W241 Problem Set 2

Yang Yang Qian

1. FE exercise 3.6

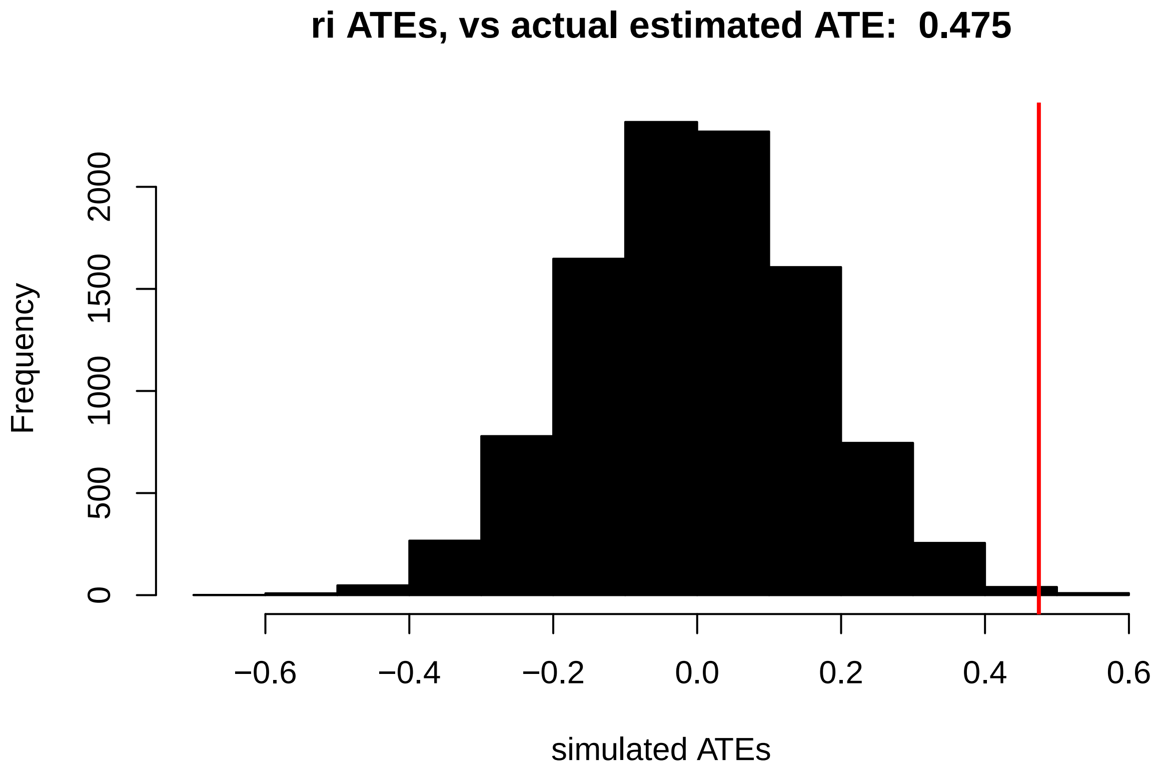
The Clingingsmith, Khwaja, and Kremer study discussed in section 3.5 may be be used to test the sharp null hypothesis that winning the visa lottery for the pilgrimage to Mecca had no effect on the views of Pakistani Muslims toward people from other countries. Assume that the Pakistani authorities assigned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

```
## Warning: package 'data.table' was built under R version 3.3.3
```

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
# 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")

# slightly more efficient code, from Alex's week3 RI code
q1$sim.study.2 <- function() {
    q1$d3.6 <- q1$d3.6[, success.q1.a := rbinom(.N,1,p=0.5)]
    q1$ate.q1.a <- q1$d3.6[,.(mu=mean(views)),by=success.q1.a][,mu[success.q1.a==1]-mu[success.q1.a==0]]
    return(q1$ate.q1.a)
}
q1$sim.results <- replicate(10000, q1$sim.study.2())
hist(q1$sim.results, xlab="simulated ATEs", col="black", main=paste("ri ATEs, vs actual estimated ATE: ",round(q1$ate,3)))
abline(v=q1$ate,lwd=2,col="red")
```


ri ATEs, vs actual estimated ATE: 0.475
Frequency
0
500
1000
1500
2000
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6
simulated ATEs

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 simulated ATEs were at least as large as the actual estimated ATE

```r
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.actua
```

```
##    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, Titiunik 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.

```r
library(foreign)
library(data.table)
q2 <- {}
q2$d3.8 <- data.table( read.dta("./data/Titiunik.2010.dta") )
# head(q2$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)]
```

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,m
# -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==


# 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 bills

for Arkansas, se is about 4 bills

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 analytical formulas
# 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) )
}


# calcuate 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$n.texas <- q2$d3.8[texas0_arkansas1==0,.N] #n for texas, 31
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$n.arkansas <- q2$d3.8[texas0_arkansas1==1,.N] #n for arkansas, 35
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$n.all <- q2$d3.8[,.N] #n for all, 66

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
```

```r
# estimate the robust se for each state, using the regression models
library(sandwich)
```

```
## Warning: package 'sandwich' was built under R version 3.3.2
```

```r
# regression for texas
summary(q2$m1)
```

```
##
## Call:
## lm(formula = bills_introduced ~ term2year)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -58.875 -16.004   0.867  15.496  54.125
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   76.875      6.588  11.670 1.78e-12 ***
## term2year    -16.742      9.470  -1.768   0.0876 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 26.35 on 29 degrees of freedom
## Multiple R-squared:  0.09728,    Adjusted R-squared:  0.06615
## F-statistic: 3.125 on 1 and 29 DF,  p-value: 0.08761
```

```r
q2$m1$vcov <- vcovHC(q2$m1, type="HC3")
q2$m1$r.se <- sqrt(diag(q2$m1$vcov))
q2$m1$r.se # term2year robust se is 9.659175
```

```
## (Intercept)   term2year
##    7.984360    9.659175
```

```r
# regression for arkansas
summary(q2$m2)
```

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

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.7059  -5.6585  -0.6111   4.2941  28.2941
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   30.706      2.400  12.793 2.43e-14 ***
## term2year    -10.095      3.347  -3.016   0.0049 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.896 on 33 degrees of freedom
## Multiple R-squared:  0.2161, Adjusted R-squared:  0.1923
## F-statistic: 9.097 on 1 and 33 DF,  p-value: 0.004899
```

```
q2$m2$vcov <- vcovHC(q2$m2, type="HC3")
q2$m2$r.se <- sqrt(diag(q2$m2$vcov))
q2$m2$r.se # term2year robust se is 3.499029
```

```
## (Intercept)   term2year
##    3.047495    3.499029
```

```
# not sure why the robust standard error doesn't quite match up with the analytical method
# though if we're measuring number of whole bills, it still matches up to Texas se 10, and Arkansas se 4. Close enough for government work
```

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
```

```
## [1] -13.2168
```

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 bills

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

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

```
# 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
```

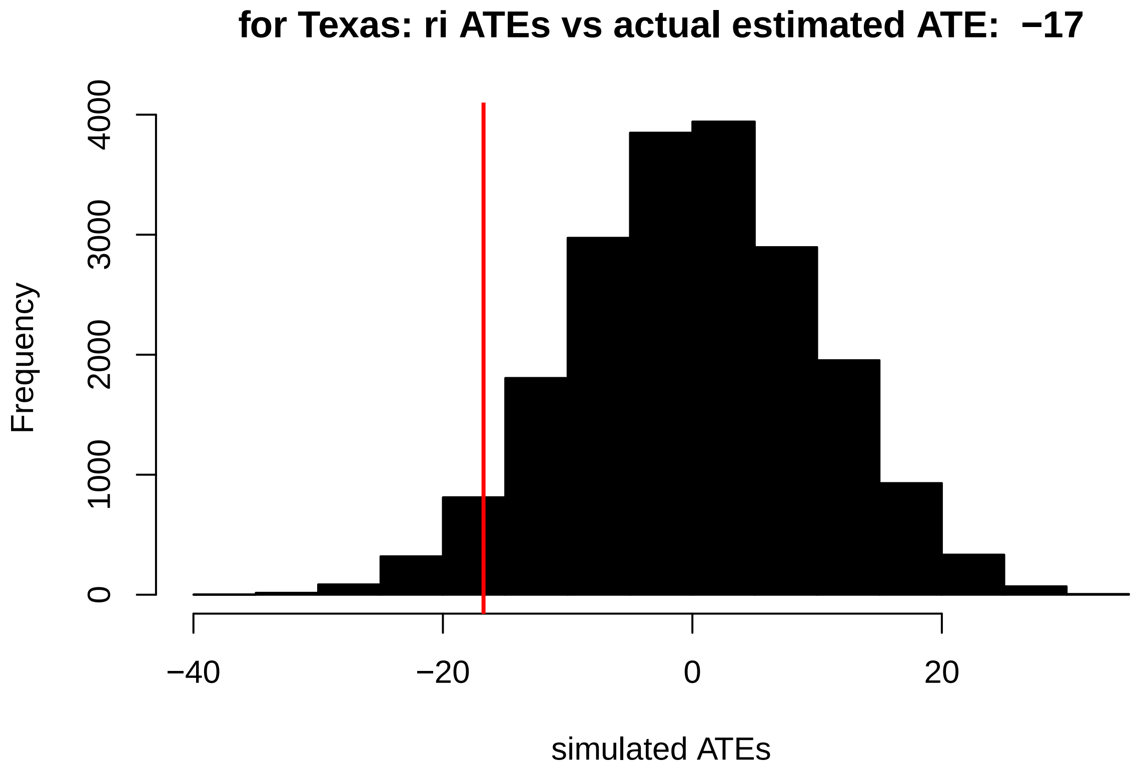
2.f.

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

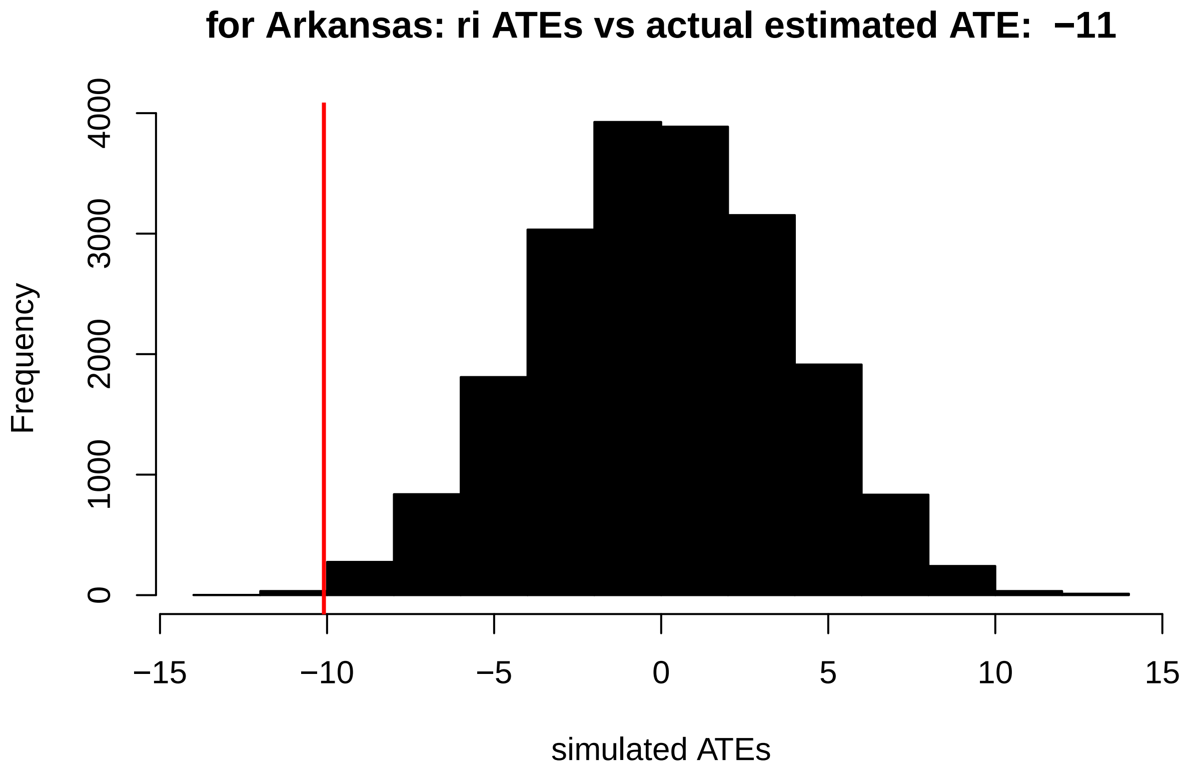
using randomization inference technique, we saw that implied p-values for both states were less than 0.05, so we can reject the sharp null hypothesis.

```
# 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,m
# # -19.48246
#
# q2$ate.arkansas.ri <- q2$d3.8[texas0_arkansas1==1 & term2yearRI==1,mean(bills_introduced)] - q2$d3.8[texas0_arkansas1==1 & term2yearRI==
# # -0.7156863





# creates a function to simulate the randomization inference
set.seed(12659)
# 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())

# more efficient code from Alex's week3 RI
q2$sim.study.2 <- function() {
    q2$d3.8 <- q2$d3.8[,term2yearRI:=rbinom(.N, 1, p = 0.5)]
    q2$ate.texas.ri <- q2$d3.8[texas0_arkansas1==0,.(mu=mean(bills_introduced)),by=term2yearRI][,mu[term2yearRI==1]-mu[term2yearRI==0]]
    q2$ate.arkansas.ri <- q2$d3.8[texas0_arkansas1==1,.(mu=mean(bills_introduced)),by=term2yearRI][,mu[term2yearRI==1]-mu[term2yearRI==0]]
    return(list(q2$ate.texas.ri, q2$ate.arkansas.ri))
}
q2$sim.results <- replicate(10000, q2$sim.study.2())
```

```r
# 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, xlab="simulated ATEs", col="black", main=paste("for Texas: ri ATEs vs actual estimated ATE: ",floor(q2$ate.2y.texas
abline(v=q2$ate.2y.texas.basic,lwd=2,col="red")
```



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



```
# calculates the implied p-value for each state
mean(q2$ate.texas.sim < q2$ate.2y.texas.basic) #p-value for texas, 0.04270214
```

```
## [1] 0.04270214
```

```
mean(q2$ate.arkansas.sim < q2$ate.2y.arkansas.basic) #p-value for arkansas, 0.00180009
```

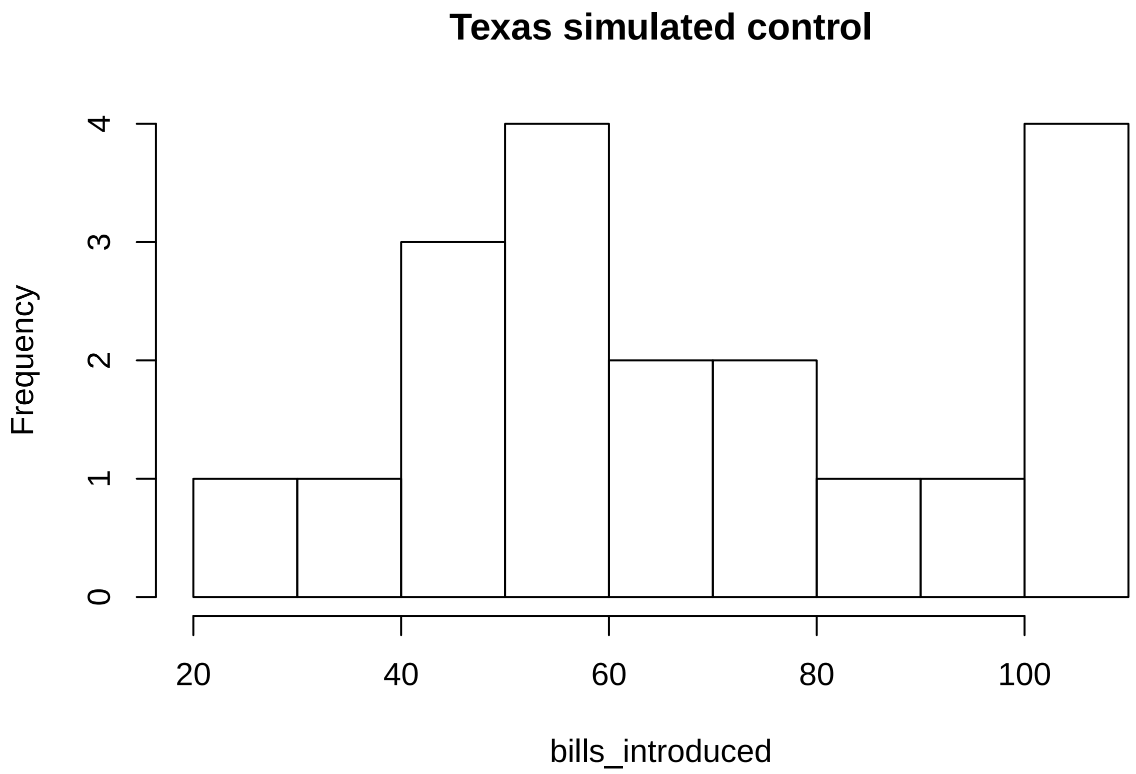
```
## [1] 0.00180009
```

```
# 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).

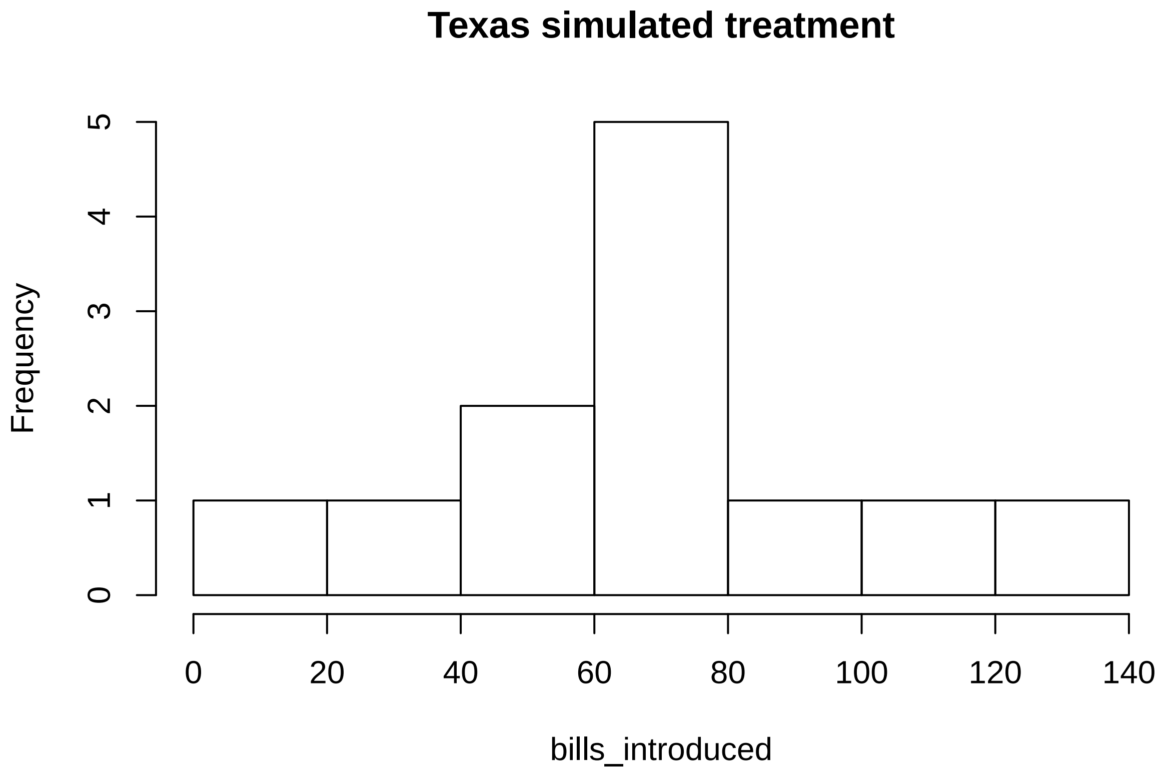
```
q2$d3.8[texas0_arkansas1==0 & term2yearRI==0,hist(bills_introduced,main="Texas simulated control")]
```



```
## $breaks
##  [1]  20  30  40  50  60  70  80  90 100 110
##
## $counts
## [1] 1 1 3 4 2 2 1 1 4
##
```

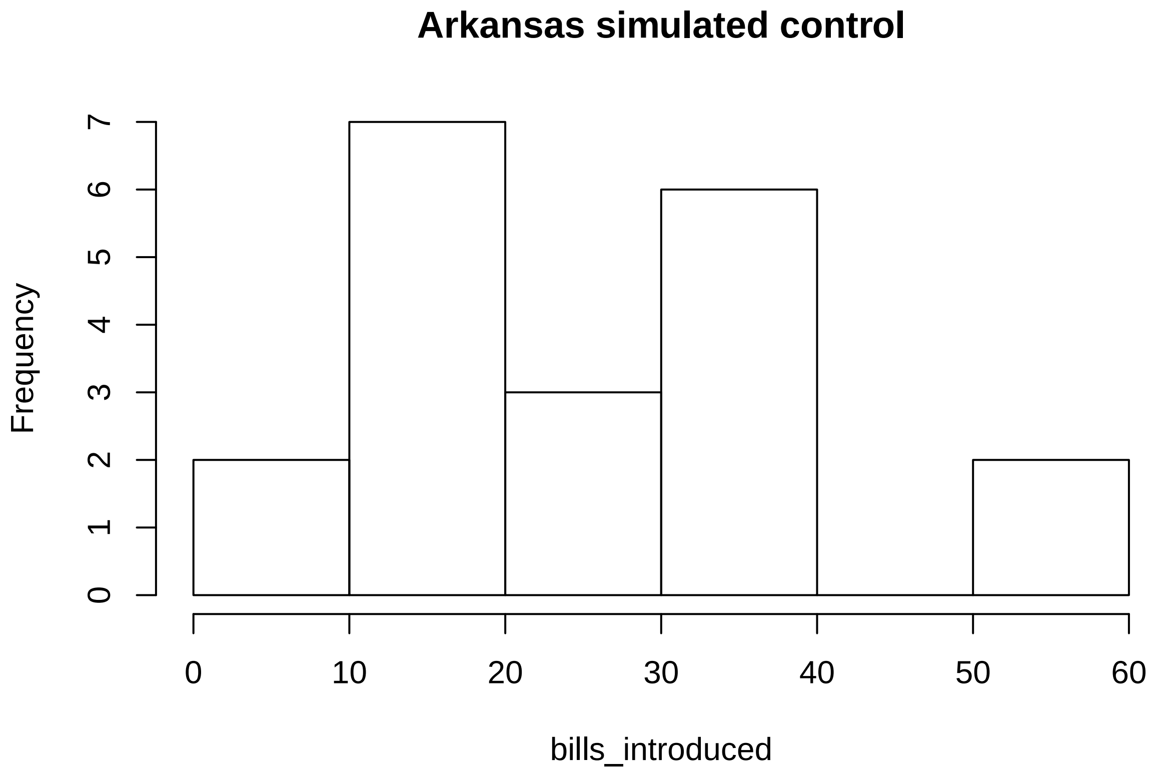
```
## $density
## [1] 0.005263158 0.005263158 0.015789474 0.021052632 0.010526316 0.010526316
## [7] 0.005263158 0.005263158 0.021052632
##
## $mids
## [1]  25  35  45  55  65  75  85  95 105
##
## $xname
## [1] "bills_introduced"
##
## $equidist
## [1] TRUE
##
## attr(,"class")
## [1] "histogram"
```

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



```
## $breaks
## [1]   0  20  40  60  80 100 120 140
##
## $counts
## [1] 1 1 2 5 1 1 1
##
## $density
## [1] 0.004166667 0.004166667 0.008333333 0.020833333 0.004166667 0.004166667
## [7] 0.004166667
##
## $mids
## [1]  10  30  50  70  90 110 130
##
## $xname
## [1] "bills_introduced"
##
## $equidist
## [1] TRUE
##
## attr(,"class")
## [1] "histogram"
```

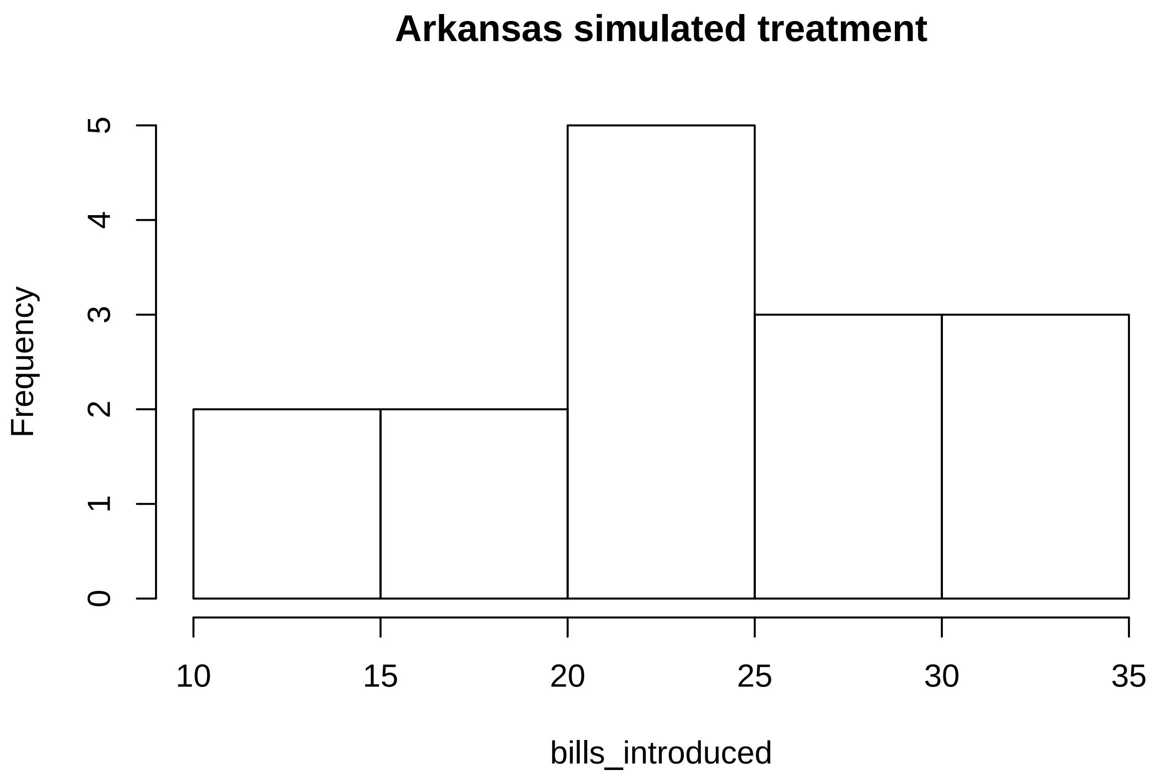
```
q2$d3.8[texas0_arkansas1==1 & term2yearRI==0,hist(bills_introduced,main="Arkansas simulated control")]
```



```
## $breaks
## [1]  0 10 20 30 40 50 60
##
## $counts
## [1] 2 7 3 6 0 2
##
## $density
## [1] 0.010 0.035 0.015 0.030 0.000 0.010
##
## $mids
## [1]  5 15 25 35 45 55
```

```
##
## $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")]
```



```
## $breaks
## [1] 10 15 20 25 30 35
##
## $counts
## [1] 2 2 5 3 3
##
## $density
## [1] 0.02666667 0.02666667 0.06666667 0.04000000 0.04000000
##
## $mids
## [1] 12.5 17.5 22.5 27.5 32.5
##
## $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.*)

```r
## load data
library(data.table)
q3 <- {}
q3$d <- data.table( read.csv("./data/ggChapter3.csv") )
# summary(q3$d)
# str(q3$d)
# head(q3$d)
# hist(q3$d$Village)
# hist(q3$d$Y)
# hist(q3$d$D)
# hist(q3$d$Block)
```

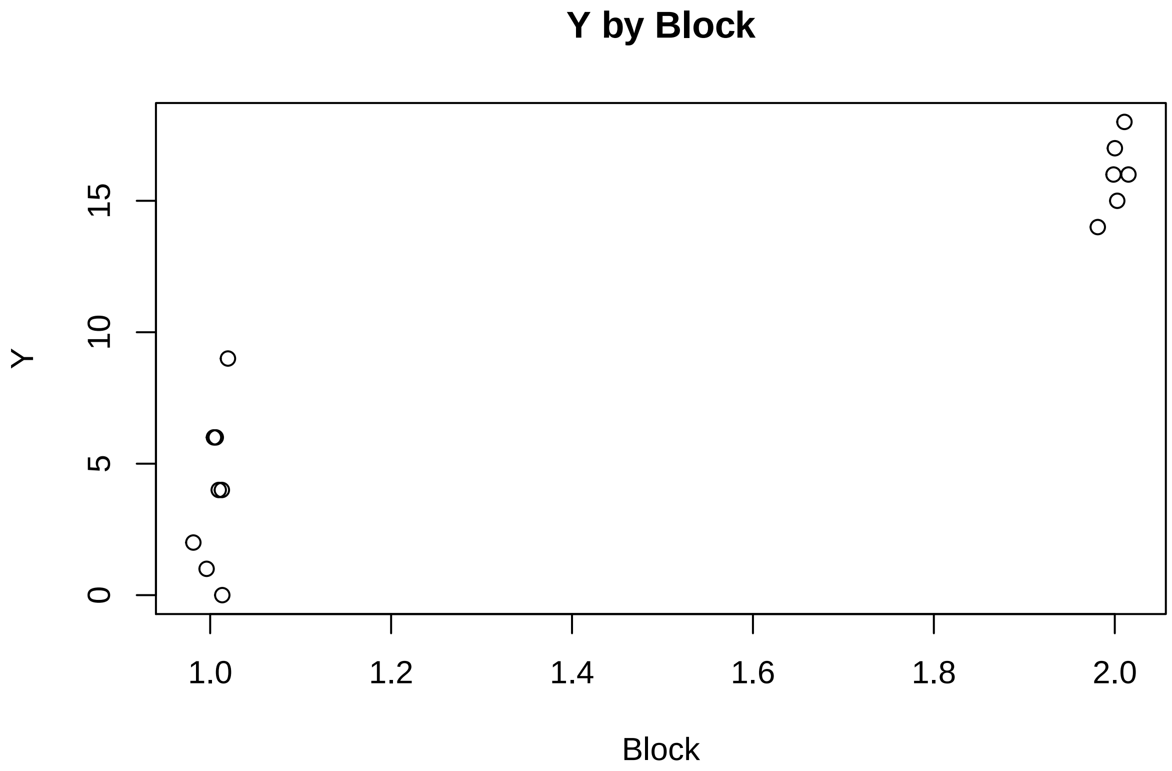
3.a.

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.

note, equation 3.22 is

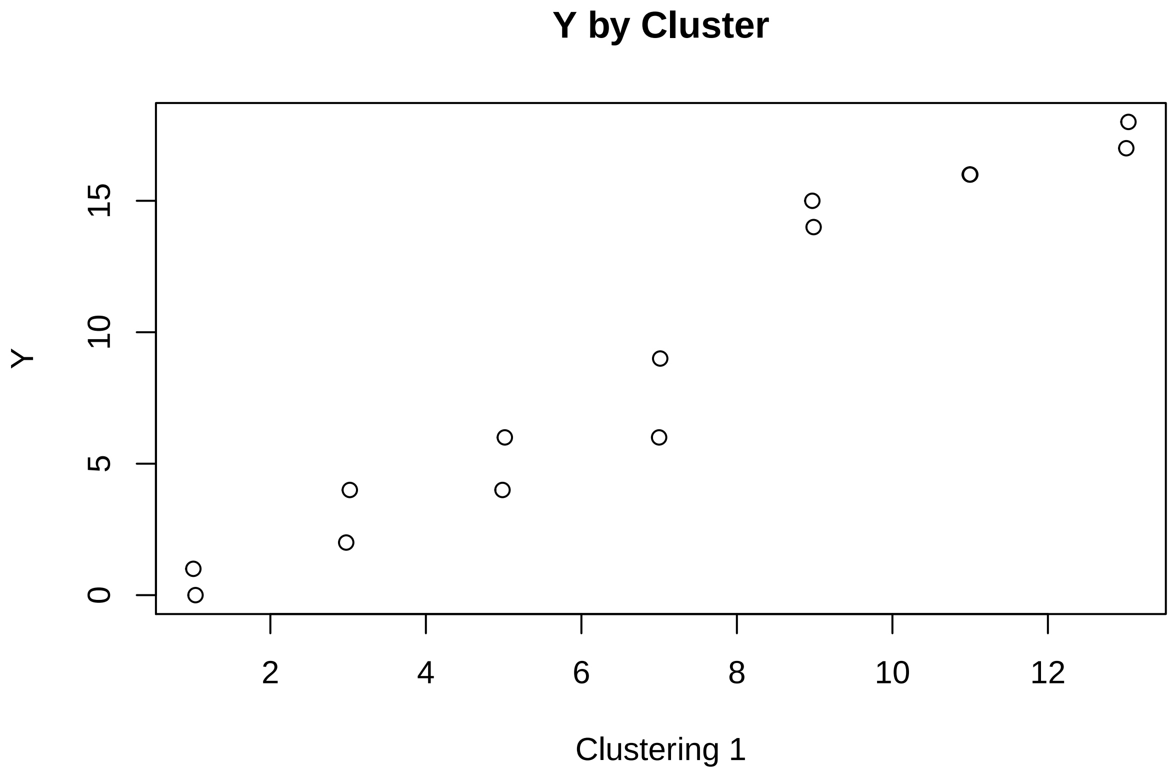
$$SE(A\hat{T}E) = \sqrt{\frac{1}{k-1}(\frac{mVar(\bar{Y}_j(0))}{N-m} + \frac{(N-m)Var(\bar{Y}_j(1))}{m} + 2Cov(\bar{Y}_j(0), \bar{Y}_j(1)))}$$

```r
# create clustering based on {1,2}, {3,4}, {5,6}, ... , {13,14}
q3$d <- q3$d[,C1:=0]
q3$make.cluster.1 <- function() {
    for (i in seq(1,nrow(q3$d),2)) {
        # q3$d[Village==i & Village==i+1, C1:=list(i,i)] #not sure why this doesn't work ...
        q3$d[Village==i, C1:=list(i)]
        q3$d[Village==i+1, C1:=list(i)]
    }
}
q3$make.cluster.1()

# head(q3$d)
q3$d[,plot(Y ~ jitter(Block,0.1), xlab="Block", main="Y by Block")]
```



```
## NULL
```

```
q3$d[,plot(Y ~ jitter(C1,0.1), xlab="Clustering 1", main="Y by Cluster")]
```



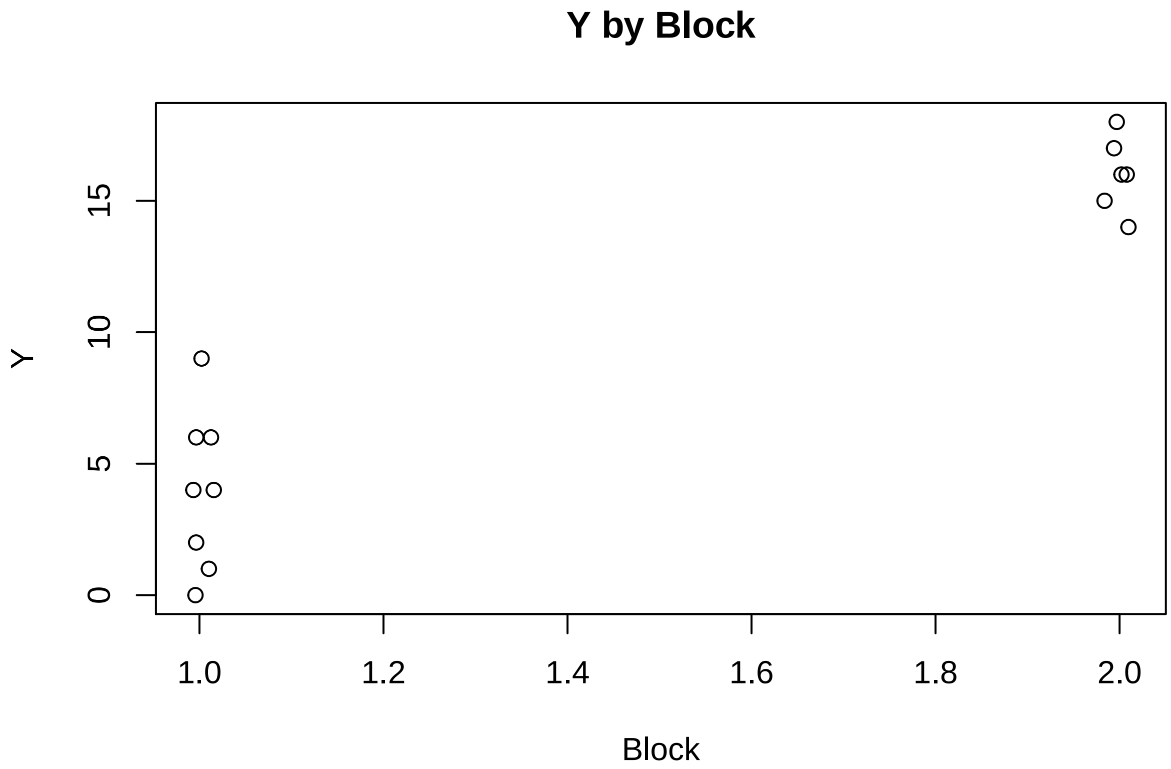
```
## NULL
```

```
# we can see from the individual values plot that the variance in our simulated clusters, C1, is relatively small compared to the variance
```

3.b.

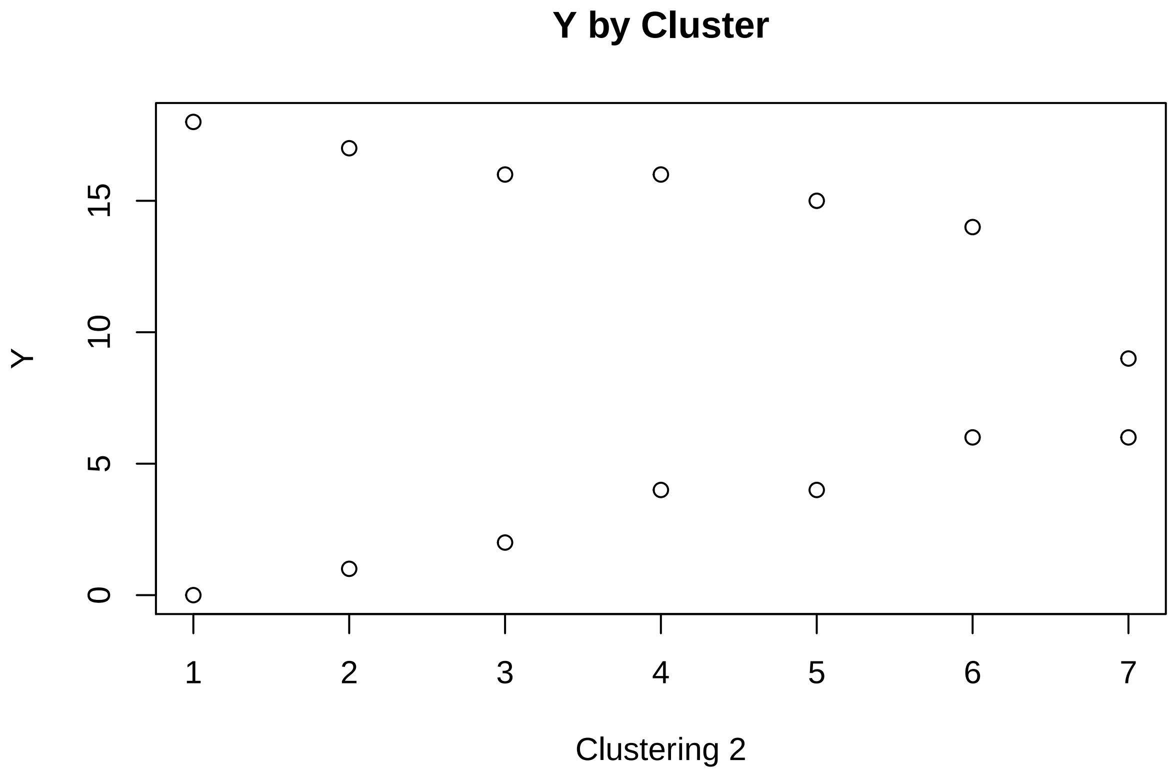
Suppose that clusters are instead formed by grouping observations {1,14}, {2,13}, {3,12}, ... , {7,8}. Use equation (3.22) to calculate the standard error assuming half of the clusters are randomly assigned to treatment.

```
# create clustering based on {1,14}, {2,13}, {3,12}, ... , {7,8}
q3$d <- q3$d[,C2:=0]
q3$make.cluster.2 <- function() {
    for (i in seq(1,nrow(q3$d)/2,1)) {
        q3$d[Village==i, C2:=list(i)]
        q3$d[Village==nrow(q3$d)+1-i, C2:=list(i)]
    }
}
q3$make.cluster.2()
# head(q3$d)
# tail(q3$d)

q3$d[,plot(Y ~ jitter(Block,0.1), xlab="Block", main="Y by Block")]
```



```
## NULL
```

```
q3$d[,plot(Y ~ C2, xlab="Clustering 2", main="Y by Cluster")]
```



```
## NULL
```

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)?

```
#
q4 <- {}
q4$sim.study <- function(baseline, effect.size, sample.size) {
    control.units <- rbinom(sample.size, 1, baseline)
    treatment.units <- rbinom(sample.size, 1, baseline + effect.size)
    all.units <- c(control.units, treatment.units)
    treatment.vector <- c(rep(0, sample.size), rep(1, sample.size))
    p.value <- summary(lm(all.units ~ treatment.vector))$coefficients[2, 4]
    effect.detected <- p.value < 0.05
    return(effect.detected)
}

# get.power(baseline = .1, effect.size = .05, sample.size = 100)
q4$sim.study(baseline = .1, effect.size = .05, sample.size = 1000000)
```

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

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).

```r
library(data.table)
q5 <- {}
q5$d2 <- data.table(read.csv("./data/listData.csv"))
# names(q5)
# head(q5$d2)
# str(q5$d2)
# q5$d2[,hist(bid)]
# q5$d2[,hist(uniform_price_auction)]
# q5$d2[,plot(jitter(bid,factor=0.1) ~ jitter(uniform_price_auction,factor=0.1))]

# fancier way to vizualize the difference in variance and means, between control vs treatment groups
# q5$d2[uniform_price_auction==0,plot(sort(bid), pch=19, col="darkorange")]
# q5$d2[uniform_price_auction==1,points(sort(bid), pch=20, col="steelblue")]

# q5$d2[,boxplot(bid ~ uniform_price_auction)] #suggests unequal variance in control vs treatment groups

# q5$d2[,.N,by=uniform_price_auction] #34 samples in each group
```

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.

95% confidence interval is about (-20.85, -3.56) USD

Note, equations for unpaired two-sample t-test, assuming unequal variances:

t statistic is

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

degrees of freedom

$$df = \frac{[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}]^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}$$

two-sample t confidence interval for $\mu_1 - \mu_2$ with a confidence interval $100(1-\alpha)\%$

$$\bar{x} - \bar{y} \pm t_{\alpha/2,v}\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}$$

```r
# computes the means for control and treatment groups
q5$bids_control <- q5$d2[uniform_price_auction==0,bid]
q5$bids_treatment <- q5$d2[uniform_price_auction==1,bid]
q5$n_control <- length(q5$bids_control) #34
q5$n_treatment <- length(q5$bids_treatment) #34
q5$mu_control <- q5$d2[uniform_price_auction==0,mean(bid)] #28.82353
q5$mu_treatment <- q5$d2[uniform_price_auction==1,mean(bid)] #16.61765
q5$sd_control<- q5$d2[uniform_price_auction==0,sd(bid)] #19.98101
q5$sd_treatment <- q5$d2[uniform_price_auction==1,sd(bid)] #15.40172
q5$ate <- q5$mu_treatment - q5$mu_control # -12.20588


# see Devore p 358 for the analytic forumla for two-sample t-test
# see https://www.statsdirect.co.uk/help/parametric_methods/utt.htm

# calculates the degrees of freedom for a two-sample test of means, with unequal variances
# note, not using, (N1 + N2) - 2 , because that assumes equal variances
```

```r
q5$df <- ( ( (q5$sd_control^2 / q5$n_control) + (q5$sd_treatment^2 / q5$n_treatment) )^2 ) / ( ( (q5$sd_control^2/q5$n_control)^2/(q5$n_co


# calculates the t statistic
q5$t <- ( q5$mu_treatment-q5$mu_control ) / ( sqrt( (q5$sd_control^2)/q5$n_control + (q5$sd_treatment^2)/q5$n_treatment ) ) # -2.821144


# uses the t statistic to calculate 95% confidence interval e.g. (mu +- 1.96 * se)
q5$ci.upper <- (q5$mu_treatment-q5$mu_control) - qt(0.05/2, q5$df) * sqrt( (q5$sd_control^2)/q5$n_control + (q5$sd_treatment^2)/q5$n_treat

q5$ci.lower <- (q5$mu_treatment-q5$mu_control) + qt(0.05/2, q5$df) * sqrt( (q5$sd_control^2)/q5$n_control + (q5$sd_treatment^2)/q5$n_treat

q5$ci <- c(q5$ci.lower, q5$ci.upper) # -20.854624  -3.557141




# sanity check: do it the way Alex does it in week3box3

# calculate ate using data.tables
q5$ate.alex <- q5$d2[,.(mu=mean(bid)),by=uniform_price_auction][,mu[1]-mu[2]] # -12.20588


# use the t.test function to do a two sample t-test on treatment vs control
q5$d2 <- q5$d2[,uniform_price_auction_inv := 1-uniform_price_auction]
# q5$d2[,var.test(bid ~ uniform_price_auction)]
q5$d2[,var.test(bid ~ uniform_price_auction_inv)]
```

```
##
##  F test to compare two variances
##
## data:  bid by uniform_price_auction_inv
## F = 0.59416, num df = 33, denom df = 33, p-value = 0.14
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.2967386 1.1896891
## sample estimates:
## ratio of variances
##          0.5941605
```

```
# actually, var.test shows that there is insufficient evidence to reject the null hypothesis that the variances of bid-by-uniform_price_au
# note to self, do this first next time ...


# t-test assuming unequal variances
# q5$d2[,t.test(bid ~ uniform_price_auction, var.equal=FALSE)] # don't use: formula flips the base group for some reason
q5$d2[,t.test(bid ~ uniform_price_auction_inv, var.equal=FALSE)] # correct formula given the negative treament effect
```

```
##
##  Welch Two Sample t-test
##
## data:  bid by uniform_price_auction_inv
## 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 in group 0 mean in group 1
##        16.61765        28.82353
```

```
# t.test(q5$bids_treatment, q5$bids_control, var.equal=FALSE) # equivalent
# t = -2.8211, df = 61.983, p-value = 0.006421
# 95 percent confidence interval: -20.854624  -3.557141

# t-test assuming equal variances
q5$d2[,t.test(bid ~ uniform_price_auction_inv, var.equal=TRUE)]
```

```
##
##  Two Sample t-test
##
## data:  bid by uniform_price_auction_inv
## t = -2.8211, df = 66, p-value = 0.006315
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -20.844162  -3.567603
## sample estimates:
## mean in group 0 mean in group 1
##        16.61765        28.82353
```

```
# t = -2.8211, df = 66, p-value = 0.006315
# 95 percent confidence interval: -20.844162  -3.567603

# as it turns out, in this case, assuming equal or unequal variances didn't seem to matter much
```

5.b.

In plain language, what does this confidence interval mean?

if we were to repeat our experiment under the same conditions many times, 95% of the time, the true ATE would fall within the range (-20.85, -3.56) USD

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.

```
# use OLS regression of bid on a binary treatment variable
q5$m1 <- q5$d2[,lm(bid ~ uniform_price_auction)]
summary(q5$m1)
```

```
## 
## Call:
## lm(formula = bid ~ 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 ***
## 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
```

```
q5$m1$ate <- unname(q5$m1$coefficients[2]) # -12.20588

# check if the OLS answers match up with the analytic answers
q5$m1$ate - q5$ate # difference is really close to zero, so probably matches
```

```
## [1] -3.552714e-15
```

5.d.

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

confidence interval from the regression is about (-20.84, -3.57) USD

```
# calculate the ci from our linear model
q5$m1$ci <- q5$d2[,t.test(bid ~ uniform_price_auction_inv, var.equal=TRUE)]$conf.int[c(1,2)] # -20.844162  -3.567603
q5$m1$ci
```

```
## [1] -20.844162  -3.567603
```

```
# confint(q5$m1) # another equivalent way
# q5$m1$ci <- confint(q5$m1)[c(2,4)] # (-20.84416, -3.567603)

# check if the OLS answers match up with the analytic answers
q5$m1$ci - q5$ci # differences are somewhat close to zero compared to the actual values (roughly 1 cent difference), so still probably mat
```

```
## [1]  0.01046204 -0.01046204
```

5.e.

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

the regression reports a p-value of about 0.006

```r
library(sandwich)

# coef(q5$m1)
summary(q5$m1)
```

```
##
## Call:
## lm(formula = bid ~ 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 ***
## 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
```

```r
q5$m1$pval <- summary(q5$m1)$coefficients[8]
q5$m1$pval
```

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

```r
# with our linear model, we have a p-value of about 0.00631 for the uniform_price_auction



# for completeness, calculate the robust se, ...
# i.e. if we can't assume constant variance of bid by uniform_price_auction
```

```
q5$m1$vcov <- vcovHC(q5$m1, type="HC3")
# q5$m1$vcov # var-cov matrix

q5$m1$r.se <- sqrt(diag(q5$m1$vcov))
q5$m1$r.se
```

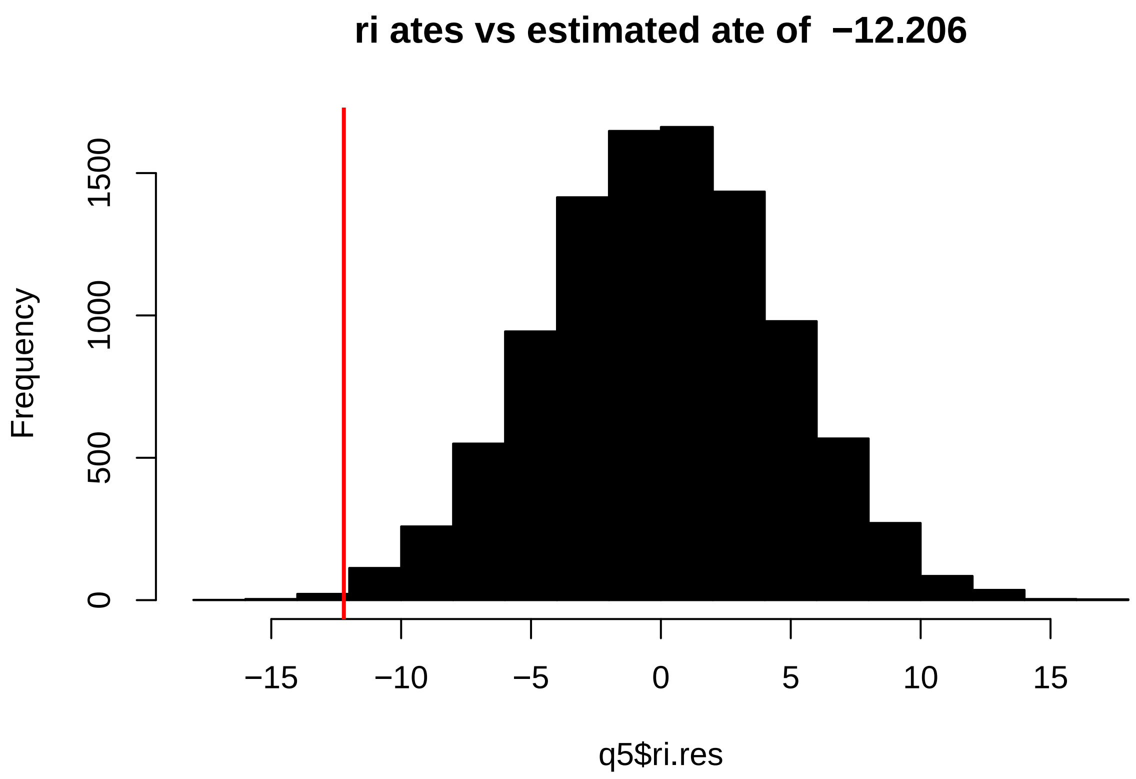
```
##           (Intercept) uniform_price_auction
##              3.478247              4.391637
```

5.f.

Now compute the same p-value using randomization inference.

randomization inference also resulted in p-value of about 0.006

```r
set.seed(2382)
q5$sim.ri <- function(numloops){
    # init empty results list, using vector makes subsequent use slightly faster
    ri.res <- vector(mode="numeric", length=numloops)
    # do the ri repeatedly, assuming the sharp null, and assign ate to our results vector
    for(i in 1:numloops) {
        # randomly assign to treatment or control by fair coin flip
        q5$d2 <- q5$d2[,ri:=rbinom(.N, 1, p = 0.5)]
        # then calculate the ate, as the difference in means between treatment vs control
        ri.res[i] <- q5$d2[,.(mu=mean(bid)),by=ri][,mu[ri==1]-mu[ri==0]]
    }
    return(ri.res)
}
q5$ri.res <- q5$sim.ri(10000)


# visualize the ri results, vs the previously calculated ate
hist(q5$ri.res, col="black", main=paste("ri ates vs estimated ate of ", round(q5$ate,3)))
abline(v=q5$ate,col="red",lwd=2)
```



```
# calculate the p-value, 2-tailed
q5$ri.pval <- mean(abs(q5$ri.res) > abs(q5$ate))
q5$ri.pval
```

```
## [1] 0.0062
```

```
# p-value from ri is smaller than 0.05, so we can reject the sharp null
```

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).)

note, formula for p-values for z tests is

$$P - value : P = \begin{cases} 1 - \Phi(z) & \text{for upper-tailed z test} \\ \Phi(z) & \text{for lower-tailed z test} \\ 2[1 - \Phi(|z|)] & \text{for two-tailed z test} \end{cases}$$

Where $\Phi(z)$ is the cumulative area under the z curve

similarly for p-values for t tests we are looking for the cumulative area under the t curve (t-curve also shaped by the degrees of freedom)

```
# see Devore p 332 and 333 for analytic formula for p-values for z tests and for t tests
# we take our t statistic and degrees of freedom from (a)
# names(q5)
# q5$t
# q5$df

# and do a lookup on the absolute value of the t statistic, given our df
2*pt(abs(q5$t), q5$df, lower.tail=FALSE) # assuming unequal variance
```

```
## [1] 0.006420778
```

```
2*pt(abs(q5$t), (q5$n_control+q5$n_treatment)-2, lower.tail=FALSE) # assuming equal variances
```

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

```
# p-values are pretty close to the ones we got from the regression and the ri
```

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?

The two p-values are very close to each other. As we do more randomization inference loops, the simulated results start matching up to the regression. Doing ri with 1k loops sometimes got us p-values of 0.005 or 0.007. But once we started doing 10k loops, the p-value from the ri seemed to stick to around 0.006. Also, we only had 34 samples in each group. Larger sample sizes might push the p-values of both ri and regression down towards zero.