# Gerber and Green (2012) Chapter 4 Problem 4

*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 4 Problem 4 three different ways: using the `ri2` package, using the `ri` package, and by hand with a loop.

## Chapter 4 Problem 4

Table 4.1 contains a column of treatment assignments that reflects a complete random assignment of 20 schools to treatment and 20 schools to control.

(a) Use equation (2.2) to generate observed outcomes based on these assigned treatments. Regress Yi on di and interpret the slope and intercept. Is the estimated slope the same as the estimated ATE based on a difference-in-means?
(b) Regress treated and untreated outcomes on Xi to see whether the condition in equation (4.6) appears to hold. What do you infer about the advisability of rescaling the dependent variable so that the outcome is a change score (i.e., Yi - Xi)?

NOT SHOWN

(c) Regress Yi on di and Xi . Interpret the regression coefficients, contrasting these results with those obtained from a regression of Yi on di alone.

SHOWN BELOW

(d) With the estimates obtained in part (a), use randomization inference (as described in Chapter 3) to evaluate the sharp null hypothesis of no effect for any school. To obtain the sampling distribution under the sharp null hypothesis, simulate 100,000 random assignments, and for each simulated sample, estimate the ATE using a regression of Yi on di . Interpret the results.

NOT SHOWN

(e) With the estimates obtained in part (c), use randomization inference to evaluate the sharp null hypothesis of no effect for any school. To obtain the sampling distribution under the sharp null hypothesis, simulate 100,000 random assignments, and for each simulated sample, estimate the ATE using a regression of Yi on di and Xi . Interpret the results.

SHOWN BELOW

(f) Use the estimated ATE in part (a) to construct a full schedule of potential outcomes for all schools, assuming that every school has the same treatment effect. Using this simulated schedule of potential outcomes, construct a 95% confidence interval for the sample average treatment effect in the following way. First, randomly assign each subject to treatment or control, and estimate the ATE by a regression of Yi on di. Repeat this procedure until you have 100,000 estimates of the ATE. Order the estimates from smallest to largest. The 2,500th estimate marks the 2.5th percentile, and the 97,501st estimate marks the 97.5th percentile. Interpret the results. (g) Use the estimated ATE in part (c) to construct a full schedule of potential outcomes for all schools, assuming that every school has the same treatment effect. Using this simulated schedule of potential outcomes, simulate the 95% con dence interval for the sample average treatment effect estimated by a regression of Yi on di and Xi. Interpret the results. Is this confidence interval narrower than one you generated in response to question (f)?

NOT SHOWN

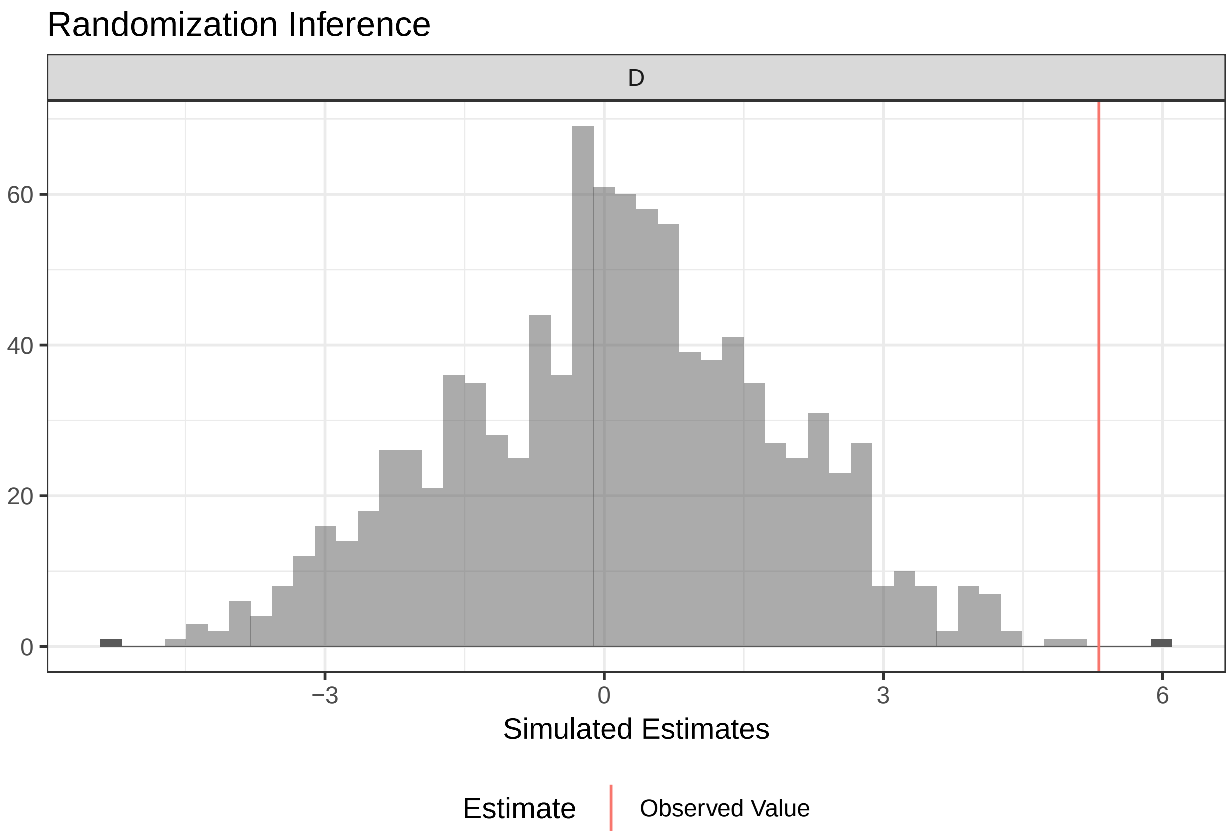
```r
# Data from http://isps.yale.edu/FEDAI
library(haven)
data4.4 <- read_dta("datasets/4.4.dta")
# Number of sims the same for all three methods
sims <- 1000
```

# In ri2

```r
library(ri2)
# Declare randomization procedure
declaration <- declare_ra(N = 40, m = 20)

# Conduct Randomization Inference
ri2_out <- conduct_ri(Y ~ D + x,
                      declaration = declaration,
                      assignment = "D",
                      sharp_hypothesis = 0,
                      sims = sims,
                      data = data4.4)

summary(ri2_out)
```

```
##   coefficient estimate two_tailed_p_value null_ci_lower null_ci_upper
## 1           D 5.315536              0.002     -3.335454      3.461804
```

```r
plot(ri2_out)
```

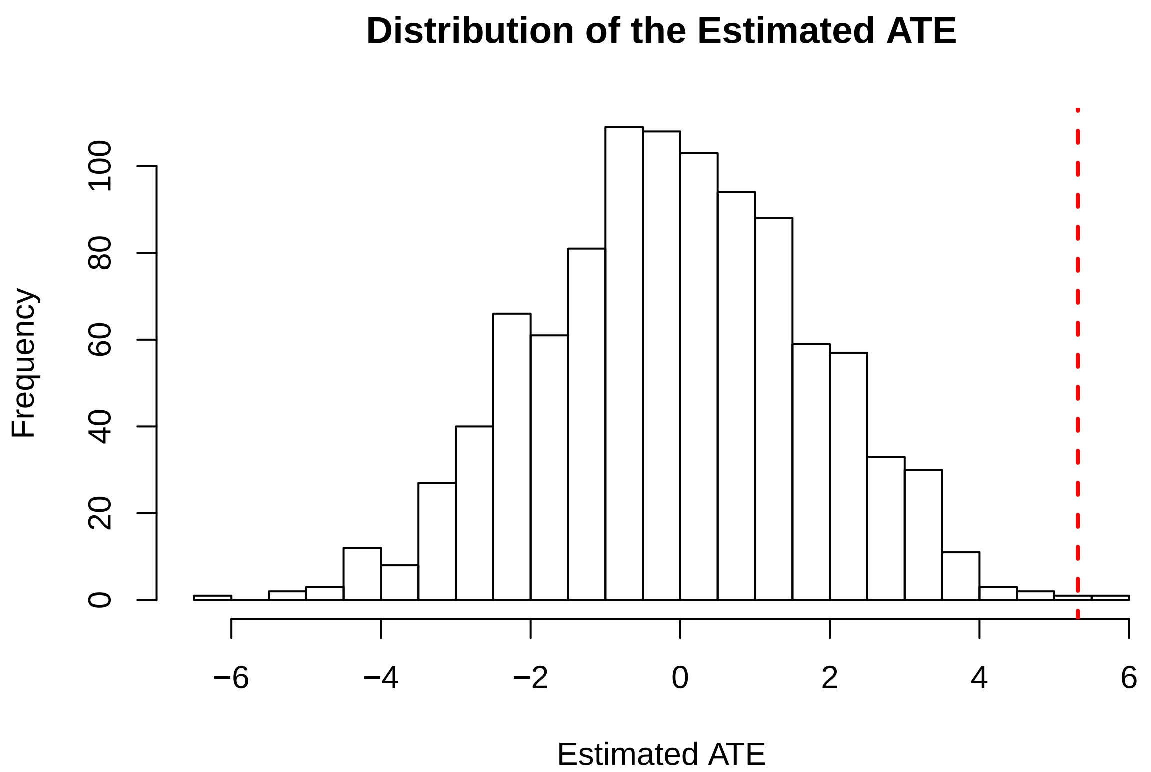



## In ri

```
library(ri)

# all possible permutations
perms <- genperms(data4.4$D, maxiter = sims)
```

```
## Too many permutations to use exact method.
## Defaulting to approximate method.
## Increase maxiter to at least 137846528820 to perform exact estimation.
```

```
# probability of treatment
probs <- genprobexact(data4.4$D)
# estimate the ATE
ate <- estate(data4.4$Y, data4.4$D, X = data4.4$x, prob = probs)

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

# generate potential outcomes under sharp null of no effect
Ys <- genouts(data4.4$Y, data4.4$D, ate = 0)
# generate sampling dist. under sharp null
distout <- gendist(Ys, perms, X = data4.4$x, 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] 0.003
##
## $greater.p.value
## [1] 0.002
##
## $lesser.p.value
## [1] 0.998
##
## $quantile
##      2.5%     97.5%
## -3.513218  3.388407
##
## $sd
## [1] 1.811943
##
## $exp.val
## [1] -0.06711396
```

## By hand

```
library(randomizr)
```

```r
fit <- lm(Y ~ D + x , data4.4)
observed_ate <- coef(fit)[2]
simulated_ates <- rep(NA, sims)

for (i in 1:sims) {
  data4.4$Z_sim <- block_ra(blocks = data4.4$x)
  fit_sim <- lm(Y ~ Z_sim + x, data4.4)
  simulated_ates[i] <- coef(fit_sim)[2]
}

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)
```

```
##        D
## 5.315536 0.007000 0.001000 0.999000
```

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