Difference in Differences

Lachlan Deer

Social Media and Web Analytics, Spring 2025

Learning Goals

1. Define and give examples of natural / quasi experiments
2. Explain three empirical strategies (cross-section, before-and-after and difference in differences) to estimate the causal effect of a treatment
3. Explain and defend the assumptions that underpin each of the three empirical designs
4. Analyze data using the three designs using R

Where we've been

In previous sessions, we have:

- Seen how **randomized control trials** and **A/B tests** can retrieve **causal effects**
- Used **linear regression** to estimate the **causal effects** of interest
- Discussed how we can use **pre-experiment data** to **minimize variance** in a A/B test design
- Addressed two importance issues for **inference**: **heteroskedasticity robust- and clustered standard errors**

Now we're going to leave the experimental ideal, to see if we can still recover causal effects of **binary** treatments

Leaving the Experimental Ideal

There are **many situations** where we either:

- **Cannot completely randomize** treatment between individuals/firms we want to study
- Or (worse), we **cannot directly manipulate treatment** assignment

Question: How can we deal with selection bias and/or omitted variable bias if we do not randomize?

Natural/Quasi Experiments

Our goal: Define a setting and hypotheses that get us **as close as possible** to a **randomized experiment**

- These methods are sometimes referred to as **"natural experiments"** or **"quasi-experiments"**

What is a **Natural Experiment**?

- Units (individuals, firms etc) are **exposed** to treatment and control conditions in a way that is **outside** of the **control** of researchers/analysts
- BUT the process that governs assignment **resembles a random assignment**
- The analyst must convincingly argue that the setting under consideration resembles an experiment

Natural/Quasi Experiments: Examples

Good natural experiments are studies in which there is a **transparent exogenous source of variation** in the explanatory variables that **determine treatment assignment**

TABLE 1: EXAMPLES OF QUASI-EXPERIMENT STUDIES IN *JM*, *JMR*, AND *MARKETING SCIENCE* IN 2018-2020

Quasi-Experimental Variation		Article	Research Question
General Category	Source		
Contractual	Timing of American-Orbitz disputes to evaluate the absence of a major airline from a popular aggregator on consumer search	Akca and Rao (2020)	Who has more market power in the airline-aggregator relationships?
	Timing of the erection of the *New York Times* paywall	Pattabhiramaiah, Sriram, and Manchanda (2019)	How does a paywall affect readership and site traffic?
Ecological	Variation in the forecast error of the pollen levels	Thomas (2020)	How much does advertising affect purchases of allergy products?
Geographical	Discontinuities in the level of advertising at the borders of DMAs	Shapiro (2020)	Does advertising affect consumer choice of health insurance?
	Discontinuities in the level of political ads at the borders of DMAs	Wang, Lewis, and Schweidel (2018)	How does political advertising source and message tone affect vote shares and turnout rates in 2010 and 2012 Senatorial elections?

Natural/Quasi Experiments: Examples

Good natural experiments are studies in which there is a **transparent exogenous source of variation** in the explanatory variables that **determine treatment assignment**

Individual	Timing of users' adoption of a music streaming service	Datta, Knox, and Bronnenberg (2018)	How does a streaming service affect total music consumption?
	Variation in national ad exposures due to the local game outcomes	Hartmann and Klapper (2018)	How do Super Bowl ads affect brand purchases?
Macroeconomic	Variation in income and wealth due to the recession between 2006 and 2009	Dube, Hitsch, and Rossi (2018)	Do income and wealth affect demand for private label products?
Organizational	Discontinuity in the rounding rule that TripAdvisor uses to convert average ratings into displayed ratings	Hollenbeck, Moorthy, and Proserpio (2019)	How do online reviews affect advertising spending in the hotel industry?
	Timing of data breach and variation whether customer information was breached in a data breach event	Janakiraman, Lim, and Rishika (2018)	How does a data breach announcement affect customer spending and channel migration?

Natural/Quasi Experiments: Examples

Good natural experiments are studies in which there is a **transparent exogenous source of variation** in the explanatory variables that **determine treatment assignment**

Organizational	Variation in timing of adoption of front-of-package nutritional labels across categories	Lim et al. (2020)	Do front-of-package nutritional labels affect nutritional quality for other brands in a category?
Regulatory	Timing of the Massachusetts open payment law	Guo, Sriram, and Manchanda (2020)	Do payment disclosure laws affect physician prescription behavior?
	Enforcement of minimum advertisement price policies	Israeli (2018)	What is the effect on violations if firms improve digital monitoring and enforcement of minimum advertised price policies?
	Timing of India's FDI liberalization reform in 1991	Ramani and Srinivasan (2019)	How do firms respond to foreign direct investment liberalization?

Research Design for Natural Experiments

We will focus in on **Difference in Differences** (DiD, or Diff-in-Diff) as a **research design** to analyze data from a natural experiment

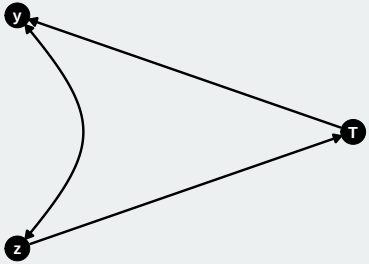
- One of the **most popular** methods used in empirical work to estimate the effect of a marketing intervention
- The idea will be **very simple** (and easy to communicate), and there are many settings where it **"fits well"** to the context we want to study
- This method "works" due to a **non-refutable assumption**: **parallel trends**
    - A lot of work thus must go in to showing this assumption is reasonable in the application under study

Confounding in Observational Data

Confounder: pre-treatment variable (Z) affecting the treatment (T) and outcome (Y)

Lead to **confounding bias** in the estimated SATE due to these differences

- $\bar{Y}_{control}$ is **not** a good proxy for $\frac{1}{n}\sum_{i=1}^{n} Y_i(0)$
- Types of bias:
  - (self-) **Selection bias** into treatment
  - **Omitted Variable** correlated to treatment and outcome



The Fundamental Problem of Causal Inference

Recall the **fundamental problem of causal inference**

Analyst must infer counterfactual outcomes

Can we find **other ways** to **estimate** the **counterfactual outcome** of a treated individual using other observations **?**

A (Fictitious) Example

Business questions: Do search engine ads increase spending?

Test setting: Google Sponsored Ads, Large Retailer

Unit: consumers

Treatments: control group - no ads, treatment group - sees ads

Response: spending in the next 30 days

Selection: all consumers who purchased in last 180 days

Assignment: Search engine ads were banned in one province by government on the first day of November of a calendar year, left on in adjacent province

Sample size: 1,000 consumers split 50/50 over provinces

Load and inspect data

```
df <- read_csv("data/diff_in_diff.csv")
head(df)
```

```
# A tibble: 6 x 5
  customer_id time_period after treatment_group sales
        <dbl>       <dbl> <dbl> <chr>           <dbl>
1           1           1     0 control          3.31
2         995           1     0 treatment        0
3         730           1     0 treatment        7.92
4         697           1     0 treatment        0
5         929           1     0 treatment        0
6         896           1     0 treatment        1.24
```

Three Questions

1. Do you expect the treatment and control group' spending to be the same on average **before** the ad ban? Why?
2. Do you expect the ad ban to be the **only** cause of changes in average spending for the treatment group between October (pre-ban) and November (post-ban)? Why?
3. Do you think that in the **absence** of the ad ban, the **change in average spending** in the treatment group would have been the **same** as in the control group? Why?

Research Designs

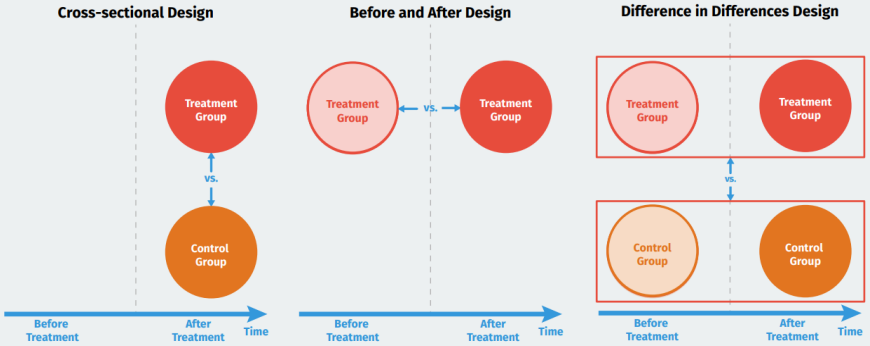
How do we **find a good comparison group**?

- Depends on the data available

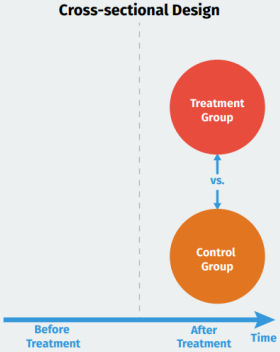
Three types observational study research designs:

1. **Cross-sectional design**: Compare outcomes of treated and control units at one point in time
2. **Before-and-after design**: Compare outcomes before and after a unit has been treated, but we need data over time for the treated group
3. **Difference-in-difference design**: Compare changes in the treatment group over time to changes in the control group over time. Needs data over time for the treated and control group

Research Designs



Cross Sectional Design



- Compare treated and control groups after the treatment happens
- **Assumption**: groups are identical on average
  - Sometimes this is called **unconfoundedness** or **as-if randomized**
- Cross-section estimate:

$$\bar{Y}^{\text{after}}_{\text{treated}} - \bar{Y}^{\text{after}}_{\text{control}}$$

- Could there be confounders?

Cross Sectional Design in R: Manually

```
# after period
cross_section <-
    df %>%
    filter(after == 1)

cross_section_est <- cross_section %>%
    # group means
    group_by(treatment_group) %>%
    summarize(
        mean_sales = mean(sales)
    ) %>%
    # difference between groups
    ungroup() %>%
    mutate(estimate = mean_sales - lag(mean_sales)) %>%
    na.omit() %>%
    purrr::pluck("estimate")

message("Cross Section Estimate: ", cross_section_est)
```

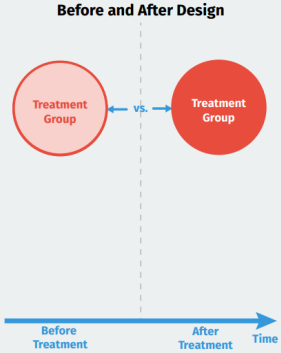
```
Cross Section Estimate: 1.31374
```

Cross Sectional Design in R: Linear Regression

```
tidy(lm(sales ~ treatment_group,
        data = cross_section))
```

```
# A tibble: 2 x 5
  term                   estimate std.error statistic  p.value
  <chr>                     <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)               2.53      0.158     16.0  1.77e-51
2 treatment_grouptreatment  1.31      0.224      5.88 5.68e- 9
```

Before-After Comparison



- Compares individuals before and after the treatment
- Advantage: all person specific features held fixed
  - "comparing within person over time"
- Before vs after estimate:

$$\bar{Y}^{\text{after}}_{\text{treated}} - \bar{Y}^{\text{before}}_{\text{treatment}}$$

- **Assumption**: No **time varying** confounders
  - i.e. no evolution of outcome over time that is not due to the treatment

Before and After in R: Manually

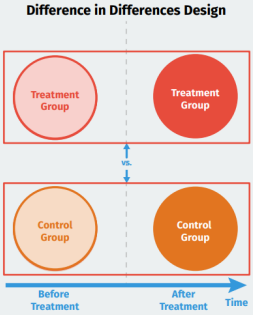
```
# treatment group, period immediately before and after
before_after <-
    df %>%
    filter(time_period %in% c(10,11),
           treatment_group == "treatment")

before_after_est <- before_after %>%
    group_by(after) %>%
    dplyr::summarize(
        mean_sales = mean(sales)
    ) %>%
    ungroup() %>%
    mutate(estimate = mean_sales - lag(mean_sales)) %>%
    na.omit() %>%
    purrr::pluck("estimate")

message("Before After Estimate: ", before_after_est)
```

```
Before After Estimate: 1.64522
```

Before and After in R: Linear Regression

```
# A tibble: 2 x 5
  term        estimate std.error statistic  p.value
  <chr>          <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)     2.20     0.155     14.2  1.14e-41
2 after           1.65     0.219      7.50 1.40e-13
```

Two differences can be better than one



- Use the before/after difference of **control group** to infer what would have happened to **treatment group** without treatment.
- DiD Estimate:

$$\underbrace{(\bar{Y}^{\text{after}}_{\text{treated}} - \bar{Y}^{\text{before}}_{\text{treatment}})}_{\text{Change in the treatment group}} - \underbrace{(\bar{Y}^{\text{after}}_{\text{control}} - \bar{Y}^{\text{before}}_{\text{control}})}_{\text{Change in the control group}}$$

i.e. the change in the treatment group above and beyond the change in the control group

- **Assumption**: **parallel trends**
  - Changes in the outcome for the treated group would have been the same as in the control group in the absence of treatment
  - Threat to inference: **non-parallel trends**

Difference in Differences in R

```
# did data -- treat and control, before and after
did_data <-
    df %>%
    filter(time_period %in% c(10,11))

did_est <- did_data %>%
    group_by(treatment_group, after) %>%
    summarise(mean_sales = mean(sales)) %>%
    ungroup() %>%
    pivot_wider(names_from = after,
                values_from = mean_sales) %>%
    mutate(across_rows_diff = `1` - `0`) %>%
    mutate(estimate = across_rows_diff - lag(across_rows_diff)) %>%
    na.omit() %>%
    purrr::pluck("estimate")

print(paste0("Diff in Diff Estimate: ", did_est))
```

```
[1] "Diff in Diff Estimate: 0.91402"
```

Estimating the Treatment Effect with Diff in Diff

Why does Diff in Diff allow us to
estimate the effect of the marketing intervention?

Let's maintain the **two groups** and **two time periods** set up

- **Groups**: Treatment and Control
- **Time Periods**: Before and after

Let's put the mean outcomes for each combination in a table

The 2 x 2 Table

	Before	After
Control	$\bar{Y}^{\text{before}}_{\text{control}}$	$\bar{Y}^{\text{after}}_{\text{control}}$
Treatment	$\bar{Y}^{\text{before}}_{\text{treatment}}$	$\bar{Y}^{\text{after}}_{\text{treatment}}$

The 2 x 2 Table

Let's update the notation:

1. Control group, before intervention starts

$$\bar{Y}^{\text{control}}_{\text{before}} = \beta_0$$

2. Control group, after intervention starts

$$\bar{Y}^{\text{control}}_{\text{after}} = \beta_0 + \beta_1$$

3. Treatment group, before intervention starts

$$\bar{Y}^{\text{treatment}}_{\text{before}} = \beta_0 + \beta_2$$

4. Treatment group, after intervention starts

$$\bar{Y}^{\text{treatment}}_{\text{after}} = \beta_0 + \beta_2 + \beta_1 + \delta$$

The 2 x 2 Table

	Before	After
Control	β_0	$\beta_0 + \beta_1$
Treatment	$\beta_0 + \beta_2$	$\beta_0 + \beta_2 + \beta_1 + \delta$

The 2 x 2 Table

Let's take differences in the following order:

- Across columns, then
- Across rows

	Before	After	After - Before
Control	β_0	$\beta_0 + \beta_1$	β_1
Treatment	$\beta_0 + \beta_2$	$\beta_0 + \beta_2 + \beta_1 + \delta$	$\beta_1 + \delta$
Treatment - Control			δ

'Double Differencing' $\implies$ estimate δ

- I call this DiD estimate using averages **simple DiD**

The 2 x 2 Table

Let's take differences in the following order:

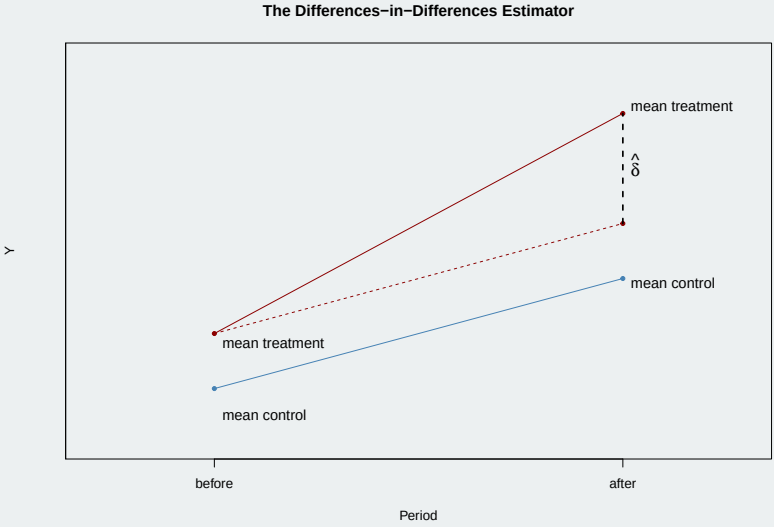
- Across columns, then
- Across rows

	Before	After	After - Before
Control	β_0	$\beta_0 + \beta_1$	
Treatment	$\beta_0 + \beta_2$	$\beta_0 + \beta_2 + \beta_1 + \delta$	
Treatment - Control	β_2	$\beta_2 + \delta$	δ

'Double Differencing' $\implies$ estimate δ

- I call this DiD estimate using averages **simple DiD**

Difference in Difference Graphically



DiD as a Regression

$$y_{it} = \beta_0 + \beta_1 After_t + \beta_2 Treated_i + \delta After_t \times Treated_i + \varepsilon_{it}$$

where:

- $After_t$ = 1 in the period after treatment occurs, zero otherwise
- $Treated_i$ = 1 if the individual is ever treated, zero otherwise

DiD Regression in R

```
tidy(lm(sales ~ treatment_group + after + treatment_group:after,
        data = did_data
        ))
```

```
# A tibble: 4 x 5
  term                           estimate std.error statistic  p.value
  <chr>                             <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)                       1.80      0.144     12.5  2.03e-34
2 treatment_grouptreatment          0.400     0.204      1.96 5.01e- 2
3 after                             0.731     0.204      3.59 3.44e- 4
4 treatment_grouptreatment:after    0.914     0.288      3.17 1.55e- 3
```

Standard Error Adjustment

- Homoskedasticity in the standard errors is likely violated
- **Cluster Robust Standard Errors** are the **default choice** for DiD designs
- What unit to cluster?
    - The variable that determines treatment assignment

Parallel Trends

We must assume that Time effects treatment and control groups equally

- Otherwise controlling for time (i.e. `after`) won't work

This is called the **parallel trends** assumption

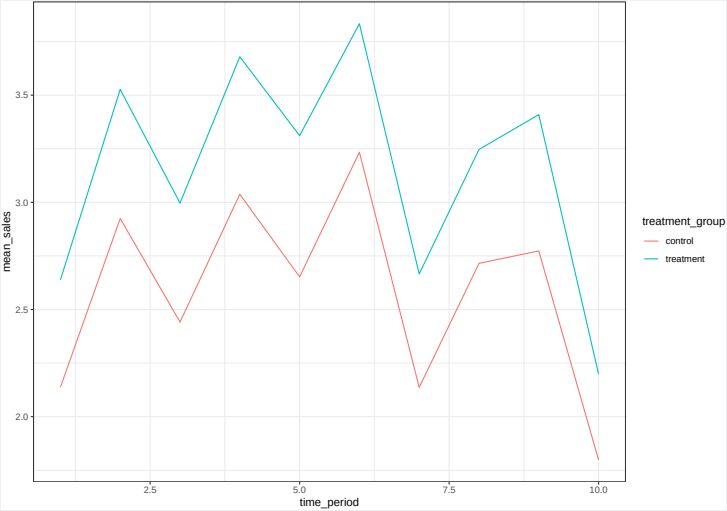
- Again, *if the Treatment hadn't happened to anyone*, the differences between the treatment and control would stay the same

Checking for Parallel Trends

Like many assumptions - its **untestable**

- Though we can **'check' whether patterns in the data are suggestive its OK**
- Here's the most popular way:
  - Are *prior trends* are the same for Treated and Control groups
  - Generally, compute average of outcome by group over time
  - (needs **multiple** pre-treatment periods)
  - Was the gap changing a lot during that period? If not, suggestive we're OK

Visualizing Parallel Trends



Isn't this just CUPED in disguise

How is this different from CUPED?

```
CUPED estimate: 0.8561
```

CUPED is equivalent to Difference in Differences under the following conditions:

- $\theta = 1$,
  - But in our example, $\hat{\theta}$ is `round(theta,4)`
- No differences in mean outcomes at the level of treatment assignment
  - In our example, the difference in average spending pre-treatment is 0.4, and statistically significant

Adding Control Variables

In applications, analysts may want to account for covariates in their DiD specification by **including covariates** in their regression.

$$y_{it} = \beta_0 + \beta_1 After_t + \beta_2 Treated_i + \delta After_t \times Treated_i + X_i\theta + \varepsilon_{it}$$

This is **additional structure** imposed by the researcher

Identifies the average treatment effect when the treatment effect is:

- Constant, and
- Additive

This is a **stronger assumption** than the control variable free approach

“As good as random”

Two assumptions for causality:

1. **Valid counterfactual outcomes**

- Control Group + parallel trends solves this one for us

2. **Conditional independence**: nothing unobserved is causing selection into treatment group

- Trickier ...
- Randomised Control Trial or A/B Test like $\rightarrow$ You’re more than likely gonna be OK
- Natural / Quasi Experiment $\rightarrow$ have you got a credible proxy for random assignment?
    - Profession’s thoughts: Large, visible, unexpected shocks

A Warning!

- DiD's popularity is relatively recent, so we're still learning a lot about it!
  - Most relevant has to do with **staggered roll out DiD**
- The regression version of DiD doesn't *necessarily* need to have treatment applied at *one* particular time
  - Treatment could be gradually implemented over time
- Nothing we've explicitly said would prevent us from using the regression DiD right!?
  - Well... that's what we thought for a long time.
  - And you'll see many of published studies doing this.
  - BUT it turns out to actually **bias results by quite a lot**
- There are more complex, newer estimators for staggered roll out case,
  - Too much for this class

Recap

- When we cannot run a randomized experiment, we want to use observational data that mimics random assignment as closely as possible
- Natural experiments (aka quasi experiments) mimic the random assignment although the treatment assignment is not controlled by the analyst
- Difference in Differences strategy allows us to estimate the causal effect of a intervention on outcome variables
- Difference in Differences relies on the parallel trends assumption, and an analyst needs to provide evidence that shows the condition is satisfied to make their research design credible

Acknowledgements

I have borrowed material and re-mixed material from the following:

- Matteo Courthoud's "Understanding CUPED"
- Matt Blackwell's Gov 50 "Data Science for the Social Sciences" lecture on "Observational Studies"

Suggested Citation:

```
@misc{smwa2025_did,
      title={"Social Media and Web Analytics: Difference in Differences"
      author={Lachlan Deer},
      year={2025},
      url = "https://tisem-digital-marketing.github.io/2025-smwa"
}
```

This course adheres to the principles of the Open Science Community of Tilburg University. This initiative advocates for transparency and accessibility in research and teaching to all levels of society and thus creating more accountability and impact.