

Problem Set 2

Gaurav Khanna

```
library(data.table)

## Warning: package 'data.table' was built under R version 3.4.4

library(dplyr)

## Warning: package 'dplyr' was built under R version 3.4.4
##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:data.table':
##
##     between, first, last
##
## The following objects are masked from 'package:stats':
##
##     filter, lag
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(experiment)

## Warning: package 'experiment' was built under R version 3.4.4
## Loading required package: boot
## Loading required package: MASS
## Warning: package 'MASS' was built under R version 3.4.4
##
## Attaching package: 'MASS'
##
## The following object is masked from 'package:dplyr':
##
##     select
##
## experiment: R Package for Designing and Analyzing Randomized Experiments
## Version: 1.1-4
```

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.

```
# Dataframe from the CSV
d <- read.csv("./data/Clingingsmith.2009.csv", stringsAsFactors = FALSE)
```

- 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.

```
head(d)
```

```
##      success views_saudi views_indonesian views_turkish views_african
## 1          0           1             1           0             0
## 2          0           1             1           0            -1
## 3          0           0             0           0             0
## 4          0           2             2           0             0
## 5          0           1             1           1             1
## 6          0           2             0           0             0
##      views_chinese views_european views
## 1                0              0    2
## 2                1             -1    1
## 3                0              0    0
## 4                1              0    5
## 5                1             -2    3
## 6                0              0    2
```

```
# Grouping data by success
# success <- group_by(d, success)
# TODO: Chaining the operations is more elegant but i'd like to see the intermediate
success <- group_by(d, success)
summarize(success, count=n(), viewSum=sum(views), viewMean=mean(views))
```

```
## # A tibble: 2 x 4
##   success count viewSum viewMean
##   <int> <int>   <int>   <dbl>
## 1     0   448    837     1.87
## 2     1   510   1195     2.34
```

The mean of views among lottery winners and losers is 2.34 and 1.86 respectively. The views for others generally look more positive among the lottery winners

```
# Estimate ATE function
est.ate <- function(result, treat) {
  mean(result[treat==1]) - mean(result[treat==0])
}
```

```
ate <- est.ate(success$views, success$success)
ate
```

```
## [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.)

```
# Sharp null allows us to complete the data, y(0) and y(1) from the data we already have
# We just need to randomize the success variable
# Creating a new data frame to try randomization on the original data
# The original is kept around intact for reference
```

```
dr <- read.csv("./data/Clingingsmith.2009.csv", stringsAsFactors = FALSE)
head(dr)
```

```
##      success views_saudi views_indonesian views_turkish views_african
## 1         0           1             1           0           0
## 2         0           1             1           0          -1
## 3         0           0             0           0           0
## 4         0           2             2           0           0
## 5         0           1             1           1           1
## 6         0           2             0           0           0
##      views_chinese views_european views
## 1              0           0      2
## 2              1          -1      1
## 3              0           0      0
## 4              1           0      5
## 5              1          -2      3
## 6              0           0      2
```

```
success1 <- group_by(dr, success)
summarize(success1, count=n(), viewSum=sum(views), viewMean=mean(views))
```

```
## # A tibble: 2 x 4
##   success count viewSum viewMean
##   <int> <int>   <int>   <dbl>
## 1     0  448    837    1.87
## 2     1  510   1195    2.34
```

```
# Randomizing the success variable
# randomize() from the expriments library, returns the "randomize" class
# randomize() creates 2 groups treatment and control which is sufficient for us in this case
# It is equivalent to us randomizing among 0 and 1 for "success" variable
# We are not doing any blocking, so subjects in a group can be variable
successr <- randomize(success1)
```

```
# Some statistics on the classes after randomization
# We're choosing to pick equal numbers in the 2 classes
head(successr$treatment)
```

```
##           1           2           3           4           5           6
## "Control"  "Treat" "Control" "Control" "Control"  "Treat"
```

```
table(success1$success)
```

```
##
##    0    1
## 448 510
```

```
table(successr$treatment)
```

```
##
## Control  Treat
##    479    479
```

```
# Modifying est.ate to use the object of "randomize" class
est.mate <- function(result, treat) {
  mean(result[treat=="Treat"]) - mean(result[treat=="Control"])
}
```

```

# ATE for 1 randomization for testing
ateR <- est.mate(successr$data$views, successr$treatment)
ateR

## [1] -0.1586639

# This is very different from our ATE

# Repeating the experiment 10,000 times
distribution.under.sharp.null <- replicate(10000, est.mate(successr$data$views, (randomize(success1))$treatment))

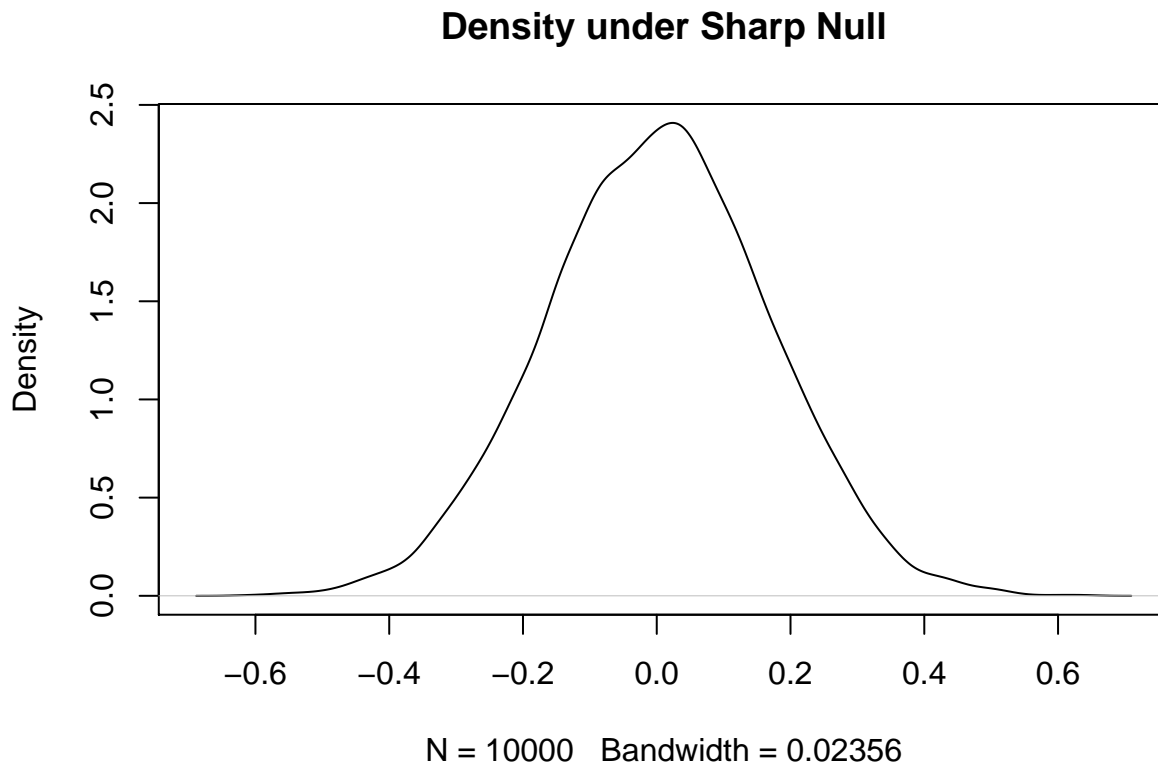
# Average estimate for the ATE
mean(distribution.under.sharp.null)

## [1] 0.00113904

# Again, very different from ATE

# Graph for the estimates
plot(density(distribution.under.sharp.null),
     main = "Density under Sharp Null")

```



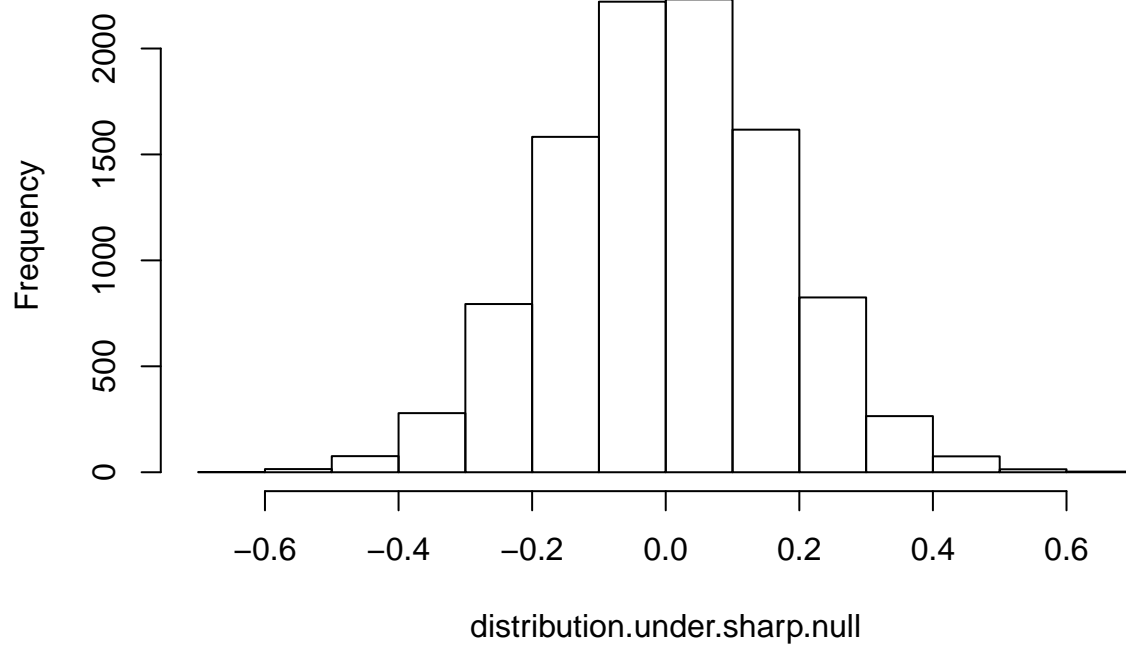
```

# Almost a normal distribution

# Histogram for the distribution
hist(distribution.under.sharp.null,
     main = "Histogram under Sharp Null")

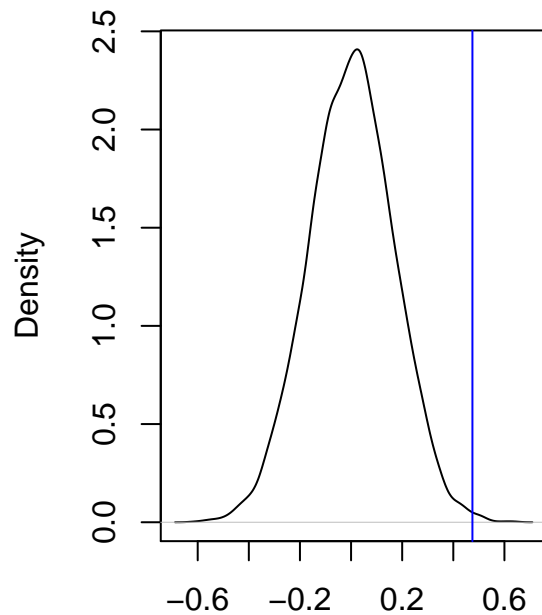
```

Histogram under Sharp Null



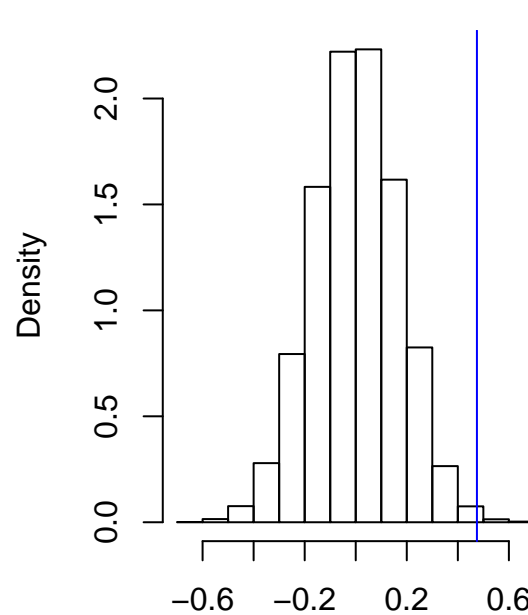
```
# More analysis on the distribution. Plotting the p value
par(mfrow = c(1,2))
plot(density(distribution.under.sharp.null),
     main = "Density Plot of ATE")
abline(v = ate, col = "blue")
hist(distribution.under.sharp.null,
     main = "Histogram of ATE",
     freq = FALSE)
abline(v = ate, col = "blue")
```

Density Plot of ATE



N = 10000 Bandwidth = 0.02356

Histogram of ATE



distribution.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?

```
# num of assignments that generate an estimated ATE at least as large as the actual
n <- sum(distribution.under.sharp.null >= ate)
n
```

```
## [1] 27
```

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

```
# p-value
m <- mean(distribution.under.sharp.null >= ate )
m
```

```
## [1] 0.0027
```

- 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?

```
# num of simulated random assignments with an estimated ATE that is at least as large *in absolute value
na <- sum(abs(distribution.under.sharp.null) >= ate)
na
```

```
## [1] 50
```

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

```
# 2 tailed p-value
ma <- mean(abs(distribution.under.sharp.null) >= ate )
ma
```

```
## [1] 0.005
```

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 “theory” 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)

## Warning: package 'foreign' was built under R version 3.4.4

d2 <- read.dta("./data/Titunik.2010.dta")
head(d2)

##   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
```

- 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?

```
# Grouping the data by state
state <- group_by(d2, texas0_arkansas1)
head(state)

## # A tibble: 6 x 3
## # Groups:   texas0_arkansas1 [1]
##   term2year bills_introduced texas0_arkansas1
##     <int>         <int>         <int>
## 1         0             18                0
## 2         0             29                0
## 3         0             41                0
## 4         0             53                0
## 5         0             60                0
## 6         0             67                0

# Mean number of bills introduced in each state
summarize(state, count=n(), meanBills=mean(bills_introduced))

## # A tibble: 2 x 3
##   texas0_arkansas1 count meanBills
##           <int> <int>     <dbl>
```

```
## 1          0    31    68.8
## 2          1    35    25.5
```

```
# Which state looks more productive
```

Texas has a mean of 68.77 compared to 25.51 for Arkansas. Texas seems to be more productive in bills

```
# Grouping by terms
```

```
terms <- group_by(d2, term2year)
head(terms)
```

```
## # A tibble: 6 x 3
## # Groups:   term2year [1]
##   term2year bills_introduced texas0_arkansas1
##   <int>          <int>          <int>
## 1         0           18             0
## 2         0           29             0
## 3         0           41             0
## 4         0           53             0
## 5         0           60             0
## 6         0           67             0
```

```
# Mean number of bills by terms
```

```
summarize(terms, meanBills=mean(bills_introduced))
```

```
## # A tibble: 2 x 2
##   term2year meanBills
##   <int>     <dbl>
## 1         0     53.1
## 2         1     38.6
```

```
# Which term is more effective
```

```
**
```

```
# Grouping by state and term term
```

```
stateTerms <- group_by(d2, texas0_arkansas1, term2year)
head(stateTerms)
```

```
## # A tibble: 6 x 3
## # Groups:   texas0_arkansas1, term2year [1]
##   term2year bills_introduced texas0_arkansas1
##   <int>          <int>          <int>
## 1         0           18             0
## 2         0           29             0
## 3         0           41             0
## 4         0           53             0
## 5         0           60             0
## 6         0           67             0
```

```
# Mean number of bills by grouping
```

```
summarize(stateTerms, meanBills=mean(bills_introduced))
```

```
## # A tibble: 4 x 3
## # Groups:   texas0_arkansas1 [?]
##   texas0_arkansas1 term2year meanBills
##   <int>          <int>     <dbl>
## 1         0         0     76.9
```



```
## 2          0          1      60.1
## 3          1          0      30.7
## 4          1          1      20.6
```

```
# Which groupings indicate more effectiveness?
# TODO
```

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

```
# Separating out the states
texas <- state[which(state$texas0_arkansas1==0),]
head(texas)
```

```
## # A tibble: 6 x 3
## # Groups:   texas0_arkansas1 [1]
##   term2year bills_introduced texas0_arkansas1
##     <int>         <int>         <int>
## 1         0          18             0
## 2         0          29             0
## 3         0          41             0
## 4         0          53             0
## 5         0          60             0
## 6         0          67             0
```

```
arkansas <- state[which(state$texas0_arkansas1==1),]
head(arkansas)
```

```
## # A tibble: 6 x 3
## # Groups:   texas0_arkansas1 [1]
##   term2year bills_introduced texas0_arkansas1
##     <int>         <int>         <int>
## 1         0          11             1
## 2         0          15             1
## 3         0          17             1
## 4         0          23             1
## 5         0          24             1
## 6         0          25             1
```

```
# ATE for Texas
texas_ate <- est.ate(texas$bills_introduced, texas$term2year)
texas_ate
```

```
## [1] -16.74167
```

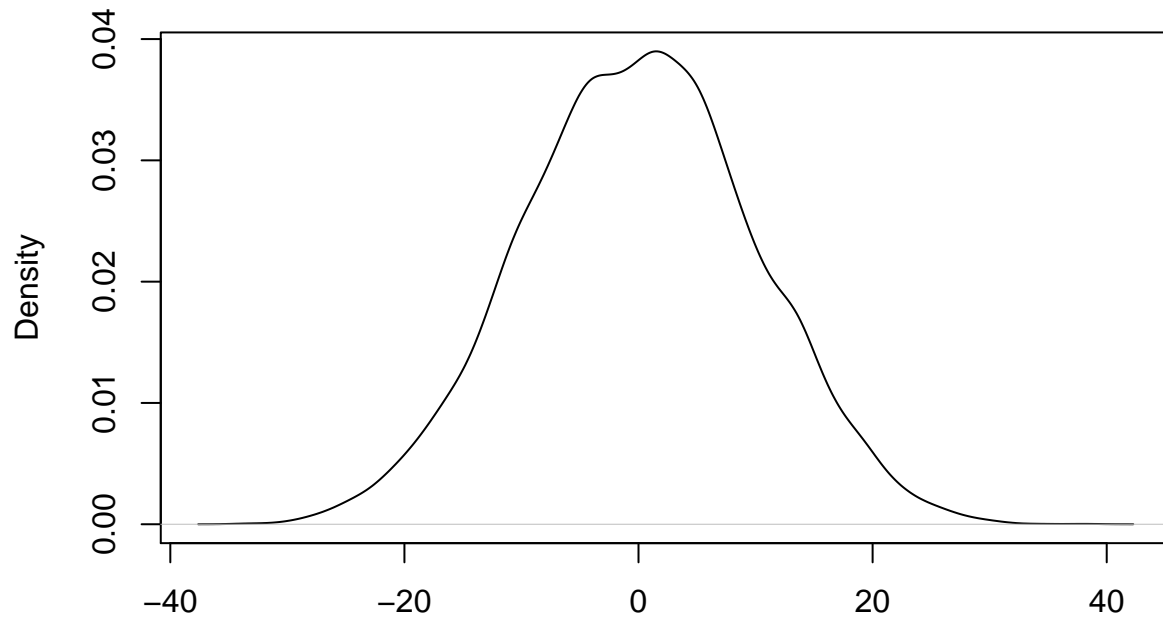
```
# Considering the distribution with other random assignments
# We'll randomize the term variable. "2 years" is our treatment
# Since the grouping is binary, randomization of the entire dataset should work
texas_r <- randomize(texas)
```

```
# Repeating the experiment 10,000 times
```

```
distribution.under.sharp.null.texas <- replicate(10000, est.mate(texas_r$data$bills_introduced, (randomize(texas_r)$term2year)))

plot(density(distribution.under.sharp.null.texas),
     main = "Density under Sharp Null for Texas")
```

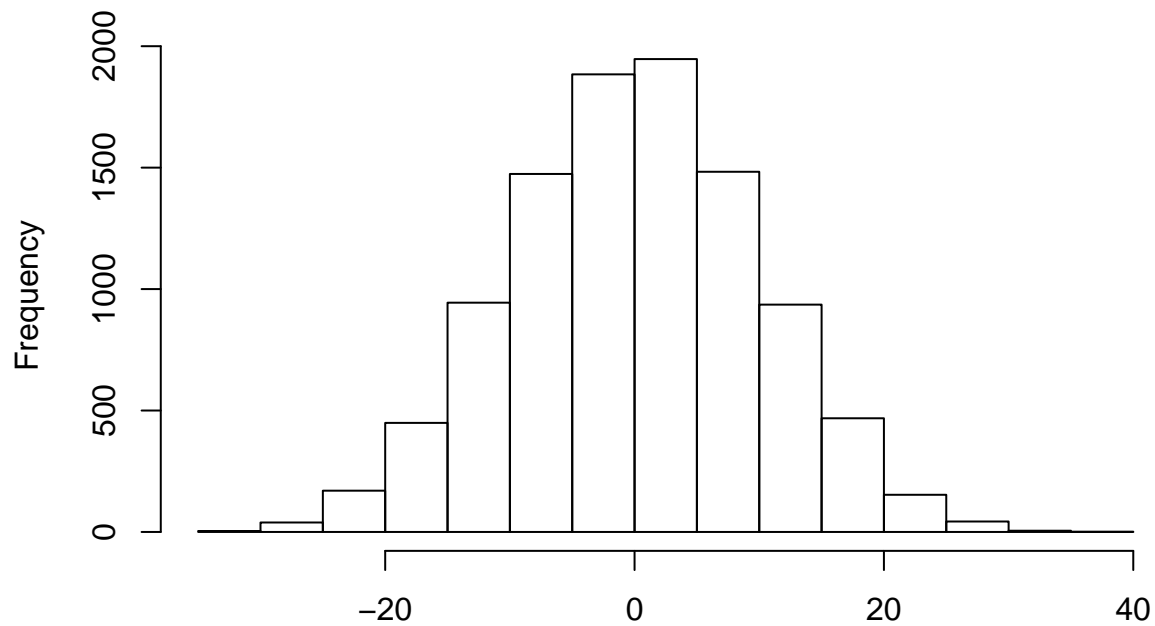
Density under Sharp Null for Texas



N = 10000 Bandwidth = 1.415

```
hist(distribution.under.sharp.null.texas,  
     main = "Histogram under Sharp Null for Texas")
```

Histogram under Sharp Null for Texas



distribution.under.sharp.null.texas

```
# average estimated ate for texas  
mean(distribution.under.sharp.null.texas)
```

```
## [1] 0.02099333
# Standard error is the square root of the averaged squared deviation (from the average EST estimate)
head(distribution.under.sharp.null.texas)

## [1] 5.8916667 1.3708333 9.8666667 -0.5958333 20.3291667 -10.2541667
se_texas <- sqrt(mean((distribution.under.sharp.null.texas - mean(distribution.under.sharp.null.texas))^2))
se_texas

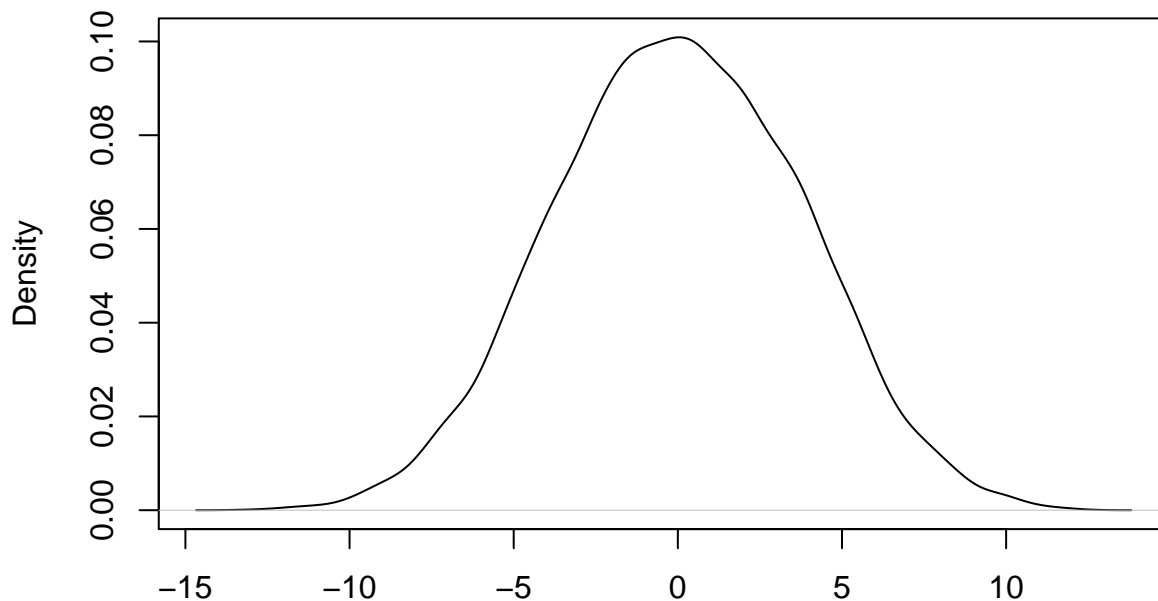
## [1] 9.921043
# ATE for arkansas
arkansas_ate <- est.ate(arkansas$bills_introduced, arkansas$term2year)
arkansas_ate

## [1] -10.09477
# Considering the distribution with other random assignments
# We'll randomize the term variable. "2 years" is our treatment
# Since the grouping is binary, randomization of the entire dataset should work
arkansas_r <- randomize(arkansas)

# Repeating the experiment 10,000 times
distribution.under.sharp.null.arkansas <- replicate(10000, est.mate(arkansas_r$data$bills_introduced, (

plot(density(distribution.under.sharp.null.arkansas),
     main = "Density under Sharp Null for arkansas")
```

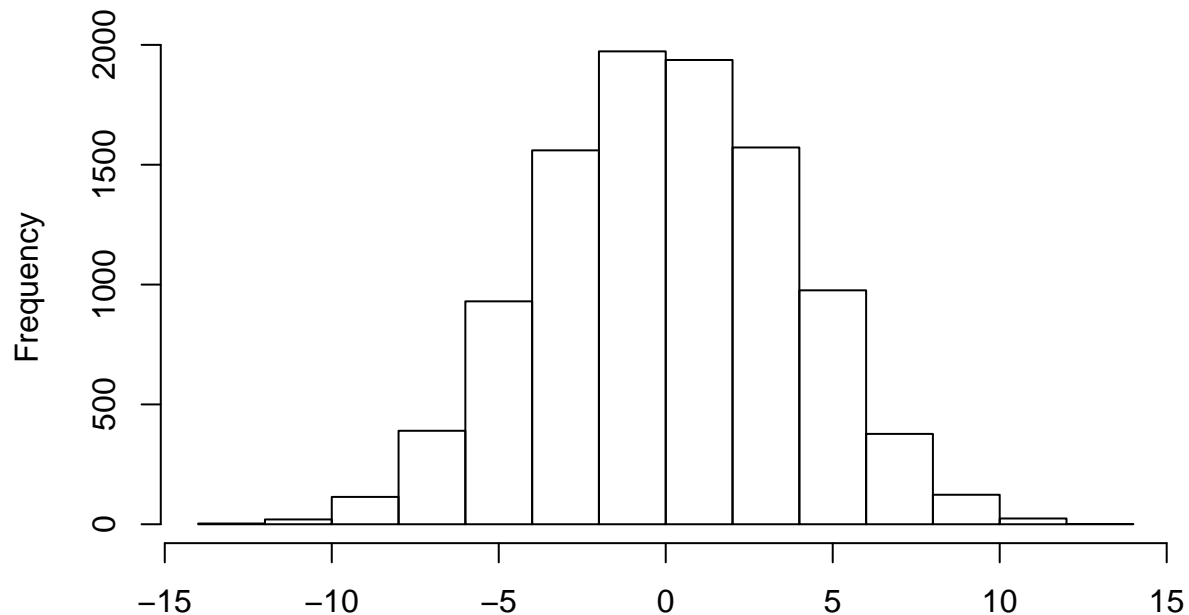
Density under Sharp Null for arkansas



N = 10000 Bandwidth = 0.5344

```
hist(distribution.under.sharp.null.arkansas,
     main = "Histogram under Sharp Null for Arkansas")
```

Histogram under Sharp Null for Arkansas



distribution.under.sharp.null.arkansas

```
# average estimated ate for arkansas
```

```
mean(distribution.under.sharp.null.arkansas)
```

```
## [1] 0.02344967
```

```
# Standard error is the square root of the averaged squared deviation (from the average EST estimate)
```

```
head(distribution.under.sharp.null.arkansas)
```

```
## [1] -5.0032680 -5.7483660 2.4869281 -0.9444444 -1.4019608 -0.1993464
```

```
se_arkansas <- sqrt(mean((distribution.under.sharp.null.arkansas - mean(distribution.under.sharp.null.arkansas))^2))
se_arkansas
```

```
## [1] 3.74618
```

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

```
# ATE for both sides combined
```

```
# Recalling the original and then using 3.10
```

```
texas_ate
```

```
## [1] -16.74167
```

```
arkansas_ate
```

```
## [1] -10.09477
```

```
nrow(texas)
```

```
## [1] 31
```

```
nrow(arkansas)
```

```
## [1] 35
```

```

nrow(d2)

## [1] 66
# Combined ATE
comb_ate <- (texas_ate*nrow(texas)/nrow(d2)) + (arkansas_ate*nrow(arkansas)/nrow(d2))
comb_ate

## [1] -13.2168

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.

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

# Recalling previous values and then putting them in 3.12 for combined SE
se_texas

## [1] 9.921043
se_arkansas

## [1] 3.74618
se_comb <- sqrt((nrow(texas)*se_texas/nrow(d2))^2 + (nrow(arkansas)*se_arkansas/nrow(d2))^2)
se_comb

## [1] 5.065682

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

# Randomization inference to test the sharp null hypothesis (treatment effect is 0 in Texas)

# We have the results of 10,000 randomizations above in distribution.under.sharp.null.texas
head(distribution.under.sharp.null.texas)

## [1] 5.8916667 1.3708333 9.8666667 -0.5958333 20.3291667 -10.2541667
# We also have the ate from the actual experiment
texas_ate

## [1] -16.74167
# One tailed p value
p_one_texas <- mean(distribution.under.sharp.null.texas >= texas_ate )
p_one_texas

## [1] 0.9541
# Two tailed p value
p_two_texas <- mean(abs(distribution.under.sharp.null.texas) >= texas_ate)
p_two_texas

## [1] 1

# Randomization inference to test the sharp null hypothesis (treatment effect is 0 in Arkansas)

# We have the results of 10,000 randomizations above in distribution.under.sharp.null.texas
head(distribution.under.sharp.null.arkansas)

```

```
## [1] -5.0032680 -5.7483660 2.4869281 -0.9444444 -1.4019608 -0.1993464
```

```
# We also have the ate from the actual experiment
arkansas_ate
```

```
## [1] -10.09477
```

```
# One tailed p value
```

```
p_one_arkansas <- mean(distribution.under.sharp.null.arkansas >= arkansas_ate )
p_one_arkansas
```

```
## [1] 0.9978
```

```
# Two tailed p value
```

```
p_two_arkansas <- mean(abs(distribution.under.sharp.null.texas) >= arkansas_ate)
p_two_arkansas
```

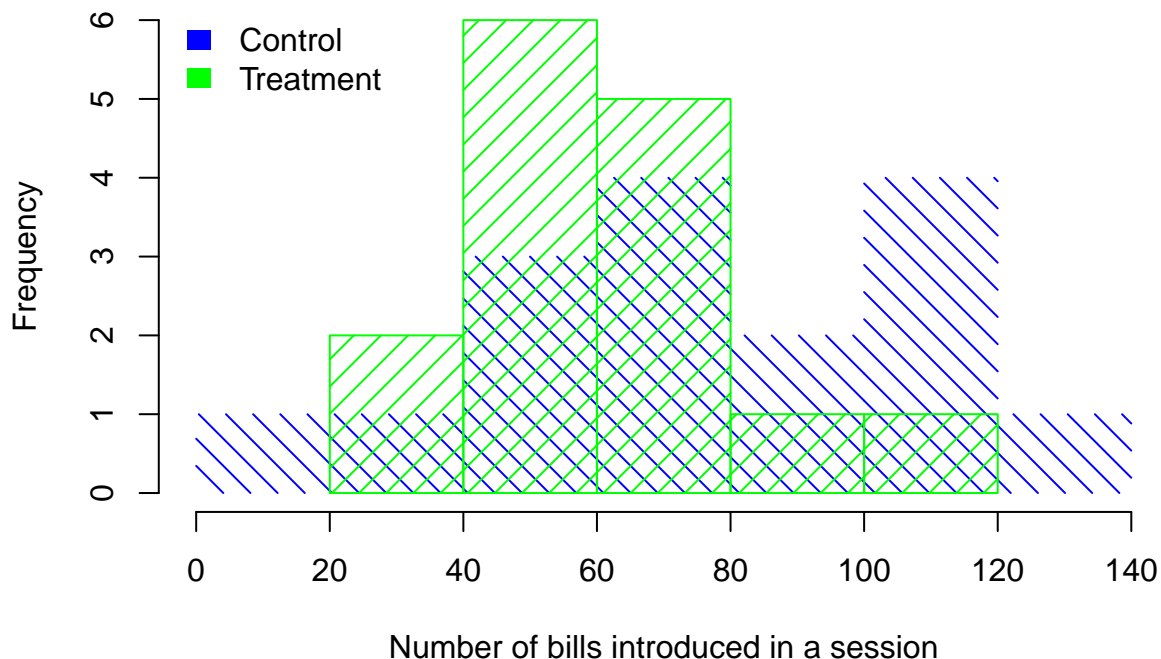
```
## [1] 1
```

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

```
# Histogram for control and Treatment in Texas
```

```
hist(texas[which(state$term2year == 0),]$bills_introduced, xlim = c(0,140), ylim = c(0,6), col = "blue")
hist(texas[which(state$term2year == 1),]$bills_introduced, xlim = c(0, 140), ylim = c(0,6), col= "green")
legend('topleft',c('Control','Treatment'),
      fill = c("blue", "green"), bty = 'n',
      border = NA)
```

Distribution of bills_introduced under control and treatment in Texa

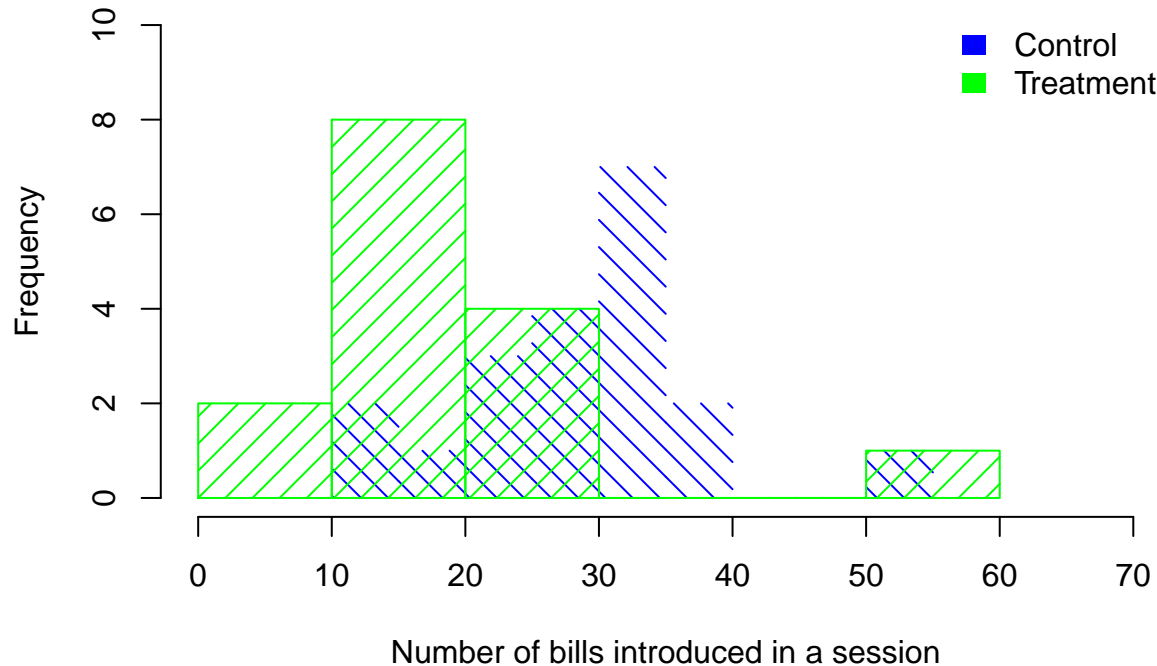


```
# Histogram for control and treatment in Arkansas
```

```
hist(arkansas[which(state$term2year == 0),]$bills_introduced, xlim = c(0,70), ylim = c(0,10), col = "blue")
hist(arkansas[which(state$term2year == 1),]$bills_introduced, xlim = c(0, 70), ylim = c(0,10), col= "green")
legend('topright',c('Control','Treatment'),
```

```
fill = c("blue", "green"), bty = 'n',
border = NA)
```

Distribution of bills_introduced under control and treatment in Arkansas



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)
```

- 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.

```
# Add the cluster ID to data
d3$cluster <- NA
# Assign the clusters for "a"
d3$cluster <- c(1,1,2,2,3,3,4,4,5,5,6,6,7,7)
head(d3)
```

```
## Village Y D Block cluster
## 1      1 0 0      1      1
## 2      2 1 0      1      1
## 3      3 2 1      1      2
## 4      4 4 2      1      2
## 5      5 4 0      1      3
```

```
## 6      6 6 0      1      3
# pull out the cluster ID's
all_clusters <- unique(d3$cluster)
all_clusters

## [1] 1 2 3 4 5 6 7
# Number of clusters in treatment
clusters.in.treatment <- 3
# Randomize based on unique cluster ID's
randomize.clusters <- function(d) {
  treat.clusters <- sample(x = all_clusters,
                           size = clusters.in.treatment,
                           replace = FALSE)
  return(as.numeric(d$cluster %in% treat.clusters))
}

# Trying out one randomization. Now we have the data prepared for the experiment
d3$treatment <- randomize.clusters(d3)

# QUESTION & TODO More randomizations to get an estimate. For now just treating it as one experiment

clusters <- group_by(d3, cluster)
clusters.summary <- summarize(clusters, count=n(), mean_y=mean(Y), mean_d=mean(D), treatment=sum(treatment))

## Warning: package 'bindrcpp' was built under R version 3.4.4
clusters.summary

## # A tibble: 7 x 5
##   cluster count mean_y mean_d treatment
##   <dbl> <int> <dbl> <dbl> <dbl>
## 1     1     2    0.5     0         0
## 2     2     2     3     1.5       1
## 3     3     2     5     0         0
## 4     4     2    7.5    2.5       0
## 5     5     2   14.5   10.5       1
## 6     6     2    16    11.5       0
## 7     7     2   17.5    11         1

# QUESTION When calculating variances, are we calculating among all the clusters or clusters in control

# Variance in y(0) clustered
variance_y <- (sum((clusters.summary$mean_y - mean(clusters.summary$mean_y))^2))/nrow(clusters.summary)
variance_y

## [1] 39.69388

# Variance in y(1) clustered
variance_d <- (sum((clusters.summary$mean_d - mean(clusters.summary$mean_d))^2))/nrow(clusters.summary)
variance_d

## [1] 25.20408

# Covariance among y(0) and Y(1)
covariance <- sum((clusters.summary$mean_y - mean(clusters.summary$mean_y))*(clusters.summary$mean_d - mean(clusters.summary$mean_d)))
covariance
```



```
## [1] 30.53061
# applying 3.22

# m is subjects in treatment (not clusters). N is count of subjects, number of clusters is 7
m <- clusters.in.treatment * 2
m

## [1] 6
N <- nrow(d3)
N

## [1] 14
nrow(clusters.summary)

## [1] 7
se <- sqrt(((m*variance_y/N-m) + ((N-m)*variance_d/m) + 2*covariance)/nrow(clusters.summary))
se

## [1] 3.885473
```

The standard error for one randomization (3 groups in treatment) is 3.885473

- 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.

```
# repeating by grouping the observations in a different way
d4 <- read.csv("./data/ggChapter3.csv", stringsAsFactors = FALSE)
# Add the cluster ID to data
d4$cluster <- NA
# Assign the clusters for "a"
d4$cluster <- c(1,2,3,4,5,6,7,7,6,5,4,3,2,1)
head(d4)

##   Village Y D Block cluster
## 1      1 0 0      1      1
## 2      2 1 0      1      2
## 3      3 2 1      1      3
## 4      4 4 2      1      4
## 5      5 4 0      1      5
## 6      6 6 0      1      6

# Trying out one randomization. Now we have the data prepared for the experiment
d4$treatment <- randomize.clusters(d4)
# Grouping by the cluster and creating the summary
clusters <- group_by(d4, cluster)
clusters.summary <- summarize(clusters, count=n(), mean_y=mean(Y), mean_d=mean(D), treatment=sum(treatment))
clusters.summary
```

```
## # A tibble: 7 x 5
##   cluster count mean_y mean_d treatment
##   <dbl> <int> <dbl> <dbl> <dbl>
## 1      1      2      9      8.5      0
## 2      2      2      9      2.5      0
## 3      3      2      9      8      1
## 4      4      2     10      5      1
```

```
## 5      5      2      9.5      4.5      1
## 6      6      2      10      6      0
## 7      7      2      7.5      2.5      0

# Variance in y(0) clustered
variance_y <- (sum((clusters.summary$mean_y - mean(clusters.summary$mean_y))^2))/nrow(clusters.summary)
variance_y

## [1] 0.622449

# Variance in y(1) clustered
variance_d <- (sum((clusters.summary$mean_d - mean(clusters.summary$mean_d))^2))/nrow(clusters.summary)
variance_d

## [1] 4.918367

# Covariance among y(0) and Y(1)
covariance <- sum((clusters.summary$mean_y - mean(clusters.summary$mean_y))*(clusters.summary$mean_d - mean(clusters.summary$mean_d)))/nrow(clusters.summary)
covariance

## [1] 0.6020408

# applying 3.22

# m is subjects in treatment (not clusters). N is count of subjects, number of clusters is 7
m <- clusters.in.treatment * 2
m

## [1] 6

N <- nrow(d4)
N

## [1] 14

nrow(clusters.summary)

## [1] 7

se <- sqrt(((m*variance_y/N-m) + ((N-m)*variance_d/m) + 2*covariance)/nrow(clusters.summary))
se

## [1] 0.5383399
```

- 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?

The methods lead to different standard errors due to differences in the cluster means (for outcomes in control and treatment) which indicate the similarities or differences among the subjects (villages in this case, that become part of a cluster).

In this problem, the first method was akin to clustering based on geography. Villages next to each other (in Sr. no), had similar outcomes and ended up in the same cluster, leading to large differences among the clusters. Method 2 mixed things up and clustered villages that had “more” different outcomes. This sharply reduced the differences among the clusters

The implication is that cluster design has to account for averages of outcomes among the subjects of the cluster or we'll get imprecise estimates for the treatment effect. The best scenario would be to bring down the variance among clusters by grouping unlike subjects (like method 2 where dissimilar villages were clustered together). This may not be possible as clustering is mostly forced (like villages, schools by geography). The next best idea could

be to increase the number of clusters and then simulate all possible random assignments of clusters to obtain the sampling distribution

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)?

```
# Each site visitor shown the ad is exposed to $.10 of ads
```

```
# There are 1,000,000 users in one week
```

```
# Profit is $100 on every phone
```

```
# y(0) - .5% of visitors to news site buy an iphone in a week (in absence of ads)
```

```
week.users <- 1000000
```

```
week.users
```

```
## [1] 1e+06
```

```
profit_per_phone <- 100
```

```
profit_per_phone
```

```
## [1] 100
```

```
week.percent_users_buy <- .5
```

```
week.percent_users_buy
```

```
## [1] 0.5
```

```
week.number_users_buy_control <- (.5/100)*1000000
```

```
week.number_users_buy_control
```

```
## [1] 5000
```

```
week.profit_control <- week.number_users_buy_control * profit_per_phone
```

```
week.profit_control
```

```
## [1] 5e+05
```

```
adspend_per_user <- .10
```

```
adspend_per_week <- adspend_per_user * week.users
```

```
adspend_per_week
```

```
## [1] 1e+05
```

```
# We should atleast make up the money spend on ads in a week. That means we have to sell +inc number of
```

```
inc <- adspend_per_week/profit_per_phone
```

```
inc
```

```
## [1] 1000
```

```
# so we have to sell 1000 additional phones
# Which is obs % points based on the original population eligible for ads and purchases
inc_per <- (inc/week.users)*100
inc_per
```

```
## [1] 0.1
```

```
# This is breakeven. Anything more would be positive ROI
```

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.

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

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. 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.
- In plain language, what does this confidence interval mean?
- 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.
- Calculate the 95% confidence interval you get from the regression.

- e. On to p-values. What p-value does the regression report? Note: please use two-tailed tests for the entire problem.
- f. Now compute the same p-value using randomization inference.
- 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).)
- 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?