Gerber and Green (2012) Chapter 5 Problem 10

Margaret Moor and Alexander Coppock, Yale University

January 30, 2018

This script shows how to conduct the randomization inference procedure in Gerber and Green (2012) Chapter 5 Problem 10 three different ways: using the `ri2` package, using the `ri` package, and by hand with a loop.

Chapter 5 Problem 10

Guan and Green report the results of a canvassing experiment conducted in Beijing on the eve of a local election. 14 Students on the campus of Peking University were randomly assigned to treatment or control groups. Canvassers attempted to contact students in their dorm rooms and encourage them to vote. No contact with the control group was attempted. Of the 2,688 students assigned to the treatment group, 2,380 were contacted. A total of 2,152 students in the treatment group voted; of the 1,334 students assigned to the control group, 892 voted. One aspect of this experiment threatens to violate the exclusion restriction. At every dorm room they visited, even those where no one answered, canvassers left a leaflet encouraging students to vote.

(a) Using the dataset at http://isps.research.yale.edu/FEDAI, estimate the ITT.

SHOWN BELOW

(b) Use randomization inference to test the sharp null hypothesis that the ITT is zero for all observations, taking into account the fact that random assignment was clustered by dorm room. Interpret your results.

SHOWN BELOW

(c) Assume that the leaflet had no effect on turnout. Estimate the CACE.
(d) Assume that the leaflet raised the probability of voting by one percentage point among both Compliers and Never-Takers. In other words, suppose that the treatment group's turnout rate would have been one percentage point lower had the leaflets not been distributed. Write down a model of the expected turnout rates in the treatment and control groups, incorporating the average effect of the leaflet.
(e) Given this assumption, estimate the CACE of canvassing.
(f) Suppose, instead, that the leaflet had no effect on Compliers (who heard the canvasser's speech and ignored the leaflet) but raised turnout among Never-Takers by 3 percentage points. Given this assumption, estimate the CACE of canvassing.

NOT SHOWN

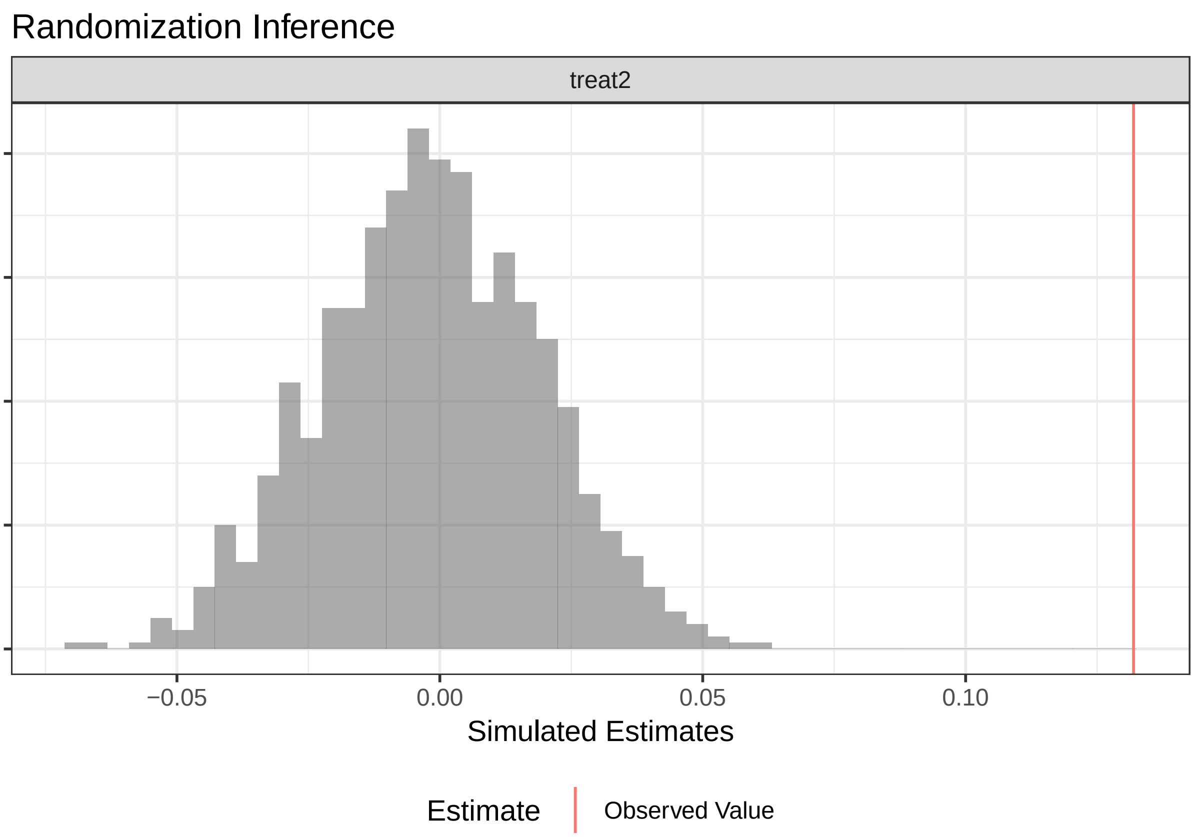
```r
# Data from http://isps.yale.edu/FEDAI
library(haven)
data5.10 <- read_dta("datasets/5.10.dta")

# Remove observations with missing outcomes
data5.10 <- na.omit(data5.10)

# Number of sims the same for all three methods
sims <- 1000
```

In ri2

```
library(randomizr)
library(ri2)

# Declare randomization procedure

declaration <- declare_ra(clusters = data5.10$dormid)

# Conduct Randomization Inference
ri2_out <- conduct_ri(turnout ~ treat2,
                      declaration = declaration,
                      assignment = "treat2",
                      sharp_hypothesis = 0,
                      data = data5.10)

summary(ri2_out)
```

```
##   coefficient  estimate two_tailed_p_value null_ci_lower null_ci_upper
## 1      treat2 0.1319296                  0   -0.04188233    0.03799655
```

```
plot(ri2_out)
```

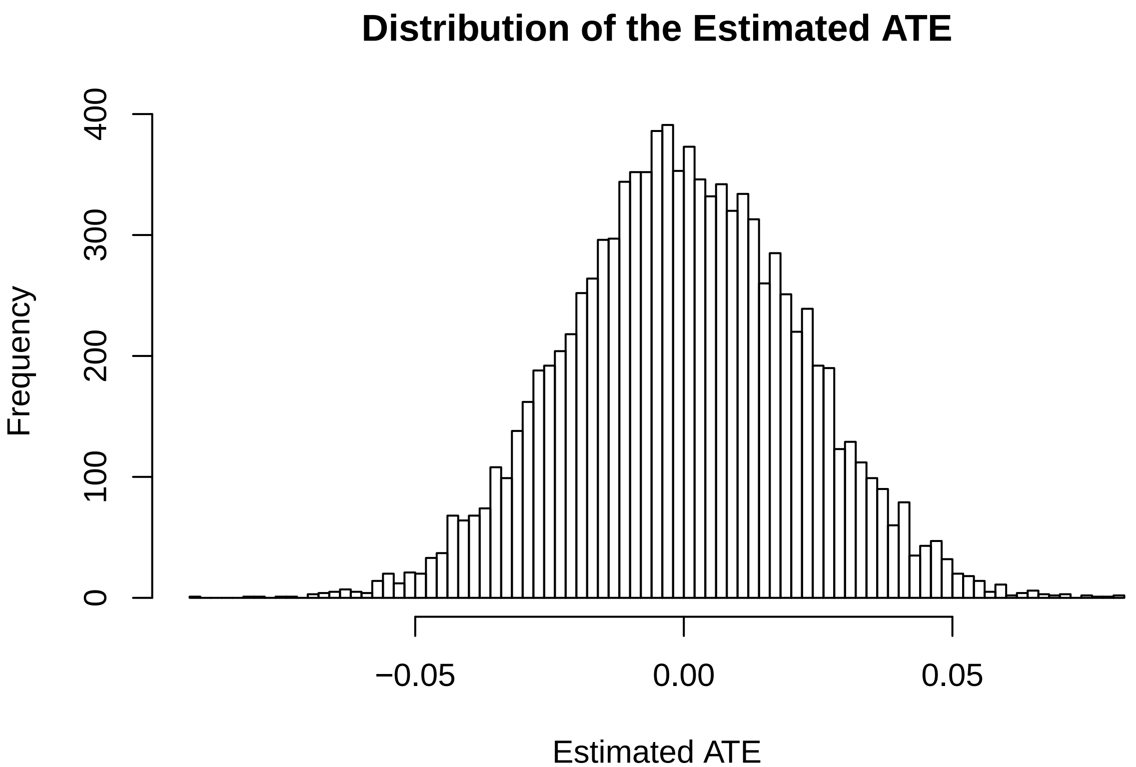


In ri

```r
library(ri)
# all possible permutations
perms <- genperms(data5.10$treat2, clustvar = data5.10$dormid)
```

```
## Too many permutations to use exact method.
## Defaulting to approximate method.
## Increase maxiter to at least 5.83377797524832e+275 to perform exact estimation.
```

```r
# probability of treatment
probs <- genprobexact(data5.10$treat2, clustvar = data5.10$dormid)
# estimate the ITTy
ate <- estate(data5.10$turnout, data5.10$treat2, prob = probs)

## Conduct Sharp Null Hypothesis Test of Zero Effect for Each Unit

# generate potential outcomes under sharp null of no effect
Ys <- genouts(data5.10$turnout, data5.10$treat2, ate = 0)
# generate sampling dist. under sharp null
distout <- gendist(Ys, perms, prob = probs)
# display characteristics of sampling dist. for inference
ri_out <- dispdist(distout, ate)
```




```r
ri_out
```

```
## $two.tailed.p.value
## [1] 0
##
## $two.tailed.p.value.abs
```

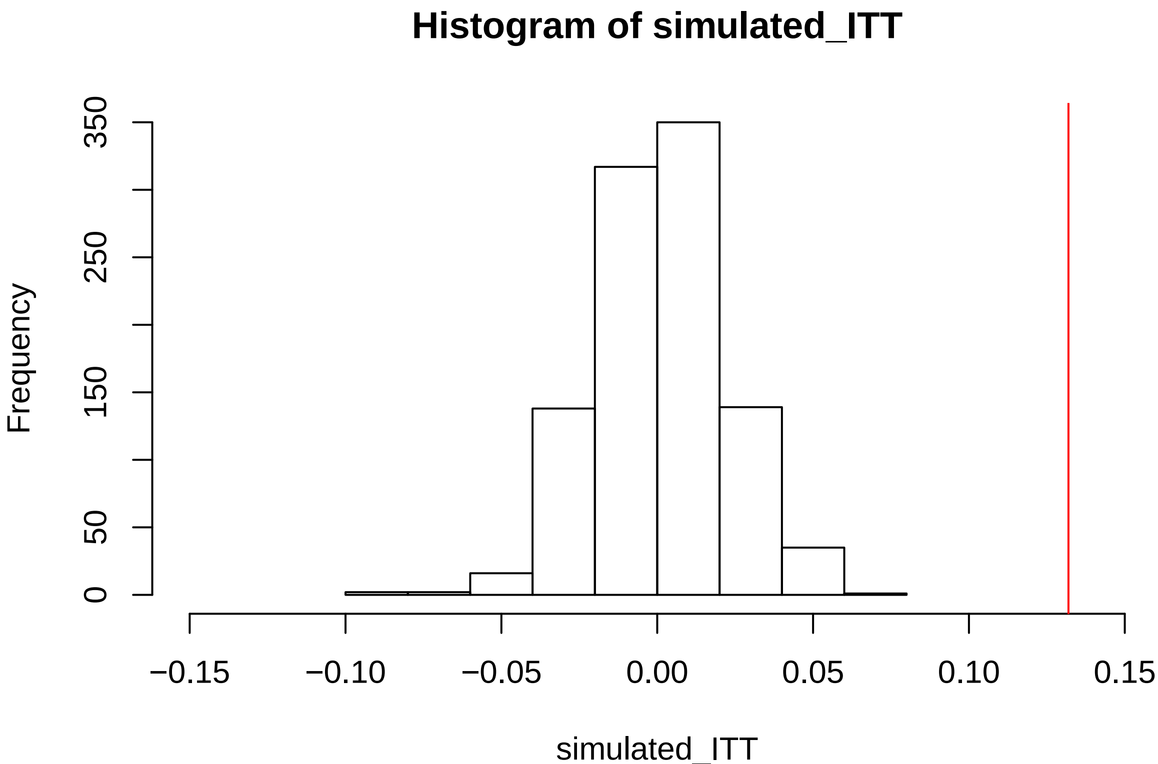
```
## [1] 0
##
## $greater.p.value
## [1] 0
##
## $lesser.p.value
## [1] 1
##
## $quantile
##         2.5%       97.5%
## -0.04218355  0.04210060
##
## $sd
## [1] 0.02159562
##
## $exp.val
## [1] -6.704101e-05
```

By hand

```r
library(randomizr)

ITTy = with(data5.10, mean(turnout[treat2 == 1]) - mean(turnout[treat2 == 0]))

simulated_ITT <- rep(NA, sims)

for (i in 1:sims){
  data5.10$Z_sim <- cluster_ra(data5.10$dormid)
  simulated_ITT[i] <- with(data5.10, mean(turnout[Z_sim == 1]) - mean(turnout[Z_sim == 0]))
}

p_two_tailed <- mean(abs(simulated_ITT) >= abs(ITTy))
p_upper <- mean(simulated_ITT >= ITTy)
p_lower <- mean(simulated_ITT <= ITTy)
c(observed_ate, p_two_tailed, p_upper, p_lower)
```

```
##        D
## 10.72431  0.00000  0.00000  1.00000
```

```r
hist(simulated_ITT, breaks = 10, xlim = c(-.15, .15))
abline(v = ITTy, col = "red")
```


Histogram of simulated_ITT
Frequency
0
50
150
250
350
−0.15
−0.10
−0.05
0.00
0.05
0.10
0.15
simulated_ITT