Problem Set 2: Randomization Inference

GOV 2003

Due at 11:59 pm (ET) on Sept 22, 2021

Instruction

Before you begin, please read the following instructions **carefully**:

- **No late submission are allowed** without a prior approval from the instructors.
- **All answers should be typed up**. We recommend the use of `Rmarkdown`. An `Rmarkdown` template for this problem set is provided. Answers to analytical solutions should also be typed up.
- **A PDF copy of your answer** including your computer code should be uploaded to `Gradescope` before the deadline. **Do not submit the markdown file itself**.
- This problem set includes a bonus question for extra credit. No deduction in the total points will be made from this question. Note that the maximum points of this problem set is 15 points. That is, if the student receives 3 points from the bonus question and 14 points from the other questions, the total points will be 15 points.

Introduction

In this problem set, we revisit the field experiments conducted by Kalla and Broockman (2020), focusing on the comparison between the placebo group and full intervention group. We will investigate two different setups — one for small samples with a binary outcome (`Part A`) and the other for large samples with a continuous outcome (`Part B`) — using randomization inference. We assume no interference between units and unit-level random assignment of the canvassing throughout. Note that we would not consider any noncompliance issue in this problem set. For the empirical questions, use each of two data frames (`small_sample` and `large_sample`) from `RI_KallaBrookman2020.RData` depending on the setup (the data is modified from the original dataset for use in the problem set).

Part A: Small samples with a binary outcome

Setup

Let D_i denote the canvassing with two categories — *placebo* conversation unrelated to immigration ($D_i = 0$) and *full intervention* including non-judgemental exchange of narratives ($D_i = 1$). Here, we assume that we are conducting **completely randomized experiment** with 6 voters where $\sum_{i=1}^{6} D_i = 3$. Let Y_i be a **binary** outcome variable indicating the support for inclusive policies for each voter i — $Y_i = 1$ if voter i agreed with the statement "I would have no problem living in areas where undocumented immigrants live" after the canvassing, and $Y_i = 0$ if disagreed.

Question 1 (5 pts; 1pt for each)

(a) Write a mathematical expression for the sharp null hypothesis of no effect (H_0) and give its interpretation in the context of the study.

(b) Using the data frame named `small_sample`, create the potential outcomes table (i.e., science table) and fill in the missing potential outcomes under the sharp null (H_0).

Hint: Use the following code as a starting point, and fill in the missing data using the sharp null. Note that the column named `canvass` is the treatment variable and `support` is the observed outcome variable.

```
load("RI_KallaBrookman2020.RData")
library(tidyverse)
table0 <- small_sample %>%
  mutate(Di = recode(canvass,
                     `Placebo` = 0, `Full Intervention` = 1),
         Yi = support) %>%
  mutate(`Yi(0)` = ifelse(Di == 0, support, NA),
         `Yi(1)` = ifelse(Di == 1, support, NA)) %>%
  select(id, Di, Yi, `Yi(0)`, `Yi(1)`)
# TODO 1. Fill in the missing potential outcomes using the sharp null of no effect
# TODO 2. Print the table
```

(c) Suppose that we chose the absolute **difference-in-means** estimator as a test statistic for our test. Write a mathematical expression for this test statistic (T_{diff}) and calculate the observed test statistic ($T_{\text{diff}}^{\text{obs}}$) using the data from (b).

(d) Consider all the possible treatment vectors ($\widetilde{\mathbf{D}}$), and calculate the corresponding test statistics ($\widetilde{T} = T_{\text{diff}}(\widetilde{\mathbf{D}}, \mathbf{Y})$) under the sharp null ($H_0$). List all the test statistics.

Hint: There exist total $\binom{6}{3} = 20$ possible treatment vectors ($\widetilde{\mathbf{D}}_1, \ldots, \widetilde{\mathbf{D}}_{20}$). Use the sample code below to calculate the test statistics ($\widetilde{T}_1, \ldots, \widetilde{T}_{20}$).

```
# This function calculates the absolute difference-in-means under the sharp null of no effect
DiffInMeans <- function(Dtilde, Y, comp_rand = T) {
  if(comp_rand & (sum(Dtilde)!=3))
    stop("Check the number of treated units in the treatment vector D.\n")
  Ttilde <- abs(sum(Dtilde*Y)/3 - sum((1-Dtilde)*Y)/3) # Yi = Yi(0) = Yi(1) under H0
```

```
  return(Ttilde)
}
# E.g. If D_tilde = (1,1,1,0,0,0), then T_diff(D_tilde, Y) is:
DiffInMeans(c(1,1,1,0,0,0), small_sample$support)

Dtilde_ls <- NULL # List of all the possible treatment vectors.
                  # Dimensions: 6 by 20 (or 20 by 6).
Ttilde_ls <- NULL # List of corresponding test statistics.
                  # Dimensions: 1 by 20 (or 20 by 1)

# TODO 1: Fill in the list of all the possible treatment vectors (Dtilde_ls).
#         You can use ri::genperms() function.
# TODO 2: Fill in the list of all corresponding test statistics (Ttilde_ls)
#         using the function DiffInMeans()
# TODO 3: Print Ttilde_ls
```

(e) Plot a histogram of the test statistics $(\widetilde{T}_1, \ldots, \widetilde{T}_{20})$ and calculate the exact p-value. Specify the alternative hypothesis and briefly discuss the results.

Hint: Review the lecture slides p.17 for the definition of the exact p-value, and p.41 for the alternative hypothesis.

Answer 1

(a)

$$H_0 : \tau_i = Y_i(1) - Y_i(0) = 0 \quad \forall i \in \{1, \ldots, 6\}$$

(b)

```
load("RI_KallaBrookman2020.RData")
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.0.4     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
table0 <- small_sample %>%
  mutate(Di = recode(canvass,
                     `Placebo` = 0, `Full Intervention` = 1),
         Yi = support) %>%
  mutate(`Yi(0)` = ifelse(Di == 0, support, NA),
         `Yi(1)` = ifelse(Di == 1, support, NA)) %>%
```

```
  select(id, Di, Yi, `Yi(0)`, `Yi(1)`)

# 1. Fill in the missing potential outcomes using the sharp null of no effect
table1 <- table0 %>%
  mutate(`Yi(0)` = ifelse(Di == 0, `Yi(0)`, paste0("(",Yi,")")),
         `Yi(1)` = ifelse(Di == 1, `Yi(1)`, paste0("(",Yi,")")),
         `Yi(1)-Yi(0)` = 0)
# 2. Print the table
table1
```

```
##   id Di Yi Yi(0) Yi(1) Yi(1)-Yi(0)
## 1  1  1  1   (1)     1           0
## 2  2  1  1   (1)     1           0
## 3  3  1  1   (1)     1           0
## 4  4  0  0     0   (0)           0
## 5  5  0  1     1   (1)           0
## 6  6  0  0     0   (0)           0
```

(c)

$$T_{\text{diff}} = \left| \frac{1}{3} \sum_{i=1}^{6} D_i Y_i - \frac{1}{3} \sum_{i=1}^{6} (1 - D_i) Y_i \right|.$$

```
Di <- recode(small_sample$canvass, `Placebo` = 0, `Full Intervention` = 1)
Yi <- small_sample$support
T_obs <- abs(sum(Di*Yi)/3 - sum((1-Di)*Yi)/3)
T_obs
```

```
## [1] 0.6666667
```

(d)

```
# This function calculates the absolute difference-in-means under the sharp null of no effect
DiffInMeans <- function(Dtilde, Y, comp_rand = T) {
  if(comp_rand & (sum(Dtilde)!=3))
    stop("Check the number of treated units in the treatment vector D.\n")
  Ttilde <- abs(sum(Dtilde*Y)/3 - sum((1-Dtilde)*Y)/3) # Yi = Yi(0) = Yi(1) under H0
  return(Ttilde)
}
# 1. Fill in the list of all the possible treatment vectors (Dtilde_ls).
Dtilde_ls <- recode(small_sample$canvass, `Placebo` = 0, `Full Intervention` = 1) %>%
  ri::genperms()
# 2: Fill in the list of all corresponding test statistics (Ttilde_ls)
Ttilde_ls <- rep(NA, ncol(Dtilde_ls))
for (i in 1:ncol(Dtilde_ls)) {
  Ttilde_ls[i] <- DiffInMeans(Dtilde_ls[,i], small_sample$support)
}
# 3: Print Ttilde_ls
Ttilde_ls
```

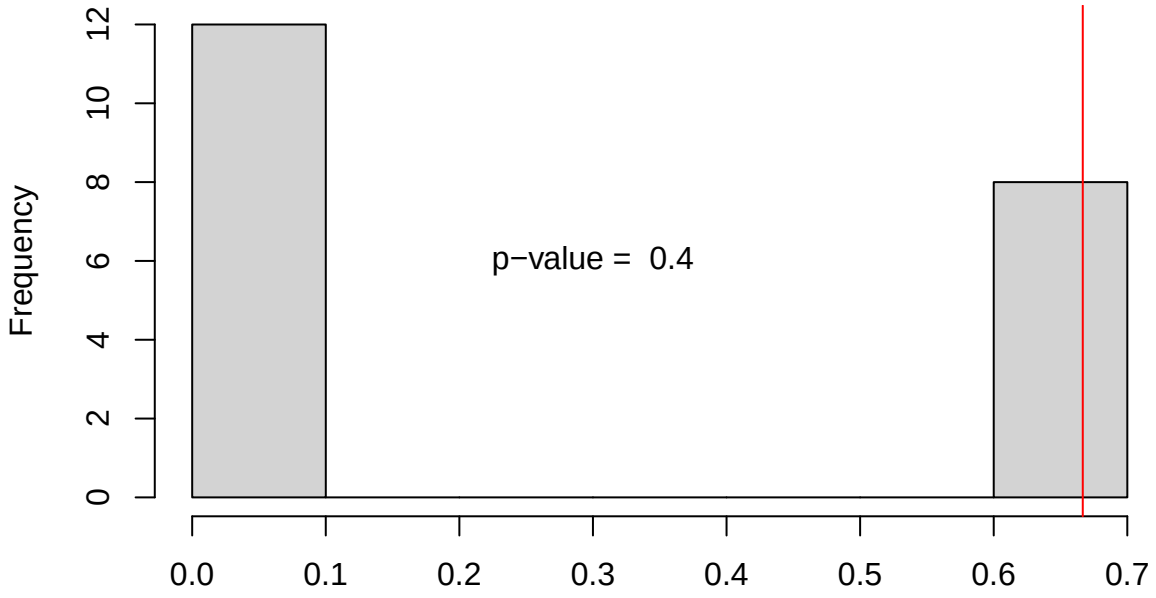
```
##  [1] 0.6666667 0.0000000 0.6666667 0.0000000 0.0000000 0.6666667 0.0000000
```

```
##  [8] 0.0000000 0.6666667 0.0000000 0.0000000 0.6666667 0.0000000 0.0000000
## [15] 0.6666667 0.0000000 0.0000000 0.6666667 0.0000000 0.6666667
```

(e)

$$H_1 : \tau_i = Y_i(1) - Y_i(0) \neq 0 \quad \text{for some } i \in \{1, \ldots, 6\}$$

```
hist(Ttilde_ls, xlab = "", main = "")
abline(v = T_obs, col = 'red')
text(0.3, 6, labels = paste("p-value = ", mean(Ttilde_ls >= T_obs)), cex = 1)
```



[Bonus] Analytical Question: Fisher's exact test (2 pts)

(This question is for extra credits. Please read the instruction above for more details.)

Suppose we have chosen the following statistics: $S = \sum_{i=1}^{6} D_i Y_i(1)$. Which distribution does the randomization distribution of S follow under the complete randomization? Write down the mathematical expressions for mean and variance.

Hint: Under the sharp null and complete randomization, $S = s$ shows that among the units who support the inclusive policies regardless of their treatment, s had indeed assigned to the full intervention group, whereas among the units who don't support the inclusive policies regardless of their treatment, $3 - s$ had assigned to the full intervention group.

[Bonus] Answer

Let $m = \sum_{i=1}^{6} Y_i$ denote the number of units who support the inclusive policies in our data. The probability of $S = s$ given the sharp null and complete randomization is same as the number of cases where s units, among m units who support the inclusive policies anyway, having been assigned to the full intervention group and $3 - s$ units, among $6 - m$ units who don't support the inclusive policies anyway, having been assigned to the same group (*numerator*) divided by the total number of cases where we randomly assign 3 units to the full intervention group among 6 units (*denominator*).

$$\Pr(S = s \mid \{Y_i(0), Y_i(1)\}_{i=1}^{6}) = \frac{\binom{m}{s}\binom{6-m}{3-s}}{\binom{6}{3}}$$

This follows the hyper-geometric distribution. Observe that $\mathbb{E}[S \mid \{Y_i(0), Y_i(1)\}_{i=1}^{6}] = \frac{m}{2}$ and $\mathbb{V}[S \mid \{Y_i(0), Y_i(1)\}_{i=1}^{6}] = \frac{3m}{10}(1 - \frac{m}{6})$.

Part B: Large samples with a continuous outcome

Setup

Now we switch to the other setup. Here, we conduct **completely randomized experiment** with the same treatment (D_i) as before. We have much larger samples (N voters) where $\sum_{i=1}^{n} D_i = n_1$, and Y_i is a **continuous** outcome variable — larger value indicates a strong support for inclusive policies.

Question 2 (5 pts; 1pt for each)

We again test the sharp null of no effect under the complete randomization using `large_sample` ($N = 1200$ and $n_1 = 600$). Let Ω = set of 2^N treatment vectors (any N-vector of 0s and 1s). We define the set of feasible assignments under the complete randomization experiment as follows: $\Omega_0 = \{\mathbf{d} \in \Omega : \sum_{i=1}^{N} d_i = n_1\}$.

(a) Suppose that we chose the **rank** statstic as a test statistic for the hypothesis testing. Specifically, we will use the absolute difference in average ranks that has been normalized to have mean 0. Note that there is no ties in the data. Write a mathematical expression for this test statistic (T_{rank}) and calculate the observed test statistic ($T_{\text{rank}}^{\text{obs}}$) using the data frame named `large_sample`.

Hint: Review the lecture slides p.35 for the definition of the rank statistics.

(b) Sample 1000 treatment vectors ($\{\widetilde{\mathbf{D}}_j\}_{j=1}^{1000}$) from Ω_0. For each treatment vector, compute the test statistic ($\{\widetilde{T}_{\text{rank},j}\}_{j=1}^{1000}$). (Please include only your code for this question. You don't have to print the test statistics.)

(c) Plot a histogram of the test statistics ($\{\widetilde{T}_{\text{rank},j}\}_{j=1}^{1000}$) and compute the p-value. Briefly discuss the results in the context of the study.

(d) Suppose we changed the assignment mechanism to **Bernoulli randomization** instead of the complete randomization. Repeat (b) and (c) under this alternative design.

Hint: Be aware to sample treatment vectors from Ω instead of Ω_0 in this case. You can ignore any Bernoulli draw that is all treated or all control.

(e) Compare the randomization distributions of the test statistic (T_{rank}) under the complete randomization and Bernoulli randomization. Similarly, compare the p-values under the complete randomization and Bernoulli randomization. Which assignment mechanism would you prefer in this context?

Answer 2

(a)

$$T_{\text{rank}} = \left| \frac{\sum_{i:D_i=1} R_i}{n_1} - \frac{\sum_{i:D_i=0} R_i}{n_0} \right|$$

where $R_i = \sum_{j=1}^{N} \mathbb{I}(Y_j \leq Y_i) - \frac{N+1}{2}$ is the normalized rank of unit i.

```r
Di <- recode(large_sample$canvass, `Placebo` = 0, `Full Intervention` = 1)
Yi <- large_sample$factor
N <- length(Yi);  n1 <- sum(Di); n0 <- sum(1-Di)
Ri <- rank(Yi) - (N+1)/2
T_obs <- abs(sum(Di*Ri)/n1 - sum((1-Di)*Ri)/n0)
T_obs
```

```
## [1] 41.71333
```

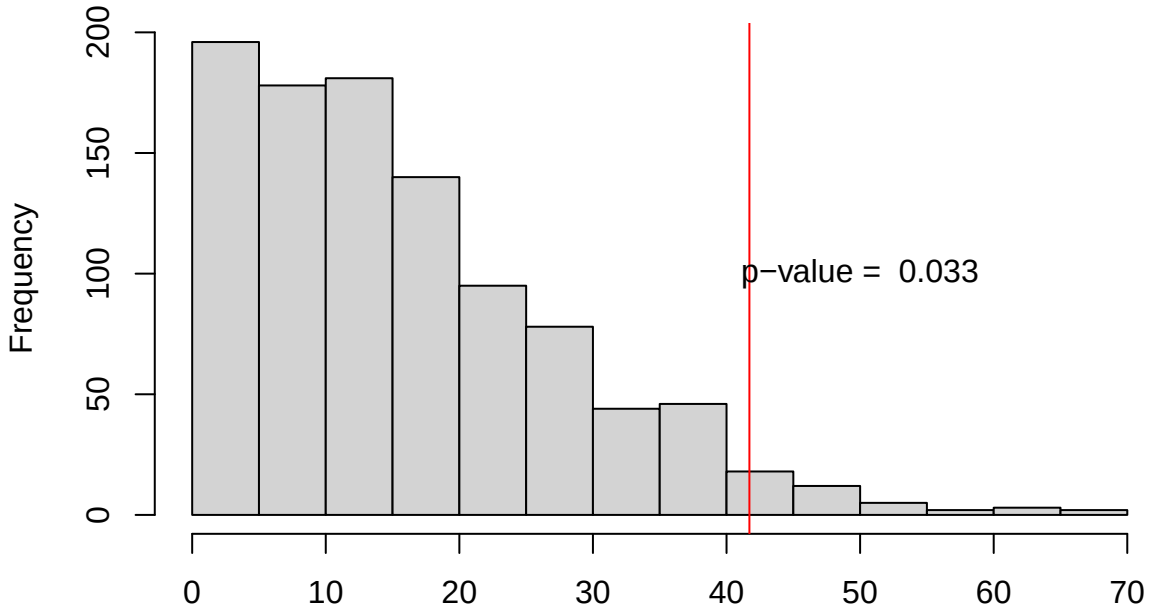
(b)

```r
# This function calculates the rank statistic under the sharp null of no effect
Rank <- function(Dtilde, Y, comp_rand = T) {
  if(comp_rand&(sum(Dtilde)!=n1))
    stop("Check the number of treated units in the treatment vector D.\n")
  Ri <- rank(Y) - (N+1)/2 # Yi = Yi(0) = Yi(1) under H0
  Ttilde <- abs(sum(Dtilde*Ri)/n1 - sum((1-Dtilde)*Ri)/n0)
  return(Ttilde)
}
n_sim <- 1000
Ttilde_ls <- rep(NA, n_sim)
set.seed(1234)
for (s in 1:n_sim) {
  Dtilde_s <- sample(Di)
  Ttilde_ls[s] <- Rank(Dtilde_s, Yi)
}
# Ttilde_ls
```

(c)

$$H_1 : \tau_i = Y_i(1) - Y_i(0) \neq 0 \quad \text{for some } i$$

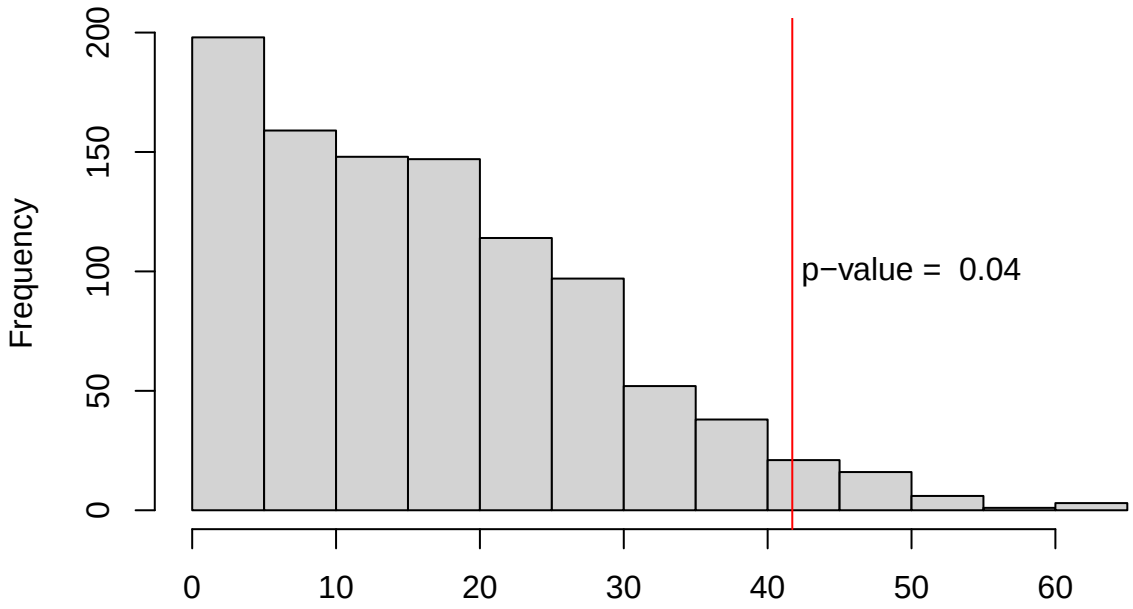
```r
hist(Ttilde_ls, xlab = "", main = "")
abline(v = T_obs, col = 'red')
text(50, 100, labels = paste("p-value = ", mean(Ttilde_ls >= T_obs)), cex = 1)
```



(d) The null and alternative hypothesis are same as before.

```
n_sim <- 1000
B_Ttilde_ls <- rep(NA, n_sim)
set.seed(1234)
for (s in 1:n_sim) {
  Dtilde_s <- rbinom(N, size=1, prob = 0.5)
  if(sum(Dtilde_s) %in% c(0,N)) cat("Dtilde_", s,": all assigned to treated/control group.\n")
  B_Ttilde_ls[s] <- Rank(Dtilde_s, Yi, comp_rand = F)
}
# Ttilde_ls
```

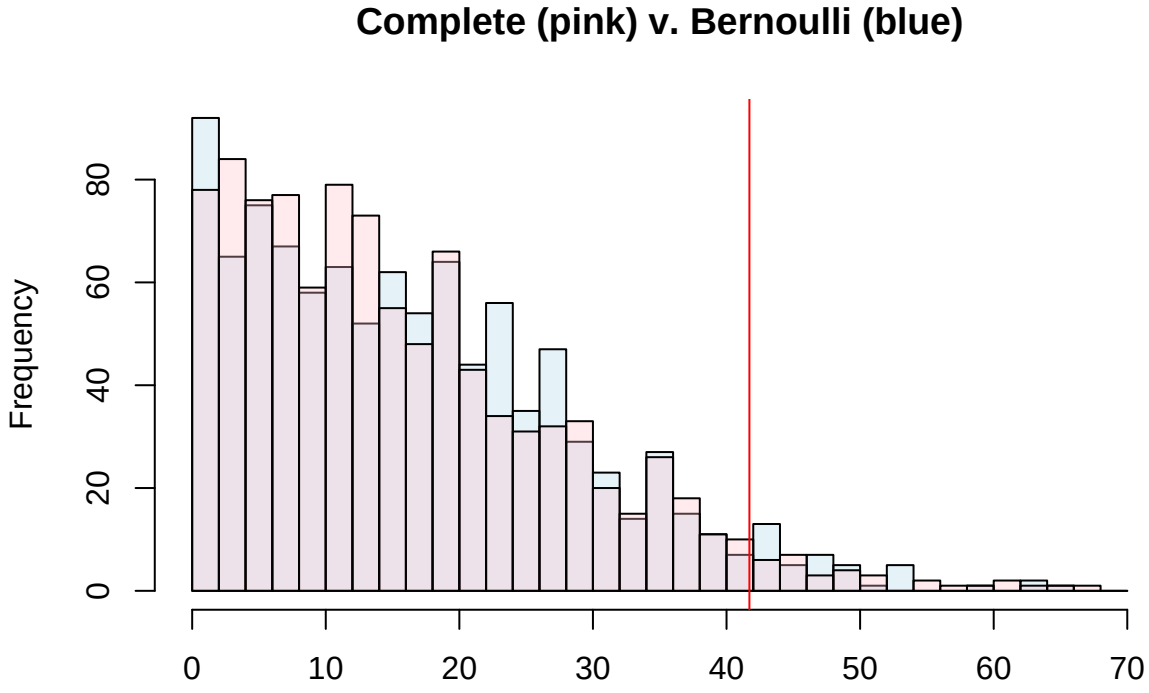
```
hist(B_Ttilde_ls, xlab = "", main = "")
abline(v = T_obs, col = 'red')
text(50, 100, labels = paste("p-value = ", mean(B_Ttilde_ls >= T_obs)), cex = 1)
```



(e)

```
Bern <- hist(B_Ttilde_ls, breaks = seq(0, 70, by=2), plot = FALSE)
Comp_Rand <- hist(Ttilde_ls, breaks = seq(0, 70, by=2), plot = FALSE)

plot(Bern, col = rgb(173,216,230,max = 255, alpha = 80, names = "lt.blue"),
     main = "Complete (pink) v. Bernoulli (blue)", xlab = NULL)
plot(Comp_Rand, col = rgb(255,192,203, max = 255, alpha = 80, names = "lt.pink"), add = TRUE)
abline(v = T_obs, col = 'red')
```



```r
sd(Ttilde_ls)
```

```
## [1] 12.07011
```

```r
sd(B_Ttilde_ls)
```

```
## [1] 12.15311
```

The randomization distribution of Bernoulli randomization has a larger variance than the complete randomization. In line with this observation, the p-value under the Bernoulli randomization has a larger value.

Also, note that the Bernoulli randomization might lead to "bad" randomization where all the units are assigned to either treated or control group which is especially a concern in small sample size.

Question 3 (5 pts)

In this question, we will conduct a test inversion to calculate the 95% confidence interval.

Assume a complete randomization experiment and choose the difference-in-means estimator as our test statistics:

$$T_{\text{diff}}(\mathbf{D}, \mathbf{Y}) = \frac{1}{n_1}\sum_{i=1}^{N} D_i Y_i - \frac{1}{1-n_1}\sum_{i=1}^{N}(1-D_i)Y_i.$$

Hint:

- Although we are **not** using the **absolute** difference-in-means, we still assume a two-sided alternative hypothesis ($H_0 : \tau_i = c$ v. $H_1 : \tau_i \neq c$ for some i) and compute the p-value accordingly (see step 2-5 of the code below).
- You may use the sample code below.
- Alternatively, you can use **adjusted outcomes** for the test (see the lecture slides p.44).

```r
# Data
Yi <- large_sample$factor # Observed outcome
Di <- recode(large_sample$canvass, `Placebo` = 0, `Full Intervention` = 1) # Observed treatmen
N <- length(Yi);  n1 <- sum(Di); n0 <- sum(1-Di)
# Pick candiate taus on a grid
tau_cand <- seq(-0.5, 0.5, by = 0.01)
save_pval <- rep(NA, length(tau_cand)) # to save the p-value below

# TODO 1. Calculate the observed statistics
T_obs <- sum(Di*Yi)/n1 - sum((1-Di)*Yi)/n0

# TODO 2: Create function for computing p-value given tau and observed statistics
your_fun <- function(tau, t_obs, n_sim = 1000) { # Input: tau, observed statistics
  # TODO 2-1: Calculate Yi(1) using Yi, Di, and tau
  Y1 <- NULL
  # TODO 2-2: Calculate Yi(0) using Yi, Ti, and tau
  Y0 <- NULL
```

```
  Ttilde_ls <- rep(NA, n_sim)
  # Simulation:
  for (s in 1:n_sim) {
    # TODO 2-3: Randomly sample treatment vectors
    Dtilde_s <- NULL
    # TODO 2-4: For each treatment vector, compute the statistics
    Ttilde_ls[s] <- NULL
  }
  # 2-5: Calculate and return the p-value
  pval <- 2 * min(mean(Ttilde_ls >= t_obs), mean(Ttilde_ls <= t_obs))
  return(pval)
}

# TODO 3: Loop over each candidate tau
set.seed(123)
for (t in 1:length(tau_cand)) {
  save_pval[t] <- your_fun(tau_t, T_obs)
}

# 4. Obtain the upper / lower bound of 95% CI
lb <- tau_cand[min(which(save_pval >= 0.05))]
ub <- tau_cand[max(which(save_pval >= 0.05))]

# TODO 5: Print the 95% CI (lb, ub)
```

Answer 3

```
# Data
Yi <- large_sample$factor # Observed outcome
Di <- recode(large_sample$canvass, `Placebo` = 0, `Full Intervention` = 1) # Observed treatmen
N <- length(Yi);  n1 <- sum(Di); n0 <- sum(1-Di)
# Pick candiate taus on a grid
tau_cand <- seq(-0.5, 0.5, by = 0.01)
save_pval <- rep(NA, length(tau_cand)) # to save the p-value below

# 1. Calculate the observed statistics
T_obs <- sum(Di*Yi)/n1 - sum((1-Di)*Yi)/n0

# 2: Create function for computing p-value given tau and observed statistics
pval_perm_test <- function(tau, t_obs, n_sim = 1000) { # Input: tau, observed statistics
  # 2-1: Calculate Yi(1) using Yi, Di, and tau
  Y1 <- ifelse(Di==1, Yi, Yi + tau)
  # 2-2: Calculate Yi(0) using Yi, Ti, and tau
  Y0 <- ifelse(Di==0, Yi, Yi - tau)
  Ttilde_ls <- rep(NA, n_sim)
  # Simulation:
```

```
  for (s in 1:n_sim) {
    # 2-3: Randomly sample treatment vectors
  Dtilde_s <- sample(Di)
    # 2-4: For each treatment vector, compute the statistics
  Ttilde_ls[s] <- sum(Dtilde_s*Y1)/n1 - sum((1-Dtilde_s)*Y0)/n0
  }
  # 2-5: Calculate and return the p-value
  pval <- 2 * min(mean(Ttilde_ls >= t_obs), mean(Ttilde_ls <= t_obs))
  return(pval)
}

# 3: Loop over each candidate tau
set.seed(123)
for (t in 1:length(tau_cand)) {
  tau_t <- tau_cand[t]
  save_pval[t] <- pval_perm_test(tau_t, T_obs)
}

# 4. Obtain the upper / lower bound of 95% CI
lb <- tau_cand[min(which(save_pval >= 0.05))]
ub <- tau_cand[max(which(save_pval >= 0.05))]

# 5: Print the 95% CI (lb, ub)
cat("95% CI: (", lb, ",", ub,")\n")
```

```
## 95% CI: ( 0.01 , 0.23 )
```

Alternatively, use the adjusted outcomes:

```
# Data
Yi <- large_sample$factor # Observed outcome
Di <- recode(large_sample$canvass, `Placebo` = 0, `Full Intervention` = 1) # Observed treatmen
N <- length(Yi);  n1 <- sum(Di); n0 <- sum(1-Di)
# Pick candiate taus on a grid
tau_cand <- seq(-0.5, 0.5, by = 0.01)
save_pval <- rep(NA, length(tau_cand)) # to save the p-value below

# 1: Create function for computing p-value given the adjusted outcomes (Ystar)
pval_perm_test <- function(ystar, t_obs, n_sim = 1000) {
  Ttilde_ls <- rep(NA, n_sim)
  # Simulation:
  for (s in 1:n_sim) {
    # Randomly sample treatment vectors
    Dtilde_s <- sample(Di)
    # For each treatment vector, compute the statistics
    Ttilde_ls[s] <- sum(Dtilde_s*ystar)/n1 - sum((1-Dtilde_s)*ystar)/n0
  }
  # Calculate and return the p-value
```

```r
  pval <- 2 * min(mean(Ttilde_ls >= t_obs), mean(Ttilde_ls <= t_obs))
  return(pval)
}

# 2: Loop over each candidate tau
set.seed(123)
for (t in 1:length(tau_cand)) {
  tau_t <- tau_cand[t]
  Ystar <- Yi - Di*tau_t
  T_obs <- sum(Di*Ystar)/n1 - sum((1-Di)*Ystar)/n0
  save_pval[t] <- pval_perm_test(Ystar, T_obs)
}

# Obtain the upper / lower bound of 95% CI
lb <- tau_cand[min(which(save_pval >= 0.05))]
ub <- tau_cand[max(which(save_pval >= 0.05))]

# 3: Print the 95% CI (lb, ub)
cat("95% CI: (", lb, ",", ub,")\n")
```

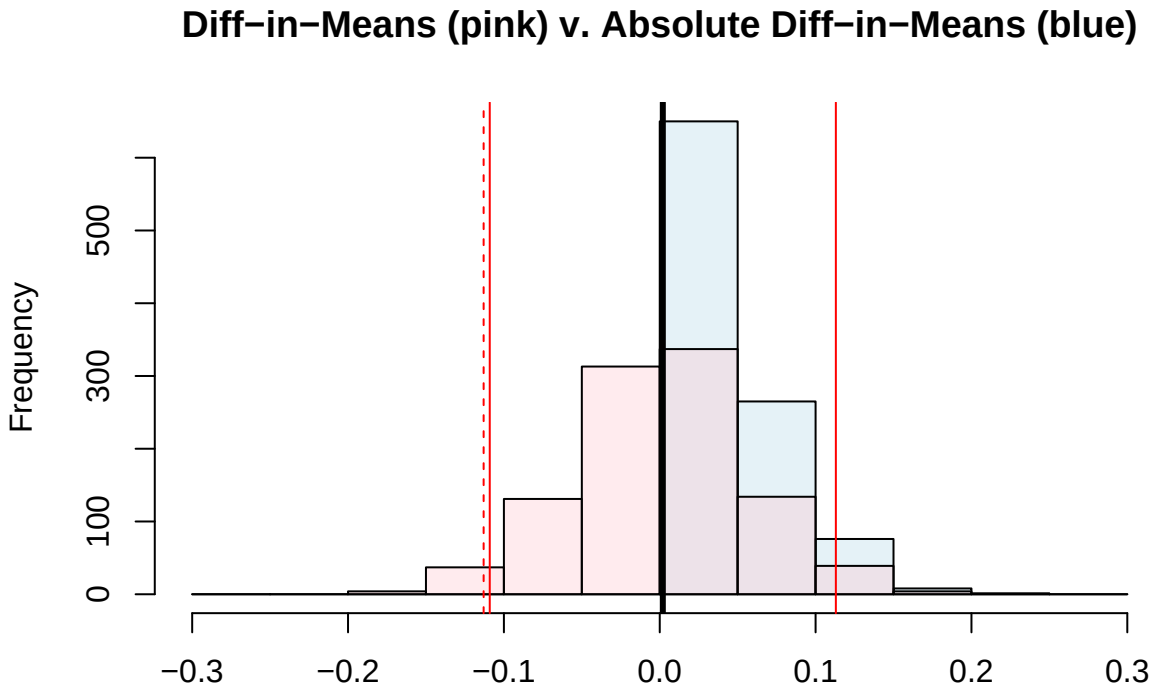
[Bonus] Question: Non-zero sharp nulls (2 pts)

If we chose the **absolute** difference-in-means estimator for the test statistics and yet did not adjust the outcomes, which part of your code above would have been changed and why?

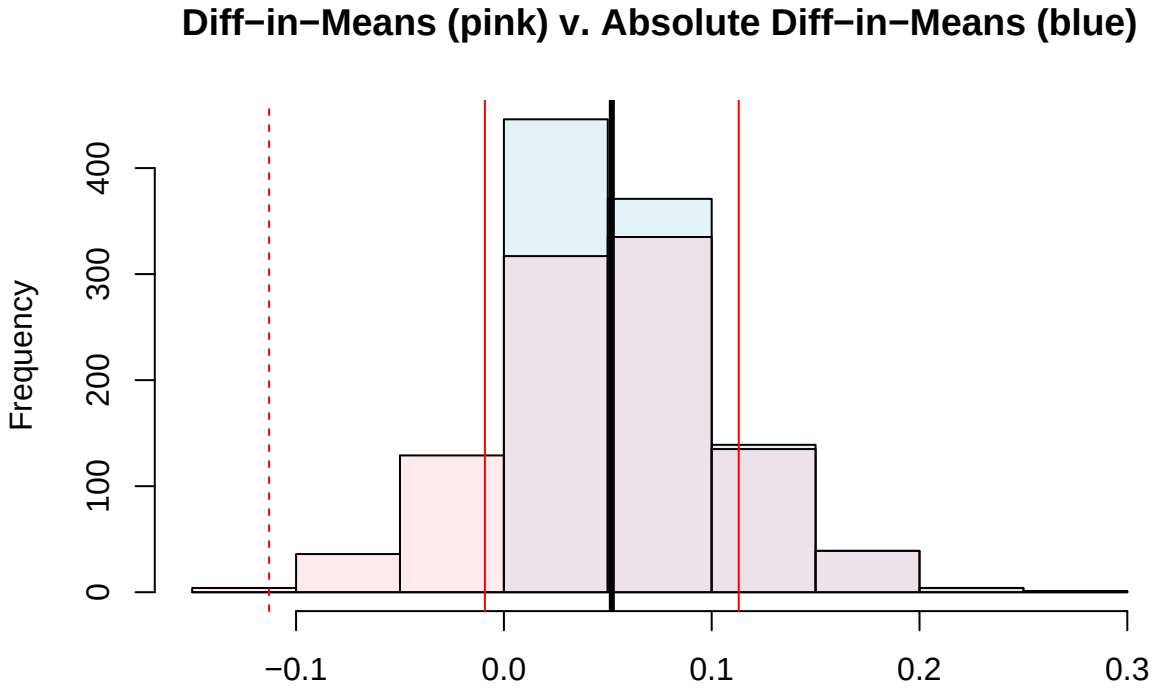
[Bonus] Answer

The way we compute the p-value would have been changed since the mean of the randomization distribution is no longer zero under the non-zero sharp null. See the examples below:

1. $H_0 : \tau_i = 0 \quad \forall i$



2. $H_0 : \tau_i = 0.05 \quad \forall i$



References

Kalla, J. and Broockman, D. (2020). Reducing exclusionary attitudes through interpersonal conversation: Evidence from three field experiments. *American Political Science Review*, 114(2):410–425.