Causal Inference: Modern Methods

Problem Set

Diogo Ribeiro
ESMAD - Escola Superior de Média Arte e Design
Lead Data Scientist, Mysense.ai
dfr@esmad.ipp.pt
ORCID: 0009-0001-2022-7072

November 12, 2025

Abstract

This problem set covers modern methods in causal inference, including potential outcomes, directed acyclic graphs (DAGs), instrumental variables, regression discontinuity, and difference-in-differences. Students will gain practical experience identifying causal effects from observational data and understanding threats to causal identification.

1 Potential Outcomes Framework

Exercise 1 (Fundamental Problem of Causal Inference)**.** *Consider a job training program where $D_i = 1$ if individual i receives training and $D_i = 0$ otherwise. Let $Y_i(1)$ be the potential outcome with training and $Y_i(0)$ without.*

(a) Define the individual treatment effect (ITE), average treatment effect (ATE), and average treatment effect on the treated (ATT).

(b) Explain why we cannot observe $Y_i(1) - Y_i(0)$ for any individual.

(c) Show that the naive comparison $E[Y_i|D_i = 1] - E[Y_i|D_i = 0]$ equals:

$$\underbrace{E[Y_i(1) - Y_i(0)|D_i = 1]}_{ATT} + \underbrace{E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0]}_{Selection\ Bias}$$

(d) Under what conditions does the naive comparison identify the ATE?

Exercise 2 (Conditional Independence Assumption)**.** *The conditional independence assumption (CIA) or unconfoundedness states:*

$$\{Y_i(0), Y_i(1)\} \perp D_i|X_i$$

(a) Explain this assumption in plain language.

(b) Why is it called "unconfoundedness"?

(c) Give an example where CIA is plausible.

(d) Give an example where CIA is clearly violated.

(e) Under CIA, show that:

$$E[Y_i(1) - Y_i(0)|X_i] = E[Y_i|D_i = 1, X_i] - E[Y_i|D_i = 0, X_i]$$

2 Directed Acyclic Graphs (DAGs)

Exercise 3 (Drawing and Interpreting DAGs)**.** *For each scenario, draw a DAG and identify the causal effect of interest:*

(a) ***Education and Earnings:***

- *Treatment: Years of education (D)*
- *Outcome: Annual earnings (Y)*
- *Confounders: Ability (A), Family background (F)*
- *Mediator: Job type (J)*

Draw the DAG and identify:

- *All paths from D to Y*
- *Which paths are backdoor paths?*
- *What variables should be controlled?*
- *What is the direct effect of D on Y (not through J)?*

(b) ***Collider Bias:*** *Consider $D \rightarrow Y \leftarrow U$ where U is unobserved.*

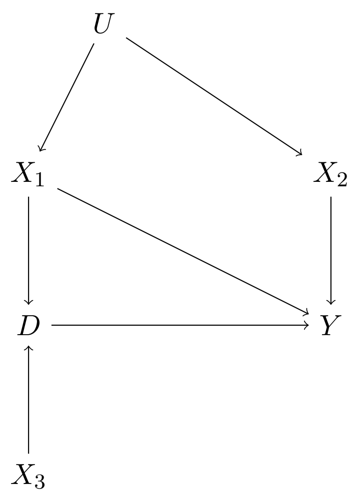
- *Explain why conditioning on Y induces association between D and U*
- *Provide a real-world example of collider bias*

(c) ***M-Bias:*** *Draw a DAG with structure: $U_1 \rightarrow D$, $U_2 \rightarrow Y$, $U_1 \rightarrow M \leftarrow U_2$.*

- *Is there an open backdoor path from D to Y?*
- *What happens if you control for M?*
- *When is controlling for M harmful?*

Problem 1 (Backdoor Criterion)**.** *Use Pearl's backdoor criterion to determine the minimal sufficient adjustment set.*

Scenario: *Consider the following DAG:*



1. *List all paths from D to Y*
2. *Identify backdoor paths*
3. *Find all valid adjustment sets*
4. *What is the minimal sufficient adjustment set?*
5. *Can we identify the causal effect if U is unobserved?*

3 Instrumental Variables

Problem 2 (Returns to Education)**.** *Estimate the returns to education using quarter of birth as an instrument.*

Background: *Angrist & Krueger (1991) used quarter of birth as an instrument for education, exploiting compulsory schooling laws.*

Dataset: *Simulate or use provided data with:*

- Y*: Log annual earnings*
- D*: Years of education*
- Z*: Quarter of birth indicator (born in Q1)*
- X*: Controls (age, region, etc.)*

Tasks:

1. ***Motivation:***
    - *Why is education endogenous?*
    - *Draw a DAG showing the endogeneity problem*
    - *Explain how quarter of birth could be a valid instrument*
2. ***Validity Checks:***
    - *Test relevance: Regress D on Z (first stage)*
    - *Compute first-stage F-statistic*
    - *Discuss exclusion restriction: Why might it fail?*
3. ***Estimation:***
    - *Estimate OLS (biased)*
    - *Estimate 2SLS*
    - *Compare coefficients*
    - *Compute heteroskedasticity-robust standard errors*
4. ***Interpretation:***
    - *What is the local average treatment effect (LATE)?*
    - *Who are the compliers?*
    - *Why might IV estimate differ from OLS?*
5. ***Sensitivity:***
    - *Add/remove controls*
    - *Test overidentification (if multiple instruments)*
    - *Discuss weak instruments problem*

Problem 3 (Weak Instruments)**.** *Investigate the consequences of weak instruments.*

Simulation:

1. *Generate data: $n = 500$, instrument strength $\gamma \in \{0.05, 0.1, 0.3, 0.5\}$*

$$D = \gamma Z + \alpha U + \epsilon_D$$
$$Y = \beta D + \alpha U + \epsilon_Y$$

   where U is unobserved confounder.

2. *For each γ:*
   - *Estimate 2SLS*
   - *Compute first-stage F-statistic*
   - *Repeat 1000 times (Monte Carlo)*

3. *Analyze:*
   - *Bias of 2SLS estimator*
   - *Coverage of 95% confidence intervals*
   - *How weak is "weak"? (F ¡ 10 rule of thumb)*

4. *Propose solutions:*
   - *Limited information maximum likelihood (LIML)*
   - *Anderson-Rubin test*
   - *Weak instrument robust inference*

4 Regression Discontinuity Design

Problem 4 (Class Size and Student Achievement)**.** *Use RDD to estimate the effect of class size on test scores.*

Setup: *Schools with enrollment above 40 students must split into two classes.*
Data:

- *Running variable: School enrollment (around cutoff of 40)*
- *Treatment: Small class (enrollment $\leq$ 40) vs. large class*
- *Outcome: Average test score*

Tasks:

1. ***Identification:***
   - *State the RDD identification assumptions*
   - *Why is continuity of potential outcomes crucial?*
   - *Draw a hypothetical RDD plot*

2. ***Validity Checks:***
   - *Density test: Check for manipulation of running variable (McCrary test)*
   - *Covariate balance: Test continuity of pre-treatment covariates*
   - *Placebo cutoffs: Test at other points*

3. ***Estimation:***

- *Local linear regression (both sides of cutoff)*
- *Optimal bandwidth selection (Imbens-Kalyanaraman or MSE-optimal)*
- *Robust inference (bias-corrected CI)*

4. ***Sensitivity:***
   - *Bandwidth sensitivity plot*
   - *Polynomial order (linear vs. quadratic)*
   - *Donut RDD (exclude observations near cutoff)*
5. ***Interpretation:***
   - *Is this effect local or global?*
   - *What is the relevant population?*
   - *External validity concerns*

Problem 5 (Fuzzy RDD)**.** *Extend to fuzzy RDD where treatment uptake is imperfect.*
Scenario: *Assignment to small class at cutoff, but:*

- *Some schools just above cutoff have small classes (non-compliance)*
- *Some schools just below have large classes*

Tasks:

1. *Explain the difference between sharp and fuzzy RDD*
2. *Show that fuzzy RDD can be seen as an IV problem*
3. *Estimate both the "first stage" (effect on treatment) and "reduced form" (effect on outcome)*
4. *Compute the fuzzy RDD estimand (ratio)*
5. *Interpret as local average treatment effect (LATE)*

5 Difference-in-Differences

Problem 6 (Minimum Wage and Employment)**.** *Replicate Card & Krueger (1994) analysis of minimum wage effects.*
Setup:

- *Treatment: New Jersey minimum wage increase (April 1992)*
- *Control: Pennsylvania (no increase)*
- *Outcome: Employment at fast-food restaurants*
- *Periods: Before (Feb 1992) and After (Nov 1992)*

Tasks:

1. ***Identification:***
   - *State parallel trends assumption*
   - *Why is it crucial?*

- *Draw hypothetical trend plots*

2. ***Estimation:***
   - *Compute 2x2 DiD estimator (manual calculation)*
   - *Estimate regression specification:*

$$Y_{it} = \alpha + \beta Treat_i + \gamma Post_t + \delta(Treat_i \times Post_t) + \epsilon_{it}$$

   - *Interpret δ (the DiD estimator)*
   - *Compute cluster-robust standard errors (at state level)*

3. ***Parallel Trends Testing:***
   - *Test for pre-treatment differential trends*
   - *Event study plot (if multiple time periods available)*
   - *Placebo tests using pre-treatment periods*

4. ***Extensions:***
   - *Add covariates*
   - *Triple differences (add another control group)*
   - *Heterogeneous treatment effects by restaurant characteristics*

Problem 7 (Staggered Adoption)**.** *Analyze staggered treatment adoption with recent methods.*
Setup: *Multiple states adopt minimum wage increases at different times.*
Problem: *Two-way fixed effects (TWFE) can give negative weights to some treatment effects (Goodman-Bacon decomposition).*
Tasks:

1. *Simulate staggered adoption data with heterogeneous effects*

2. *Estimate naive TWFE:*

$$Y_{it} = \alpha_i + \lambda_t + \delta D_{it} + \epsilon_{it}$$

3. *Show that TWFE can be biased with:*
   - *Treatment effect heterogeneity*
   - *Dynamic effects*

4. *Implement modern alternatives:*
   - *Callaway & Sant'Anna (2021)*
   - *Sun & Abraham (2021)*
   - *Borusyak et al. (2021)*

5. *Compare all estimates and discuss*

6 Matching and Propensity Scores

Problem 8 (Propensity Score Matching)**.** *Estimate treatment effects using propensity score methods.*

Scenario: *Effect of college degree on earnings.*

Data:

- *Treatment: College degree (vs. high school)*
- *Outcome: Annual earnings*
- *Covariates: Age, gender, parents' education, test scores, etc.*

Tasks:

1. ***Propensity Score Estimation:***
   - *Fit logistic regression:* $P(D = 1|X)$
   - *Check for common support*
   - *Plot propensity score distributions by treatment group*
2. ***Matching Methods:***
   - *Nearest neighbor matching (1:1, 1:k)*
   - *Caliper matching*
   - *Kernel matching*
3. ***Balance Checking:***
   - *Standardized mean differences before/after matching*
   - *Love plot*
   - *Variance ratios*
4. ***Estimate ATT:***
   - *Using matched samples*
   - *IPW (inverse probability weighting)*
   - *Doubly robust estimation*
5. ***Sensitivity Analysis:***
   - *Rosenbaum bounds*
   - *How strong would hidden bias need to be?*

7 Synthetic Control Methods

Problem 9 (Policy Evaluation)**.** *Use synthetic control method to evaluate a policy intervention.*

Example: *German reunification effect on West German GDP (Abadie et al. 2015).*

Tasks:

1. *Explain the synthetic control idea*

2. *Construct synthetic control as weighted average of donor units:*

$$\hat{Y}_{0t} = \sum_{j=2}^{J+1} w_j Y_{jt}$$

   where weights minimize pre-treatment fit

3. *Estimate treatment effect:* $\hat{\tau}_t = Y_{1t} - \hat{Y}_{0t}$
4. *Placebo tests: Apply method to control units*
5. *Inference: Permutation-based p-values*
6. *Sensitivity: Leave-one-out analysis*

8 Submission Guidelines

8.1 Format

- Python (Jupyter notebook) or R (R Markdown)
- Include all code, results, and visualizations
- Clear explanations and interpretations
- Well-documented code

8.2 Evaluation

- **Identification (30%):** Understanding of causal assumptions
- **Implementation (30%):** Correct estimation procedures
- **Interpretation (25%):** Thoughtful discussion of results
- **Validity Checks (15%):** Robustness and sensitivity analyses

9 Resources

Code: `code/causal_inference/`
Bibliography: `bibliographies/causal_inference_references.bib`
Key Papers:

- Angrist & Pischke (2008). *Mostly Harmless Econometrics*
- Pearl (2009). *Causality*
- Imbens & Rubin (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences*

Contact: Diogo Ribeiro, dfr@esmad.ipp.pt