Section: Module 1

Potential Outcome

Kentaro Nakamura

GOV 2003

September 13th, 2026

Logistics

- Section Time: Friday 1:00pm-2:00pm
- Office Hour: Thursday 4:30pm-6:00pm

- Problem Set 1 is due September 14th 10am
  - We use the highest score **8** Problems out of 10

Simple Example for Lecture Review

- School i either gives tablets ($T_i = 1$) or does not ($T_i = 0$).
- Define $Y_i(0)$ and $Y_i(1)$ as its test score under each decision.
- Suppose, purely for illustration, that both potential outcomes were visible:

i	$Y_i(0)$	$Y_i(1)$	T_i	Y_i	$Y_i(1) - Y_i(0)$
1	500	450	0	500	-50
2	600	600	0	600	0
3	800	600	1	600	-200
4	700	750	1	750	50

- We can calculate ATE as the average of the individual effects:

$$\text{ATE} = \frac{1}{4}\sum_{i=1}^{4}\{Y_i(1) - Y_i(0)\} = \frac{-50 + 0 - 200 + 50}{4} = -50.$$

The Fundamental Problem of Causal Inference

What the data contain

i	$Y_i(0)$	$Y_i(1)$	T_i	Y_i
1	500	?	0	500
2	600	?	0	600
3	?	600	1	600
4	?	750	1	750

Observed difference in means

$$\widehat{\Delta}_{\text{obs}} = \frac{600 + 750}{2} - \frac{500 + 600}{2} = 125.$$

In the illustrative *complete* potential-outcome schedule,

$$\text{ATE} = -50.$$

Why can the signs reverse?

We observe only $Y_i = Y_i(T_i)$. Tablet and non-tablet schools may differ systematically, so selection bias can make the observed association positive even when the causal effect is negative.

Potential Outcomes and Consistency

- **Consistency** links the potential-outcome framework to the observed data.

$$Y_i = Y_i(T_i) = T_i Y_i(1) + (1 - T_i) Y_i(0).$$

  - Example of Violation: Interference between units (e.g., spillover effects)

- **Unit-level determinism**:

$$\underbrace{i,\ Y_i(0),\ Y_i(1),\ X_i}_{\text{fixed before treatment}} \quad \longrightarrow \quad \underbrace{T_i}_{\text{assign treatment}} \quad \longrightarrow \quad \underbrace{Y_i = Y_i(T_i)}_{\text{observe outcome}}$$

  - Treatment is only source of randomness
  - potential outcomes and covariates can be random due to sampling

Principal Strata

Classify binary response types

$Y_i(0)$	$Y_i(1)$	Principal stratum
0	0	Never-survivor
0	1	Responder
1	0	Harmed
1	1	Always-survivor

Membership is defined by potential outcomes, so it does not change with the realized treatment assignment.

One observation leaves two possibilities

T_i	Y_i	Possible strata
1	1	Responder / Always
1	0	Harmed / Never
0	1	Always / Harmed
0	0	Never / Responder

The missing potential outcome prevents us from learning a person's stratum from (T_i, Y_i) alone.

Takeaway

The individual effect $\tau_i = Y_i(1) - Y_i(0)$ is not identified from one observation per unit. Causal inference therefore targets average effects under assumptions that make treated and control units comparable.

General tips to solve question

- Understand the setting / notation
- Understand the key statistical tools
  - **Key Concepts**: conditional probability, statistical independence
- Apply these tools
  - If you struggle, always check lectures!

Review Question: Setting

- A judge decides whether to **detain** an arrestee while considering the risk of a new crime.
- Variables:
  - $D_i = 1$ if detained; $D_i = 0$ if released
  - $Y_i = 1$ if a crime is committed; $Y_i = 0$ otherwise
  - $X_i = 1$ for Group A; $X_i = 0$ for Group B

	Group A ($X_i = 1$)		Group B ($X_i = 0$)	
	Detained	Released	Detained	Released
Crime ($Y_i = 1$)	150	100	100	100
No crime ($Y_i = 0$)	100	150	120	180

Goal

Does the decision rule treat the groups fairly—and what does "fairly" mean when detention itself can change the outcome?

Review Question 1: Observational Fairness

- **Accuracy (equalized odds):**

$$D_i \perp\!\!\!\perp X_i \mid Y_i$$

  Among people with the same observed outcome, is the detention rate the same across groups?

- **Calibration (sufficiency):**

$$Y_i \perp\!\!\!\perp X_i \mid D_i$$

  Among people receiving the same decision, is the observed crime rate the same across groups?

Task

Evaluate each conditional independence using the observed table.

Question 1 Tool: Conditional Independence

- For binary X_i, $A_i \perp\!\!\!\perp X_i \mid B_i$ means comparing groups **within every value of** B_i:

$$\Pr(A_i = 1 \mid X_i = 1, B_i = b) = \Pr(A_i = 1 \mid X_i = 0, B_i = b) \quad \text{for every } b.$$

How to calculate conditional probabilities from a table

$$\Pr(A_i = 1 \mid X_i = x, B_i = b) = \frac{\#\{A_i = 1, X_i = x, B_i = b\}}{\#\{X_i = x, B_i = b\}}.$$

Hold fixed the variables to the right of the conditioning bar, then divide the target cell by the corresponding subtotal.

- Accuracy: compare detention rates within $Y_i = 1$ and within $Y_i = 0$.
- Calibration: compare crime rates within $D_i = 1$ and within $D_i = 0$.

Question 1 (Answer): Accuracy

	$X_i = 1$ (Group A)	$X_i = 0$ (Group B)
$Y_i = 1$	$\frac{150}{150+100} = .60$	$\frac{100}{100+100} = .50$
$Y_i = 0$	$\frac{100}{100+150} = .40$	$\frac{120}{120+180} = .40$

Conclusion: accuracy does not hold

Among those who committed a crime, Group A has a 10 percentage-point higher detention rate than Group B (.60 versus .50).

- For those who did not commit a crime, the detention rates are equal.
- The full conditional-independence criterion fails if equality fails at any outcome level.

Question 1 (Answer): Calibration

	$X_i = 1$ (Group A)	$X_i = 0$ (Group B)
$D_i = 1$	$\frac{150}{150+100} = .60$	$\frac{100}{100+120} \approx .45$
$D_i = 0$	$\frac{100}{100+150} = .40$	$\frac{100}{100+180} \approx .36$

Conclusion: calibration does not hold

Within both decisions, observed crime rates differ across the groups.

- Among detained arrestees: .60 for Group A versus .45 for Group B.
- Among released arrestees: .40 for Group A versus .36 for Group B.

Review Question 2: Potential Outcomes

- Let $Y_i(d)$ be whether arrestee i would commit a crime under decision d. Each person belongs to one **principal stratum**.

Stratum	$Y_i(1)$	$Y_i(0)$	Group A $D_i = 1$	Group A $D_i = 0$	Group B $D_i = 1$	Group B $D_i = 0$
Dangerous	1	1	120	30	80	20
Backlash	1	0	30	30	20	20
Preventable	0	1	70	70	80	80
Safe	0	0	30	120	40	160

- Evaluate **principal accuracy** and **principal calibration**:

$$D_i \perp\!\!\!\perp X_i \mid \{Y_i(1), Y_i(0)\}, \qquad \{Y_i(1), Y_i(0)\} \perp\!\!\!\perp X_i \mid D_i.$$

Question 2 Tool: Principal Stratification

1. Translate the types

	$Y_i(1)$	$Y_i(0)$
Dangerous	1	1
Backlash	1	0
Preventable	0	1
Safe	0	0

2. Match the criterion to a denominator

- Principal accuracy: within each **stratum**, compare detention rates across X_i.
- Principal calibration: within each **decision**, compare the distribution of strata across X_i.

Connection to observed outcomes

$$Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0).$$

Only one potential outcome is observed, so principal strata are generally latent even though the exercise supplies them.

Question 2 (Answer): Principal Accuracy

- Within each stratum, calculate $\Pr(D_i = 1 \mid X_i = x, \text{stratum})$.

	Group A	Group B
Dangerous	$120/(120+30) = .80$	$80/(80+20) = .80$
Backlash	$30/(30+30) = .50$	$20/(20+20) = .50$
Preventable	$70/(70+70) = .50$	$80/(80+80) = .50$
Safe	$30/(30+120) = .20$	$40/(40+160) = .20$

Conclusion: principal accuracy holds

For people who would respond to detention in the same way, the probability of detention is identical across Groups A and B.

Question 2 (Answer): Principal Calibration

- Within each decision, compare the **entire distribution** of strata.

Decision	Group	(Dangerous, Backlash, Preventable, Safe)
Detained	A	(.48, .12, .28, .12)
	B	(.36, .09, .36, .18)
Released	A	(.12, .12, .28, .48)
	B	(.07, .07, .29, .57)

Conclusion: principal calibration does not hold

For the same detention decision, the distribution of causal types differs between Groups A and B.

Why Potential Outcomes Change the Question

- Observed $Y_i = 1$ combines different causal types:
  - among detained: Dangerous and Backlash;
  - among released: Dangerous and Preventable.
- A policy objective is naturally stated using **counterfactuals**:
  - detain Preventable arrestees;
  - release Backlash and Safe arrestees;
  - how to treat Dangerous arrestees depends on the costs involved.
- Therefore, observational fairness and principal fairness can give different answers.

Identification challenge

We never observe both $Y_i(1)$ and $Y_i(0)$ for the same person. The observed table alone does not identify the principal-stratum table without additional assumptions or information.