Difference in Differences

Saumya Seth

2023-12-08

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v forcats   1.0.0     v readr     2.1.4
## v ggplot2   3.4.4     v stringr   1.5.0
## v lubridate 1.9.2     v tibble    3.2.1
## v purrr     1.0.1     v tidyr     1.3.0

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(gt)

set.seed(123)
```

1. Scenario

I would use this method when say I want to determine whether an intervention to reduce bullying in schools actually did. Here, I would want to compare and contrast the level of bullying in schools which received the intervention versus the rates of bullying in schools which did not. I would need the rates of bullying of all the schools at both time points, before and after the intervention was introduced. Then, assuming DID assumptions are satisfied, it would be possible to estimate the ATC, ATE and ATT. In this report, we will focus on estimating the ATT. Thus, w.r.t. this estimand, we would be able to estimate the causal effect of the intervention on the schools where the intervention was introduced. If it is reasonable to assume that the change in bullying in schools which did not receive the intervention is the same as what it would have been if they did, estimating the ATT using DID would be a smart choice.

2. DGP

World A

$$\text{class size} \sim \text{N}(15, 2^2)$$

$$\text{bullying}_{pre} \sim \text{N}(20, 2^2)$$

$$Z = \begin{cases} 1, & \text{if bullying}_{pre} \geq 22 \\ 0, & \text{otherwise} \end{cases}$$

$$\text{Y}(0) \sim \text{N}(2 + \text{bullying}_{pre}, 2^2)\text{Y}(1) \sim \text{N}(2 + \text{bullying}_{pre} - 4, 2^2)$$

$$\text{bullying}_{post} = \begin{cases} Y(0), & \text{if } Z = 0 \\ Y(1), & \text{otherwise} \end{cases}$$

World B

$$\text{class size} \sim \text{N}(15, 2^2)$$

$$\text{bullying}_{pre} \sim \text{N}(20, 2^2)$$

$$Z = \begin{cases} 1, & \text{if bullying}_{pre} \geq 22 \\ 0, & \text{otherwise} \end{cases}$$

$$\text{Y}(0) \sim \text{N}(2 + 4 * \text{bullying}_{pre}, 2^2)\text{Y}(1) \sim \text{N}(2 + 4 * \text{bullying}_{pre} - 4, 2^2)$$

$$\text{bullying}_{post} = \begin{cases} Y(0), & \text{if } Z = 0 \\ Y(1), & \text{otherwise} \end{cases}$$

For all worlds: - class size = average class size of the school - bullying_pre = observed bullying rate before the intervention - bullying_post = observed bullying rate after the intervention - Z = exposure group (treatment assignment) - Y(0) = potential outcome of bullying rate at schools which did not receive the intervention before and after the time of the intervention being implemented - Y(1) = potential outcome of bullying rate at schools which received the intervention before and after the time of the intervention being implemented

3. Methods and Estimands

a)

DID is an approach which assumes parallel trends. This means that it assumes that a change in the outcome variable is the same across control and treatment groups after the introduction of an intervention to which the treatment group was exposed to while the control group was not. It would not make sense to compare the outcomes of the treatment and control groups after the introduction of the intervention ONLY as the treatment and control groups might be systematically different in how their characteristics effect the outcome. Eg. it might be the case that 'control' schools just have a higher rate of bullying. Comparing this rate to that of the treatment group after the intervention might mislead the researcher to think that the intervention worked, when instead, their comparison group, by chance, has a higher bullying rate. So, if DID assumptions hold, we would be able to compare the changes in the reports of bullying across time periods (i.e. before and after the intervention) in affected groups (schools exposed to the intervention) with changes observed across the same time periods for unaffected groups (schools unexposed to the intervention).

The estimand that I have chosen to estimate using DID is the ATT. Since we wish to identify how effective the intervention is, it would make sense to estimate how effective the intervention was on the schools which received the intervention. We can estimate the ATC and ATE too but I have chosen to estimate the ATT in this report.

b)

```
n <- 1000
class_size <- rnorm(n, mean = 15, sd = 2)
bullying_pre <- rnorm(n, mean = 20, sd = 2)
exposed <- numeric(n)
exposed[bullying_pre > 22] <- 1

y0 <- rnorm(n, 2 + bullying_pre, 2)
y1 <- rnorm(n, 2 + bullying_pre - 4, 2)
y <- y0 * (1 - exposed) + y1 * exposed
world <- data.frame(
  y0 = y0, y1 = y1, bullying_post = y,
  bullying_pre = bullying_pre,
  exposed = exposed, class_size
)
treat_bullying_t0 <- mean(world$bullying_pre[world$exposed == 1])
treat_bullying_t1 <- mean(world$bullying_post[world$exposed == 1])
control_bullying_t0 <- mean(world$bullying_pre[world$exposed == 0])
control_bullying_t1 <- mean(world$bullying_post[world$exposed == 0])
countrol_group_y1 <- mean(world$y1[world$exposed == 0])

world$bullying_change <- world$bullying_post - world$bullying_pre
print(paste0(
  "The estimated ATT where all assumptions hold using DID is: ",
  summary(glm(bullying_change ~ exposed + class_size, data = world))$coef[2, 1],
  ". This estimate is very close to the true ATT of -4."
))
```

4. Assumptions

Assumption 1:

D(0) is independent of Z (parallel trends assumption). This means that the change in bullying rates for schools which did not receive the intervention is equal to the change in bullying rates in the same schools IF THEY HAD received the intervention. This assumption is satisfied in world A where the changes in bullying are not dependent on the covariate of pre-treatment bullying which has a coefficient other than 1. With pre-treatment bullying having an intercept of 1 for treatment and control schools at t1, we would be able to ensure that the change in bullying rates for control schools and treatment schools being the same is a reasonable assumption to make statistically. When the coefficient is 1, we implicitly satisfy the assumption that the change over time (represented as a difference in means, not a percentage) is the same across groups. However, practically, there may be many other factors between time points t0 and t1 which may result in bullying rates in schools which did not receive the intervention to have very different bullying rates at t1 than those which received the intervention - factors which only affected the control schools and not the treatment schools (or factors which affected both, but differently). Eg. Say the control schools were randomly located geographically in an area in which bullying policies changed from t0 to t1 - thus not affecting the treatment schools in this time period. Such a situation may result in our assumption of ignorability i.e. D(0) being independent of Z being violated. Thus, a secondary assumption (or a good practice) is to ensure that the controls ‘look like’ the treated through matching, weighting, etc. before doing a DID analysis to estimate the ATT.

We violated the assumption of ignorability in world B by assigning the change in bullying rates for schools which did not receive the intervention to not equal the change in bullying rates in the same schools IF THEY HAD received the intervention. This was done by changing the coefficient of pre-treatment bullying when defining the bullying rates at time t1 for the control and treatment schools.

Assumption 2:

Parametric assumptions must be met if covariates are present in the model which determine bullying rates across the time periods. Since the DID analysis is highly dependent on comparing the difference in differences of the means of the outcomes of the control and treated across time periods, it highly depends on the true model being linear. Both worlds A and B satisfy this assumption as the models in both the worlds are linear. This assumption is not necessarily plausible in real life but if the aim is to determine the effectiveness of an intervention, I think it may not be of paramount importance (in comparison to the parallel trends assumption) that this assumption be satisfied. However, if many covariates are included in the model, I think this assumption would be important to satisfy too.

Assumption 3:

We cannot observe if SUTVA has been satisfied from the data. We need our study design to be made in such a way that SUTVA doesn’t end up biasing our results. SUTVA would be satisfied in our example if the bullying rates in a school due to getting the intervention do not effect the bullying rates in another. This may not always be the case, though. Bullying rates in one school can greatly effect those in another, especially if the schools are in close proximity and students from different schools interact with one another.

5. Code

World A

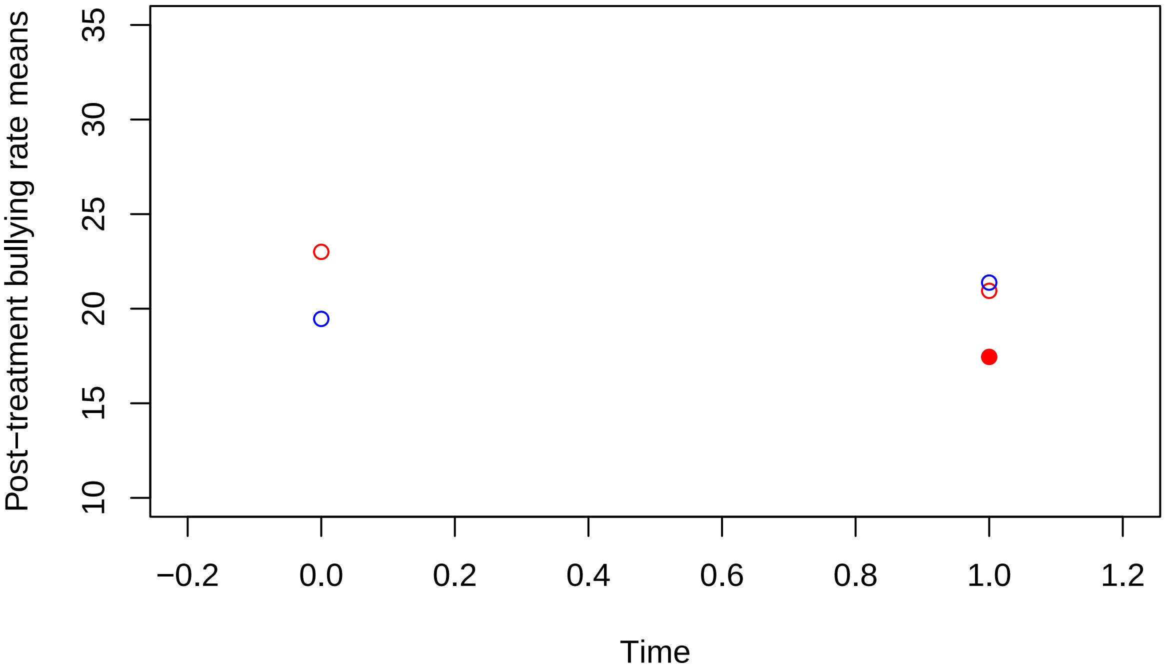
```r
n <- 1000
class_sizeA <- rnorm(n, mean = 15, sd = 2)
bullying_preA <- rnorm(n, mean = 20, sd = 2)
exposedA <- numeric(n)
exposedA[bullying_preA > 22] <- 1

y0A <- rnorm(n, 2 + bullying_preA, 2)
y1A <- rnorm(n, 2 + bullying_preA - 4, 2)
yA <- y0A * (1 - exposedA) + y1A * exposedA
worldA <- data.frame(
  y0A = y0A, y1A = y1A, bullying_postA = yA,
  bullying_preA = bullying_preA,
  exposedA = exposedA, class_sizeA
)
treat_bullying_t0A <- mean(worldA$bullying_pre[worldA$exposedA == 1])
treat_bullying_t1A <- mean(worldA$bullying_post[worldA$exposedA == 1])
control_bullying_t0A <- mean(worldA$bullying_pre[worldA$exposedA == 0])
control_bullying_t1A <- mean(worldA$bullying_post[worldA$exposedA == 0])
countrol_group_y1A <- mean(worldA$y1A[worldA$exposedA == 0])

head(worldA)
```

```
##        y0A      y1A bullying_postA bullying_preA exposedA class_sizeA
## 1 18.98520 15.70779       18.98520      18.00840        0    13.87905
## 2 20.39397 15.26458       20.39397      17.92009        0    14.53965
## 3 20.88086 15.06771       20.88086      19.96404        0    18.11742
## 4 24.17411 16.34108       24.17411      19.73565        0    15.14102
## 5 17.24959 18.09829       17.24959      14.90131        0    15.25858
## 6 22.85061 20.00632       20.00632      22.08115        1    18.43013
```

```
plot(
  x = c(0, 1), y = c(treat_bullying_t0A, treat_bullying_t1A), col = "red",
  ylim = c(10, 35),
  ylab = "Post-treatment bullying rate means", xlab = "Time", xlim = c(-.2, 1.2)
)
points(x = c(0, 1), y = c(control_bullying_t0A, control_bullying_t1A), col = "blue")
points(x = 1, y = countrol_group_y1A, col = "red", pch = 19)
```



```
# assuming parallel trends, the slope from control in time t0 to t1
# should be the same as the slope for treated in time t0 to t1
worldA$bullying_changeA <- worldA$bullying_postA - worldA$bullying_preA
print(paste0(
  "The estimated ATT in world A where all assumptions hold using DID is: ",
  summary(glm(bullying_changeA ~ exposedA + class_sizeA, data = worldA))$coef[2, 1],
  ". This estimate is very close to the true ATT of -4."
))
```

```
## [1] "The estimated ATT in world A where all assumptions hold using DID is: -3.96843706963792. This e
```

World B

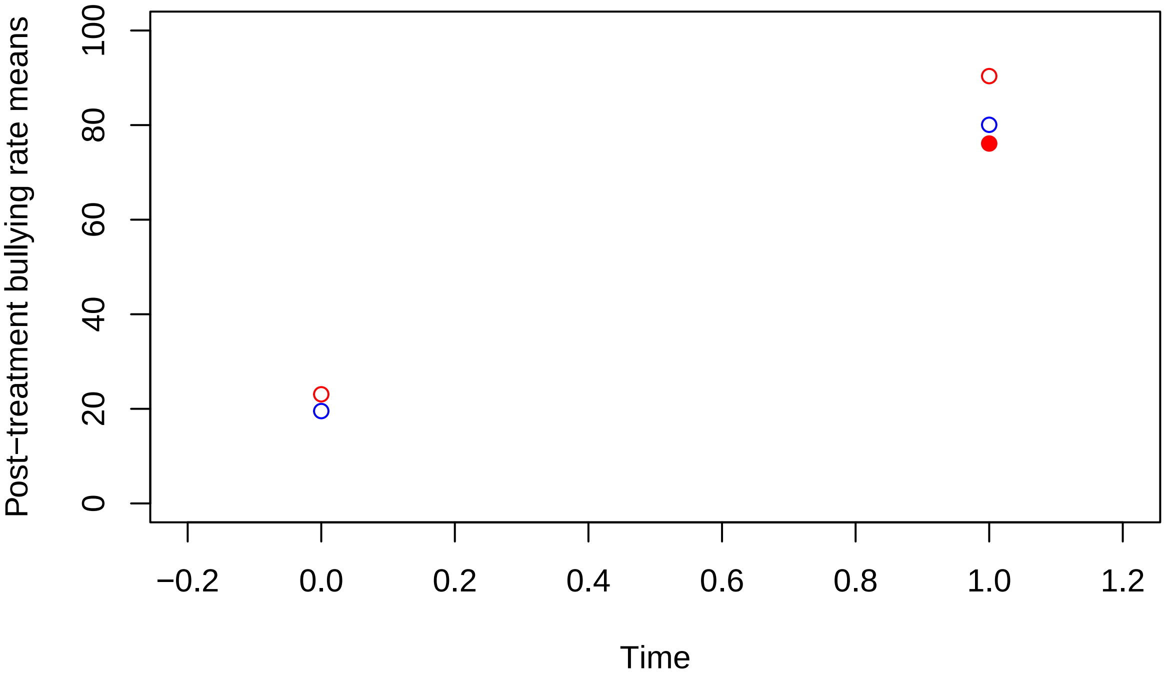
```r
n <- 1000
class_sizeB <- rnorm(n, mean = 15, sd = 2)
bullying_preB <- rnorm(n, mean = 20, sd = 2)
exposedB <- numeric(n)
exposedB[bullying_preB > 22] <- 1

y0B <- rnorm(n, 2 + 4 * bullying_preB, 2)
y1B <- rnorm(n, 2 + 4 * bullying_preB - 4, 2)
yB <- y0B * (1 - exposedB) + y1B * exposedB
worldB <- data.frame(
  y0B = y0B, y1B = y1B, bullying_postB = yB,
  bullying_preB = bullying_preB,
  exposedB = exposedB, class_sizeB
)
treat_bullying_t0B <- mean(worldB$bullying_pre[worldB$exposedB == 1])
treat_bullying_t1B <- mean(worldB$bullying_post[worldB$exposedB == 1])
control_bullying_t0B <- mean(worldB$bullying_pre[worldB$exposedB == 0])
control_bullying_t1B <- mean(worldB$bullying_post[worldB$exposedB == 0])
countrol_group_y1B <- mean(worldB$y1B[worldB$exposedB == 0])

head(worldB)
```

```
##        y0B      y1B bullying_postB bullying_preB exposedB class_sizeB
## 1 76.64815 70.81054       76.64815      19.01165        0    15.39310
## 2 93.01365 87.77911       87.77911      22.25519        1    16.30023
## 3 71.43891 72.62890       71.43891      17.70610        0    16.34201
## 4 93.64118 91.05190       91.05190      22.96204        1    12.43168
## 5 90.53726 88.79423       90.53726      21.83238        0    10.94778
## 6 83.46496 80.38731       83.46496      20.67026        0    19.41065
```

```r
plot(
  x = c(0, 1), y = c(treat_bullying_t0B, treat_bullying_t1B), col = "red",
  ylim = c(0, 100),
  ylab = "Post-treatment bullying rate means", xlab = "Time", xlim = c(-.2, 1.2)
)
points(x = c(0, 1), y = c(control_bullying_t0B, control_bullying_t1B), col = "blue")
points(x = 1, y = countrol_group_y1B, col = "red", pch = 19)
```



```r
# assuming parallel trends, the slope from control in time t0 to t1
# should be the same as the slope for treated in time t0 to t1
worldB$bullying_changeB <- worldB$bullying_postB - worldB$bullying_preB
print(paste0(
  "The estimated ATT in world B where the parallel trends assumption has been violated using DID is: ",
  summary(glm(bullying_changeB ~ exposedB + class_sizeB, data = worldB))$coef[2, 1],
  ". This is far off from the true ATT."
))
```

```
## [1] "The estimated ATT in world B where the parallel trends assumption has been violated using DID i
```

6. Causal Estimate and Interpretation

World A

The causal estimate for the intervention on the schools which received it in world A was:

```r
att_didA_estimate <- summary(glm(bullying_changeA ~ exposedA + class_sizeA,
  data = worldA
))$coef[2, 1]
att_didA_estimate
```

```
## [1] -3.968437
```

With a standard error of:

```r
att_didA_se <- summary(glm(bullying_changeA ~ exposedA + class_sizeA,
  data = worldA
))$coef[2, 2]
att_didA_se
```

```
## [1] 0.1603682
```

This means that if DID assumptions hold, especially the parallel trends assumption (which says that the change in the observed bullying rates for schools which did not receive the intervention equals that for the same schools HAD THEY received the intervention), then, we can say that the intervention led to a 4 point decrease in bullying rates in schools which received the intervention as compared to the situation if they had they not received it.

```
estimates_and_se <- data.frame(
  world = "world A",
  estimate = att_didA_estimate,
  standard_error = att_didA_se
)
```

World B

The causal estimate for the intervention on the schools which received it in world B was:

```
att_didB_estimate <- summary(glm(bullying_changeB ~ exposedB + class_sizeB,
  data = worldB
))$coef[2, 1]
att_didB_estimate
```

```
## [1] 6.765757
```

With a standard error of:

```
att_didB_se <- summary(glm(bullying_changeB ~ exposedB + class_sizeB,
  data = worldB
))$coef[2, 2]
att_didB_se
```

```
## [1] 0.4342495
```

```
estimates_and_se <- estimates_and_se %>%
  add_row(
    world = "world B",
    estimate = att_didB_estimate,
    standard_error = att_didB_se
  )
```

Since the DID assumptions were not satisfied in this world, we can say that if we compared 2 groups of schools with similar class sizes on average, one which were exposed to the intervention and one which weren't, then the group which was exposed to the intervention has a rate of bullying 6.765757 units more than those which did not receive the intervention, on average.

```
# final table
estimates_and_se <-
  estimates_and_se %>%
  rename(
    "World" = "world",
    "Causal Estimate" = "estimate",
    "Standard Error" = "standard_error"
  ) %>%
  gt() %>%
  tab_style(
    style = cell_text(weight = "bold"),
    locations = cells_column_labels(
      columns = c(
        "World",
        "Causal Estimate",
        "Standard Error"
      )
    )
  )

estimates_and_se
```

World	Causal Estimate	Standard Error
world A	-3.968437	0.1603682
world B	6.765757	0.4342495

7. Bias

```
iter <- 1000
true_value <- -4
```

World A: Linear Regression

```
did_attA <- rep(NA, 1000)

for (i in 1:iter) {
  n <- 1000
  class_sizeA <- rnorm(n, mean = 15, sd = 2)
  bullying_preA <- rnorm(n, mean = 20, sd = 2)
  exposedA <- numeric(n)
  exposedA[bullying_preA > 22] <- 1
  y0A <- rnorm(n, 2 + bullying_preA, 2)
  y1A <- rnorm(n, 2 + bullying_preA - 4, 2)
  yA <- y0A * (1 - exposedA) + y1A * exposedA
  worldA <- data.frame(
    y0A = y0A, y1A = y1A, bullying_postA = yA,
    bullying_preA = bullying_preA,
    exposedA = exposedA, class_sizeA
  )
  treat_bullying_t0A <- mean(worldA$bullying_pre[worldA$exposedA == 1])
  treat_bullying_t1A <- mean(worldA$bullying_post[worldA$exposedA == 1])
  control_bullying_t0A <- mean(worldA$bullying_pre[worldA$exposedA == 0])
  control_bullying_t1A <- mean(worldA$bullying_post[worldA$exposedA == 0])
  countrol_group_y1A <- mean(worldA$y1A[worldA$exposedA == 0])

  worldA$bullying_changeA <- worldA$bullying_postA - worldA$bullying_preA
  did_attA[i] <- summary(glm(bullying_changeA ~ exposedA +
    class_sizeA, data = worldA))$coef[2, 1]
}

bias_worldA <- true_value - mean(as.numeric(did_attA))
abs(bias_worldA)
```

```
## [1] 0.0003708259
```

World B: Linear Regression

```
did_attB <- rep(NA, 1000)

for (i in 1:iter) {
  n <- 1000
  class_sizeB <- rnorm(n, mean = 15, sd = 2)
  bullying_preB <- rnorm(n, mean = 20, sd = 2)
  exposedB <- numeric(n)
  exposedB[bullying_preB > 22] <- 1
  y0B <- rnorm(n, 2 + 4 * bullying_preB, 2)
  y1B <- rnorm(n, 2 + 4 * bullying_preB - 4, 2)
  yB <- y0B * (1 - exposedB) + y1B * exposedB
  worldB <- data.frame(
    y0B = y0B, y1B = y1B, bullying_postB = yB,
    bullying_preB = bullying_preB,
    exposedB = exposedB, class_sizeB
  )
  treat_bullying_t0B <- mean(worldB$bullying_pre[worldB$exposedB == 1])
  treat_bullying_t1B <- mean(worldB$bullying_post[worldB$exposedB == 1])
  control_bullying_t0B <- mean(worldB$bullying_pre[worldB$exposedB == 0])
  control_bullying_t1B <- mean(worldB$bullying_post[worldB$exposedB == 0])
  countrol_group_y1B <- mean(worldB$y1B[worldB$exposedB == 0])

  worldB$bullying_changeB <- worldB$bullying_postB - worldB$bullying_preB
  did_attB[i] <- summary(glm(bullying_changeB ~ exposedB +
    class_sizeB, data = worldB))$coef[2, 1]
}

bias_worldB <- true_value - mean(as.numeric(did_attB))
abs(bias_worldB)
```

```
## [1] 10.87379
```

- In world A, I did not violate any assumptions; thus, the estimates from the difference in means/linear regression were very close to the true causal effect at the cutoff. Thus, bias was small.

- In world B, I violated the assumption of parallel trends - this led to inaccurate estimates as given by difference in means/linear regression as the change in bullying rates for schools which did not receive the intervention did not equal the change in bullying rates in the same schools IF THEY HAD received the intervention. When we change the coefficient on pre-treatment bullying to be 1 we implicitly satisfy the assumption that the change over time (represented as a difference in means, not a percentage) is the same across groups. In world B, however, we violated this assumption by assigning the pre-treatment bullying rate coefficient to be 4. Thus, bias was much larger in this case.