Data Assignment 4

Jongrak Jeong

2022-12-01

Default Setting and check the data

Load the packages

```
library(foreign)
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──
## ✓ ggplot2 3.3.6     ✓ purrr   0.3.4
## ✓ tibble  3.1.7     ✓ dplyr   1.0.9
## ✓ tidyr   1.2.0     ✓ stringr 1.4.0
## ✓ readr   2.1.2     ✓ forcats 0.5.1
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
```

```
library(rempsyc)
```

```
## Suggested APA citation: Thériault, R. (2022). rempsyc: Convenience functions for p
sychology
## (R package version 0.1.0) [Computer software]. https://rempsyc.remi-theriault.com
```

Load and check the data set

```
data <- read.dta("Hyde_2007_Armenia.dta")

names(data)
```

```
##  [1] "pollingstation" "subregion"      "mon_voting_r1"  "mon_voting_r2"
##  [5] "avagian"        "geghamian"      "demirchian"     "karapetian"
##  [9] "harutyunian"    "manukian"       "margarian"      "sargsian"
## [13] "kocharian"      "demirchian_r2"  "kocharian_r2"   "turnout"
## [17] "nearNagorno"
```

Q1.

(a)

The difference in the vote share: about '6%'. (decrease) And this result is statistically significant in that t value of this result is big enough.

(b)

The difference in the vote share: about '2%'. (decrease) And this result has statistical significance in that t value of this result is bigger than the critical value.

(c)

In both of round 1 and round 2, the averages of the vote share decrease in monitored group. And these results have statistical significance. According to these results, we can expect the presence of international observers reduces the effect of fraud on the election.

If the fraud affect on the result of election, it can be expected that the incumbent party will have more vote share. The presence of observers affect on the result and the vote share in monitored group decreased. In other words, the observers reduced the effect of fraud. These observers are independent from the incumbent party in that they are international observers. Therefore, the influences of fraud by the incumbent party can be decreased.

They were randomly assigned to the polling station without the incumbent party knowing it. The incumbent cannot expect it and take preemptive action, hence, the effect of fraud can be decreased. Also, it can control potential confounding variables so that we can focus on and figure out the relationship between the presence of international observation and the effect of fraud.

In round 2, the difference in the vote share get closer than before. Still, the average vote share of monitored group is lower than the other. This may be because the regions where the Kocharian has strong support are less represented in round 1 than round 2. Therefore, we need to examine which regions are selected or do more randomly assigning experiment in order to obtain larger number of observed results.

t-Test | Row 1: Round 1

```
t.test(kocharian ~ mon_voting_r1, data = data, var.equal = TRUE)

##
##  Two Sample t-test
##
## data:  kocharian by mon_voting_r1
## t = 5.918, df = 1762, p-value = 3.908e-09
## alternative hypothesis: true difference in means between group 0 and group 1 is no
t equal to 0
## 95 percent confidence interval:
##  0.03922980 0.07812222
## sample estimates:
## mean in group 0 mean in group 1
##       0.5419031       0.4832271

nice_t_test(data = data,
            response = "kocharian",
            group = "mon_voting_r1",
            var.equal = TRUE) |>
  nice_table()

## [97mUsing Student t-test.
##  [97m
```

Dependent Variable	*t*	*df*	*p*	*d*	95% CI
kocharian	5.92	1,762	< .001	0.29	[0.19, 0.38]

t-Test | Row 2: Round 2

```
t.test(kocharian_r2 ~ mon_voting_r2, data = data, var.equal = TRUE)

##
##  Two Sample t-test
##
## data:  kocharian_r2 by mon_voting_r2
## t = 2.4669, df = 1761, p-value = 0.01373
## alternative hypothesis: true difference in means between group 0 and group 1 is no
t equal to 0
## 95 percent confidence interval:
##  0.00406605 0.03561577
## sample estimates:
## mean in group 0 mean in group 1
##       0.6925111       0.6726701

nice_t_test(data = data,
            response = "kocharian_r2",
            group = "mon_voting_r2",
            var.equal = TRUE) |>
  nice_table()

## [97mUsing Student t-test.
##  [97m

## Warning: Missing values detected. NAs dropped.
```

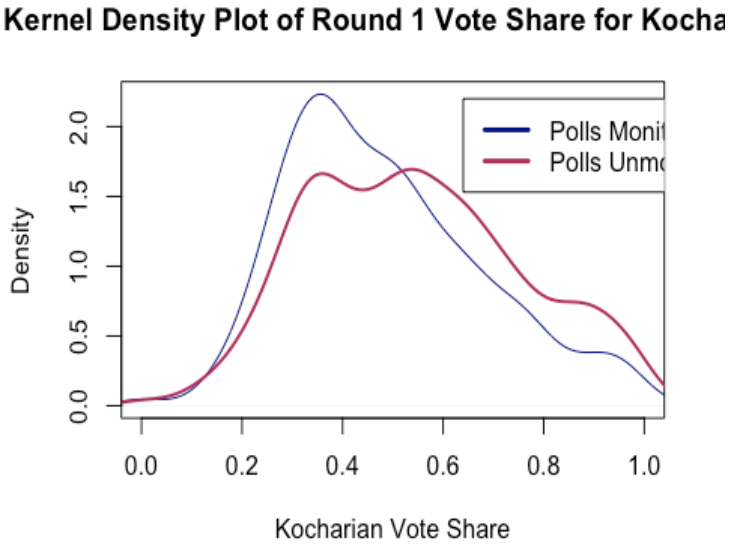
Dependent Variable	*t*	*df*	*p*	*d*	95% CI
kocharian_r2	2.47	1,761	.014	0.12	[0.03, 0.22]

Q2. Visualize the difference in Kocharian's round 1 vote share between monitored and unmonitored polling stations.

(a)

plot the density graph - Figure 1

```
plot(density(data$kocharian[data$mon_voting_r1 == 1]),
     xlim = c(0,1), col="navy",
     xlab = "Kocharian Vote Share",
     main = "Kernel Density Plot of Round 1 Vote Share for Kocharian")
lines(density(data$kocharian[data$mon_voting_r1 == 0]),
      col = "maroon", lwd = 2)
legend(.64, 2.2,legend = c("Polls Monitored","Polls Unmonitored"),
       col = c("navy","maroon"), lty = c(1,1), lwd = c(3,3))
```



(b)

Based on the plot, we can expect international observers reduce election-day fraud. The average vote share of Kocharian decreased among monitored group. Also, considering overall distribution, we can see that the density of about 40% share increased and the variance decreased in monitored polling stations. In the case of unmonitored stations, the density of 20-50% share is lower and the one of 50-100% share is higher than monitored group. This fact can be enough to suspect fraud. On the contrary, international monitoring seems to have affect the election result, which means it reduce election-day fraud.

Check the values describing the plot

```
median(data$kocharian[data$mon_voting_r1 == 1])

## [1] 0.4523

median(data$kocharian[data$mon_voting_r1 == 0])

## [1] 0.52715

mean(data$kocharian[data$mon_voting_r1 == 1])

## [1] 0.4832271

mean(data$kocharian[data$mon_voting_r1 == 0])

## [1] 0.5419031

var(data$kocharian[data$mon_voting_r1 == 1])

## [1] 0.0384595

var(data$kocharian[data$mon_voting_r1 == 0])

## [1] 0.04521667
```

Q3.

(a)

They were not less likely to visit the regions near Nagorno-Karabakh in round 1. They monitored in similar numbers of stations near this region in both Round 1 and Round 2, with 35 and 36, respectively. And the proportion of monitored stations near this region in each round is about 1.98% and 2.04%, respectively.

Subset and mutate the dataset

```
near <- select(data, mon_voting_r1, mon_voting_r2, nearNagorno, kocharian) %>%
  mutate(near_mon1 = ifelse(nearNagorno == 1 & mon_voting_r1 == 1, 1, 0)) %>%
  mutate(near_mon2 = ifelse(nearNagorno == 1 & mon_voting_r2 == 1, 1, 0))
```

The observation numbers and the proportion of monitored stations near Nagorno-Karabakh in each round

```
nrow(near[near$near_mon1 == 1,])

## [1] 35

100 * mean(near$near_mon1)

## [1] 1.984127

nrow(near[near$near_mon2 == 1,])

## [1] 36

100 * mean(near$near_mon2)

## [1] 2.040816
```

(b)

Our findings are not driven by the result near Nogorno-Karabakh. There is no regional difference in the number of observation(near Nagorno-Karabakh). Also, if we divide the stations as two group - 'Near Nagorno-Karabakh' and 'Not near Nagorno-Karabakh', the t-test results say there is no statistical significance in 'Near Nagorno-Karabakh' group. Therefore, our findings in round 1 are not driven by the result near Nogorno-Karabakh and it is consistent to the result in round 2.

Difference of means tests Kocharian vote share in round 1 (not near Nagorno-Karabakh)

```
near0 <- select(near, mon_voting_r1, nearNagorno, kocharian) %>%
  filter(nearNagorno == 0)

t.test(kocharian ~ mon_voting_r1, data = near0, var.equal = TRUE)

##
##  Two Sample t-test
##
```

```
## data:  kocharian by mon_voting_r1
## t = 6.1389, df = 1668, p-value = 1.036e-09
## alternative hypothesis: true difference in means between group 0 and group 1 is no
t equal to 0
## 95 percent confidence interval:
##  0.04158133 0.08062748
## sample estimates:
## mean in group 0 mean in group 1
##       0.5315228       0.4704184

nice_t_test(data = near0,
            response = "kocharian",
            group = "mon_voting_r1",
            var.equal = TRUE) |>
  nice_table()

## [97mUsing Student t-test.
##  [97m
```

Dependent Variable	*t*	*df*	*p*	*d*	95% CI
kocharian	6.14	1,668	< .001	0.30	[0.21, 0.40]

Difference of means tests Kocharian vote share in round 1 (near Nagorno-Karabakh)

```
near1 <- select(near, mon_voting_r1, nearNagorno, kocharian) %>%
  filter(nearNagorno == 1)

t.test(kocharian ~ mon_voting_r1, data = near1, var.equal = TRUE)

##
##  Two Sample t-test
##
## data:  kocharian by mon_voting_r1
## t = -0.90621, df = 92, p-value = 0.3672
## alternative hypothesis: true difference in means between group 0 and group 1 is no
t equal to 0
## 95 percent confidence interval:
##  -0.10814506  0.04037761
## sample estimates:
## mean in group 0 mean in group 1
##       0.7102763       0.7441600

nice_t_test(data = near1,
            response = "kocharian",
            group = "mon_voting_r1",
            var.equal = TRUE) |>
  nice_table()

## [97mUsing Student t-test.
##  [97m
```

Dependent Variable	*t*	*df*	*p*	*d*	95% CI
kocharian	-0.91	92	.367	-0.19	[-0.61, 0.23]

Q4.

This research design is good example of natural experiment in political science. In the process of deriving causality, how to prevent the effect of confounding variables is important. In order to overcome this problem, randomly assigning the monitored polling station, the researcher conducted a natural experiment. This can lower or control the effect of confounders. Furthermore, randomly assigning stations to be monitored is efficient in that the research target it fraud problem. Lastly, the researcher anticipated the expected critical points in advance and examine them through t-test - regional differences - in order to increase the consistency of the case in Round 1. This point is also good example.