Gerber and Green Chapter 8 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 8 Problem 10 two different ways: using the `ri2` package and by hand with a loop. We can't use the `ri` package for this problem.

Chapter 8 Problem 10

A doctoral student conducted an experiment in which she randomly varied whether she ran or walked 40 minutes each morning. In the middle of each afternoon over a period of 26 days, she measured the following outcome variables: (1) her weight (minus a constant, for privacy's sake), (2) her score in a game of Tetris, (3) her mood on a 0–5 scale, with 5 being the most pleasant, (4) her energy level on a 0–5 scale, with 5 being the most energetic, and (5) whether she answered correctly a randomly selected problem from the math section of the GRE. Outcomes are missing for days 13 and 17.

(a) Suppose you were seeking to estimate the average effect of running on her Tetris score. Explain the assumptions needed to identify this causal effect based on this within-subjects design. Are these assumptions plausible in this instance? What special concerns arise due to the fact that the subject was conducting the study, undergoing the treatments, and measuring her own outcomes?

NOT SHOWN

(b) Estimate the effect of running on Tetris score. Use randomization inference to test the sharp null hypothesis that running has no immediate or lagged effect on Tetris scores.

SHOWN BELOW

(c) One way to lend credibility to within-subjects results is to verify the no-anticipation assumption. Use the variable Run to predict the Tetris score on the preceding day. Presumably, the true effect is zero. Does randomization inference confirm this prediction?

(d) If Tetris responds to exercise, one might suppose that energy levels and GRE scores would as well. Are these hypotheses borne out by the data?

NOT SHOWN

```
# Data from http://isps.yale.edu/FEDAI

library(haven)
data8.10 <- read_dta("datasets/8.10.dta")

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

In ri2

```
library(ri2)

# Declare randomization procedure
declaration <- declare_ra(N = 26, m = 14)
```

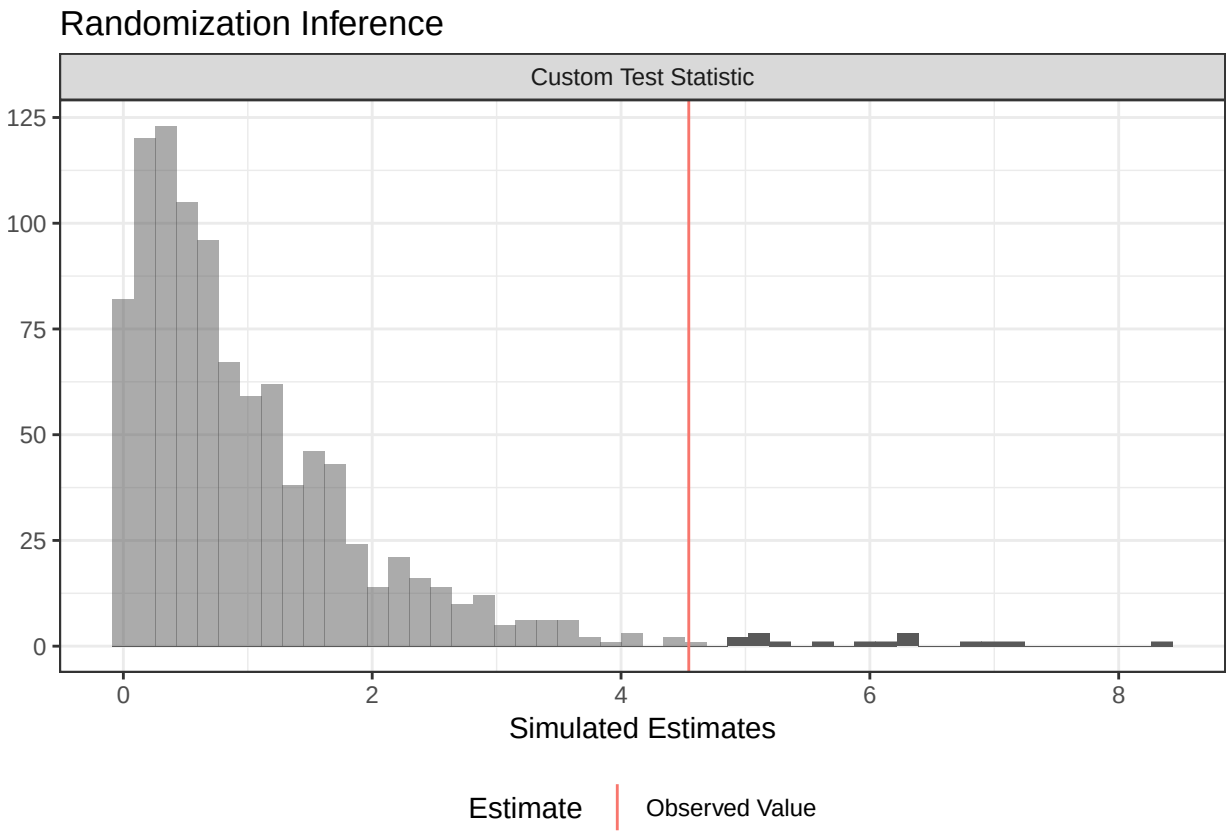
```r
# Create function
f_fun <- function(data) {
  data$run_lag <- with(data, lag(run))
  fit <- lm(tetris ~ run + run_lag, data = data)
  f_stat <- summary(fit)$fstatistic[1]
  names(f_stat) <- NULL
  return(f_stat)
}


ri2_out <- conduct_ri(test_function = f_fun,
                      declaration = declaration,
                      assignment = "run",
                      outcome = "tetris",
                      sharp_hypothesis = 0,
                      data = data8.10)

summary(ri2_out)
```

```
##             coefficient estimate two_tailed_p_value null_ci_lower
## 1 Custom Test Statistic 4.544592              0.016     0.0203098
##   null_ci_upper
## 1      3.624429
```

```r
plot(ri2_out)
```



By hand

```
data8.10$lag_run <- with(data8.10, lag(run))
fit <- lm(tetris ~ run + lag_run, data8.10)
f_stat <- summary(fit)$fstatistic[1]

simulated_f <- rep(NA, sims)

for (i in 1:sims){
  data8.10$Z_sim <- complete_ra(N = 26, m = 14)
  data8.10$Z_sim_lag <- with(data8.10, lag(Z_sim))
  fit <- lm(tetris ~ Z_sim + Z_sim_lag, data8.10)
  simulated_f[i] <- summary(fit)$fstatistic[1]
}

p_two_tailed <- mean(abs(simulated_f) >= abs(f_stat))
p_upper <- mean(simulated_f >= f_stat)
p_lower <- mean(simulated_f <= f_stat)

c(p_two_tailed, p_upper, p_lower)
```

```
## [1] 0.009 0.009 0.991
```

```
hist(simulated_f, breaks = 10)
abline(v = f_stat, col = "red")
```