Permutation Test

Problem Set 2

Gov2001

Due Wednesday February 7, 12:00 pm

Class size Experiment

We will analyze the data from the STAR project. The data file is available as `STAR_Student_dist.csv`. Specifically, we are interested in the effect of small class size on the math score. Assume throughout that treatment assignment is by complete randomization and hence, unconfounded and for the sake of the exercise, ignore any rows with missing data for the outcome `g8tmathss`.

variable name	variable description
`gkclasstype`	kindergarten class type
`gender`	student's gender
`race`	student's race
`gkfreelunch`	student's free lunch status at the kindergarten
`gksurban`	the urbanicity of kindergarten
`g8tmathss`	the 8th grade math score
`g8treadss`	the 8th grade reading score
`hsgrdcol`	highschool graduation status

Question 1

a. Use the Wilcoxon rank-sum test and produce the p-value as on Slide 12 only now for the math score outcome instead of the reading score outcome. Use the `wilcox.test()` function. Briefly interpret the results.

```
library(tidyverse)
library(knitr)

star <- read.csv("STAR_Students_dist.csv")

star %>%
  filter(!is.na(g8tmathss)) %>%
  group_by(gkclasstype) %>%
  summarise(median = median(g8tmathss),
            mean = mean(g8tmathss),
            sd = sd(g8tmathss),
            n = n()) %>%
  kable()
```

gkclasstype	median	mean	sd	n
REGULAR CLASS	789.0	790.7185	45.05366	1183
SMALL CLASS	794.5	793.2815	45.94114	1016

```
wilcox.test(g8tmathss ~ gkclasstype, data = star)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  g8tmathss by gkclasstype
## W = 574462, p-value = 0.0742
## alternative hypothesis: true location shift is not equal to 0
```

The results of the Wilcoxon rank-sum test indicate that the median math score is not significantly higher at the 95% level in the small class type compared to the regular class type ($p > 0.05$).

b. Consider the sharp null hypothesis of no treatment effect. Choose a test statistic based on your favorite regression model that incorporates the pre-treatment covariates. Conduct a permutation test using Monte Carlo approximation. Briefly interpret the results.

```r
set.seed(2001)

# Regression model
original_model <- lm(g8tmathss ~ gkclasstype + gender +
                                 race + gkfreelunch + gksurban,
                     data = star)

stargazer::stargazer(
  original_model, type = "text", title = "Regression Results",
  dep.var.labels = "8th Grade Math Score",
  covariate.labels = "Small Class Type",
  omit = c('gender','race', 'gkfreelunch', 'gksurban'),
  omit.labels = c('Gender Fixed Effects', 'Race', 'Free Lunch Status', 'Urbanicity'),
  report=('vc*p'), align = TRUE, digits = 3, font.size = "small")
```

```
##
## Regression Results
## ================================================
##                          Dependent variable:
##                      ---------------------------
##                         8th Grade Math Score
## ------------------------------------------------
## Small Class Type                2.129
##                               p = 0.255
##
## Constant                     806.461***
##                               p = 0.000
##
## ------------------------------------------------
## Gender Fixed Effects            Yes
## Race                            Yes
## Free Lunch Status               Yes
## Urbanicity                      Yes
## ------------------------------------------------
## Observations                   2,194
## R2                             0.092
## Adjusted R2                    0.087
## Residual Std. Error      43.477 (df = 2182)
## F Statistic           20.114*** (df = 11; 2182)
## ================================================
## Note:                *p<0.1; **p<0.05; ***p<0.01
```

```r
estimate <- original_model$coefficients[2]

# Monte Carlo Permutation Test
n_permutations <- 1000
permuted_estimate <- numeric(n_permutations)

for (i in 1:n_permutations) {
```

```r
  shuffled_classtype <- sample(star$gkclasstype)

  permuted_model <- lm(g8tmathss ~ shuffled_classtype + gender +
                                  race + gkfreelunch + gksurban,
                      data = star)

  permuted_estimate[i] <- permuted_model$coefficients[[2]]

}

p_value <- mean(abs(permuted_estimate) >= abs(estimate))

print(p_value)
```

```
## [1] 0.256
```

The permutation test, consistent with the original regression analysis, yielded a p-value of 0.256, indicating no statistically significant effect of class size on 8th grade math scores. This result suggests that any observed differences in math scores by class size are likely due to chance, supporting the null hypothesis of no treatment effect.

c. Using the Monte Carlo approximation, compute the Hodges-Lehmann estimate and 95% confidence interval by inverting the Wilcoxon rank-sum test.

```r
hl_estimate <- wilcox.test(g8tmathss ~ gkclasstype,
                           data = star, conf.int = T)

hl_estimate
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  g8tmathss by gkclasstype
## W = 574462, p-value = 0.0742
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
##  -6.999993e+00  1.880217e-05
## sample estimates:
## difference in location
##              -3.000027
```

```r
print(paste("Hodges-Lehmann Estimate:", hl_estimate$estimate))
```

```
## [1] "Hodges-Lehmann Estimate: -3.0000267087378"
```

```r
hl_estimates <- numeric(100)

for (i in 1:100) {

  star$shuffled_scores <- sample(star$g8tmathss)

  # Calculate HL estimate for shuffled data
  test_shuffled <- wilcox.test(shuffled_scores ~ gkclasstype,
                               data = star,
                               exact = F, conf.int = T)

  hl_estimates[i] <- test_shuffled$estimate
}

# Compute 95% CI from the permutation distribution
ci_lower <- quantile(hl_estimates, probs = 0.025)
ci_upper <- quantile(hl_estimates, probs = 0.975)
print(paste("95% CI: [", ci_lower, ", ", ci_upper, "]"))
```

```
## [1] "95% CI: [ -3.00000382371393 ,  3.52501575339365 ]"
```

Social Network Experiment

We will analyze the data from a social network experiment conducted by Paluk *et al.* (2016) *PNAS*. The authors conducted a randomized experiment by leveraging the existing friendship network in schools. The experiment involved a total of 56 schools. Out of these, 28 schools are randomly selected for the anticonflict intervention to "encourage anticonflict norms and behavior." Our analysis will focus on these treatment schools. In each of these selected schools, some students are non-randomly selected as "seeds". One half of these seeds (26 students per school on average) are randomly assigned to a year-long anticonflict program. Finally, various outcomes are measured. In this exercise, we focus on a particular outcome, representing whether or not student wore anti-conflict wristband. We are interested in the spillover effects among friends.

The data set `PNASsocial.csv` contains the following basic variables. We note that within a given school, the researchers used stratified randomization, in which the complete randomization of treatment assignment was used across eligible students within each block. Each school has four blocks of eligible students indicated by the `student_block` variable where `0` represents non-eligible students. The data set also contains the information about friendship network. Each student lists up to 10 friends within their school.

variable name	description
`studentID`	student ID
`eligible`	indicator for being eligible to receive treatment (i.e., seeds)
`school_id`	school ID
`student_block`	blocking variable used for treatment randomization
`treatment`	indicator for treatment assignment
`woreAnticonflictWristband`	outcome variable: indicator of whether a student wore anti-conflict wristband
`friendX`	student ID of 10 friends: `X` $= 1, \ldots, 10$

In addition, the data set contains several pre-treatment covariates such as gender, race, grade, and whether a student has older siblings.

Question 2

We would like to examine the existence of spillover effects in this experiment. In particular, we conjecture that non-eligible students may be affected by the treatment assignment of eligible students who are their friends. Note that some friends are not part of the data and are indicated by `NA`.

We consider the following null hypothesis

$$H_0 : Y_{ij} \perp\!\!\!\perp \{T_{i'j} : i' \notin \mathcal{F}_j, E_{i'} = 1\} \mid \{T_{i'j} : i' \in \mathcal{F}_j\} \text{ for any student } i \text{ with } E_i = 0 \text{ at school } j.$$

where $\mathcal{F}_j$ represents the set of eligible students who have no non-eligible friends in school j and E_i represents the eligibility indicator. By this conditioning, we are focusing on the treatment assignments of eligible students who can affect non-eligible students through their friendship.

Propose a test statistic that are based on the friendship data. Conduct the conditional randomization test and briefly interpret the results. There are many ways to do this, but you should justify your proposal.

```r
social <- read.csv("PNASsocial.csv")

students_to_drop <- social[is.na(social$woreAnticonflictWristband),]$studentID
social <- social[!is.na(social$woreAnticonflictWristband),]
```

```r
social[,7:16] <- lapply(social[,7:16], function(x) ifelse(x %in% students_to_drop, NA, x))
social[,7:16] <- lapply(social[,7:16], function(x) ifelse(x %in% social$studentID, x, NA))


social$friends <- rowSums(!is.na(social[,7:16]))

social$friend_treated <- apply(
  social[,7:16], 1,
  function(x) sum(social$treatment[match(x, social$studentID)], na.rm = T))

social$friend_treated_frac <- ifelse(social$friends > 0, social$friend_treated / social$friends, NA)
```

```r
library(lfe)

model <- felm(woreAnticonflictWristband ~ friend_treated_frac |
               school_id + student_block | 0 | school_id, data = subset(social, eligible==0))

stargazer::stargazer(model, type = "text")
```

```
##
## ===============================================
##                         Dependent variable:
##                     ---------------------------
##                      woreAnticonflictWristband
## -----------------------------------------------
## friend_treated_frac          0.204***
##                               (0.047)
##
## -----------------------------------------------
## Observations                   8,239
## R2                             0.058
## Adjusted R2                    0.055
## Residual Std. Error      0.486 (df = 8210)
## ===============================================
## Note:               *p<0.1; **p<0.05; ***p<0.01
```