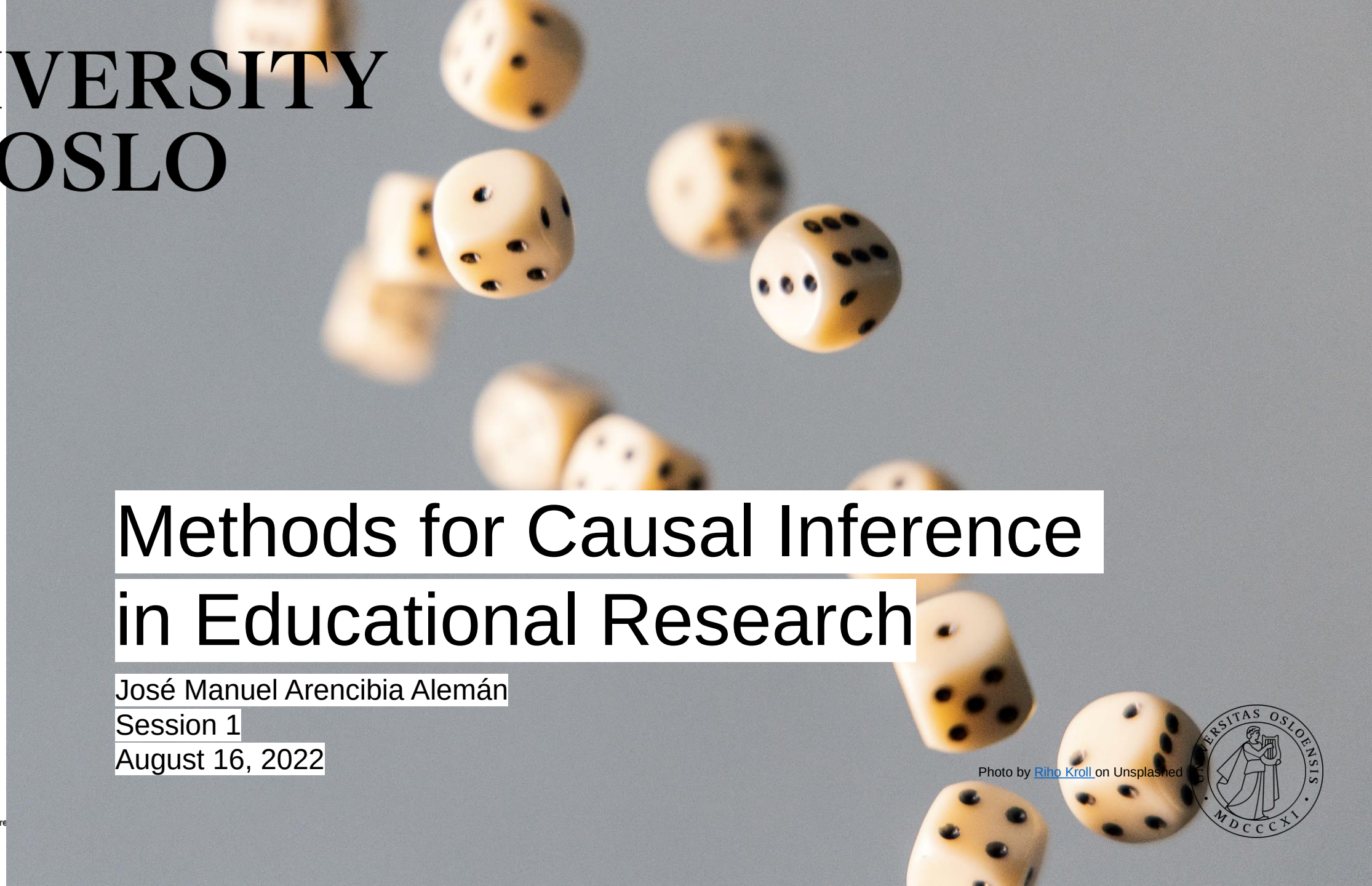
UNIVERSITY OF OSLO



Methods for Causal Inference in Educational Research

José Manuel Arencibia Alemán

Session 1

August 16, 2022

Photo by Riho Kroll on Unsplashed

UiO : Centre for Educational Measure
Det utdanningsvitenskapelige fakultet

Roadmap

1. Getting to know each other
2. Workshop aims and goals
3. Formalities & Schedule
4. Oral Exam requirements
5. Introduction to Causal Inference
   - ✓ Anatomy of Causal Relationships
   - ✓ Potential Outcome Framework

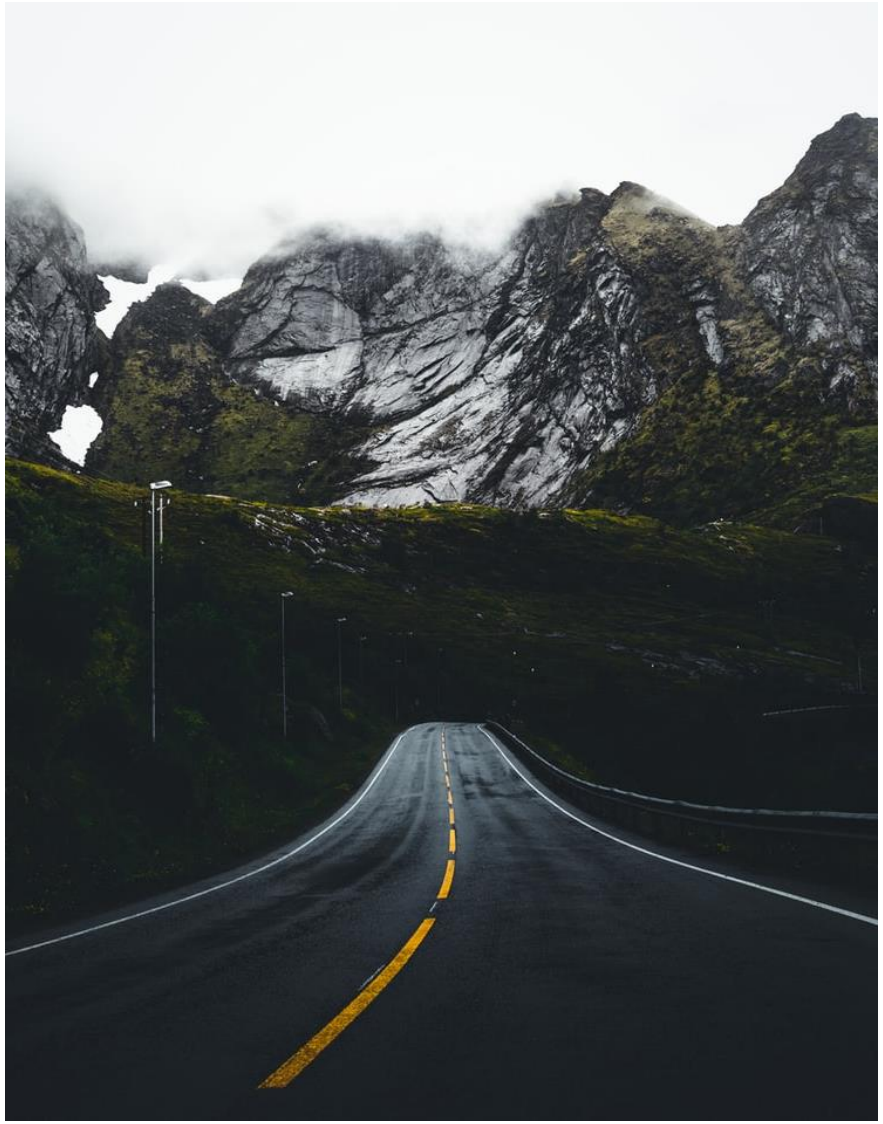


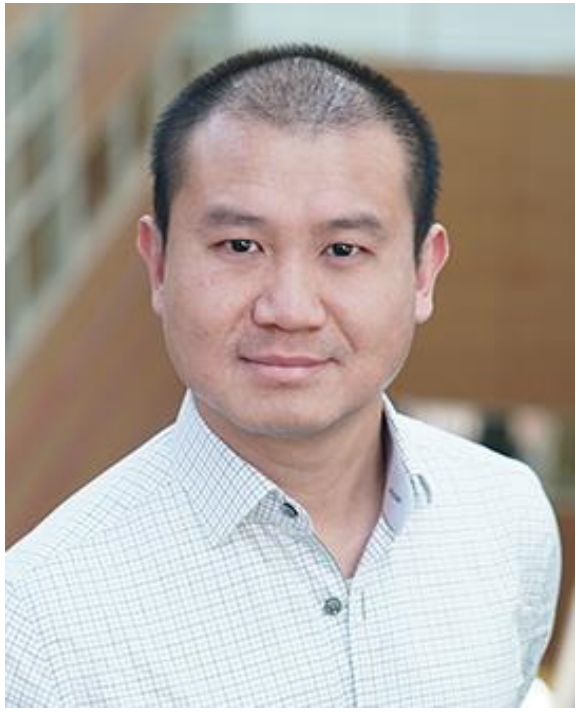
Photo by Taneli Lahtinen on Unsplash

Getting to know each other

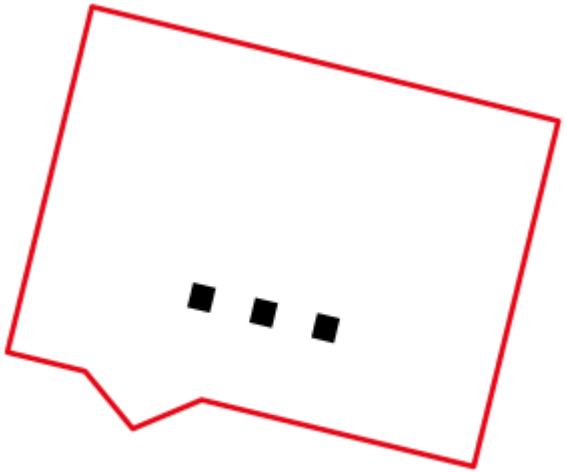
José

j.m.a.aleman@cemo.uio.no

Tony



tony.tan@cemo.uio.no



- You?
- Thesis Project?
- Previous knowledge of causal inference methods/ theory?

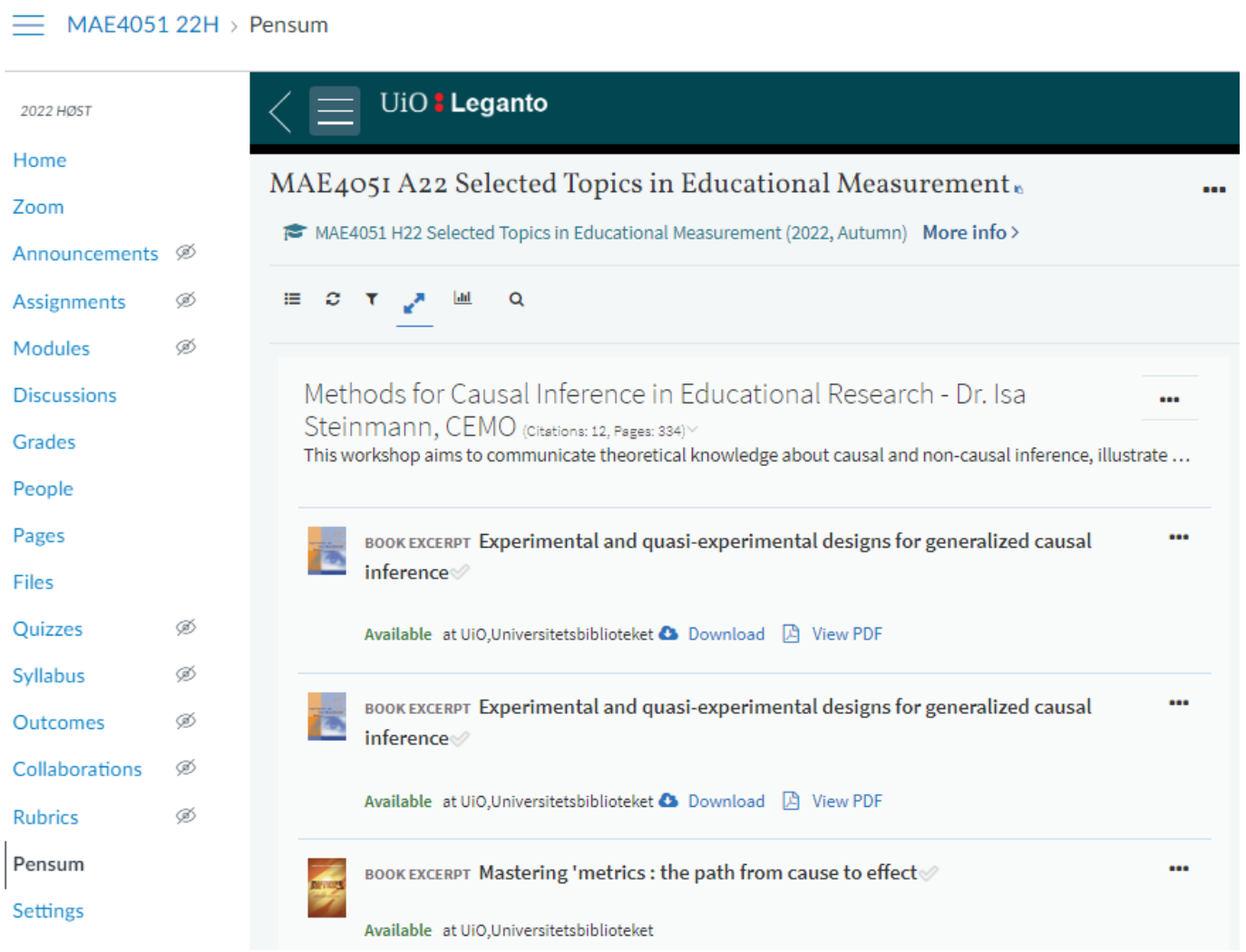
Workshop aims

1. Communicating theoretical knowledge about causal and non-causal inference
   - Causal questions
   - Experimental designs
   - Methods
2. Illustrating methods for causal inference from experimental and observational data
   - From previous research
3. Exemplifying applications of the methods on R

Workshop learning outcomes

1. Understanding of the potential outcomes framework
2. Broad knowledge about assumptions, prerequisites, and procedures of randomized control trials, regression models, fixed effect models, instrumental variable approaches, regression discontinuity designs and differences in differences.
3. Ability to critically read and interpret results of causal inference studies
4. Applying methodological peculiarities of the causal inference methods in own analyses
5. Develop Causal inference research questions and approaches

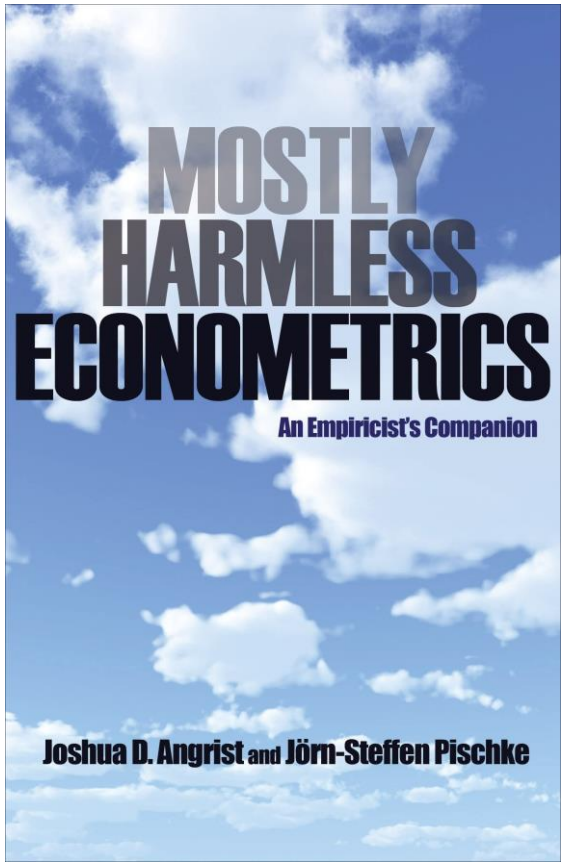
Formalities and Schedule



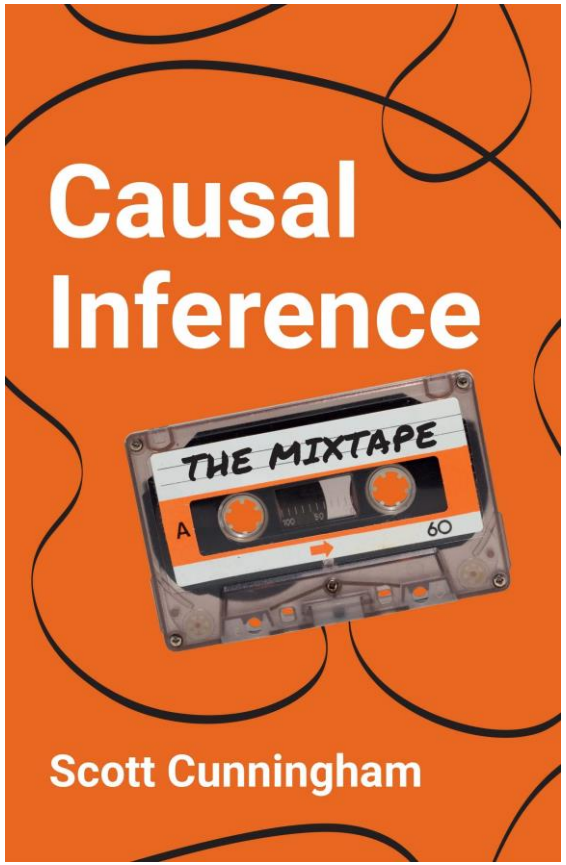
- Required readings (in Canvas / Leganto)
- Slides and Materials (in Canvas after each session)
- Dates and Place (in Canvas)

Other recommendations for the specially interested

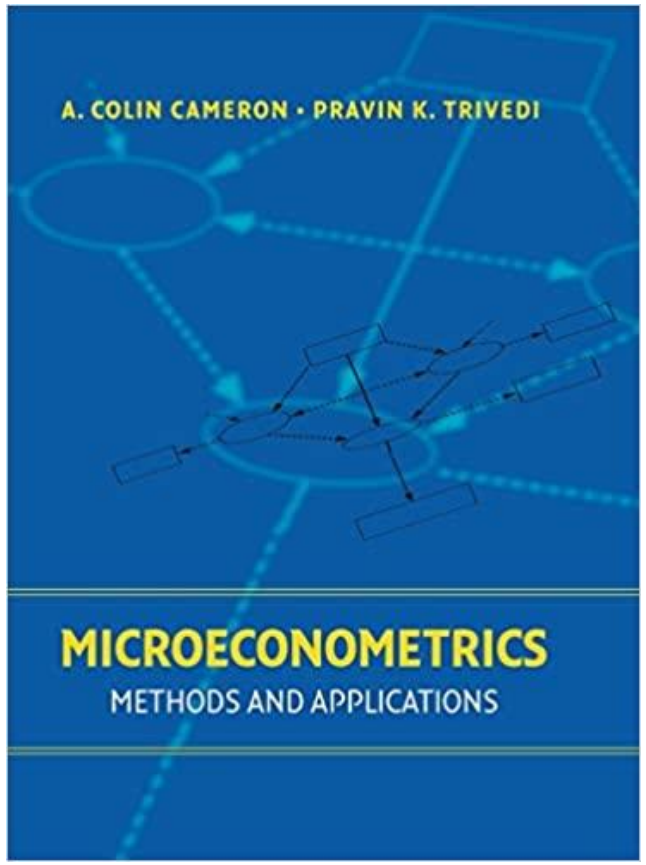
Angrist, & Pischke, J.-S. (2009)



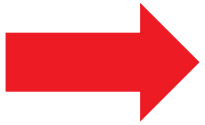
Cunningham. (2021)



Cameron, & Trivedi, P. K. (2005)



Formalities and Schedule



1	Introduction - Time: 16 August 2022, 12:15-14:00h - Main instructor: José Manuel Arencibia Alemán - Required reading: Shadish, Cook, & Campbell (2002), chapter 1
2	Designing Experiments - Time: 19 August 2022, 12:15-14:00h - Main instructor: Tony Tan - Required reading: Shadish, Cook, & Campbell (2002), chapter 8
3	Randomized Trials - Time: 23 August 2022, 12:15-14:00h - Main instructor: José Manuel Arencibia Alemán - Required reading: Angrist & Pischke (2015), chapter 1
4	Regression Models - Time: 26 August 2022, 10:15-12:00h - Main instructor: Tony Tan - Required reading: Angrist & Pischke (2015), chapter 2

Formalities and Schedule

5	Fixed Effects - Time: 30 August 2022, 12:15-14:00h - Main instructor: José Manuel Arencibia Alemán - Required reading: Huntington-Klein (2022), chapter 16
6	Fixed Effects / Instrumental Variables - Time: 2 September 2022, 10:15-12:00h - Main instructor: José Manuel Arencibia Alemán - Required reading: H-K (2022), chapter 16; A. & P., chapter 3
7	Instrumental Variables - Time: 6 September 2022, 12:15-14:00h - Main instructor: José Manuel Arencibia Alemán - Required reading: Angrist & Pischke (2015), chapter 3
8	Regression Discontinuity I - Time: 9 September 2022, 10:15-12:00h - Main instructor: José Manuel Arencibia Alemán - Required reading: Angrist & Pischke (2015), chapter 4

Formalities and Schedule

9	Regression Discontinuity II - Time: 13 September 2022, 12:15-14:00h - Main instructor: Tony Tan - Required reading: Steinman & Olsen (2002)
10	Differences-in-Differences I - Time: 16 September 2022, 10:15-12:00h - Main instructor: José Manuel Arencibia Alemán - Required reading: Angrist & Pischke (2015), chapter 5
11	Differences-in-Differences II - Time: 20 September 2022, 12:15-14:00h - Main instructor: Tony Tan - Required reading: Strello, Strietholdt, Steinmann & Siepmann (2021)
12	Lessons Learned - Time: 23 September 2022, 10:15-12:00h - Main instructor: José Manuel Arencibia Alemán - Required reading: Rutkowski & Delandshere (2016)

Assignment requirement (Old)

- Written assignment has to be submitted by November 9
- Task is to discuss three published empirical articles which use methods for causal inference from non-experimental data
- Three articles must be chosen from list

Falch, T., Nyhus, O. H., Strøm, B. (2014). Causal effects of mathematics. *Labour Economics, 31*, 174-187. **(RM)**

Fergusson, D. M., McLeod, G. F., & Horwood, L. J. (2014). Unemployment and psychosocial outcomes to age 30: a fixed-effects regression analysis. *Australian & New Zealand Journal of Psychiatry, 48*(8), 735-742. **(FE)**

Hanandita, W., & Tampubolon, G. (2014). Does poverty reduce mental health? An instrumental variable analysis. *Social Science & Medicine, 113*, 59-67. **(IV)**

Lesik, S. A. (2007). Do developmental mathematics programs have a causal impact on student retention? An application of discrete-time survival and regression-discontinuity analysis. *Research in Higher Education, 48*(5), 583-608.**(RD)**

Nye, B., Hedges, L. V., & Konstantopoulos, S. (2000). The effects of small classes on academic achievement: The results of the Tennessee class size experiment. *American Educational Research Journal, 37*(1), 123–151. **(RCT)**

Rosén, M. & Gustafsson, J. E. (2016). Is computer availability at home causally related to reading achievement in grade 4? A longitudinal difference in differences approach to IEA data from 1991 to 2006. *Large-scale Assessment in Education, 4*(5), 1–19. **(Diff-in-Diff)**

Assignment requirements (Old)

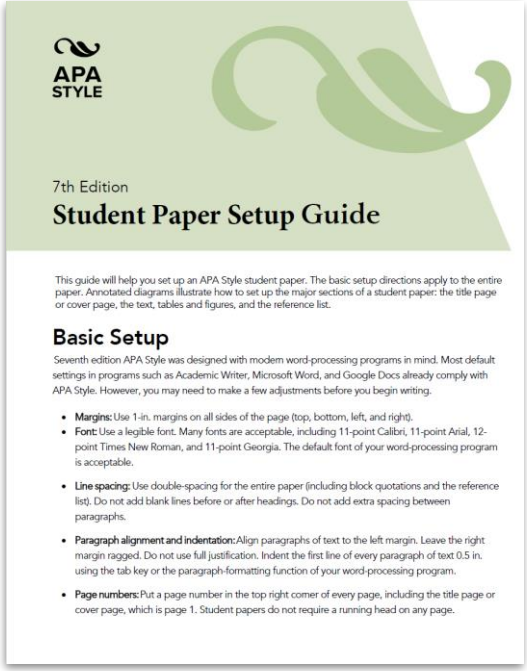
For each study, the report contains:

- Summary of the background and motivation of the study (i.e., why was the study conducted, what was the theoretical background, what did the literature review show)
- The identification of the causal question investigated in the study (i.e., what are treatment, outcome, and assumed causal mechanism)
- Critical discussion of the causal identification approach against the design principles of the respective causal inference method (i.e., what was the statistical approach, in how far were assumptions of the methods met)
- Critical summary of the findings of the study (i.e., what were the findings that directly relate to the research question, what were important other findings)
- Concluding discussion about the extent to which the causal question was appropriately answered by the study (i.e., were the statistical conclusions valid, were the conclusions internally and externally valid, are there other explanations for the findings, which limitations were and were not discussed by the authors)
- Recommendation on how this or future studies could be improved

Assignment requirements (Old)

The written assignment will be evaluated (pass vs. fail) against following criteria:

- Structure: The overall report contains a title page and 3 reports on empirical studies from the list below. The 3 reports contain the subsections (1) background, (2) research question, (3) methods, (4) findings, (5) discussion, and (6) literature list.
- Scope: Each of the three reports must be limited to 4-5 pages. The overall report is therefore limited to 12-15 pages in addition to the title page.
- Readability: The text is in correct and clear, scientific English. Sections have appropriate headings and all abbreviations are introduced.
- Format: The format follows the APA standard (see https://apastyle.apa.org/instructional-aids/student-paper-setup-guide.pdf)
- File format: The written assignment is submitted as one PDF.


APA STYLE

7th Edition

Student Paper Setup Guide

This guide will help you set up an APA Style student paper. The basic setup directions apply to the entire paper. Annotated diagrams illustrate how to set up the major sections of a student paper: the title page or cover page, the text, tables and figures, and the reference list.

Basic Setup

Seventh edition APA Style was designed with modern word-processing programs in mind. Most default settings in programs such as Academic Writer, Microsoft Word, and Google Docs already comply with APA Style. However, you may need to make a few adjustments before you begin writing.

- **Margins:** Use 1-in. margins on all sides of the page (top, bottom, left, and right).
- **Font:** Use a legible font. Many fonts are acceptable, including 11-point Calibri, 11-point Arial, 12-point Times New Roman, and 11-point Georgia. The default font of your word-processing program is acceptable.
- **Line spacing:** Use double-spacing for the entire paper (including block quotations and the reference list). Do not add blank lines before or after headings. Do not add extra spacing between paragraphs.
- **Paragraph alignment and indentation:** Align paragraphs of text to the left margin. Leave the right margin ragged. Do not use full justification. Indent the first line of every paragraph of text 0.5 in. using the tab key or the paragraph-formatting function of your word-processing program.
- **Page numbers:** Put a page number in the top right corner of every page, including the title page or cover page, which is page 1. Student papers do not require a running head on any page.

Assignment requirements (Old)

- Feedback
  - Pro:
    - Great exercise
  - Con:
    - Writing was too time consuming; simultaneous with other written assignments
- This year proposal (oral exam):
  - Pro:
    - Less writing (only presentation slides)
    - Same content to the exam
    - Faster feedback
  - Con:
    - You **have to** be at UiO for the exam (written assignment could be sent from anywhere with internet connection)

Oral Exam Requirements

November 9

- Goal:
  - Enable students to **demonstrate scientific rigor** and **solid knowledge** about the **principles** of causal inference and the **assumptions, prerequisites, and procedures** of methods for causal inference.
- Core material:
  - 1 (chosen by the instructors) of 3 papers (chosen by the students) from the list in the course description (in Canvas)
  - **Notice that you still have to draft three presentations (but only make one)!!**

- Presentation (hard 20 minutes):
  - Background and motivation
  - Causal question
  - Critical assessment of methods
  - Critical assessment of results and discussion
  - Future studies recommendations
- Questions (10 minutes)
  - Clarification questions
  - Principles, assumptions, prerequisites and procedures

Introduction to Causal Inference

What is the causal effect of something on something else?

$$D \xrightarrow[Causal\ effect]{} Y$$

- Cause (variations in D)
  - Can be insufficient (i.e., require additional conditions)
  - Can be unnecessary (i.e., other causes would lead to the same effect)
  - Can occur only under some conditions (i.e., inference at issue)
- Effect (variations in Y)
  - Difference between what happens and *what did not happen*
- Causal relationship (defines the causal effect)
  - Requires cause to precede the effect
  - Requires cause to be related to the effect
  - Requires that no plausible alternative explanation (we rarely know → not deterministic)

Campfire example

Introduction to Causal Inference

What is an experiment?

- Deliberate variations of something, D, so as to discover what happens (*causal effect*), on something else later, Y.
- Experimentation involve:
    1. Manipulation of the presumed cause and observation of the outcome afterward
    2. Observation of whether the variation in the cause is related to variation in the effect
    3. Use of specialized methods to reduce the plausibility of other explanations for the effect

Photo by Akram Huseyn on Unsplash

Introduction to Causal Inference

What is an experiment?

- Deliberate variations of something, D, so as to discover what happens (*causal effect*), on something else later, Y.
- Experimentation involve:
  1. Manipulation of the presumed cause and observation of the outcome afterward
  2. Observation of whether the variation in the cause is related to variation in the effect
  3. Use of specialized methods to reduce the plausibility of other explanations for the effect

- Causal relationship (defines the causal effect)
  - Requires cause to precede the effect
  - Requires cause to be related to the effect
  - Requires that no plausible alternative explanation (we rarely know → not deterministic)

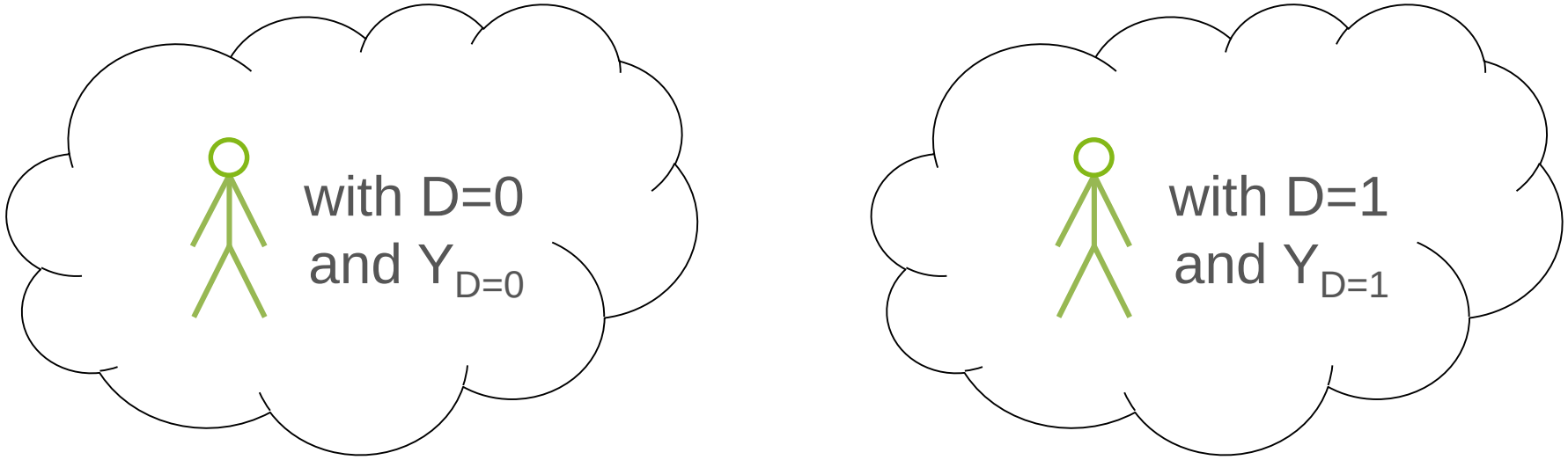
Rubin's Potential Outcome Framework

- Say we are interested on whether a treatment, D, changes an outcome, Y.
- D can take two values (before the outcome):
  - $D = 0$ (no treatment / control / comparison)
  - $D = 1$ (treatment).
- If there is a causal effect
  - $Y_{D=0} \neq Y_{D=1}$

What does that mean?

What is the causal effect?

Rubin's potential outcome framework



- The impossible, ideal experiment
  - In parallel universes where the same individual is once treated and once not treated, nothing differs except the treatment and outcome (ceteris paribus/other things equal)
  - What would have been (potential outcome or counterfactual)?
- Then, the difference $Y_{D=0} - Y_{D=1}$ is the causal effect of D on Y

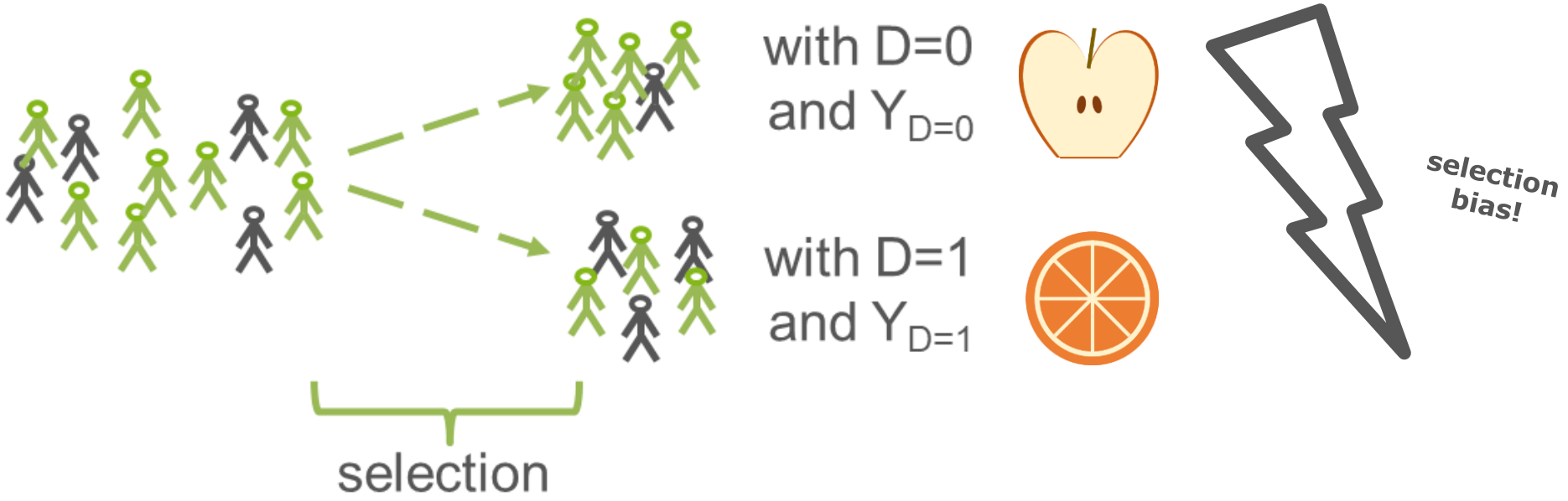
Rubin, 1974

Rubin's potential outcome framework

- In reality, we need to compare groups
  - Difference between treatment and control/comparison group only reflects effect of D, if other things are equal (*ceteris paribus*)
  - If selection into groups relates to something that is also relevant for the outcome, difference between groups reflects causal effect and *selection bias*

Rubin, 1974

Rubin’s potential outcome framework

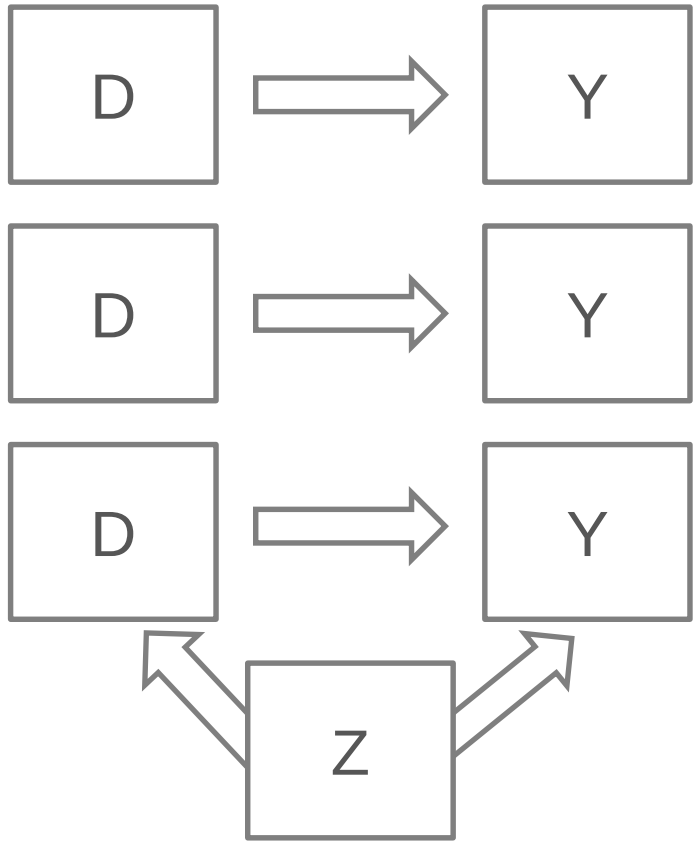


Example:

Negative correlation between private tutoring and mathematics achievement

Correlation ≠ Causation (!)

- Why does correlation not prove causation?
  - maybe we don't know which variable came first (e.g., *reverse causation*)
  - maybe, there is no true association, at all (e.g., *spurious correlation*)
  - maybe, there are other explanations (e.g., *third-variable effect*, *confounding*, *selection bias*)



Summing up

Take away messages

- Causal Inference is possible if
  - ✓ Plausible causal mechanism
  - ✓ Treatment before outcome
  - ✓ Comparison with counterfactual
  - ✓ Ceteris paribus
- Very good counterfactuals and ceteris paribus are difficult to establish
- Multiple issues prohibit causal inferences, for example:
  - ✓ Selection bias
  - ✓ Reverse causation
  - ✓ Third-variable bias (or confounding effects)

Formalities and Schedule

1	Introduction - Time: 16 August 2022, 12:15-14:00h - Main instructor: José Manuel Arencibia Alemán - Required reading: Shadish, Cook, & Campbell (2002), chapter 1
2	Designing Experiments - Time: 19 August 2022, 12:15-14:00h - Main instructor: Tony Tan - Required reading: Shadish, Cook, & Campbell (2002), chapter 8
3	Randomized Trials - Time: 23 August 2022, 12:15-14:00h - Main instructor: José Manuel Arencibia Alemán - Required reading: Angrist & Pischke (2015), chapter 1
4	Regression Models - Time: 26 August 2022, 10:15-12:00h - Main instructor: Tony Tan - Required reading: Angrist & Pischke (2015), chapter 2

References

Angrist, J. & Pischke, J.-S. (2015). *Mastering `Metrics: The Path from Cause to Effect*. Princeton University Press

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

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Mifflin and Company