HW5_Natural Experiments

Annie Chen

```
rm(list = ls()) # Clear the workspace
library(dplyr)
library(ggplot2)
library(fixest)  # use for regression approach + cluster-robust robust standard errors
```

Part 1 set up

Set up

```
data.sales <- read.csv("sales_DID.csv")
summary(data.sales)
```

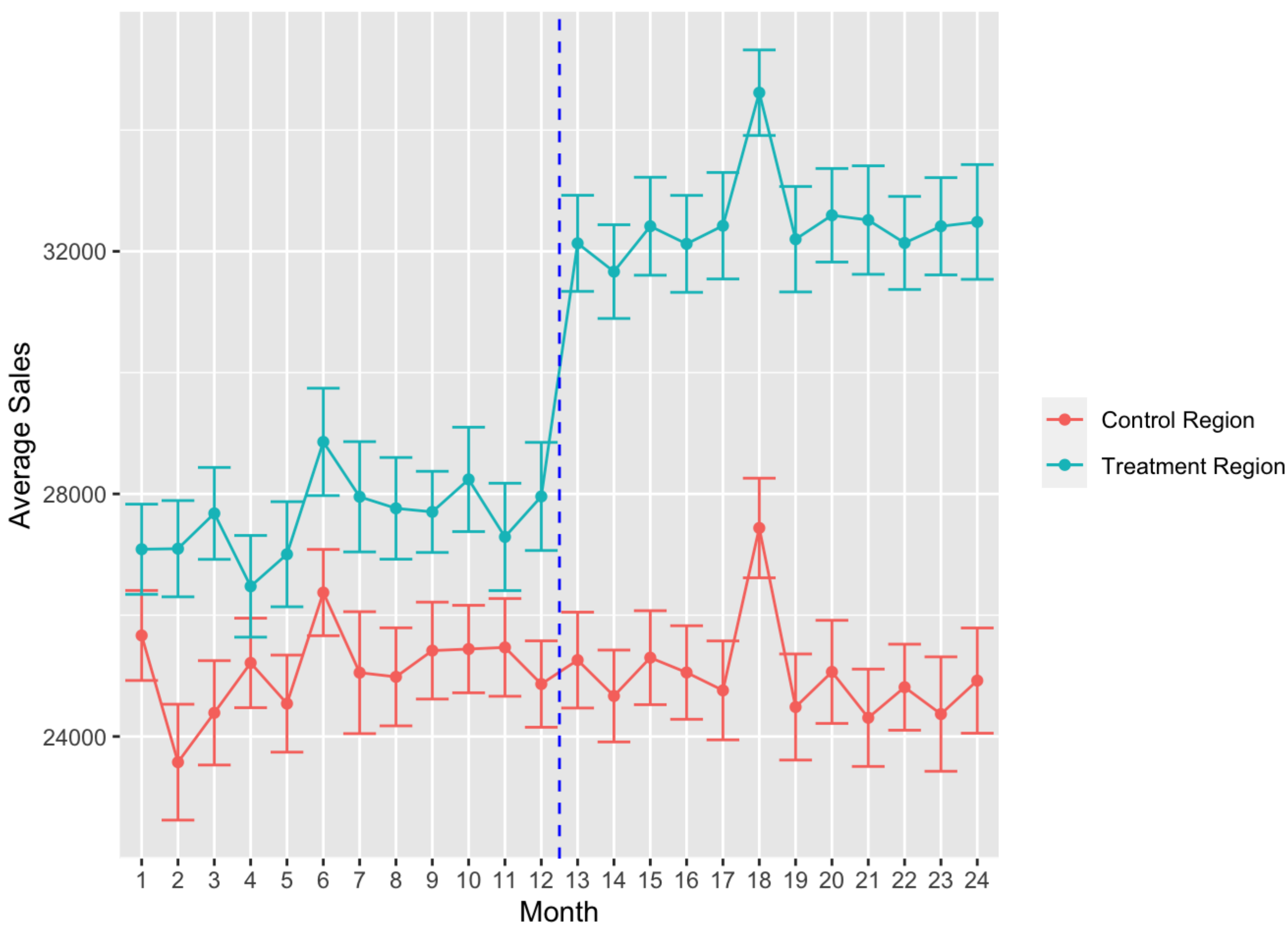
```
##      month          storenum       treatregion      sales      
##  Min.   : 1.00   Min.   :  1.00   Min.   :0.0   Min.   :16284  
##  1st Qu.: 6.75   1st Qu.: 25.75   1st Qu.:0.0   1st Qu.:24423  
##  Median :12.50   Median : 50.50   Median :0.5   Median :27123  
##  Mean   :12.50   Mean   : 50.50   Mean   :0.5   Mean   :27547  
##  3rd Qu.:18.25   3rd Qu.: 75.25   3rd Qu.:1.0   3rd Qu.:30424  
##  Max.   :24.00   Max.   :100.00   Max.   :1.0   Max.   :41940
```

Make a DID plot

We start by creating a graph of the average sales in each region by month (averaged across the 50 stores in that region for that month).Notice that the last bit of code adds a vertical line separating the pre- and post-periods.

```
summary <- data.sales %>%
  mutate(treatregion = as.factor(treatregion)) %>%
  mutate(month = as.factor(month)) %>%
  group_by(treatregion,month) %>%  #group the data by treatregion and month
  summarise (n = length(storenum),
             m.sales= mean(sales),
             error = sd(sales)/sqrt(n),
             lci = m.sales - 1.96*error,
             uci = m.sales + 1.96*error)


summary %>%
  ggplot(aes(x=month, y = m.sales, group = treatregion, color = treatregion)) +
  geom_point()+
  geom_line()+
  geom_errorbar(aes(ymin = lci, ymax = uci))+
  ylab("Average Sales") +
  xlab("Month") +
  scale_color_discrete(name=NULL, labels = c("Control Region", "Treatment Region"))+
  #  add a vertical line
  geom_vline(xintercept=12.5, linetype="dashed", color = "blue")
```



Q1. What does the plot show?

Describe the plot and note (i) whether it appears consistent with the parallel trends assumption and (ii) whether we should expect to see a positive, negative, or zero DID estimate. (Bonus (iii): use the plot to get a rough guess of the DID estimate. Explain your reasoning.)

Q1 The above plot shows a event studies in treatment region where there's an abrupt change after the month 12.

i. Before the threshold, there are parallel trends in both treatment and control region. To be more specific, before implementing the treatment and no one got treated, treatment region and control region are moving in the parallel trends.

ii. After the month 12 (implementing the treatment), the plot shows that there is a positive Difference-In-Difference estimate because the average sales of the treatment region spike abruptly.

iii. My estimation of the DID in the average sales is approximantely $3800.

- Change in treatment region = post-avg sales in treatment - pre-avg sales in treatment = 32000 - 28000 = 4000
- Change in control region = post-avg sales in control - pre-avg sales in control = 25000 - 24800 = 200
- DID = change in treatment region - change in control region = 4000 - 200 = 3800

2. Answer your smart colleague

Suppose your smart and experienced colleague after seeing this graph stated: "Well, I guess it is helpful to have the control region as a comparison to validate that the changes we see in the treatment region were not due to other factors at that time. But ultimately it looks like the difference-in-difference approach will give us basically the same estimate as just a simple event-study analysis for the treatment region." Briefly explain what your colleague means by that statement and what she is looking at in the graph that leads her to that conclusion.

Q2 The colleague means that there seemed to be no difference between using DID method and only analyzing the changes in the treatment region. Well, I think what she means by this statement is because there is no trend, either trending upward or downward, within the control group. The control group seemed to be a horizontal line. Therefore, when comparing the treatment and control region in the graph, we will not have the DID that will adjusted for the trend. I think that's the reason why she drawn her conclusion.

3. Regression Analysis

The code below makes the "post" and "post times treat" variables we ened for the DID regression, and then estimates the regression. Notice that we use the fixest package to get correct standard errors. You'll need to install this package if you don't have it.

```
#generate a dummy variable named post to indicate post-period (after month 12)
# and make post*treat
data.sales = data.sales %>% mutate(
  post = ifelse(month>12,1,0) ,
  post_treat = post*treatregion )

# estimate model - use feols to get correct standard errors
fit.sales = feols(sales~treatregion+post+post_treat, data = data.sales)

# correct standard errors
summary(fit.sales, cluster~storenum)
```

```
## OLS estimation, Dep. Var.: sales
## Observations: 2,400
## Standard-errors: Clustered (storenum)
##               Estimate Std. Error    t value  Pr(>|t|)
## (Intercept) 25082.1850    124.012 202.255550 < 2.2e-16 ***
## treatregion  2510.1546    176.354  14.233639 < 2.2e-16 ***
## post          -45.3107    145.184  -0.312091   0.75563
## post_treat   4929.5619    210.270  23.443939 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 3,008.1   Adj. R2: 0.502703
```

```
# calculate confidence intervals
n <- 2400
treatregion_lb <- 2510.1546 - 1.96 * 176.354 / sqrt(n)
treatregion_ub <- 2510.1546 + 1.96 * 176.354 / sqrt(n)
post_lb <- -45.3107 - 1.96 * 145.184 / sqrt(n)
post_ub <- -45.3107 + 1.96 * 145.184 / sqrt(n)
post_treat_lb <- 4929.5619 - 1.96 * 210.270 / sqrt(n)
post_treat_ub <- 4929.5619 + 1.96 * 210.270 / sqrt(n)
cat("treatregion CI: (", treatregion_lb, treatregion_ub, ")")
```

```
## treatregion CI: ( 2503.099 2517.21 )
```

```
cat("post CI: (", post_lb, post_ub, ")")
```

```
## post CI: ( -51.11927 -39.50213 )
```

```
cat("post_treat CI: (", post_treat_lb, post_treat_ub, ")")
```

```
## post_treat CI: ( 4921.149 4937.974 )
```

The table reports the estimated coefficient on each term (treat dummy, post dummy, interaction) in the first column and the standard error in the second.

i. Briefly interpret each of the coefficients in the regression and explain how they related to the graph. Focus on the magnitude. DO they seem large? (ii) Calculate and report confidence intervals (you'll need to transform these results)

Q3 According to the estimated coefficient on each term, we can get the regression for DID.

i. regression result for estimating DID:

- The coefficient for the treat dummy (2510.1546) means the pre-existing difference in treatment region and control region. It is the difference between two lines before month 12 in the line chart.
- The coefficient for the post dummy (-45.3107) means the trends common in both group. If we related back to the chart, we can see the slightly trend in both treatment and control region and the trend might be downward because the coefficient is negative.
- The coefficient for the interaction (post * treat) is 4929.5619, and this number represents the post time treat, which is the data before and after treatment period. Therefore, we can estimate that the number of 4929 is the Difference-In-Difference estimate.

ii. Confidence Intervals:

- treatregion CI: (2503.10, 2517.21)
- post CI: (-51.12, -39.50)
- post_treat CI: (4921.15, 4937.97)

Part 2.

```
# code for loading data
data.rating <- read.csv("RatingData.csv")
summary(data.rating)
```

```
##        id             month           rating          treat       
##  Min.   :    75   Min.   : 1.00   Min.   :1.000   Min.   :0.0000  
##  1st Qu.:  4224   1st Qu.: 7.00   1st Qu.:3.000   1st Qu.:0.0000  
##  Median : 16820   Median :14.00   Median :4.000   Median :0.0000  
##  Mean   : 29071   Mean   :13.39   Mean   :3.742   Mean   :0.2341  
##  3rd Qu.: 47383   3rd Qu.:20.00   3rd Qu.:5.000   3rd Qu.:0.0000  
##  Max.   :104668   Max.   :24.00   Max.   :5.000   Max.   :1.0000  
```

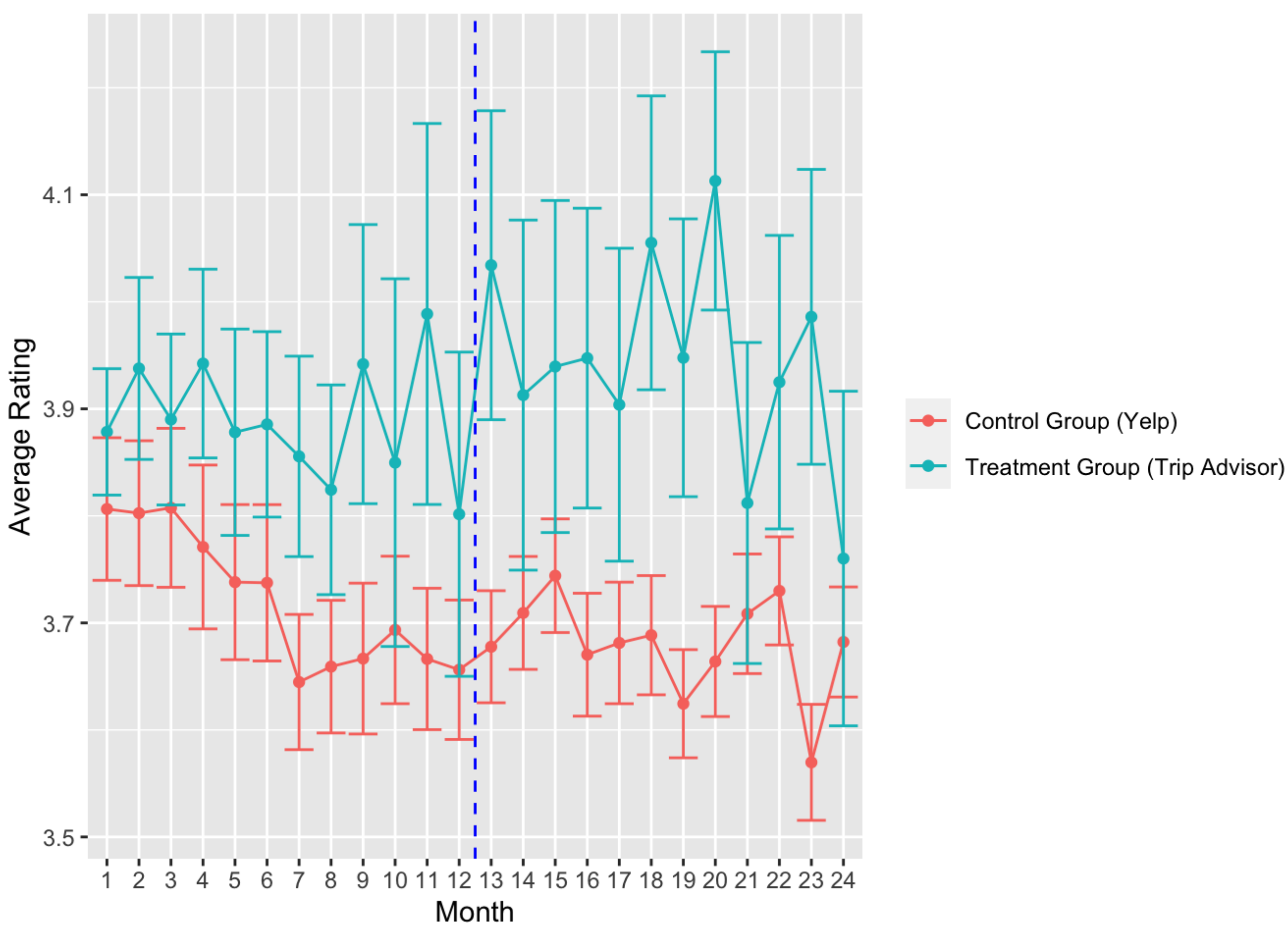
Q4 New DID plot

Plot the average rating for each website by month and include 95% confidence intervals on the average rating, similar to the example case. Make sure to label the axes and legend appropriately so that it is easy to process the information in the graph.

Interpret what you are seeing in the graph. Does the "parallel trends" assumption between Trip Advisor ratings and Yelp ratings appear valid here? Is there a discernable effect of the treatment at Trip Advisor that you can see in the graph?

```
# code for q4
summary_rating <- data.rating %>%
                 mutate(treat = as.factor(treat)) %>%
                 mutate(month = as.factor(month)) %>%
                 group_by(treat,month) %>%  #group the data by treatment and month
                 summarise (n = length(id),
                            m.rating= mean(rating),
                            error = sd(rating)/sqrt(n),
                            lci = m.rating - 1.96*error,
                            uci = m.rating + 1.96*error)


summary_rating %>%
                 ggplot(aes(x = month, y = m.rating, group = treat, color = treat)) +
                 geom_point()+
                 geom_line()+
                 geom_errorbar(aes(ymin = lci, ymax = uci))+
                 ylab("Average Rating") +
                 xlab("Month") +
                 scale_color_discrete(name=NULL, labels = c("Control Group (Yelp)", "Tr
eatment Group (Trip Advisor)"))+
                 #  add a vertical line
                 geom_vline(xintercept=12.5, linetype="dashed", color = "blue")
```



Q4 In the graph above, we can see that the average rating of treatment group (Trip Advisor) increase after implementing the multi-dimensional rating treatment. However, the plot shows that the parallel trends assumption did not appear valid here because there starts to have discrepancy around month 7 and the divergence seems to continue in the post period.

Q5 Regression estimate

Now run a regression (as in the example case) to quantify the difference-in-difference estimate of the effect of the switch to multi-attribute rating on average restaurant ratings. What is the DID estimate and its 95% confidence interval? Is the estimate statistically significant? Business significant?

```
# code for q5
data.rating = data.rating %>% mutate(
                                    post_r = ifelse(month > 12,1,0) ,
                                    post_r_treat = post_r * treat )

# estimate model - use feols to get correct standard errors
fit.rating = feols(rating ~ treat + post_r + post_r_treat, data = data.rating)

# correct standard errors
summary(fit.rating, cluster ~ id)
```

```
## OLS estimation, Dep. Var.: rating
## Observations: 38,004
## Standard-errors: Clustered (id)
##               Estimate Std. Error   t value   Pr(>|t|)
## (Intercept)   3.713904   0.024881 149.26665  < 2.2e-16 ***
## treat         0.172921   0.029289   5.90402 5.1461e-09 ***
## post_r       -0.036300   0.017546  -2.06880 3.8870e-02 *
## post_r_treat  0.094929   0.036589   2.59442 9.6402e-03 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 1.07243   Adj. R2: 0.007578
```

```
# calculate confidence intervals
N <- 38004
treat_lb <- 0.1729 - 1.96 * 0.0293 / sqrt(N)
treat_ub <- 0.1729 + 1.96 * 0.0293 / sqrt(N)
post_r_lb <- -0.0363 - 1.96 * 0.0175 / sqrt(N)
post_r_ub <- -0.0363 + 1.96 * 0.0175 / sqrt(N)
post_r_treat_lb <- 0.0949 - 1.96 * 0.0366 / sqrt(N)
post_r_treat_ub <- 0.0949 + 1.96 * 0.0366 / sqrt(N)
cat("treat CI: (", treat_lb, treat_ub, ")")
```

```
## treat CI: ( 0.1726054 0.1731946 )
```

```
cat("post_r CI: (", post_r_lb, post_r_ub, ")")
```

```
## post_r CI: ( -0.03647595 -0.03612405 )
```

```
cat("post_r_treat CI: (", post_r_treat_lb, post_r_treat_ub, ")")
```

```
## post_r_treat CI: ( 0.09453202 0.09526798 )
```

Q5 The DID estimate is 0.0949 and its 95% CI is between 0.0945 and 0.0952. In this case, the 95% confidence interval of the DID estimate does not contain the value 0, therefore, we can state that it is not statistically significant. To evaluate whether this multi-dimensional rating treatment is business significant or not, we should consider many factors at the same time such as the cost of implementing the treatment and the scale of it. Since the DID estimation is not statistically significant in this case, I would say it is not business significant at all.

Q6 Regression discussion

Focusing on the regression estimates and ignoring graphical evidence for the moment, what do you conclude about the effect of the switch to the multivariate rating system on average ratings at Trip Advisor? Please be specific about the evidence you are using for your conclusion.

Q6 According to the regression estimates, we can conclude that the Difference-In-Difference effect of switching to multivariate rating system is only approximately 0.095 increase in the average rating.

Q7 Regression vs. plot

Briefly discuss how the regression results compare to your visual analysis of the graph you created. Do you reach a similar conclusion or different? Which approach would you rely on more in a business situation for this case, and why?

Q7 I got 0.095 increase in the average rating according to the regression estimates; however, if I calculate the DID using the graph, I will get 0.21 for the DID estimate of implementing the multi-dimensional system.

- Change in treat (Trip Advisor) = post-avg rating in treatment - pre-avg rating in treatment = 4.03 - 3.80 = 0.23
- Change in control (Yelp) = post-avg rating in control - pre-avg rating in control = 3.68 - 3.66 = 0.02
- DID = change in treat (Trip Advisor) - change in control (Yelp) = 0.23 - 0.02 = 0.21

The DID estimate reach a difference from the graph result (0.21) and the regression result (0.095). In the future, I would rely more on the graph result in a business situation because it is more straightforward to explain the logic using the graph. If I'm going to present in the business situation, the graph is definitely more efficient to explain to the stakeholders and convince them to invest in the mechanism.

Q8 generative AI

Did you use generative AI in this homework? If yes, please explain how you used it.

Q8 No, I did not use generative AI in this homework.