PLSC 30600

Week 1: Course orientation, potential outcomes, missing data, bounds

Molly Offer-Westort

Department of Political Science,
University of Chicago

Winter 2026

Overview

This course:

- Objectives.
- Course structure and meetings.
- Coding and Claude Code.
- Assignments.

Link to course repo.

Where we're going

Definition 6.0.1. *Point Identification*
Suppose we have complete knowledge of the joint CDF, F, of some observable random vector. Then a statistical functional of some unobservable random vector with joint CDF F_u, $T(F_u)$, is *point identified* by a set of assumptions A_u if there exists one and only one value of $T(F_u)$ that is logically compatible with F given A_u.[1]

1 The notation F_u and A_u will not be used again.

Aronow and Miller (2019)

Sampling setup

- We treat observations as draws from a joint generating process.
- Observations are i.i.d.: *independent and identically distributed*.
- We index realized values with i—the generating distribution is the same for each unit, the realized values may not be.
- Intuition: a large population, we randomly draw individuals one at a time.

Manski and Nagin (1998) Utah juvenile court data

- How should judges sentence convicted juvenile offenders?
- Juvenile offenders in Utah may be assigned to residential or non-residential treatment programs.

Science table: observed outcomes only

i	D_i	Y_i
1	0	0
2	1	1
3	0	1
4	1	1
5	0	0
6	1	1
7	0	1
8	1	0
9	0	0
10	1	1

Treatment (sentencing): $D = 1$ residential, $D = 0$ non-residential

Outcome (recidivism): $Y = 1$ recidivates, $Y = 0$ does not recidivate

Possible quantities of interest

- What would recidivism rates be if all juveniles were assigned to non-residential treatment?
- What would recidivism rates be if all juveniles were assigned to residential treatment?
- (maybe) What is the average difference in recidivism rates if all residents were given residential vs. non-residential treatment?

Notation: missing data

- $\mathrm{Supp}\,[R] = \{0, 1\}$; $R_i = 1$ if Y_i is observed, 0 otherwise.

$$Y_i^* = \begin{cases} -99 & : R_i = 0 \\ Y_i & : R_i = 1 \end{cases}$$

  equivalent to the *switching equation*:

$$= Y_i R_i + (-99)(1 - R_i).$$

Notation: potential outcomes

- $\mathrm{Supp}\,[D] = \{0, 1\}$; $D_i = 1$ if Y_i is treated, 0 otherwise.

$$Y_i = \begin{cases} Y_i(0) & : D_i = 0 \\ Y_i(1) & : D_i = 1 \end{cases}$$

  equivalent to the *switching equation*:

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

Potential outcomes

- What assumptions are embedded in the *stable outcome model* and *stable unit treatment value* assumptions?

- These are *logical/structural assumptions.*

SUTVA: stable unit treatment value assumption

- Two core pieces: Rubin (1974)
  - No interference: one unit's treatment does not affect another unit's outcome.
  - No hidden versions of treatment: each treatment level is well-defined.
- These assumptions let us write $Y_i(d)$ without ambiguity.

Science table: missing data view

What would recidivism rates be if all juveniles were assigned to non-residential treatment?

Target: $\mathbb{E}[Y_i(0)]$

i	$Y_i(0)$	R_i	$Y_i^*(0)$
1	0	1	0
2	?	0	-99
3	1	1	1
4	?	0	-99
5	0	1	0
6	?	0	-99
7	1	1	1
8	?	0	-99
9	1	1	1
10	?	0	-99

$$Y_i^*(0) = Y_i(0)R_i + (-99)(1 - R_i)$$

Missing data as the identification problem

- We observe (Y_i^*, R_i), not Y_i.[†]

- Without assumptions on (Y_i, R_i), $\mathrm{E}[Y_i]$ is not identified.

- A parameter $\theta(P)$ is identified if all data-generating processes consistent with the observed data imply the same value of θ.

† Note target.

Bounds under minimal assumptions

- If $Y_i \subseteq [a, b]$, we can bound $\mathrm{E}[Y_i]$.
- Lower bound: set missing values to a.
- Upper bound: set missing values to b.
- Tighter bounds require stronger assumptions.

Sharp bounds for the expected value

- Let Y_i and R_i be random variables with $\mathrm{Supp}\,[Y_i] \subseteq [a, b]$ and $\mathrm{Supp}\,[R_i] = \{0, 1\}$.
- Let $Y_i^* = Y_i R_i + (-99)(1 - R_i)$.
- Then:

$$\mathrm{E}\,[Y_i] \in \Big\{ \mathrm{E}\,[Y_i^* \mid R_i = 1] \Pr[R_i = 1] + a \Pr[R_i = 0], \\ \big(\mathrm{E}\,[Y_i^* \mid R_i = 1] \Pr[R_i = 1] + b \Pr[R_i = 0] \big) \Big\}.$$

Estimating sharp bounds for the expected value

- Given n i.i.d. observations of (Y_i^*, R_i):
- the plug-in estimator for the upper bound is:

$$\frac{1}{n}\sum_{i=1}^{n}\Big(Y_i^* R_i + a(1 - R_i)\Big)$$

- and the plug-in estimator for the lower bound is:

$$\frac{1}{n}\sum_{i=1}^{n}\Big(Y_i^* R_i + b(1 - R_i)\Big).$$

- These estimators are unbiased and consistent for their respective bounds.

Science table: bounds (lower bound)

What would recidivism rates be if all juveniles were assigned to non-residential treatment?

Target: $\mathrm{E}[Y_i(0)]$

i	$Y_i(0)$	R_i	$\hat{Y}_i(0)$
1	0	1	0
2	?	0	0
3	1	1	1
4	?	0	0
5	0	1	0
6	?	0	0
7	1	1	1
8	?	0	0
9	1	1	1
10	?	0	0

$\overline{Y_i^L}$ = mean of $Y_i^*(0)$ with missing set to $0 = 0.3$

Science table: bounds (upper bound)

What would recidivism rates be if all juveniles were assigned to non-residential treatment?

Target: $\mathbb{E}[Y_i(0)]$

i	$Y_i(0)$	R_i	$\hat{Y}_i(0)$
1	0	1	0
2	?	0	1
3	1	1	1
4	?	0	1
5	0	1	0
6	?	0	1
7	1	1	1
8	?	0	1
9	1	1	1
10	?	0	1

$$\overline{Y_i^U}(0) = \text{mean of } Y_i^*(0) \text{ with missing set to } 1 = 0.8$$

Interval identification

- Our bounds are $[0.3, 0.8]$ for this example.
- Without further assumptions, the bounds are the best we can do.
- $\mathrm{E}[Y_i]$ is **partially identified** (set identified), not point identified.
- Even with full knowledge of (Y_i^*, R_i), the exact value of $\mathrm{E}[Y_i]$ is not identified.

The Law of Decreasing Credibility Manski (2003)

The Law of Decreasing Credibility: The credibility of inference decreases with the strength of the assumptions maintained.

MCAR: Missing Completely at Random

- Let Y_i and R_i be random variables with $\mathrm{Supp}\,[R_i] = \{0, 1\}$.
- Let $Y_i^* = Y_i R_i + (-99)(1 - R_i)$.
- Y_i is MCAR if:
  - $Y_i \perp R_i$ (independence of outcome and response).
  - $\Pr[R_i = 1] > 0$ (nonzero probability of response).
- Note: implicitly, MCAR is with respect to *all* data, observed and unobserved. So, $R_i \perp (Y_i, X_i, Z_i, \dots)$.

Point identification

- Stronger assumptions can deliver **point identification**.
- Under MCAR, $\mathrm{E}[Y_i]$ is point identified.
- The key tradeoff: tighter conclusions require stronger, (possibly) less credible assumptions.

Expected value under MCAR

- Under MCAR, the expected value is identified from observed outcomes:

$$\mathrm{E}[Y_i] = \mathrm{E}[Y_i^* \mid R_i = 1]$$

- Plug-in estimator:

$$\widehat{\mathrm{E}}[Y_i] = \frac{\sum_{i=1}^{n} Y_i^* R_i}{\sum_{i=1}^{n} R_i}$$

Science table: MCAR plug-in imputation

Target: $\mathrm{E}[Y_i(0)]$

i	$Y_i(0)$	R_i	$\widehat{Y}_i(0)$
1	0	1	0
2	?	0	0.6
3	1	1	1
4	?	0	0.6
5	0	1	0
6	?	0	0.6
7	1	1	1
8	?	0	0.6
9	1	1	1
10	?	0	0.6

$$\widehat{\mathrm{E}}[Y_i(0)] = \frac{0+1+0+1+1}{5} = 0.6$$

$$\text{Imputed sample mean} = \frac{0+1+0+1+1+0.6+0.6+0.6+0.6+0.6}{10} = 0.6$$

MAR: Missing at Random

- Let Y_i and R_i be random variables with $\mathrm{Supp}\,[R_i] = \{0, 1\}$.
- Let $Y_i^* = Y_i R_i + (-99)(1 - R_i)$, and let X_i be a random vector.
- Y_i is MAR conditional on X_i if:
  - $Y_i \perp R_i \mid X_i$ (independence of outcome and response conditional on X_i).
  - $\exists \varepsilon > 0$ such that $\Pr[R_i = 1 \mid X_i] > \varepsilon$ (nonzero probability of response conditional on X_i).
- This is a *statistical assumption.*

Positivity conditions under MCAR vs. MAR

Missing Completely at Random (MCAR)

$$\Pr[R_i = 1] > 0.$$

- Interpretation: we must observe *some* outcomes.

Missing at Random (MAR)

$$\Pr[R_i = 1 \mid X_i = x] \geq \varepsilon > 0 \quad \forall x \in \mathrm{Supp}\,[X].$$

- Interpretation: we must observe outcomes for *every covariate stratum*.

- MCAR asks: “Do we observe *anyone*?”
- MAR asks: “Do we observe *someone like everyone*?”

Expected value under MAR

- Under MAR, the expected value is identified by conditioning on X_i:

$$
\begin{aligned}
\mathrm{E}[Y_i] &= \sum_x \mathrm{E}[Y_i^* \mid R_i = 1, X_i = x] \Pr[X_i = x] \\
&= \mathrm{E}\left[\mathrm{E}[Y_i \mid R_i = 1, X_i]\right]
\end{aligned}
$$

MAR plug-in estimator

- Sample analogue (plug-in):

$$\widehat{\mathrm{E}}[Y_i] = \sum_{x \in \mathrm{Supp}[X_i]} \widehat{\mathrm{E}}[Y_i \mid R_i = 1, X_i = x] \widehat{\Pr}[X_i = x]$$

$$\widehat{\mathrm{E}}[Y_i \mid R_i = 1, X_i = x] = \frac{\sum_{i=1}^{n} Y_i^* R_i \mathbb{1}(X_i = x)}{\sum_{i=1}^{n} R_i \mathbb{1}(X_i = x)}$$

Science table: MAR plug-in imputation

Target: $\mathrm{E}[Y_i(0)]$

i	X_i	$Y_i(0)$	R_i	$\widehat{Y}_i(0)$
1	A	0	1	0
2	A	?	0	1/3
3	B	1	1	1
4	B	?	0	1
5	A	0	1	0
6	A	?	0	1/3
7	B	1	1	1
8	B	?	0	1
9	A	1	1	1
10	B	?	0	1

$$\widehat{\mathrm{E}}[Y_i \mid R_i = 1, X_i = \mathrm{A}] = 1/3, \quad \widehat{\mathrm{E}}[Y_i \mid R_i = 1, X_i = \mathrm{B}] = 1$$

$$\widehat{\mathrm{E}}[Y_i(0)] = \frac{5}{10}\widehat{\mathrm{E}}[Y_i \mid R_i = 1, X_i = \mathrm{A}] + \frac{5}{10}\widehat{\mathrm{E}}[Y_i \mid R_i = 1, X_i = \mathrm{B}] = 0.667$$

$$\text{Imputed sample mean} = \frac{0 + 1/3 + 1 + 1 + 0 + 1/3 + 1 + 1 + 1 + 1}{10} = 0.667$$

Potential outcomes are missing data

- We observe $Y_i(1)$ only if $D_i = 1$ and $Y_i(0)$ only if $D_i = 0$.
- The unobserved potential outcome is missing data.
- Identification requires assumptions about assignment.

Science table: potential outcomes

i	$Y_i(0)$	$Y_i(1)$	D_i	Y_i
1	0	?	0	0
2	?	1	1	1
3	1	?	0	1
4	?	1	1	1
5	0	?	0	0
6	?	1	1	1
7	1	?	0	1
8	?	0	1	0
9	1	?	0	0
10	?	1	1	1

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

Average treatment effect

- Individual effect: $\tau_i = Y_i(1) - Y_i(0)$.
- Average treatment effect: $\tau = \mathrm{E}[\tau_i]$.

$$\begin{aligned}\mathrm{E}[\tau_i] &= \mathrm{E}\left[Y_i(1) - Y_i(0)\right] \\ &= \mathrm{E}\left[Y_i(1)\right] - \mathrm{E}\left[Y_i(0)\right].\end{aligned}$$

- We need to make some assumptions to link observed data to $\mathrm{E}[Y_i(d)]$.

Sharp bounds for the ATE

- Let $Y_i(0)$, $Y_i(1)$, and D_i be random variables such that, for all $d \in \{0,1\}$, $\mathrm{Supp}\,[Y_i(d)] \subseteq [a,b]$ and $\mathrm{Supp}\,[D_i] = \{0,1\}$

- Let $Y_i = Y_i(1)D_i + Y_i(0)(1-D_i)$ and $\tau_i = Y_i(1) - Y_i(0)$.

- Then:

$$
\begin{aligned}
\mathrm{E}\,[\tau_i] \in \Big\{ & \mathrm{E}\,[Y_i \mid D_i = 1]\Pr[D_i = 1] + a\Pr[D_i = 0] - \\
& \qquad \big(\mathrm{E}\,[Y_i \mid D_i = 0]\Pr[D_i = 0] + b\Pr[D_i = 1]\big), \\
& \mathrm{E}\,[Y_i \mid D_i = 1]\Pr[D_i = 1] + b\Pr[D_i = 0] - \\
& \qquad \big(\mathrm{E}\,[Y_i \mid D_i = 0]\Pr[D_i = 0] + a\Pr[D_i = 1]\big) \Big\}.
\end{aligned}
$$

Estimating sharp bounds for the ATE

- Given n i.i.d. observations of (Y_i, D_i):
- The plug-in estimator for the upper bound is:

$$\frac{1}{n}\sum_{i=1}^{n}\Big(Y_i D_i + a(1-D_i) - Y_i(1-D_i) - bD_i\Big)$$

- and the plug-in estimator for the lower bound is:

$$\frac{1}{n}\sum_{i=1}^{n}\Big(Y_i D_i + b(1-D_i) - Y_i(1-D_i) - aD_i\Big).$$

- These estimators are unbiased and consistent for the lower and upper bounds respectively for $\mathrm{E}[\tau_i]$.

Science table: bounds for $\mathbb{E}[Y_i(1)]$ (lower bound)

What would recidivism rates be if all juveniles were assigned to residential treatment?

Target: $\mathbb{E}[Y_i(1)]$

i	$Y_i(1)$	R_i	$\hat{Y}_i(1)$
1	?	0	0
2	1	1	1
3	?	0	0
4	1	1	1
5	?	0	0
6	1	1	1
7	?	0	0
8	0	1	0
9	?	0	0
10	1	1	1

$$\overline{Y_i^L}(1) = \text{mean of } Y_i^*(1) \text{ with missing set to } 0 = 0.4$$

Science table: bounds for $\mathbb{E}[Y_i(1)]$ (upper bound)

What would recidivism rates be if all juveniles were assigned to residential treatment?

Target: $\mathbb{E}[Y_i(1)]$

i	$Y_i(1)$	R_i	$\hat{Y}_i(1)$
1	?	0	1
2	1	1	1
3	?	0	1
4	1	1	1
5	?	0	1
6	1	1	1
7	?	0	1
8	0	1	0
9	?	0	1
10	1	1	1

$\overline{Y_i^U}(1) =$ mean of $Y_i^*(1)$ with missing set to $1 = 0.9$

ATE bounds from potential outcomes

We already have $\widehat{\mathrm{E}}\left[Y_i(0)\right] \in [0.3, 0.8]$. From the tables above,

$$\widehat{\mathrm{E}}\left[Y_i(1)\right] \in [0.4, 0.9].$$

Therefore,

$$\begin{aligned}
\widehat{\mathrm{E}}\left[\tau_i\right] = \widehat{\mathrm{E}}\left[Y_i(1)\right] - \widehat{\mathrm{E}}\left[Y_i(0)\right] \in & \left[\overline{Y_i^L}(1) - \overline{Y_i^U}(0),\ \overline{Y_i^U}(1) - \overline{Y_i^L}(0)\right] \\
&= [0.4 - 0.8,\ 0.9 - 0.3] \\
&= [-0.4, 0.6].
\end{aligned}$$

Alternative estimands

- Other causal targets include:
    - ATT: $\mathbb{E}[\tau_i \mid D_i = 1]$.
    - ATC: $\mathbb{E}[\tau_i \mid D_i = 0]$.
- These are often of policy interest (e.g., ATT speaks to the effect among those who actually receive treatment).
- ATT and ATC can be point identified under weaker assumptions than the ATE.

References I

Aronow, P. M. and Miller, B. T. (2019). *Foundations of agnostic statistics*. Cambridge University Press.

Manski, C. F. (2003). *Partial identification of probability distributions*. Springer.

Manski, C. F. and Nagin, D. S. (1998). Bounding disagreements about treatment effects: A case study of sentencing and recidivism. *Sociological methodology*, 28(1):99–137.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701.