Impact Evaluation: Assignment 4

Mukhammed Zainidinov

Impact Evaluation and Policy Inherence

Violence and insurgency

```
chechen = read.csv('data/chechen.csv')
```

1. By running the following code, we can see the number of observations for each value of the variable *fire*. There are 318 observations, 159 of which were shelled by Russians, the rest of them were not.

```
table(chechen$fire)
```

2. First, let's compare means of lethality of attacks on Groznyy and villages outside of Groznyy.

```
library(tidyverse)

chechen %>%

  group_by(groznyy) %>%

  summarize(mean(deaths, na.rm = TRUE))
```

The average number of killed individuals in Groznyy is 3.71 while outside of Groznyy it is 1.57.

Next, we compare lethality based on medians.

```
chechen %>%

  group_by(groznyy) %>%

  summarize(median(deaths, na.rm = TRUE))
```

The median for Groznyy is 3 while for other villages is 0. This means that the distribution of death in both Groznyy and villages outside of Griznyy is skewed to the left.

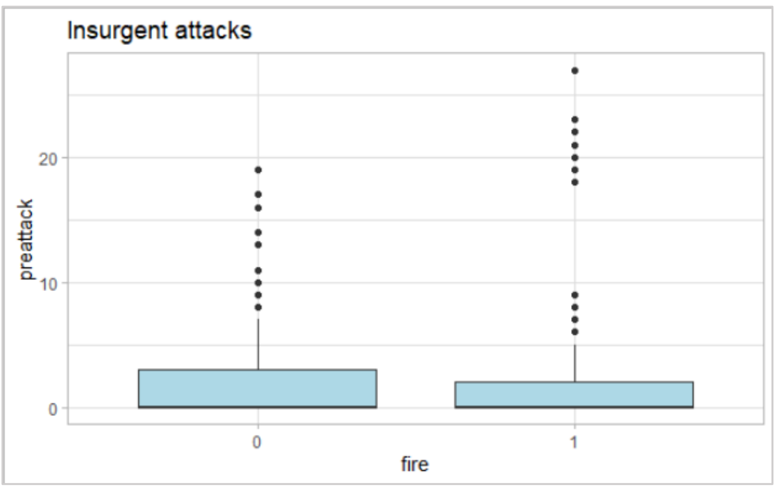
3. Number of insurgent attacks in pre- and post-shelling period:

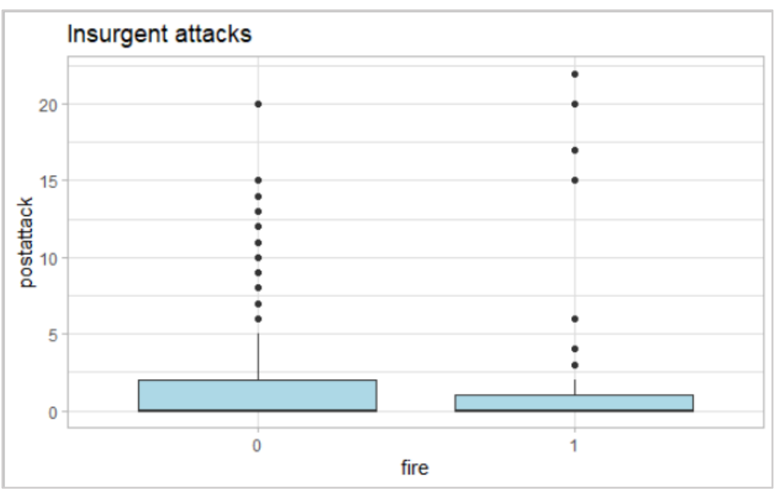
```
chechen %>%

  group_by(fire)%>%

  summarise(mean(preattack, na.rm = TRUE))

chechen %>%

  group_by(fire)%>%

  summarise(mean(postattack, na.rm = TRUE))
```

The average number of insurgent attacks in shelled villages decreased from 2.11 in the pre-shelling period to 1.5 in the post-shelling period while this indicator in not shelled villages decreased from 2.15 to only 2.05.

Let’s compare the quartiles:

```
chechen$fire = as.factor(chechen$fire)
chechen %>%
  ggplot(aes(x = fire, y = preattack)) + theme_light()+
  geom_boxplot(fill = 'lightblue') +
  ggtitle('Insurgent attacks')
chechen %>%
  ggplot(aes(x = fire, y = postattack)) + theme_light()+
  geom_boxplot(fill = 'lightblue') +
  ggtitle('Insurgent attacks')
```





The boxplots show us the distribution of insurgent attacks in villages before and after artillery fire. In not shelled villages, the interquartile range decreased but the number of outliers increased. But in the shelled villages both interquartile range and number of outliers decreased. If we apply the difference-in-differences model to this situation and parallel trend assumption holds, we can conclude that indiscriminate violence reduces insurgent attacks. If PTA doesn't hold, there might be other factors causing to decrease insurgent attacks in the shelled villages.

4. Pre-shelling period

```
shelled = chechen %>%
  subset(fire == '1') %>%
  summarise(mean(preattack, na.rm = TRUE))
not_shelled = chechen %>%
  subset(fire == '0') %>%
  summarise(mean(preattack, na.rm = TRUE))
pre_diff = shelled - not_shelled
pre_diff
```

In the pre-shelling period, the average number of insurgent attacks in the shelled villages is less by 0.038 than in not shelled villages. We can compare two groups even if they have different means and get valid results if the parallel trend assumption holds.

5. Difference in the number of insurgent attack before and the after the shelling period:

```
chechen$diffattack = chechen$postattack - chechen$preattack
chechen %>%
  subset(fire == '1') %>%
  summarise(mean(diffattack, na.rm = TRUE))
```

In the shelled villages, the average number of insurgent attacks decreased by 0.616 after shelling.

6. Difference-in-differences:

```
chechen %>%
  group_by(fire) %>%
  summarise(mean(diffattack, na.rm = TRUE)
```

Mean difference = -0.616 – (-0.101) = -0.515. This tells us that the average number of insurgent attacks in shelled villages decreased by 0.515 comparing to not shelled villages. These results support the claim that indiscriminate violence reduces insurgency attacks because we computed difference-in-differences. The validity of this analysis doesn't improve over the analyses conducted in the previous questions because we could estimate difference-in-differences in question 3 by calculating:

diff-in-diff = (average postattack in shelled – average postattack in not shelled) – (average preattack in shelled – average preattack in not shelled) = (1.50 – 2.05) – (2.11 – 2.15) = -0.51

This analysis additionally estimates difference-in-differences comparing to the analyses conducted in the previous questions.

Exploiting variation in policy changes

7. If we use cross-sectional data, we don't know about driving fatalities before the law was implemented. There might be a difference between the states where the law was implemented and not implemented before treatment. Consequently, this leads us to inaccurate estimation and low validity of the results.

8. If there is data for the period before and after the law was implemented, I would use the difference-in-differences method to estimate the causal inference. This method would give us causal inference comparing to other methods if there is a parallel trend in driving fatalities between treated and non-treated groups before the treatment period.

```
library(wooldridge)

data("traffic1")
```

9. Effect of the open container law on the number of traffic deaths per 100 million miles driven in 1985:

```
m1 = lm(dthrte85 ~ open85, data = traffic1)

summary(m1)
```

The coefficient of the independent variable is not statistically significant. Even if it was statistically significant, as I mentioned in question 7, it wouldn't give us a causal. There might be other factors affecting traffic death such as speed, condition of roads, condition of a car, fastened seat belts, etc.

10. Effect of the open container law on the number of traffic deaths per 100 million miles driven in 1990:

```
m2 = lm(dthrte90 ~ open90, data = traffic1)

summary(m2)
```

Again the coefficient is not statistically significant. As mentioned above it doesn't give a causal effect.

11. Difference-in-differences estimate based on previous regressions in questions 9-10:

```
library(texreg)

screenreg(list(m1,m2))
```

```
==================================
               Model 1    Model 2
----------------------------------
(Intercept)     2.78 ***   2.12 ***
               (0.11)     (0.10)
open85         -0.21
               (0.17)
open90                     0.07
                          (0.15)
----------------------------------
R^2             0.03       0.00
Adj. R^2        0.01      -0.02
Num. obs.      51         51
==================================
*** p < 0.001; ** p < 0.01; * p < 0.05
```

We can estimate difference-in-differences based on regression coefficients of the above models. To get the treatment effect, we subtract the coefficient of *open85* from the coefficient of *open90:* 0.07 – (-0.21) = 0.28. Since we have data before and after the treatment, this is a better estimate of the causal effect. The treatment effect is 0.28 but not statistically significant meaning open container laws don't change the traffic death. That's why we cannot rely on the results.

12. The difference-in-differences estimate by including the interaction term.

First, we need to reshape the data from wide to long format:

```
library(reshape2)

traffic = traffic1
traffic$cdthrte = NULL
traffic$cadmn = NULL
traffic$copen = NULL
traffic$cspeed = NULL

traffic = traffic %>%
  rename(admn_1990 = admn90, admn_1985 = admn85,
         open_1990 = open90, open_1985 = open85,
         dthrte_1990 = dthrte90, dthrte_1985 = dthrte85,
         speed_1990 = speed90, speed_1985 = speed85)
```

```
long = reshape(traffic, varying = c('admn_1990', 'admn_1985',
                                    'open_1990', 'open_1985',
                                    'dthrte_1990', 'dthrte_1985',
                                    'speed_1990', 'speed_1985'),
               direction = 'long', idvar = 'state', sep = '_')
```

Now we run regression including the interaction term:

```
long$year = ifelse(long$time == '1990', 1, 0)
m3 = lm(dthrte ~ open + year + open * year, data = long)
screenreg(list(m3))
```

```
=======================
             Model 1
-----------------------
(Intercept)    2.78 ***
              (0.10)
open          -0.21
              (0.16)
year          -0.65 ***
              (0.14)
open:year      0.28
              (0.23)
-----------------------
R^2            0.21
Adj. R^2       0.18
Num. obs.    102
=======================
*** p < 0.001; ** p < 0.01; * p
```

The treatment effect is 0.28 that is the same as what we got in question 11. The effect is not statistically significant that's why we can not that the open container laws increased or decreased the number of traffic death.

13. Administrative per se laws

```
m4 = lm(dthrte85 ~ admn85, data = traffic1)
m5 = lm(dthrte90 ~ admn90, data = traffic1)
m6 = lm(dthrte ~ admn + year + admn*year, data = long)
screenreg(list(m4,m5,m6))
```

```
===============================================
               Model 1     Model 2     Model 3
-----------------------------------------------
(Intercept)     2.61 ***    2.11 ***    2.61 ***
               (0.11)      (0.11)      (0.10)
admn85          0.23
               (0.17)
admn90                      0.07
                           (0.15)
admn                                    0.23
                                       (0.16)
year                                   -0.49 **
                                       (0.16)
admn:year                              -0.15
                                       (0.22)
-----------------------------------------------
R^2             0.03        0.00        0.21
Adj. R^2        0.02       -0.02        0.19
Num. obs.      51          51         102
===============================================
*** p < 0.001; ** p < 0.01; * p < 0.05
```

The effect of administrative per se laws on traffic deaths is -0.15 but statistically insignificant. We can say administrative per se laws whether increased or decreased the traffic deaths.

14. ATE of open container and administrative per se laws together

```
m7 = lm(dthrte ~ admn + open + year + open*year + admn * year, data = long)

screenreg(list(m7))
```

```
=======================
               Model 1
-----------------------
(Intercept)     2.68 ***
               (0.12)
admn            0.23
               (0.16)
open           -0.21
               (0.16)
year           -0.59 **
               (0.18)
open:year       0.27
               (0.23)
admn:year      -0.17
               (0.23)
-----------------------
R^2             0.23
Adj. R^2        0.19
Num. obs.     102
=======================
*** p < 0.001; ** p < 0.01; * p < 0.05
```

Now treatment effect of open container laws is 0.27 while the administrative per laws treatment effect is -0.17. They are still statistically significant. We can conclude that these laws don't change the number of traffic deaths.