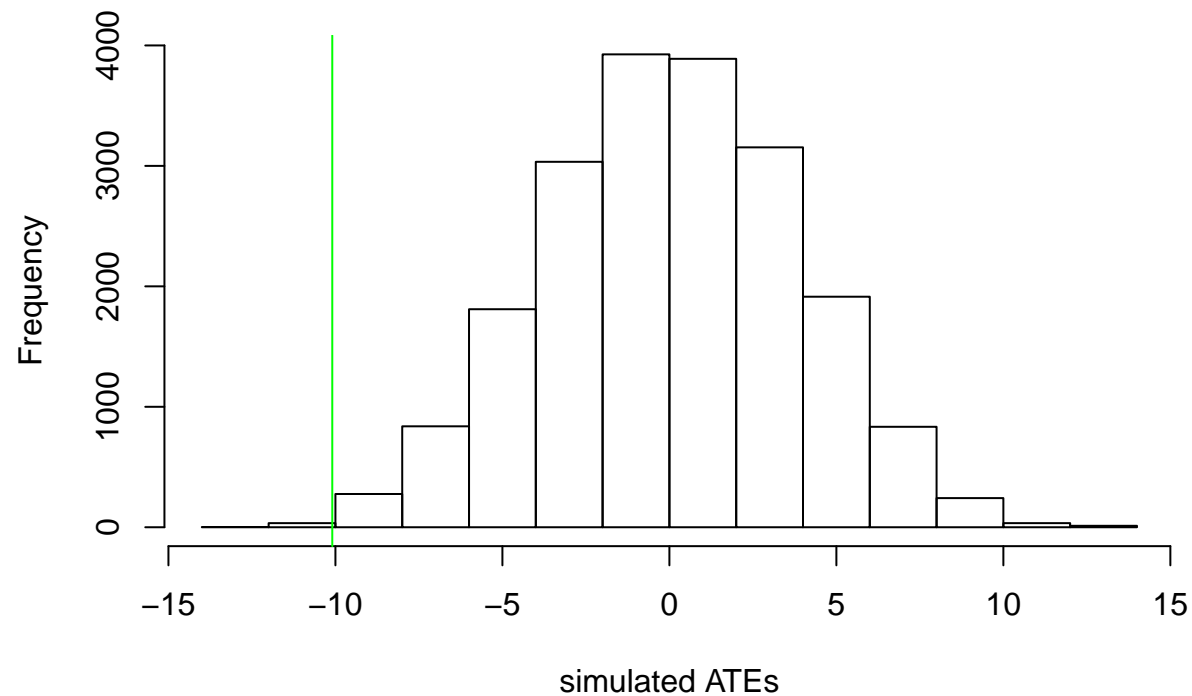
# extracts the simulated ATEs for texas and arkansas
q2$ate.texas.sim <- unlist( q2$sim.results[1,1:length(q2$sim.results)/2] )
hist(q2$ate.texas.sim, main="simulated estimated ATE for Texas, vs actual estimated ATE", xlab="simulated ATEs")
abline(v=q2$ate.2y.texas.basic,col="green")
```

simulated estimated ATE for Texas, vs actual estimated ATE



```
q2$ate.arkansas.sim <- unlist( q2$sim.results[2,1:length(q2$sim.results)/2] )  
hist(q2$ate.arkansas.sim, main="simulated estimated ATE for Arkansas, vs actual estimated ATE", xlab="simulated ATEs")  
abline(v=q2$ate.2y.arkansas.basic,col="green")
```

simulated estimated ATE for Arkansas, vs actual estimated ATE



```
# calculates the implied p-value for each state  
mean(q2$ate.texas.sim < q2$ate.2y.texas.basic) #p-value for texas, 0.04470224
```

```
## [1] 0.04260213
```

```
mean(q2$ate.arkansas.sim < q2$ate.2y.arkansas.basic) #p-value for arkansas, 0.002100105
```

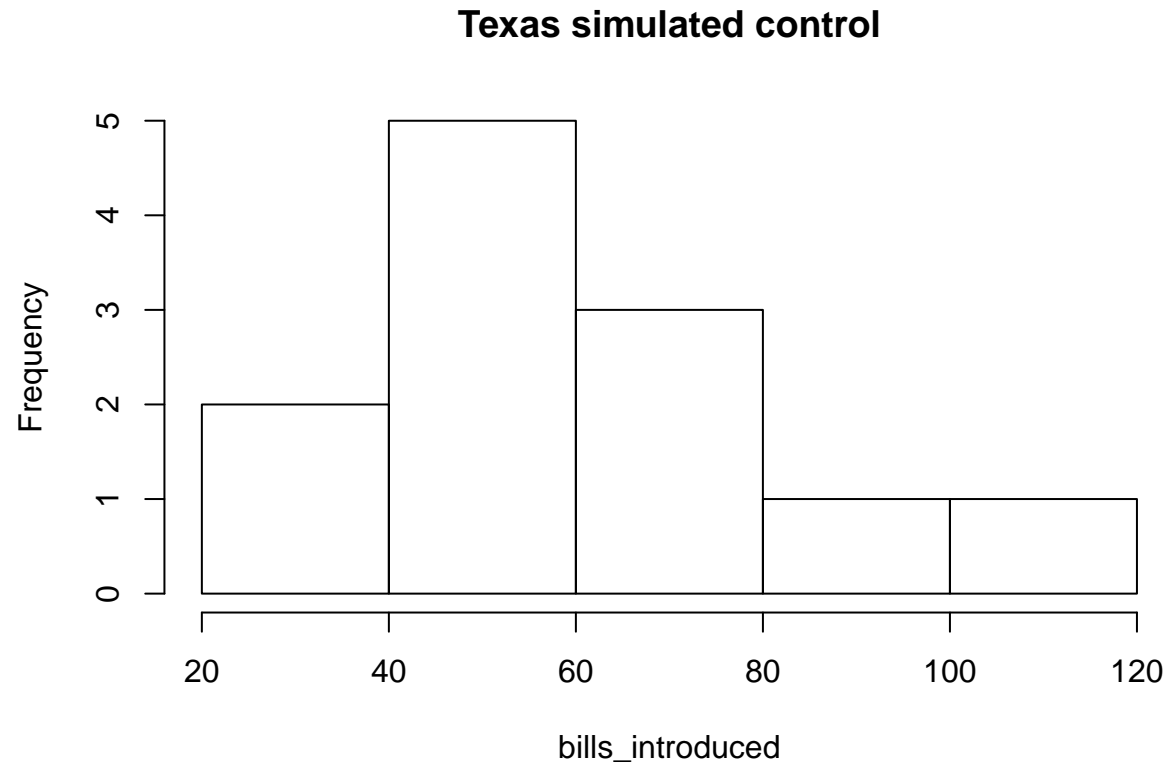
```
## [1] 0.00180009
```

```
# both p-values are < than 0.05, so we can reject the sharp null hypothesis
```

2.g.

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

```
q2$d3.8[texas0_arkansas1==0 & term2yearRI==0,hist(bills_introduced,main="Texas simulated control")]
```



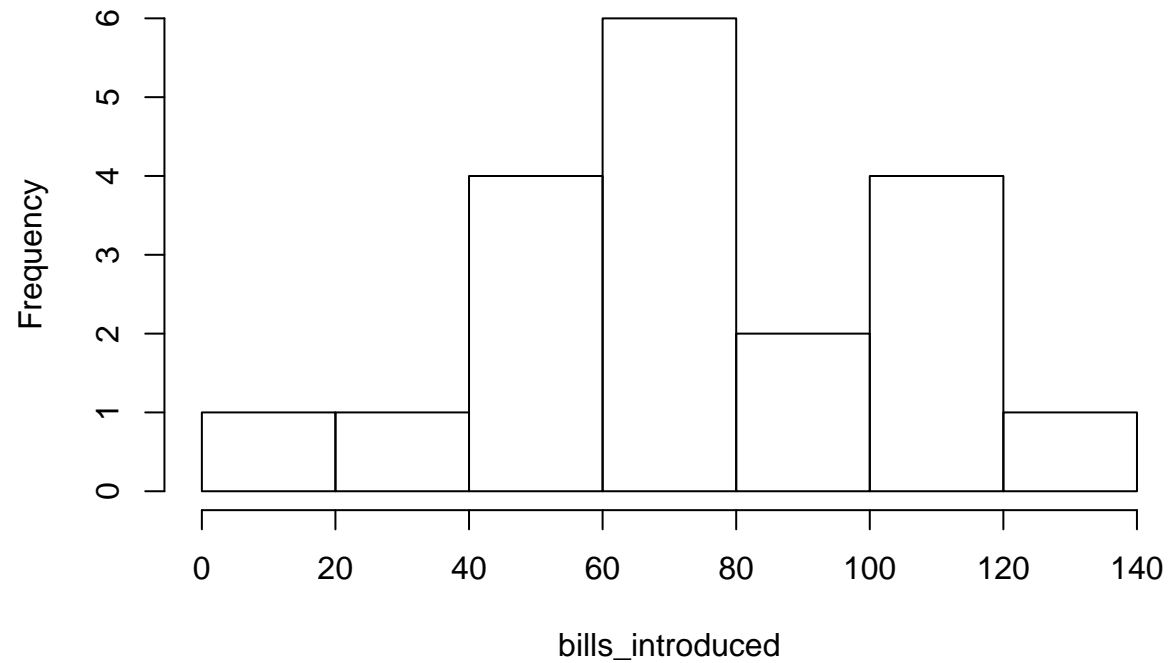
```
## $breaks
## [1] 20 40 60 80 100 120
##
## $counts
## [1] 2 5 3 1 1
##
```

```

## $density
## [1] 0.008333333 0.020833333 0.012500000 0.004166667 0.004166667
##
## $mids
## [1] 30 50 70 90 110
##
## $xname
## [1] "bills_introduced"
##
## $equidist
## [1] TRUE
##
## attr("class")
## [1] "histogram"
q2$d3.8[texas0_arkansas1==0 & term2yearRI==1,hist(bills_introduced,main="Texas simulated treatment")]

```

Texas simulated treatment

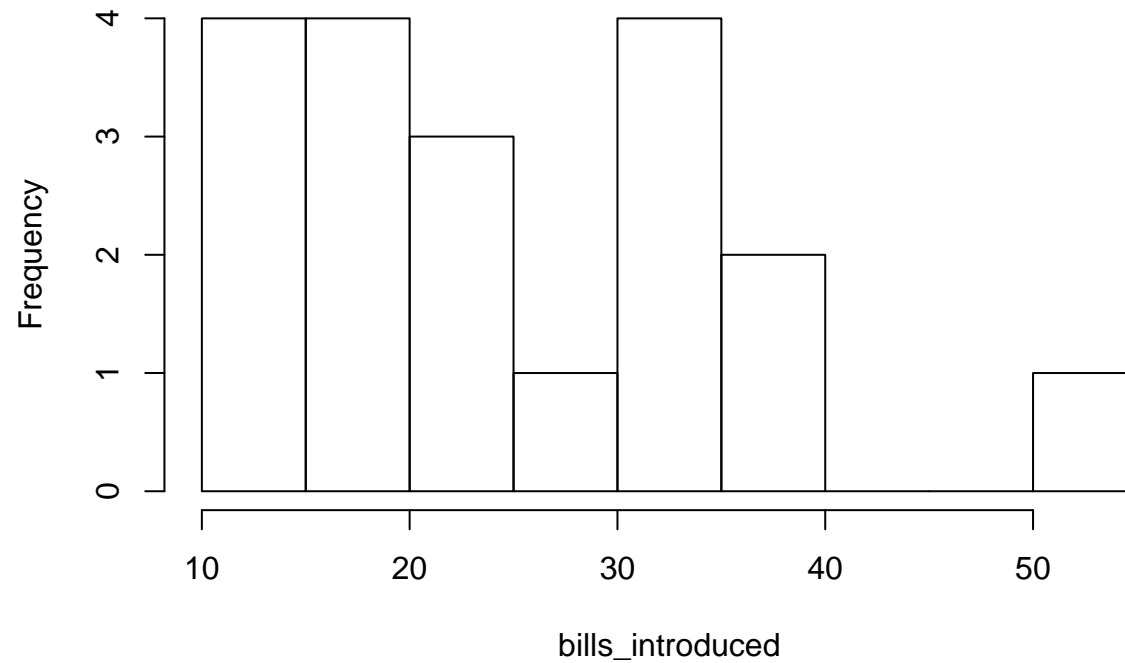


```
## $breaks
## [1]  0  20  40  60  80 100 120 140
##
## $counts
## [1] 1 1 4 6 2 4 1
##
## $density
## [1] 0.002631579 0.002631579 0.010526316 0.015789474 0.005263158 0.010526316
## [7] 0.002631579
##
## $mids
```

```
## [1] 10 30 50 70 90 110 130
##
## $xname
## [1] "bills_introduced"
##
## $equidist
## [1] TRUE
##
## attr("class")
## [1] "histogram"
```

```
q2$d3.8[texas0_arkansas1==1 & term2yearRI==0,hist(bills_introduced,main="Arkansas simulated control")]
```


Arkansas simulated control



```
## $breaks
## [1] 10 15 20 25 30 35 40 45 50 55
##
## $counts
## [1] 4 4 3 1 4 2 0 0 1
##
## $density
## [1] 0.04210526 0.04210526 0.03157895 0.01052632 0.04210526 0.02105263
## [7] 0.00000000 0.00000000 0.01052632
##
## $mids
```

```
## [1] 12.5 17.5 22.5 27.5 32.5 37.5 42.5 47.5 52.5
```

```
##
```

```
## $xname
```

```
## [1] "bills_introduced"
```

```
##
```

```
## $equidist
```

```
## [1] TRUE
```

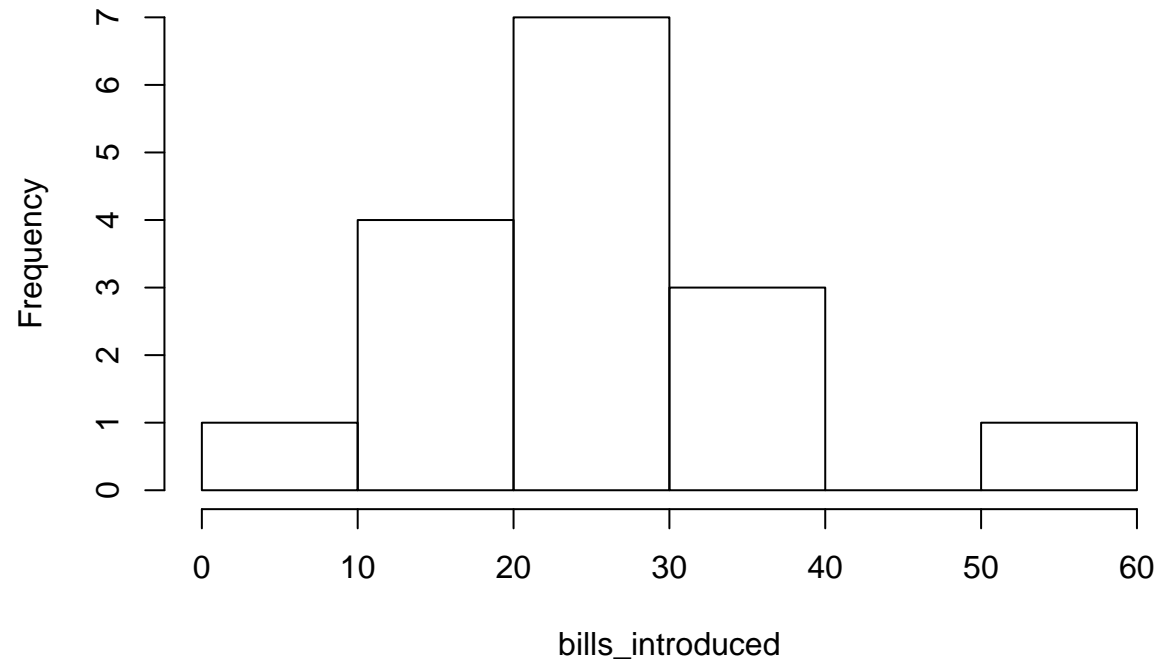
```
##
```

```
## attr("class")
```

```
## [1] "histogram"
```

```
q2$d3.8[texas0_arkansas1==1 & term2yearRI==1,hist(bills_introduced,main="Arkansas simulated treatment")]
```

Arkansas simulated treatment



```
## $breaks
## [1]  0 10 20 30 40 50 60
##
## $counts
## [1] 1 4 7 3 0 1
##
## $density
## [1] 0.00625 0.02500 0.04375 0.01875 0.00000 0.00625
##
## $mids
## [1]  5 15 25 35 45 55
```

```
##  
## $xname  
## [1] "bills_introduced"  
##  
## $equidist  
## [1] TRUE  
##  
## attr("class")  
## [1] "histogram"
```

3. FE exercise 3.11

Use the data in 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  
d <- read.csv("../data/ggChapter3.csv")
```

3.a.

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

3.b.

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

3.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?

4. More Practice #1

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.

4.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)?

4.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?

- **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.

4.c.

Is this confidence interval precise enough that you would recommend running this experiment? Why or why not?

4.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?

5. More Practice #2

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")  
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.

5.a.

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.

5.b.

In plain language, what does this confidence interval mean?

5.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.

5.d.

Calculate the 95% confidence interval you get from the regression.

5.e.

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

5.f.

Now compute the same p-value using randomization inference.

5.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).)

5.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?