Gerber and Green (2012) Chapter 3 Problem 8

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 3 Problem 8 three different ways: using the `ri2` package, using the `ri` package, and by hand with a loop.

Chapter 3 Problem 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 Texas and Arkansas serve two-year or four-year terms in the aftermath of decennial redistricting. 25 of 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.

(a) For each state, estimate the effect of having a two-year term on the number of bills introduced.
(b) For each state, estimate the standard error of the estimated ATE.
(c) Use equation (3.10) to estimate the overall ATE for both states combined.
(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 estimates of the overall ATE.
(e) Insert the estimated standard errors into equation (3.12) to estimate the standard error for the overall ATE.

NOT SHOWN

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

SHOWN BELOW

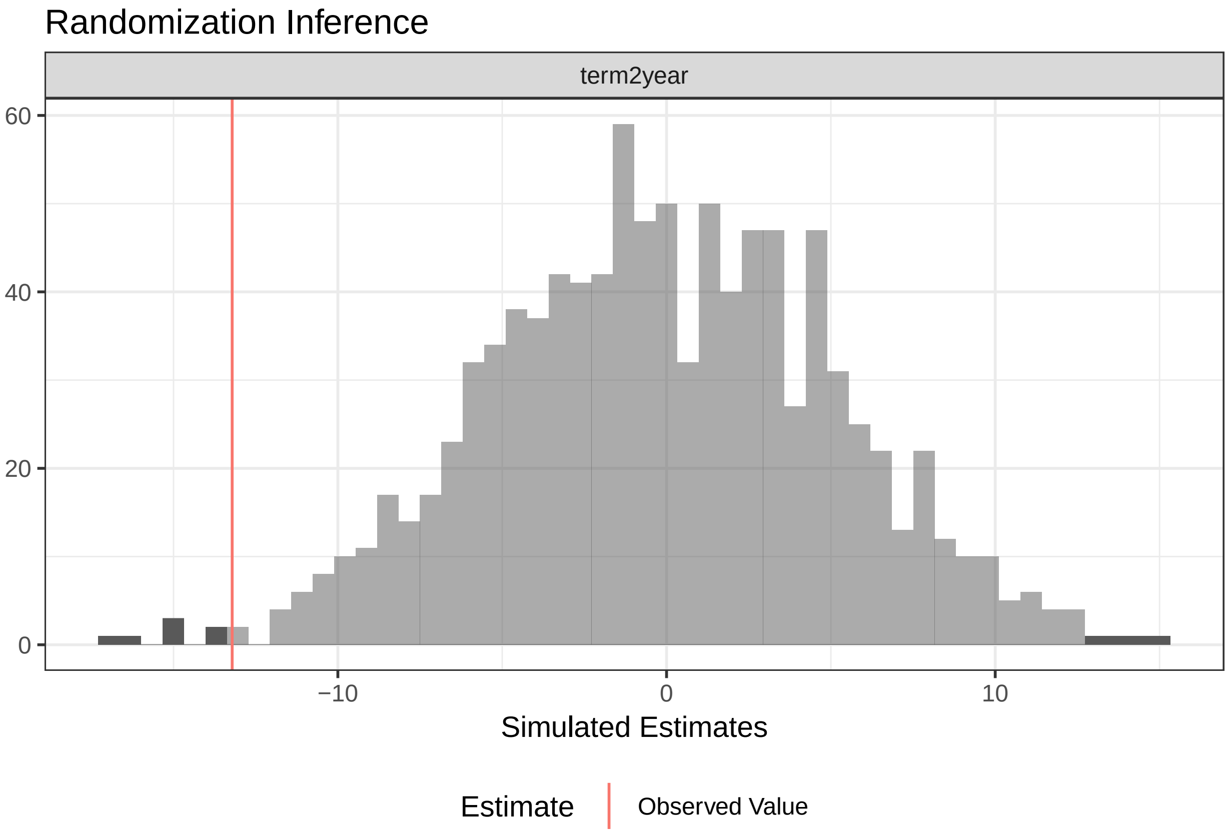
```r
set.seed(343)

# Data from http://isps.yale.edu/FEDAI
data3.8 <- read.csv("datasets/3.8.csv")
# Number of sims the same for all three methods
sims <- 1000
```

In ri2

```r
library(ri2)

# Declare randomization procedure
declaration <- declare_ra(N = 66, blocks = data3.8$texas0_arkansas1, block_m = c(15,18))

# Conduct Randomization Inference
ri2_out <- conduct_ri(bills_introduced ~ term2year,
                      declaration = declaration,
                      assignment = "term2year",
                      sharp_hypothesis = 0,
                      sims = sims,
                      data = data3.8)

summary(ri2_out)
```

```
##   coefficient estimate two_tailed_p_value null_ci_lower null_ci_upper
## 1   term2year -13.2168              0.011     -10.12361      10.01769
```

```
plot(ri2_out)
```



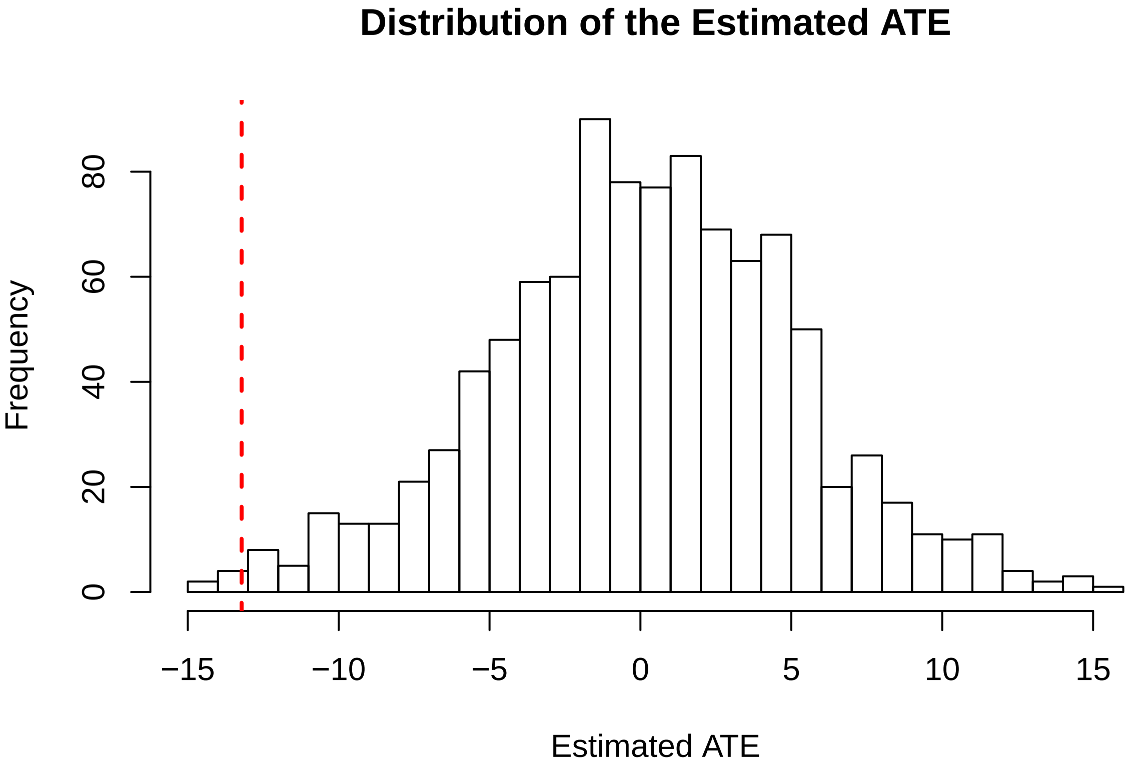
In ri

```
library(ri)

# all possible permutations
perms <- genperms(data3.8$term2year, blockvar = data3.8$texas0_arkansas1, maxiter = sims)
```

```
## Too many permutations to use exact method.
## Defaulting to approximate method.
## Increase maxiter to at least 1363721466356691712 to perform exact estimation.
```

```
# probability of treatment
probs <- genprobexact(data3.8$term2year, blockvar = data3.8$texas0_arkansas1)
# estimate the ATE
```

```
ate <- estate(data3.8$bills_introduced, data3.8$term2year, prob = probs)

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

# generate potential outcomes under sharp null of no effect
Ys <- genouts(data3.8$bills_introduced, data3.8$term2year, 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)
```



```
ri_out
```

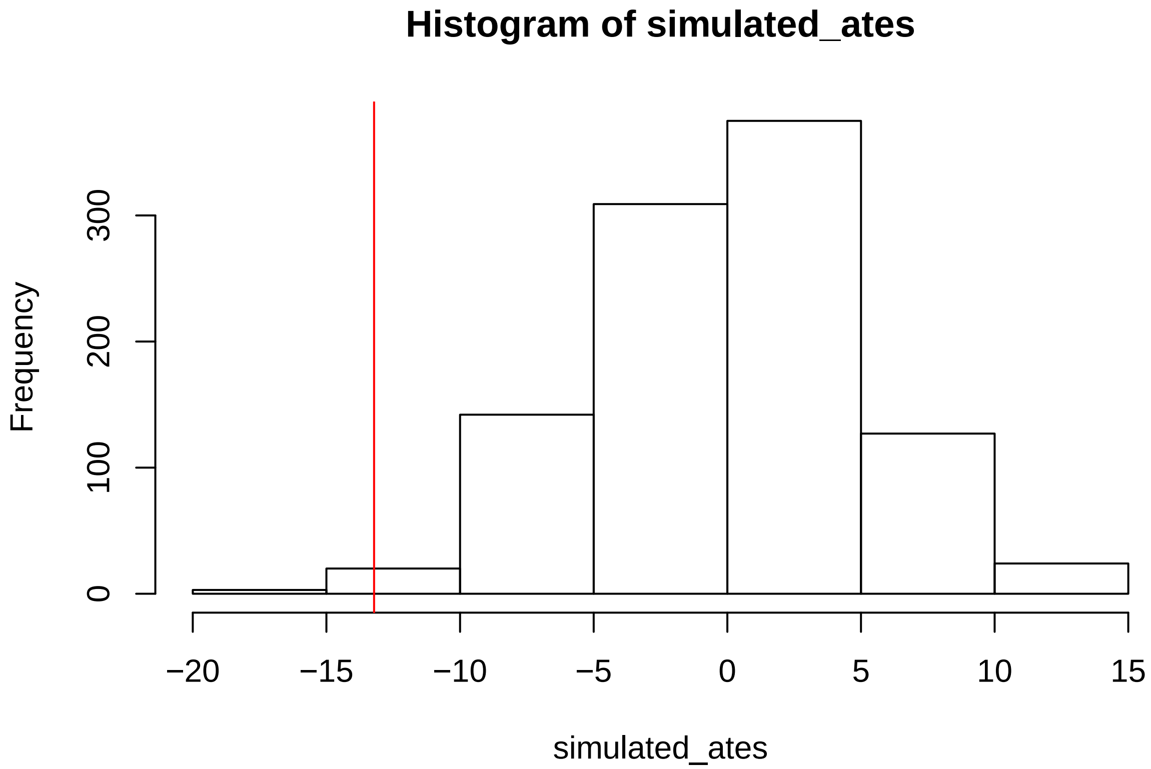
```
## $two.tailed.p.value
## [1] 0.004
##
## $two.tailed.p.value.abs
## [1] 0.008
##
## $greater.p.value
## [1] 0.998
##
## $lesser.p.value
## [1] 0.002
##
## $quantile
##      2.5%     97.5%
## -10.60992  10.33142
##
## $sd
## [1] 5.083438
##
## $exp.val
## [1] 0.1238158
```

By hand

```
library(randomizr)
library(dplyr)

CATEs <-
data3.8 %>%
  group_by(texas0_arkansas1) %>%
  summarize(CATE = mean(bills_introduced[term2year == 1]) - mean(bills_introduced[term2year == 0]),
            proportion = n()/nrow(data3.8))

observed_ate <-  CATEs %>% summarise(ATE = sum(CATE*proportion)) %>% .$ATE

simulated_ates <- rep(NA, sims)

for (i in 1:sims) {
  data3.8$Z_sim <- block_ra(blocks = data3.8$texas0_arkansas1, block_m = c(15, 18))

  CATEs <-
    data3.8 %>%
    group_by(texas0_arkansas1) %>%
    summarize(CATE = mean(bills_introduced[Z_sim == 1]) - mean(bills_introduced[Z_sim == 0]),
              proportion = n()/nrow(data3.8))

  simulated_ates[i] <-  CATEs %>% summarise(ATE = sum(CATE*proportion)) %>% .$ATE
}

p_two_tailed <- mean(abs(simulated_ates) >= abs(observed_ate))
p_upper <- mean(simulated_ates >= observed_ate)
p_lower <- mean(simulated_ates <= observed_ate)

c(observed_ate, p_two_tailed, p_upper, p_lower)
```

```
## [1] -13.2168   0.0110   0.9900   0.0100
```

```
hist(simulated_ates, breaks = 10)
abline(v = observed_ate, col = "red")
```


Histogram of simulated_ates
Frequency
0
100
200
300
−20
−15
−10
−5
0
5
10
15
simulated_ates