

Causal Inference for Data Science
Instructor: Adam Kelleher
Syllabus and Tentative Schedule

General Information

- **Meeting time:** Monday, 7:00p-9:30p
- **Contact:** Adam Kelleher, ak4063@columbia.edu
- **Office Hours:** TBA
- **Textbook:** Morgan and Winship, Counterfactuals and Causal Inference, 2nd Ed.
- **Grading:** Homework 40%, Mid-term 30%, Final exam 30%

Pre-requisites

- **Math:** Undergrad probability theory; Some experience with regression analysis will be useful; some knowledge of information theory will be useful, but not required. Some knowledge of bayesian networks will be useful but not required.
- **CS:** Knowledge of a programming language, preferably Python.

Homework

- Homework will be due by midnight two weeks after they are assigned. There will be assignments approximately every two weeks.
- Late assignments will be reduced as follows:
 - 0+ - 24 hours late: 25% of points deducted
 - 24 - 48 hours late: 50% of points deducted
 - More than 48 hours: no credit
- Exceptions will be made for medical emergencies or other exceptional circumstances discussed in advance.

Collaboration

- Collaboration is strongly encouraged. Everyone must write up their assignments on their own. Copying collaborators' work or copying work from other sources (textbooks, the internet) is prohibited.

Programming Assignments

- Programming assignments will be completed with Jupyter notebooks. Install and familiarize yourselves with Jupyter notebooks as soon as possible.

Tentative Schedule:

Causal Frameworks

- *Jan 23:* Context: causality in data science; causality and ML; computational social science
- *Jan 30:* Intro to Counterfactuals and Potential Outcomes; HW 1
- *Feb 6:* The Pearlman Framework
- *Feb 13:* Confounding; Berkson's Paradox; simpson's paradox; HW 2

Effect Estimation with Conditioning

- *Feb 20:* Identifying variables for conditioning: back-door criterion; matching effect estimators; HW 3
- *Feb 27:* Regression estimators of causal effects 1
- *Mar 6:* Mid Term
- *Mar 13:* Spring Break
- *Mar 20:* Regression estimators of causal effects 2; the data processing inequality; HW 4

Effect Estimation without Conditioning

- *Mar 27:* Conditioning to reduce entropy; Conditioning to remove bias; Self-selection
- *Apr 3:* Instrumental variables; recommender system example; HW 5
- *Apr 10:* Mechanisms of action and the Front-door criterion
- *Apr 17:* Regression Discontinuity Design; HW 6

Experiment Design for Estimating Causal Effects

- *Apr 24:* AB testing and interventions; Multifactorial design
- *May 1:* Intent to Treat analysis; an (observational) statistical criterion for causation

- *May 8: Final Exam*

Tentative Homework Schedule

HW	Assigned	Due
HW1	Jan 30	Feb 13
HW2	Feb 13	Feb 27
HW3	Feb 20	Mar 20
HW4	Mar 20	Apr 3
HW5	Apr 3	Apr 17
HW6	Apr 17	May 1