DID

## Setup

setwd("C:/Users/22700/Desktop")  
library(data.table)  
  
library(ggplot2)  
library(stargazer)

##   
## Please cite as:

## Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.

## R package version 5.2.2. https://CRAN.R-project.org/package=stargazer

library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v tibble 3.0.4 v dplyr 1.0.2  
## v tidyr 1.1.2 v stringr 1.4.0  
## v readr 1.4.0 v forcats 0.5.0  
## v purrr 0.3.4

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::between() masks data.table::between()  
## x dplyr::filter() masks stats::filter()  
## x dplyr::first() masks data.table::first()  
## x dplyr::lag() masks stats::lag()  
## x dplyr::last() masks data.table::last()  
## x purrr::transpose() masks data.table::transpose()

library(plm)

##   
## Attaching package: 'plm'

## The following objects are masked from 'package:dplyr':  
##   
## between, lag, lead

## The following object is masked from 'package:data.table':  
##   
## between

library(car)

## Loading required package: carData

##   
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':  
##   
## recode

## The following object is masked from 'package:purrr':  
##   
## some

library(gplots)

##   
## Attaching package: 'gplots'

## The following object is masked from 'package:stats':  
##   
## lowess

library(tseries)

## Registered S3 method overwritten by 'quantmod':  
## method from  
## as.zoo.data.frame zoo

library(lmtest)

## Loading required package: zoo

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

library(sandwich)

## Loading required package: sandwich

library(GGally)

## Loading required package: GGally

## Registered S3 method overwritten by 'GGally':  
## method from   
## +.gg ggplot2

library(Metrics)  
  
#load dataset  
load("txtbans.RData")  
  
#explore data  
head(dt.txtbans)

## state time after treated accsvobyinhab laccsvobyinhab txmsban callban  
## 1: 1 1 0 0 0.005148116 0.005134909 0 0  
## 2: 1 2 0 0 0.005791630 0.005774923 0 0  
## 3: 1 3 0 0 0.005362620 0.005348293 0 0  
## 4: 1 4 0 0 0.005148116 0.005134909 0 0  
## 5: 1 5 0 0 0.007722173 0.007692510 0 0  
## 6: 1 6 0 0 0.006649649 0.006627638 0 0  
## bantime pop lpop unemp lunemp permale rgastax lrgastax  
## 1: 0 4661900 15.35493 3.3 1.193922 48.42248 0.001798277 -6.320926  
## 2: 0 4661900 15.35493 3.3 1.193922 48.42248 0.001788707 -6.326262  
## 3: 0 4661900 15.35493 3.3 1.193922 48.42248 0.001772566 -6.335327  
## 4: 0 4661900 15.35493 3.3 1.193922 48.42248 0.001761126 -6.341802  
## 5: 0 4661900 15.35493 3.3 1.193922 48.42248 0.001750429 -6.347894  
## 6: 0 4661900 15.35493 3.3 1.193922 48.42248 0.001747043 -6.349831

summary(dt.txtbans)

## state time after treated   
## Min. : 1.00 Min. : 1.00 Min. :0.0 Min. :0.0000   
## 1st Qu.:18.00 1st Qu.: 6.75 1st Qu.:0.0 1st Qu.:0.0000   
## Median :29.50 Median :24.50 Median :0.5 Median :0.0000   
## Mean :29.16 Mean :24.50 Mean :0.5 Mean :0.4211   
## 3rd Qu.:42.00 3rd Qu.:42.25 3rd Qu.:1.0 3rd Qu.:1.0000   
## Max. :54.00 Max. :48.00 Max. :1.0 Max. :1.0000   
## accsvobyinhab laccsvobyinhab txmsban callban   
## Min. :0.000000 Min. :0.000000 Min. :0.0000 Min. :0.00000   
## 1st Qu.:0.001900 1st Qu.:0.001898 1st Qu.:0.0000 1st Qu.:0.00000   
## Median :0.002965 Median :0.002961 Median :0.0000 Median :0.00000   
## Mean :0.003393 Mean :0.003385 Mean :0.2105 Mean :0.07895   
## 3rd Qu.:0.004559 3rd Qu.:0.004548 3rd Qu.:0.0000 3rd Qu.:0.00000   
## Max. :0.015660 Max. :0.015539 Max. :1.0000 Max. :1.00000   
## bantime pop lpop unemp   
## Min. : 0 Min. : 639715 Min. :13.37 Min. : 2.400   
## 1st Qu.: 0 1st Qu.: 2041928 1st Qu.:14.53 1st Qu.: 4.200   
## Median : 0 Median : 4475968 Median :15.31 Median : 5.650   
## Mean :12 Mean : 6527008 Mean :15.22 Mean : 6.524   
## 3rd Qu.:31 3rd Qu.: 6554090 3rd Qu.:15.70 3rd Qu.: 8.800   
## Max. :36 Max. :37349364 Max. :17.44 Max. :14.900   
## lunemp permale rgastax lrgastax   
## Min. :0.8755 Min. :48.30 Min. :0.001501 Min. :-6.502   
## 1st Qu.:1.4351 1st Qu.:48.70 1st Qu.:0.001764 1st Qu.:-6.340   
## Median :1.7316 Median :49.23 Median :0.001934 Median :-6.248   
## Mean :1.7817 Mean :49.31 Mean :0.002014 Mean :-6.220   
## 3rd Qu.:2.1748 3rd Qu.:49.80 3rd Qu.:0.002221 3rd Qu.:-6.110   
## Max. :2.7014 Max. :50.91 Max. :0.003000 Max. :-5.809

View(dt.txtbans)  
dt.bf.aft <- data.table(dt.txtbans) # new data table  
dt.bf.aft <- dt.bf.aft[, list( # Create a list of the columns of your new table   
mean\_accsvobyinhab = mean(accsvobyinhab ,na.rm=TRUE),  
mean\_laccsvobyinhab = mean(laccsvobyinhab , na.rm=TRUE)  
), by=list(after, treated)] # Specifiy the list of grouping variables  
  
t.test( dt.txtbans[after==0 & treated==0, accsvobyinhab])

##   
## One Sample t-test  
##   
## data: dt.txtbans[after == 0 & treated == 0, accsvobyinhab]  
## t = 27.779, df = 263, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## 0.004053524 0.004672002  
## sample estimates:  
## mean of x   
## 0.004362763

mean(dt.txtbans[after==0 & treated==0, accsvobyinhab])

## [1] 0.004362763

lm1 <- lm( accsvobyinhab ~ after + treated + after\*treated, data = dt.txtbans)   
summary(lm1) #after:treatd = 0.0003974

##   
## Call:  
## lm(formula = accsvobyinhab ~ after + treated + after \* treated,   
## data = dt.txtbans)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.0043628 -0.0013653 -0.0004519 0.0011961 0.0112971   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.0043628 0.0001249 34.931 < 2e-16 \*\*\*  
## after -0.0010336 0.0001766 -5.852 6.79e-09 \*\*\*  
## treated -0.0012736 0.0001925 -6.617 6.27e-11 \*\*\*  
## after:treated 0.0003947 0.0002722 1.450 0.147   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.002029 on 908 degrees of freedom  
## Multiple R-squared: 0.1048, Adjusted R-squared: 0.1018   
## F-statistic: 35.43 on 3 and 908 DF, p-value: < 2.2e-16

lm2 <- lm( laccsvobyinhab ~ after + treated + after\*treated, data = dt.txtbans)  
summary(lm2) #after:treatd = 0.0003915

##   
## Call:  
## lm(formula = laccsvobyinhab ~ after + treated + after \* treated,   
## data = dt.txtbans)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.0043501 -0.0013571 -0.0004497 0.0011933 0.0111885   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.0043501 0.0001243 34.994 < 2e-16 \*\*\*  
## after -0.0010283 0.0001758 -5.849 6.89e-09 \*\*\*  
## treated -0.0012672 0.0001916 -6.615 6.35e-11 \*\*\*  
## after:treated 0.0003915 0.0002709 1.445 0.149   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.00202 on 908 degrees of freedom  
## Multiple R-squared: 0.1048, Adjusted R-squared: 0.1019   
## F-statistic: 35.44 on 3 and 908 DF, p-value: < 2.2e-16

1) Explore the data set. How is this case different from the two cases we have seen until now?

The ban was initially implemented for 13 months, and the longest is 36 months. Some states just ban calls, some just ban text messages, and some ban them. In order to overcome this further, calling and texting bans were implemented at different times.

2) Define the control and treatment group and write down your regression model in equation form.

Control: States that have not implemented a text ban. Treatment: A state that has indeed implemented a text ban. The regression will be some coefficients before and after the ban (in time series), and the processed coefficients in time series (whether or not the ban)

Then the equation would be:

3) Create a table with the means of the variables of interest for the treatment and control groups before

and after treatment.

Deaths before treatment in untreated: 0.00436

Deaths after treatment in untreated: 0.00333

Deaths before treatment in treated: 0.00309

Deaths treatment in treated after: 0.00245

4) Use the table created in 3) to calculate the difference-in-differences estimator.

(Deaths treatment in treated after- Deaths before treatment in treated)-(Deaths after treatment in untreated- Deaths before treatment in untreated)= (0.00245-0.00309)-( 0.00333-0.00436)=0.00039

5) As seen in class, run a simple difference-in-differences regression model in order to obtain the same

estimator as in 4).

summary(lm1) #after:treatd = 0.0003974

summary(lm2) #after:treatd = 0.0003915

6) Interpret all the coefficients in the model. Was the text message ban effective in preventing car

accidents?

Tu sum up, = -0.0010336 proves that whether enforced the ban, the number of deaths decreased in all states. The shows that regardless of whether before or after the ban, there were fewer deaths in the states banned, indicates a very low variance after treatment. According to their p-value are all strongly significant, therefore, we are not able to say the text message ban is effective in preventing car accidents.

7) On top of the standard linear regression model assumptions, what is the additional assumption needed

for difference-in-differences to estimate the average treatment effect?

Parallel trend assumption. In this case, we need to assume that if no action is taken, the changes in accidents in states that do not implement the ban will be the same as the changes in incidents in the states that implement the ban.