

Week 11: Causation

POP88162 Introduction to Quantitative Research Methods

Tom Paskhalis

Department of Political Science, Trinity College Dublin

Research Design

- Approximately 1-2 pages and no more than 500 words (references excluded).
- Due **08:59 Tuesday, 8 April 2025**
- Key components:
 - Research question;
 - Dependent and independent variables;
 - Data;
 - Statistical test to be used;
- Narrative how the last three help answer the RQ.
- More details in Research Design Guidelines.

Topics for Today

- Causation
- Potential outcomes
- Average treatment effect (ATE)
- Average treatment effect on the treated (ATET)
- Bias
- Randomised experiments
- Difference-in-difference (DiD)

Previously...

Review: F-Test

- While t-test can be used to test the null hypothesis that the population coefficient of a single explanatory variable is 0, F-test is used to test multiple-coefficient hypotheses where:
 - Null hypothesis: $H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$ - in the population all coefficients $\beta_1, \beta_2, \dots, \beta_k$ are 0.
 - Alternative hypothesis: H_a : at least one of $\beta_1, \beta_2, \dots, \beta_k$ coefficients is not 0 in the population.
- Particularly useful for testing a related group of variables (e.g. dummies for different categories of a single variable).

Review: Interaction Term

- The simple model we have been studying assumes ‘constant associations’ (i.e. the relationship between X and Y does not depend on other X ’s).

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i$$

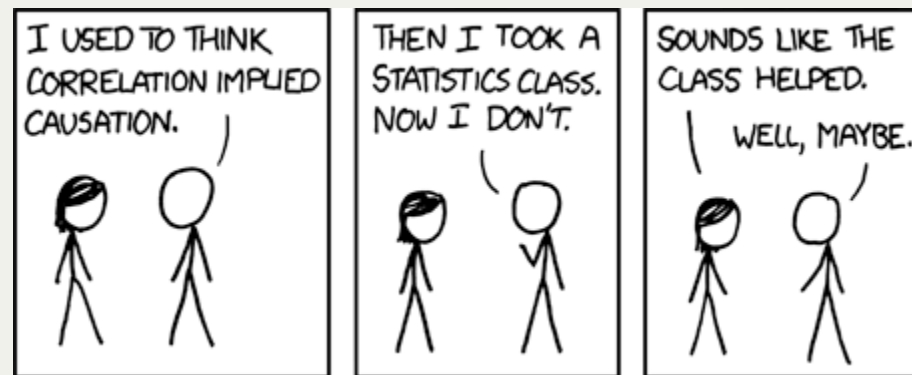
- We can relax the assumption of constant association by adding the product of explanatory variables to a model:

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + \epsilon_i$$

Causation

Why Causation?

- Cause-and-effect relationships are at the heart of some of the most interesting social science theories.
- We can often express important research topics as a simple question:
Does X cause Y ?
- Establishing causation is difficult.



xkcd

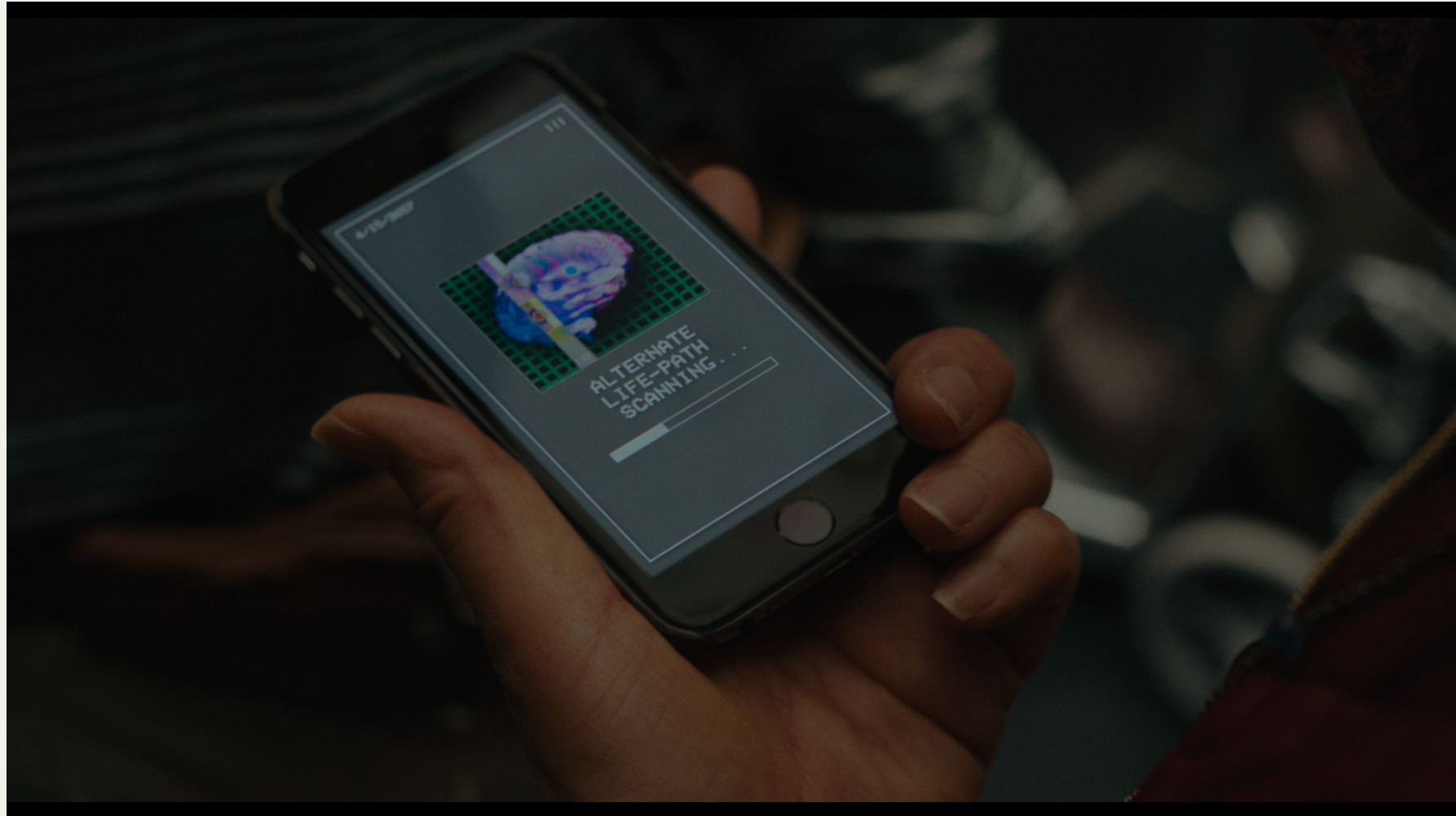
What is Causation?

- Causes of an effect
 - What causes lung cancer?
 - What caused a war?
 - What are the causes of inequality?
 - Why did Brexit happen?
- Effects of causes
 - Does smoking have an effect on lung cancer development?
 - What mobilization effect did protests have?
 - What is the effect of housing prices on inequality?
 - What was the effect of globalization on Brexit vote?

Parallel Worlds I



Parallel Worlds II



Counterfactual

Thus, if a person eats of a particular dish, and dies in consequence, that is, would not have died if he had not eaten of it, people would be apt to say that eating of that dish was the cause of his death.

John Stuart Mill (1843)

- The principles of establishing causality are based on **counterfactual comparison**.
- What would be the state of the world had some event not taken place?

Treatment and Outcome

- We will think about causal relationships in terms of effects of **treatments** on **outcomes**:
 - Treatment: where is the source of change
 - Outcome: what is affected by change

Potential Outcomes

- We are interested in the effect on some dependent variable (outcome) Y .
- We also are focussed on the effect of some independent variable (cause) X .
- To make this causal variable look more special let's denote it as T (treatment)
- Then our causal effect (unit treatment effect) is:

$$Y_{T=1,i} - Y_{T=0,i} = Y_{1i} - Y_{0i}$$

where:

- Observations $i = 1, \dots, n$
- Y is the dependent variable (outcome)
- $T = 1$ treatment condition (one state of the world)
- $T = 0$ control condition (alternative state of the world)
- Both Y_{1i} and Y_{0i} are **potential outcomes**, but we only get to observe one state of the world and, hence one of them!

Fundamental Problem of Causal Inference

It is impossible to observe the value of Y_{1i} and Y_{0i} on the same unit and, therefore, it is impossible to observe the effect of T on Y .

Paul W. Holland (1986)

- We cannot observe outcome of interest both under treatment Y_{1i} and under control Y_{0i} .
- Thus, the quantity $Y_{1i} - Y_{0i}$ (unit-level causal effect) is fundamentally unobservable.
- But it does not mean that **causal inference** is impossible!
- First, we would need to make some assumptions
- Second, we need to focus on alternative quantities.

Stable Unit Treatment Value Assumption (SUTVA)

- The observed outcomes Y_i are realised from potential outcomes:

$$Y_i = T_i Y_{1i} + (1 - T_i) Y_{0i} \text{ so } Y_i = \begin{cases} Y_{1i} & \text{if } T_i = 1 \\ Y_{0i} & \text{if } T_i = 0 \end{cases}$$

1. This implies that potential outcomes for a unit must be unaffected by treatment status of any other unit.
 2. It also implies that the received treatment is the same for all units.
- There are a number of cases where it is conceivable that *Stable Unit Treatment Value Assumption (SUTVA)* is violated.
 - Potential effect of Get out the vote (GOTV) campaign on spouse's turnout.
 - Individuals are less likely to get the disease if those around them are already vaccinated.
 - Exposing some students to educational interventions may affect outcomes for their classmates.

Average Treatment Effect (ATE)

- Individual-level causal effects are fundamentally unidentifiable.
- Thus, in most situations we focus on averages.
- In particular, one such quantity is **average treatment effect (ATE)**:

$$ATE = \frac{\sum_{i=1}^n Y_{1i} - Y_{0i}}{n}$$

- Equivalently, we can express ATE as:

$$E(Y_{1i} - Y_{0i})$$

- This formulation emphasises that what we are modelling is the *mean* or *expected value* of the difference between potential outcomes.

Average Treatment Effect on the Treated (ATET)

- Frequently, we also focus on subpopulation that was exposed to treatment.
- The effect calculated on such subpopulation would be **average treatment effect on the treated (ATET)**.
- We can express it as:

$$\frac{\sum_{i=1}^n T_i(Y_{1i} - Y_{0i})}{n_1} \text{ where } n_1 = \sum_{i=1}^n T_i$$

- Here, N_1 is the size of the treatment group.
- Equivalently, we can express ATET as:

$$E(Y_{1i} - Y_{0i} | T_i = 1)$$

ATE and ATET

- Why ATE would not be equal to ATET?
- Because $E(Y_{1i}) \neq E(Y_{1i}|T_i = 1)$ and, likewise for $E(Y_{0i})$.
- That is treatment status T_i and potential outcomes Y_{ti} are associated.

Bias

- So far, we looked at quantities (ATE and ATET) that we do not usually observe.
- The quantity that we can actually observe would be:

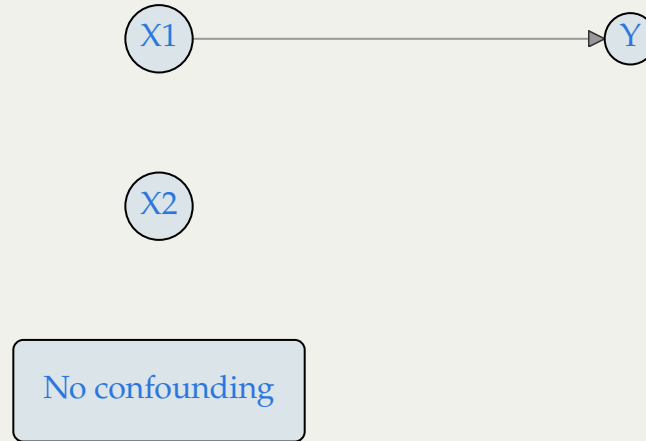
$$E(Y_i|T_i = 1) - E(Y_i|T_i = 0)$$

- We can further re-write it (note that we just add and subtract $E(Y_{0i}|T_i = 1)$):

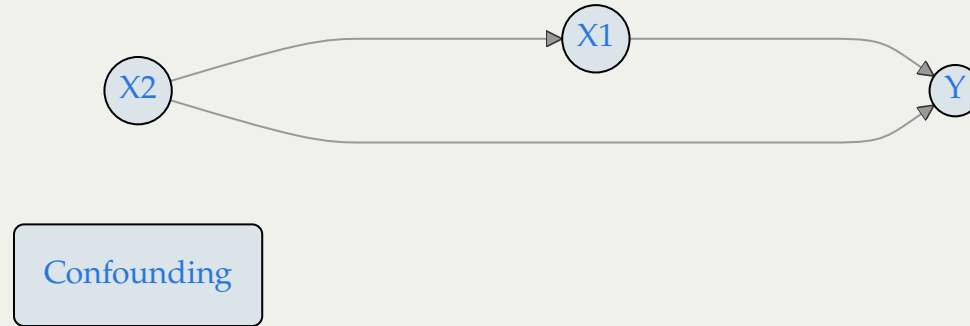
$$\begin{aligned} E(Y_i|T_i = 1) - E(Y_i|T_i = 0) &= \\ E(Y_{1i}|T_i = 1) - E(Y_{0i}|T_i = 0) &= \\ E(Y_{1i}|T_i = 1) - E(Y_{0i}|T_i = 1) + E(Y_{0i}|T_i = 1) - E(Y_{0i}|T_i = 0) &= \\ \underbrace{E(Y_{1i} - Y_{0i}|T_i = 1)}_{ATET} + \underbrace{E(Y_{0i}|T_i = 1) - E(Y_{0i}|T_i = 0)}_{Bias} &= \end{aligned}$$

- In other words, if $Bias \neq 0$ it says that there would have been a difference between untreated units and treated units even if they were untreated (counterfactual).
- This might be seen as some *baseline differences* between two groups.

Omitted Variable



Omitted Variable Bias



Randomised Experiments

- Randomised experiments typically are seen as the “gold standard” in causal identification.
- By assigning units to treatment and control group at random, we make sure that two groups do not differ in their baseline characteristics.
- Recall that we have this decomposition:

$$\underbrace{E(Y_{1i} - Y_{0i} | T_i = 1)}_{ATET} + \underbrace{E(Y_{0i} | T_i = 1) - E(Y_{0i} | T_i = 0)}_{Bias}$$

- Random assignment would eliminate the bias term by making:

$$E(Y_{0i} | T_i = 1) = E(Y_{0i} | T_i = 0)$$

- And, thus, making bias go to 0.
- Experiments and field experiments are becoming more common in political science.
 - E.g. program/policy evaluation, political behaviour (mobilisation/persuasion).

Internal vs External Validity

- **Internal validity** is how much a study captures actual concept that it intends to capture.
- **External validity** is the extent to which conclusions can be generalized beyond a specific study.
- Randomised experiments (aka **randomised control trials (RCT)**) are often conducted in settings (e.g. laboratory) that maximize internal validity at the expense of external validity.

Difference-in- Difference

Example: Minimum Wage and Employment

- In 1992 the state of New Jersey raised the minimum wage from \$4.25 to \$5.05.
- RQ: Did an increase in minimum wage reduce employment?
- Card & Krueger (1994) looked at fast-food restaurants, their employment practices and wages before and after the change.

Source

Card & Krueger (1994)

Cross-section Comparison Design

- Let's select another state, which did not have a commensurate increase in minimum wage, as a comparison.
- Here Pennsylvania, a neighbouring state with, arguably, similar economy can provide such contrast.
- Compare percentages of full-time employers.
- We can then treat:
 - NJ: treatment group
 - PA: control group
- (+) Time-varying confounders are held constant
- (-) State-specific confounders can bias causal inference

Example: Cross-section Comparison Design

```
1 minwage <- read.csv("../data/minwage.csv")
2 minwageNJ <- subset(minwage, subset = (location != "PA"))
3 minwagePA <- subset(minwage, subset = (location == "PA"))
4 t.test(
5   minwageNJ$fullAfter/(minwageNJ$fullAfter + minwageNJ$partAfter),
6   minwagePA$fullAfter/(minwagePA$fullAfter + minwagePA$partAfter)
7 )
```

Welch Two Sample t-test

```
data: minwageNJ$fullAfter/(minwageNJ$fullAfter + minwageNJ$partAfter) and minwagePA$fullAfter/
(minwagePA$fullAfter + minwagePA$partAfter)
t = 1.4322, df = 99.761, p-value = 0.1552
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.01854186  0.11477959
sample estimates:
mean of x mean of y
0.3204010 0.2722821
```

Before-and-After Comparison Design

- But this absence of difference might be just an artifact of state-specific confounders.
- E.g. relatively high pre-minimum wage employment rate in NJ relative to PA
- So, now instead of comparing NJ restaurant to PA let's compare them to themselves pre-change.
- We can then treat:
 - NJ post-change: treatment group
 - NJ pre-change: control group
- (+) State-specific confounders are held constant
- (-) Time-varying confounders can bias causal inference

Example: Before-and-After Comparison Design

```
1 t.test(  
2   minwageNJ$fullAfter/(minwageNJ$fullAfter + minwageNJ$partAfter),  
3   minwageNJ$fullBefore/(minwageNJ$fullBefore + minwageNJ$partBefore)  
4 )
```

Welch Two Sample t-test

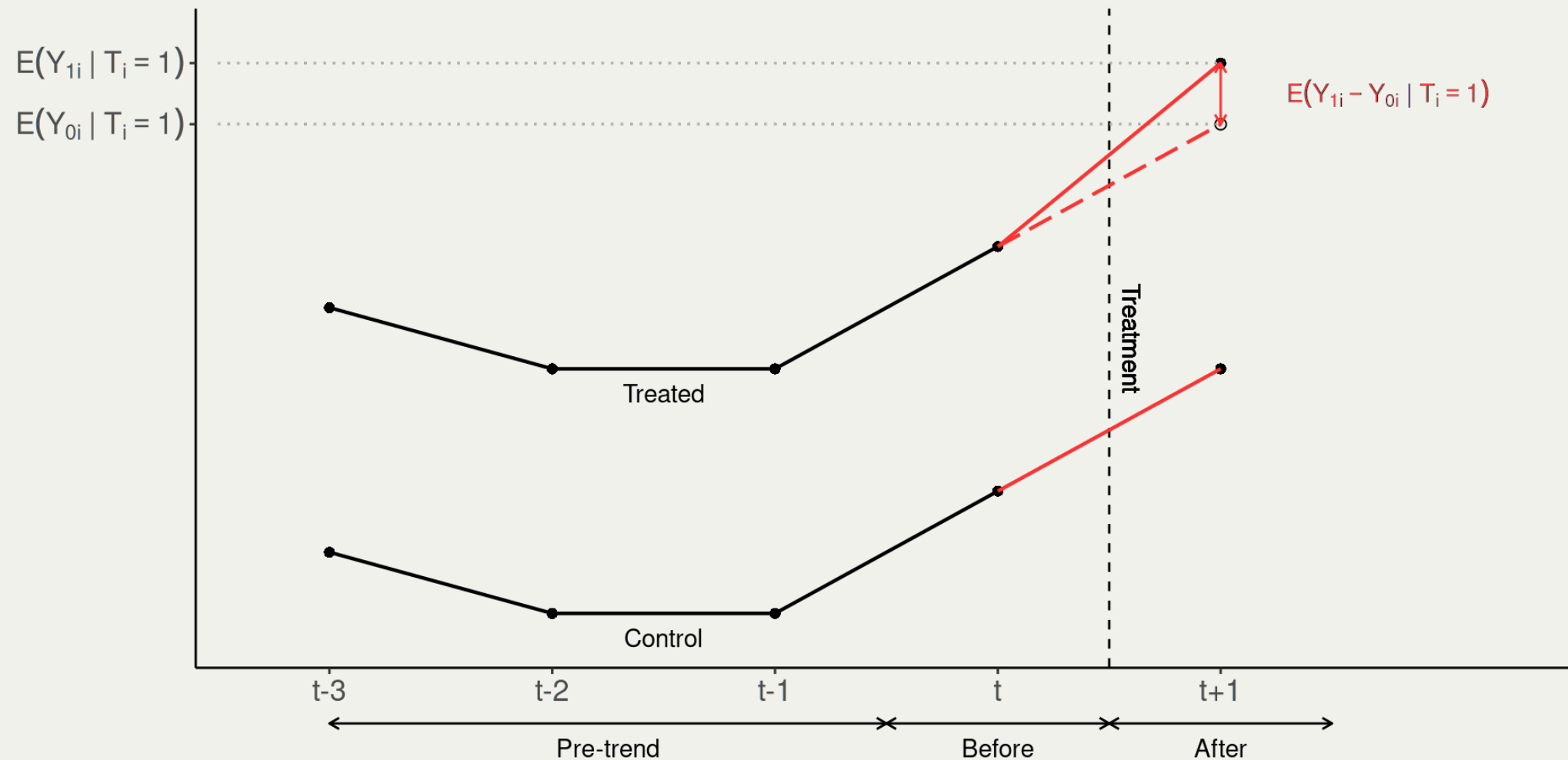
```
data: minwageNJ$fullAfter/(minwageNJ$fullAfter + minwageNJ$partAfter) and minwageNJ$fullBefore/  
(minwageNJ$fullBefore + minwageNJ$partBefore)  
t = 1.1952, df = 575.82, p-value = 0.2325  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
 -0.01535869  0.06310817  
sample estimates:  
mean of x mean of y  
0.3204010 0.2965262
```

Difference-in-difference (DiD) Design

- Difference-in-difference design makes use of both pre-treatment and post-treatment measurements for both treatment and control group.
- We calculate the counterfactual by supposing that our treatment group would have followed the same time trend as control.

$$(E(Y_{1i}|T_i = 1) - E(Y_{0i}|T_i = 1)) - (E(Y_{1i}|T_i = 0) - E(Y_{0i}|T_i = 0))$$

Geometric Interpretation of DiD



Example: Difference-in-difference (DiD) Design

```
1 t.test(  
2   # NJ: FTE After  
3   minwageNJ$fullAfter/(minwageNJ$fullAfter + minwageNJ$partAfter) -  
4   # NJ: FTE Before  
5   minwageNJ$fullBefore/(minwageNJ$fullBefore + minwageNJ$partBefore),  
6   # PA: FTE After  
7   minwagePA$fullAfter/(minwagePA$fullAfter + minwagePA$partAfter) -  
8   # PA: FTE Before  
9   minwagePA$fullBefore/(minwagePA$fullBefore + minwagePA$partBefore)  
10  )
```

Welch Two Sample t-test

```
data: minwageNJ$fullAfter/(minwageNJ$fullAfter + minwageNJ$partAfter) - minwageNJ$fullBefore/  
(minwageNJ$fullBefore + minwageNJ$partBefore) and minwagePA$fullAfter/(minwagePA$fullAfter +  
minwagePA$partAfter) - minwagePA$fullBefore/(minwagePA$fullBefore + minwagePA$partBefore)  
t = 1.3526, df = 90.777, p-value = 0.1796  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
-0.02884997  0.15196659  
sample estimates:  
mean of x    mean of y  
0.02387474 -0.03768357
```

Parallel Trends Assumption

- Parallel trends assumption is critical!
- This assumption is impossible to verify as counterfactual time trend for treatment groups is unobserved.
- However, this can be mitigated by comparing multiple time periods pre-treatment.

Next

- Workshop:
 - RQ Presentations IV
- Research design due:
 - **08:59 Tuesday, 8 April**
- Next week:
 - Logistic Regression