Applying Causal Inference to Study Questions in Educational Policies

Ye Wang

Wilf Family Department of Politics, New York University

10/20/2020

Introduction

- Ye Wang, sixth-year PhD candidate at the politics department, NYU.

Introduction

- Ye Wang, sixth-year PhD candidate at the politics department, NYU.
- Areas of interest: political methodology and authoritarianism.

Introduction

- Ye Wang, sixth-year PhD candidate at the politics department, NYU.
- Areas of interest: political methodology and authoritarianism.
- Causal inference, experimental design, panel data analysis, machine learning. . .

Introduction

- Ye Wang, sixth-year PhD candidate at the politics department, NYU.
- Areas of interest: political methodology and authoritarianism.
- Causal inference, experimental design, panel data analysis, machine learning...
- What do political methodologists do?

Introduction

- Ye Wang, sixth-year PhD candidate at the politics department, NYU.
- Areas of interest: political methodology and authoritarianism.
- Causal inference, experimental design, panel data analysis, machine learning. . .
- What do political methodologists do?
- We develop and introduce statistical tools for political scientists to analyze questions of interest.

Roadmap

- A brief introduction of causal inference and experimental design.
- What is causal inference and why do we need experimental design?
- We work through an educational experiment implemented in Afghanistan.
- It is an example to show the power of randomized controlled trial (RCT).
- We discuss how to design and analyze an experiment.
- We extend the results to general datasets in social sciences.

Social sciences and causality

- As social scientists, we are always interested in causal relationships.

Social sciences and causality

- As social scientists, we are always interested in causal relationships.
- In particular, we want to know what will happen to an outcome of interest, Y, when the value of a factor D changes.

Social sciences and causality

- As social scientists, we are always interested in causal relationships.
- In particular, we want to know what will happen to an outcome of interest, Y, when the value of a factor D changes.
  - Does economic development (D) leads to democratization (Y)?
  - Do political ads (D) change the ideology of voters (Y)?
  - Does college education (D) increase your wage on the labor market (Y)?

Social sciences and causality

- As social scientists, we are always interested in causal relationships.
- In particular, we want to know what will happen to an outcome of interest, Y, when the value of a factor D changes.
  - Does economic development (D) leads to democratization (Y)?
  - Do political ads (D) change the ideology of voters (Y)?
  - Does college education (D) increase your wage on the labor market (Y)?
- Why do we only study the effect of a cause, not the causes of an effect?

Social sciences and causality

- As social scientists, we are always interested in causal relationships.
- In particular, we want to know what will happen to an outcome of interest, Y, when the value of a factor D changes.
  - Does economic development (D) leads to democratization (Y)?
  - Do political ads (D) change the ideology of voters (Y)?
  - Does college education (D) increase your wage on the labor market (Y)?

- Why do we only study the effect of a cause, not the causes of an effect?
- The latter has too many possibilities.

A concrete example

- In Afghanistan, young children, especially young girls, lack the access to formal education.

A concrete example

- In Afghanistan, young children, especially young girls, lack the access to formal education.
- A potential solution is to set up community-based education (CBE) schools.

A concrete example

- In Afghanistan, young children, especially young girls, lack the access to formal education.
- A potential solution is to set up community-based education (CBE) schools.

A concrete example

- But we do not want to waste money...

A concrete example

- But we do not want to waste money. . .
- Do children and their parents trust these schools?

A concrete example

- But we do not want to waste money...
- Do children and their parents trust these schools?
- Will they attend classes when having the opportunity?

A concrete example

- But we do not want to waste money...
- Do children and their parents trust these schools?
- Will they attend classes when having the opportunity?
- Does CBE (D) really improve their capability and performance in exams (Y)?

Correlations and causations

- If we observe that students who attend the CBE schools have higher scores in exams, does it imply that the CBE has generated positive effects?

Correlations and causations

- If we observe that students who attend the CBE schools have higher scores in exams, does it imply that the CBE has generated positive effects?
- Maybe not: correlations do not imply causation.

Correlations and causations

- If we observe that students who attend the CBE schools have higher scores in exams, does it imply that the CBE has generated positive effects?
- Maybe not: correlations do not imply causation.
- Usually there are two potential problems:

Correlations and causations

- If we observe that students who attend the CBE schools have higher scores in exams, does it imply that the CBE has generated positive effects?
- Maybe not: correlations do not imply causation.
- Usually there are two potential problems:
  - Omitted variables: it is the CBE or their own acumen?
  - Reverse causality: are better students more likely to attend the CBE?

Correlations and causations

- If we observe that students who attend the CBE schools have higher scores in exams, does it imply that the CBE has generated positive effects?
- Maybe not: correlations do not imply causation.
- Usually there are two potential problems:
  - Omitted variables: it is the CBE or their own acumen?
  - Reverse causality: are better students more likely to attend the CBE?
- Causal identification is usually hard!
- That's why we have an academic field for it: causal inference.

The fundamental problem of causal inference

- Ideally, causal relationships are identified by using a time machine.

The fundamental problem of causal inference

- Ideally, causal relationships are identified by using a time machine.
- You travel back in time, setting D to a different value and observing how Y changes accordingly.

The fundamental problem of causal inference

- Ideally, causal relationships are identified by using a time machine.
- You travel back in time, setting D to a different value and observing how Y changes accordingly.
- Suppose the *treatment* D takes two values, 0 and 1.
- The corresponding values of Y are denoted as $Y(0)$ and $Y(1)$.

The fundamental problem of causal inference

- Ideally, causal relationships are identified by using a time machine.
- You travel back in time, setting D to a different value and observing how Y changes accordingly.
- Suppose the *treatment* D takes two values, 0 and 1.
- The corresponding values of Y are denoted as $Y(0)$ and $Y(1)$.
- The causal effect of changing D from 0 to 1 equals $\tau = Y(1) - Y(0)$.

The fundamental problem of causal inference

- Ideally, causal relationships are identified by using a time machine.
- You travel back in time, setting D to a different value and observing how Y changes accordingly.
- Suppose the *treatment* D takes two values, 0 and 1.
- The corresponding values of Y are denoted as $Y(0)$ and $Y(1)$.
- The causal effect of changing D from 0 to 1 equals $\tau = Y(1) - Y(0)$.
- τ is the parameter of interest, the *treatment effect*.
- $Y(0)$ and $Y(1)$ are denoted as *potential outcomes* of Y.

The fundamental problem of causal inference

- If we observe both $Y(0)$ and $Y(1)$, then we can infer the value of τ easily.

The fundamental problem of causal inference

- If we observe both $Y(0)$ and $Y(1)$, then we can infer the value of τ easily.
- But we don't have a time machine!

The fundamental problem of causal inference

- If we observe both $Y(0)$ and $Y(1)$, then we can infer the value of τ easily.
- But we don't have a time machine!
- In reality, we observe either $Y(0)$ or $Y(1)$, but never both of them.

The fundamental problem of causal inference

- If we observe both $Y(0)$ and $Y(1)$, then we can infer the value of τ easily.
- But we don't have a time machine!
- In reality, we observe either $Y(0)$ or $Y(1)$, but never both of them.
- This is called "the fundamental problem of causal inference" (Holland, 1986)
- The unobserved potential outcome is called the "counterfactual."

The fundamental problem of causal inference

"*Two roads diverged in a wood, and I—*
I took the one less traveled by,
And that has made all the difference."
— *The road not taken*, Robert Frost

The Rubin model

- The framework we have introduced to describe causal relationships is called the Rubin model in statistics.
- It was first proposed by the Harvard statistician Donald Rubin in the 70s.

The Rubin model

- For each individual i, we observe an outcome Y_i and a treatment D_i.
- Suppose D_i is binary, then,

$$Y_i = \begin{cases} Y_i(1) \text{ if } D_i = 1, \\ Y_i(0) \text{ if } D_i = 0. \end{cases}$$

- $\tau_i = Y_i(1) - Y_i(0)$ is the treatment effect for individual i, or the "individualistic treatment effect."
- The average of τ_i, $\frac{1}{N}\sum_{i=1}^{N} \tau_i$, is called the average treatment effect (ATE).

The Rubin model

- For each individual i, we observe an outcome Y_i and a treatment D_i.
- Suppose D_i is binary, then,

$$Y_i = \begin{cases} Y_i(1) \text{ if } D_i = 1, \\ Y_i(0) \text{ if } D_i = 0. \end{cases}$$

- $\tau_i = Y_i(1) - Y_i(0)$ is the treatment effect for individual i, or the "individualistic treatment effect."
- The average of τ_i, $\frac{1}{N}\sum_{i=1}^{N} \tau_i$, is called the average treatment effect (ATE).
- How to interpret the ATE?

Two solutions

- The light goes on after I turn it on— is it causation?

Two solutions

- The light goes on after I turn it on— is it causation?
- Yes, under the assumption of temporal invariance:

Two solutions

- The light goes on after I turn it on— is it causation?
- Yes, under the assumption of temporal invariance:
- Denote the time when I turn on the light as t and the time when it is on as $t + 1$.
- We observe $Y_{t+1}(1)$ (on) and $Y_t(0)$ (off) but not $Y_{t+1}(0)$.

Two solutions

- The light goes on after I turn it on— is it causation?
- Yes, under the assumption of temporal invariance:
- Denote the time when I turn on the light as t and the time when it is on as $t + 1$.
- We observe $Y_{t+1}(1)$ (on) and $Y_t(0)$ (off) but not $Y_{t+1}(0)$.
- The difference between $Y_{t+1}(1)$ and $Y_t(0)$ is causal only when $Y_t(0) = Y_{t+1}(0)$.
- When will this assumption be violated?

Two solutions

- The light goes on after I turn it on— is it causation?
- Yes, under the assumption of temporal invariance:
- Denote the time when I turn on the light as t and the time when it is on as $t + 1$.
- We observe $Y_{t+1}(1)$ (on) and $Y_t(0)$ (off) but not $Y_{t+1}(0)$.
- The difference between $Y_{t+1}(1)$ and $Y_t(0)$ is causal only when $Y_t(0) = Y_{t+1}(0)$.
- When will this assumption be violated?
- Causal inference relying on this assumption is called the "scientific solution."

Two solutions

- The assumption of temporal invariance is hard to satisfy in social sciences.
- If we find that the average test score becomes highers after students attend the CBE schools, can we claim any causation?

Two solutions

- The assumption of temporal invariance is hard to satisfy in social sciences.
- If we find that the average test score becomes highers after students attend the CBE schools, can we claim any causation?
- Now D is attending the CBE schools and Y is test score.
- The assumption implies that the average test score before they attend the schools is the same as the average test score if they do not attend the schools.

Two solutions

- The assumption of temporal invariance is hard to satisfy in social sciences.
- If we find that the average test score becomes highers after students attend the CBE schools, can we claim any causation?
- Now D is attending the CBE schools and Y is test score.
- The assumption implies that the average test score before they attend the schools is the same as the average test score if they do not attend the schools.
- Most outcomes in social sciences change with time.
- Hence, we need a different solution.

Two solutions

- Suppose we can find two groups of students who are on average the same in all dimensions except for whether having attended the CBE schools.

Two solutions

- Suppose we can find two groups of students who are on average the same in all dimensions except for whether having attended the CBE schools.
- Now, the difference in average test scores between the two groups can be solely attributed to the influence of the schools.

Two solutions

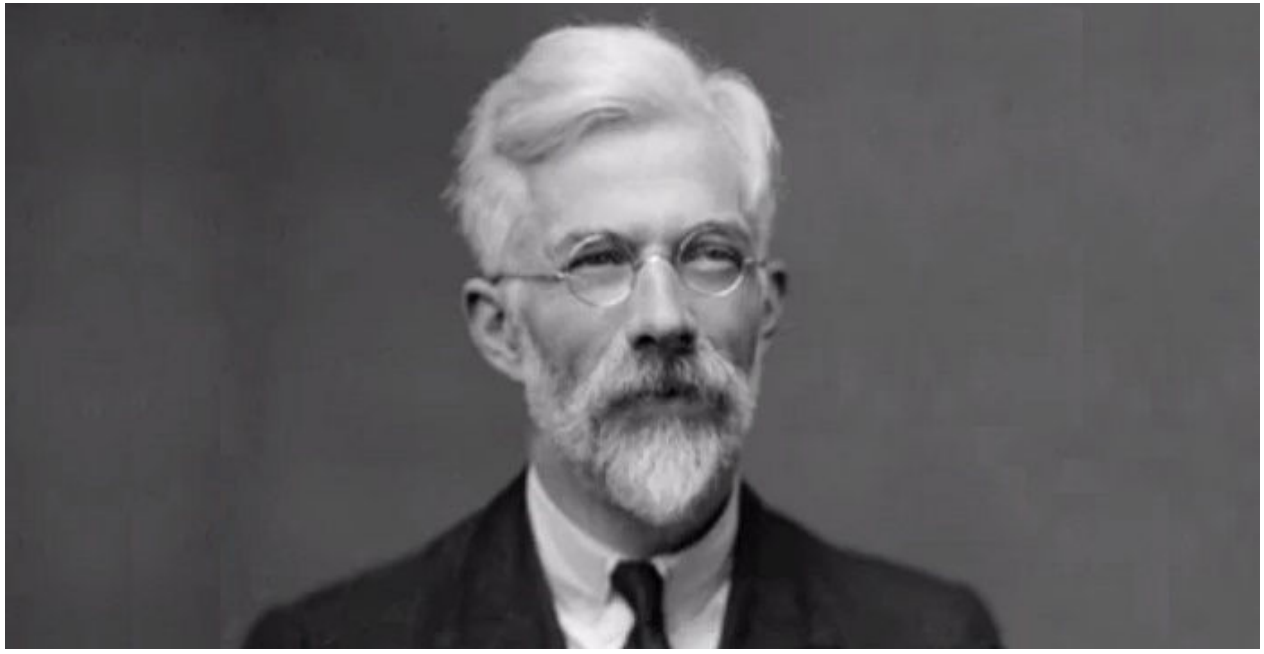
- Suppose we can find two groups of students who are on average the same in all dimensions except for whether having attended the CBE schools.
- Now, the difference in average test scores between the two groups can be solely attributed to the influence of the schools.
- This is what Holland calls a "statistical solution."
- It is the most common solution in social sciences.

The golden standard: randomized controlled trial

- Yet it is hard to find two groups that are identical in all the dimensions.

The golden standard: randomized controlled trial

- Yet it is hard to find two groups that are identical in all the dimensions.
- Ronald Fisher: randomization can fix the problem.



The golden standard: randomized controlled trial

- Suppose there are N students.
- Under randomization, whether D_i equals 0 or 1 is uncorrelated with the characteristics of student i.

The golden standard: randomized controlled trial

- Suppose there are N students.
- Under randomization, whether D_i equals 0 or 1 is uncorrelated with the characteristics of student i.
- In expectation, the two groups (with $D_i = 0$ or $D_i = 1$) are identical in all the dimensions.
- The group with $D_i = 1$ is called the treatment group and the one with $D_i = 0$ is called the control group.

Design randomized controlled trials

- In practice, though, we may have "bad draws."
- The two groups may look quite different under a particular draw.

Design randomized controlled trials

- In practice, though, we may have "bad draws."
- The two groups may look quite different under a particular draw.
- If we are able to re-randomize, do it.
- Otherwise, we can divide subjects into homogeneous "blocks" first and randomize within each block.

Design randomized controlled trials

- In practice, though, we may have "bad draws."
- The two groups may look quite different under a particular draw.
- If we are able to re-randomize, do it.
- Otherwise, we can divide subjects into homogeneous "blocks" first and randomize within each block.
- Blocking experiments are usually more efficient and popular.

Design randomized controlled trials

- Sometimes randomization at individual level is unfeasible/impractical.
- We can also randomize at a higher level (class/village).
- This is called a clustering experiment.
- Subjects in the same cluster will receive the same treatment.

Design randomized controlled trials

- Sometimes randomization at individual level is unfeasible/impractical.
- We can also randomize at a higher level (class/village).
- This is called a clustering experiment.
- Subjects in the same cluster will receive the same treatment.
- There are more complicated designs (dynamic experiment, network experiment, etc.).
- By "design," we mean how the treatment is assigned to subjects.

Estimate the causal effect from randomized controlled trials

- After obtaining results from the experiment, we can estimate the ATE by comparing the difference in means between the two groups:

$$\widehat{ATE} = \frac{1}{N_{tr}} \sum_{i=1}^{N} D_i Y_i - \frac{1}{N_c} \sum_{i=1}^{N} (1 - D_i) Y_i$$

Estimate the causal effect from randomized controlled trials

- After obtaining results from the experiment, we can estimate the ATE by comparing the difference in means between the two groups:

$$\widehat{ATE} = \frac{1}{N_{tr}} \sum_{i=1}^{N} D_i Y_i - \frac{1}{N_c} \sum_{i=1}^{N} (1 - D_i) Y_i$$

- In statistics, randomization guarantees that the treatment D_i is independent to the potential outcomes $Y_i(0)$ and $Y_i(1)$.
- By independence, we can further prove that $\widehat{ATE}$ is 1. unbiased and 2. consistent.

Estimate the causal effect from randomized controlled trials

- Unbiasedness means that if we run the same experiment for many times, the average of the estimates will be very close to the true ATE.
- Consistency means that if we have a very large sample, then the average of the estimates will be very close to the true ATE.

Estimate the causal effect from randomized controlled trials

- Unbiasedness means that if we run the same experiment for many times, the average of the estimates will be very close to the true ATE.
- Consistency means that if we have a very large sample, then the average of the estimates will be very close to the true ATE.
- As we only run the experiment once, getting a larger sample always helps.

Pros and cons of RCT

- Estimates from RCTs are unbiased and consistent for the true ATE.
- RCT has the highest internal validity.
- It is the foundation of natural sciences.
- It has become increasingly popular in social sciences.

Pros and cons of RCT

- It is often more expensive.
- Many problems cannot be studies by RCT.
  - Unrealistic
  - Unethical
- The external validity of RCT is not always very high.
- Do RCTs prevent people from getting the treatment?

The Afghanistan experiment: design

- We ran two experiments in different areas of Afghanistan, one in 2008 and one in 2015.
- Why twice?

The Afghanistan experiment: design

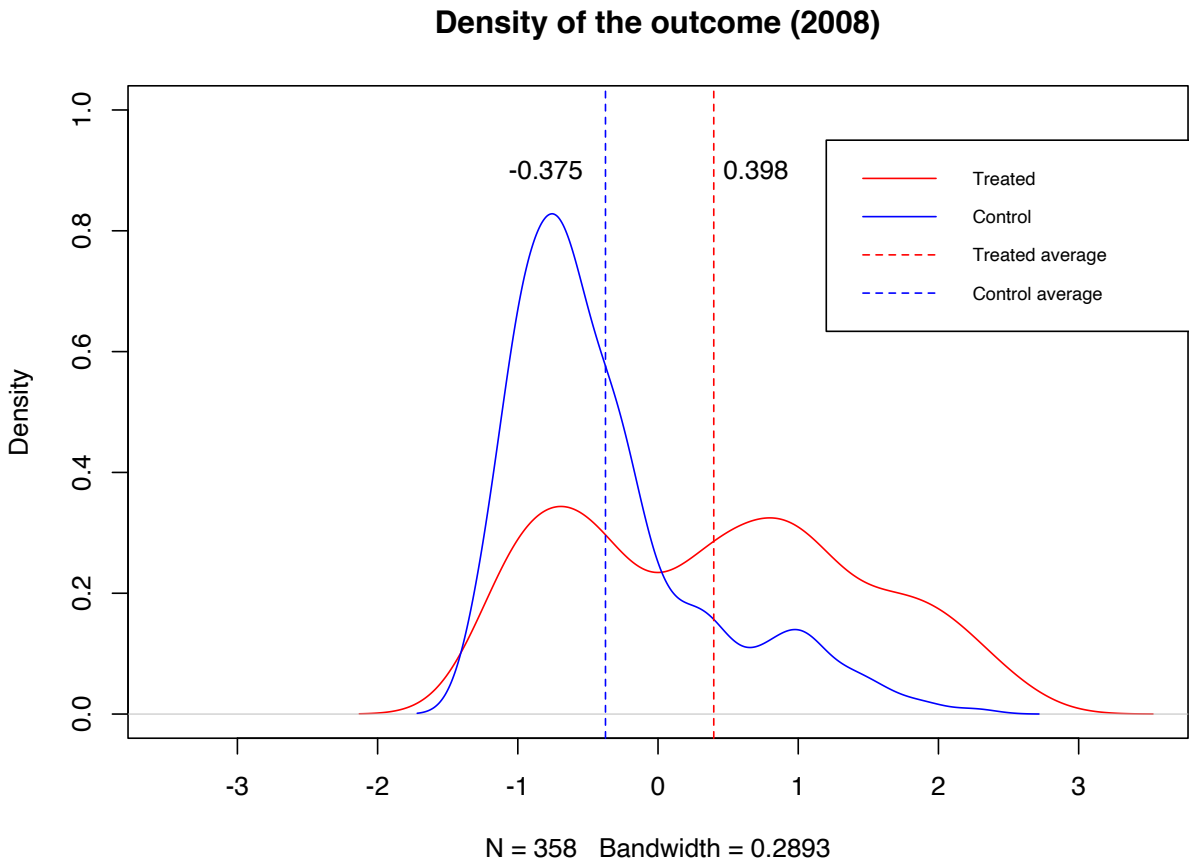
- We ran two experiments in different areas of Afghanistan, one in 2008 and one in 2015.
- Why twice?
- We first choose five districts in Afghanistan.
- Within each district, we provide young girls in some randomly selected villages the access to CBE schools.

The Afghanistan experiment: design

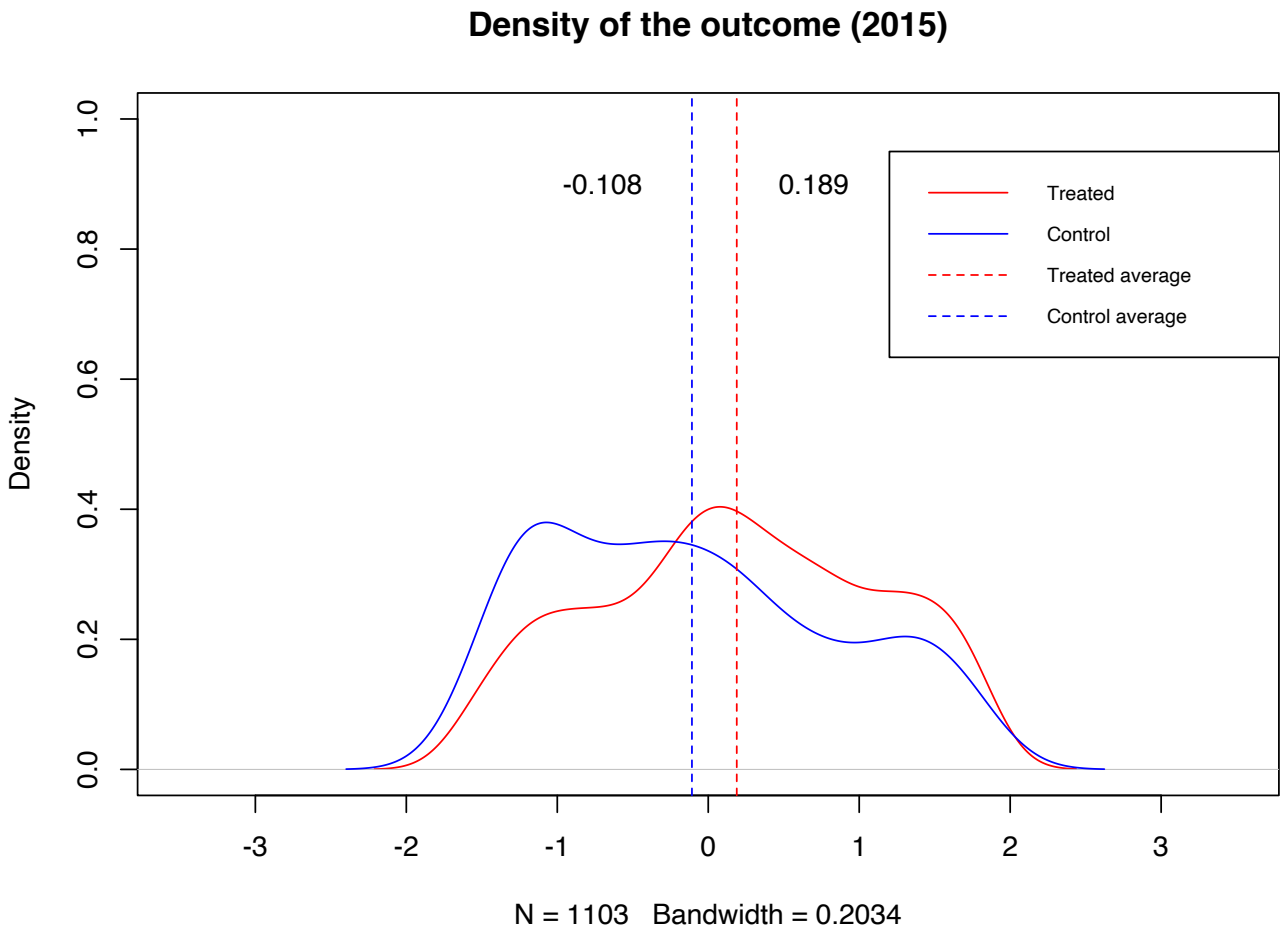
- We ran two experiments in different areas of Afghanistan, one in 2008 and one in 2015.
- Why twice?
- We first choose five districts in Afghanistan.
- Within each district, we provide young girls in some randomly selected villages the access to CBE schools.
- Each district is a block and each village is a cluster.

The Afghanistan experiment: result

- We have 358 treated students and 331 students under control in the 2008 experiment.
- In the 2015 experiment, the numbers are 909 and 309, respectively.



The Afghanistan experiment: result



The Afghanistan experiment: result

	t=1 (2008)	t=2 (2015)
ITT Effect	0.73	0.35
	(0.11)	(0.13)
Control mean	-0.36	-0.17
	(0.09)	(0.12)
N	689	1218

Table 1: Least squares regression estimates of the intention-to-treat (ITT) effects and control group means. Outcome is combined math-verbal test score, standardized. Standard errors accounting for village-level clustering in parentheses.

- Why is the effect much smaller in 2015?

Non-compliance in an experiment

- Is it the treatment effect we want?

Non-compliance in an experiment

- Is it the treatment effect we want?
- Yes and no.
- A severe problem in this experiment is non-compliance.
- Students with the access may not attend the schools, while students without the access may attend.
- They do not comply with the treatment assignment.

Non-compliance in an experiment

- We need to distinguish treatment assignment (Z) and treatment exposure (D).
- ATE is the effect caused by D rather than Z.

Non-compliance in an experiment

- We need to distinguish treatment assignment (Z) and treatment exposure (D).
- ATE is the effect caused by D rather than Z.
- Now the difference between the treatment group and the control group is no longer the ATE.
- It is called the intention to treat (ITT) effect.

Principal strata

- How to estimate the ATE in this case?
- We need a concept called principal strata.
- Let's take a closer look at non-compliance.

Principal strata

- There are four types of individuals in the experiment.
  - Always-taker: attend the CBE no matter the value of Z
  - Never-taker: do not attend the CBE no matter the value of Z
  - Complier: attend the CBE only when $Z = 1$
  - Defier: attend the CBE only when $Z = 0$

Principal strata

- There are four types of individuals in the experiment.
  - Always-taker: attend the CBE no matter the value of Z
  - Never-taker: do not attend the CBE no matter the value of Z
  - Complier: attend the CBE only when $Z = 1$
  - Defier: attend the CBE only when $Z = 0$
- Those types are called "Principal strata."
- Usually we assume that there is no defier.

Estimate the ATE with non-compliance

- The ATE equals to the effect on the compliers.
- We do not know who is a complier.
- D can also be written in the form of potential outcomes.

$$D_i = \begin{cases} D_i(1) \text{ if } Z_i = 1, \\ D_i(0) \text{ if } Z_i = 0. \end{cases}$$

Estimate the ATE with non-compliance

- Always-taker: $D_i(1) = 1$ and $D_i(0) = 1$.
- Never-taker: $D_i(1) = 0$ and $D_i(0) = 0$.
- Complier: $D_i(1) = 1$ and $D_i(0) = 0$.
- Defier: $D_i(1) = 0$ and $D_i(0) = 1$.

Estimate the ATE with non-compliance

- We don't know the value of D when Z becomes different.
- Again, the fundamental problem of causal inference.

Estimate the ATE with non-compliance

- We don't know the value of D when Z becomes different.
- Again, the fundamental problem of causal inference.
- But we can infer the proportion of compliers.

Estimate the ATE with non-compliance

- When $D = 1$ and $Z = 1$: always-takers + compliers.
- When $D = 1$ and $Z = 0$: always-takers.
- When $D = 0$ and $Z = 1$: never-takers.
- When $D = 0$ and $Z = 0$: never-takers + compliers.
- Randomization means the proportion of each principal strata should be similar in the treated group and the control group.

Estimate the ATE with non-compliance

- Remember that in the 2015 experiment, there are 909 treated subjects and 309 subjects under control.
- $D = 1$ and $Z = 1$: $0.54 =$ always-takers $+$ compliers.
- When $D = 1$ and $Z = 0$: $0.17 =$ always-takers.
- When $D = 0$ and $Z = 1$: $0.46 =$ never-takers.
- When $D = 0$ and $Z = 0$: $0.83 =$ never-takers $+$ compliers.

Estimate the ATE with non-compliance

- Always-takers: 0.17, never-takers: 0.46, compliers: 0.37.
- The ATE estimate equals $0.35/0.37 = 0.946$.
- Similarly, in the 2008 experiment, we have:
- Always-takers: 0, never-takers: 0.31, compliers: 0.38.
- The ATE estimate equals $0.73/0.69 = 1.058$.

Estimate the ATE with non-compliance

- The difference is much smaller!
- We further divide compliers into two strata: true-compliers and substitutors.
- The latter are those who transfer from public schools.
- The effect is smaller for them and but their proportion is much higher in 2015.

Design-based perspective

- It is easy to identify causal relationships via experiments.
- But many questions cannot be answered by running RCT.

Design-based perspective

- It is easy to identify causal relationships via experiments.
- But many questions cannot be answered by running RCT.
- How can we establish causality in observational studies?

Design-based perspective

- It is easy to identify causal relationships via experiments.
- But many questions cannot be answered by running RCT.
- How can we establish causality in observational studies?
- We should take the "design-based perspective."

Design-based perspective

- Instead of assuming a correct model for the outcome, we focus on how the treatment is assigned.
- In other words, we should infer the hypothetical experiment that generates the data at hand.
- These hypothetical experiments are called "natural experiments."
- If we want to claim causality, we must have an experiment conducted by either researchers or Mother Nature.
- Our job in observational studies is to find how the experiment is implemented by Mother Nature using our substantive knowledge.

Design-based perspective

- Let's go back to the relationship between college education and wage.
- We should think about how the admission into colleges is decided and when it could be seen as a randomized assignment.

Design-based perspective

- Let's go back to the relationship between college education and wage.
- We should think about how the admission into colleges is decided and when it could be seen as a randomized assignment.
- For example, in some countries, it is decided by a threshold in the test score.
- We can compare those who are just above the threshold with those who are just below it.

Design-based perspective

- Let's go back to the relationship between college education and wage.
- We should think about how the admission into colleges is decided and when it could be seen as a randomized assignment.
- For example, in some countries, it is decided by a threshold in the test score.
- We can compare those who are just above the threshold with those who are just below it.
- Or we can interview the admission committee to see whether there is any randomization in the process.
- How do they choose between two candidates who are similar in all the aspects they care about?

Design-based perspective

- Causal inference is not magic.
- It helps you find experiments in your own field and analyze it in rigorous ways.
- But it requires your deep understanding of the subject to find an experiment.
- Always keep your theory in mind and talk to your subjects!