Causal Inference in Introductory Statistics Courses

Kevin Cummiskey, Bryan Adams, James Pleuss, Dusty Turner, Nicholas Clark, Krista Watts

Department of Mathematical Sciences, West Point, NY

Introduction

- Despite advances in causal inference and changes in introductory courses, causality is rarely discussed in introductory courses except for warning students “correlation does not imply causation” or mentioning the special case of randomized experiments.
- We argue investigating causal applications and introducing topics associated with causal inference, especially causal diagrams, support statistics education guidelines for introductory courses.
- In addition, we provide an activity appropriate for introductory courses.

GAISE Guidelines

- Develop statistical thinking (especially multivariable thinking).
- Give students experience with the investigative process.
- Use real data with research questions of interest to students.
- Focus less on calculations, mathematical derivations, and probability theory.

What is causal inference?

- The goal of causal inference is to perform inference under changing conditions, such as those induced by treatments or external interventions.
- Causal inference requires assumptions about the causal relationships between variables.
- Investigators specify these causal assumptions a priori using domain knowledge.
- They frequently use graphical aids called causal diagrams to facilitate this process.

How does causal inference support GAISE?

- Students must explicitly state their assumptions about the causal relationships among variables in a study.
- This process gives them valuable experience with multivariable thinking and exposure to the complex relationships that often exist in real-world data.
- Causal diagrams provide a visual tool for structuring their multivariable thinking.

What topics to discuss in intro courses?

- Difference between associational and causal relationships
- Confounding
- Causal diagrams
- Methods for confounding adjustment

Causal diagrams

- In causal inference, researchers specify assumptions about the causal relationships between variables a priori. Causal diagrams are useful tools for depicting these assumptions.
- Causal diagrams are directed, acyclic graphs. Each node is a variable. An edge between nodes indicates the parent node is a cause of the child.
- No edge between nodes indicates the variables are independent.
- In the diagrams below, A is the treatment and Y is the outcome.

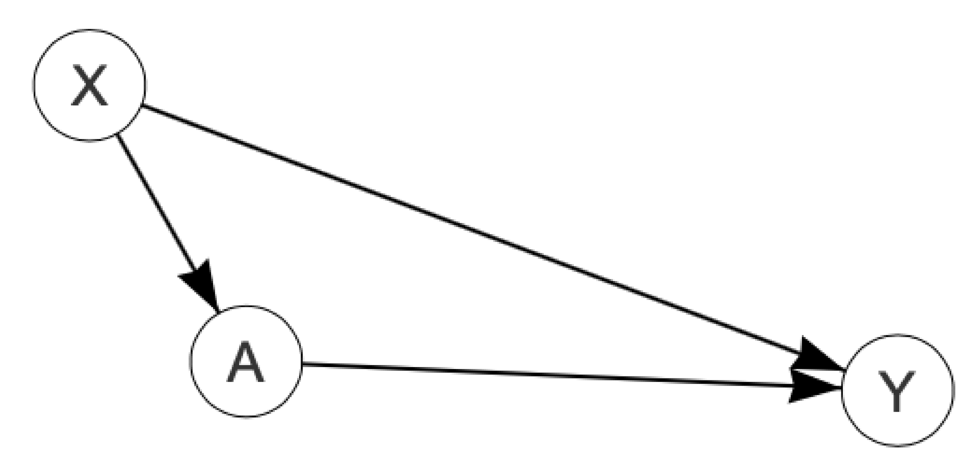

Figure: Confounding. A and Y with common cause X.

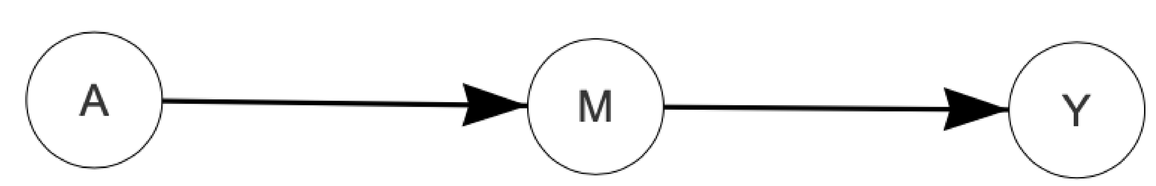

Figure: Mediation. The effect of A on Y is mediated by M.

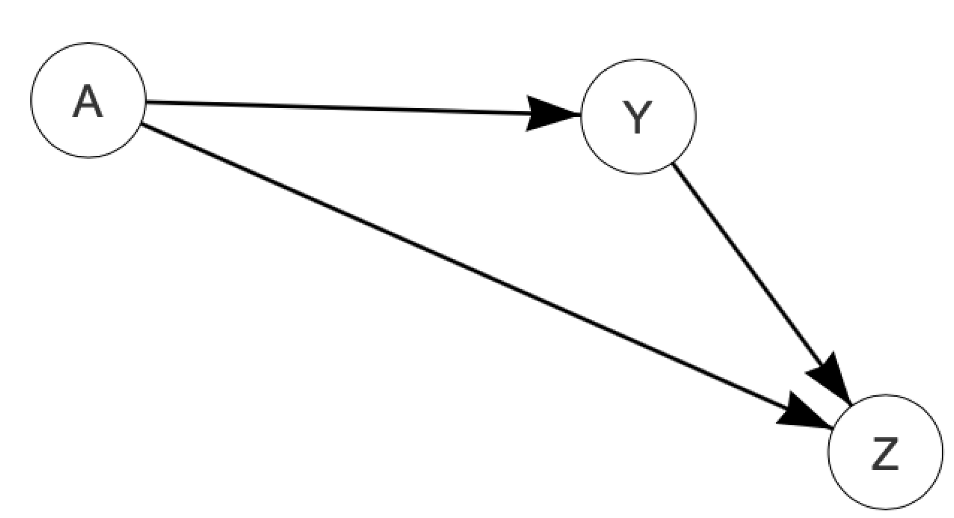

Figure: Collider. A and Y are both causes of Z.

Student activity

- Our student activity and instructor resources are available at `https://github.com/kfcaby/causalLab`.
- Investigates the health effects of youth smoking using data from Kahn 2005.
- The guided lab requires students to develop a causal diagram like the one below, defend their assumptions, fit an appropriate statistical model based on their diagram, and discuss their results.

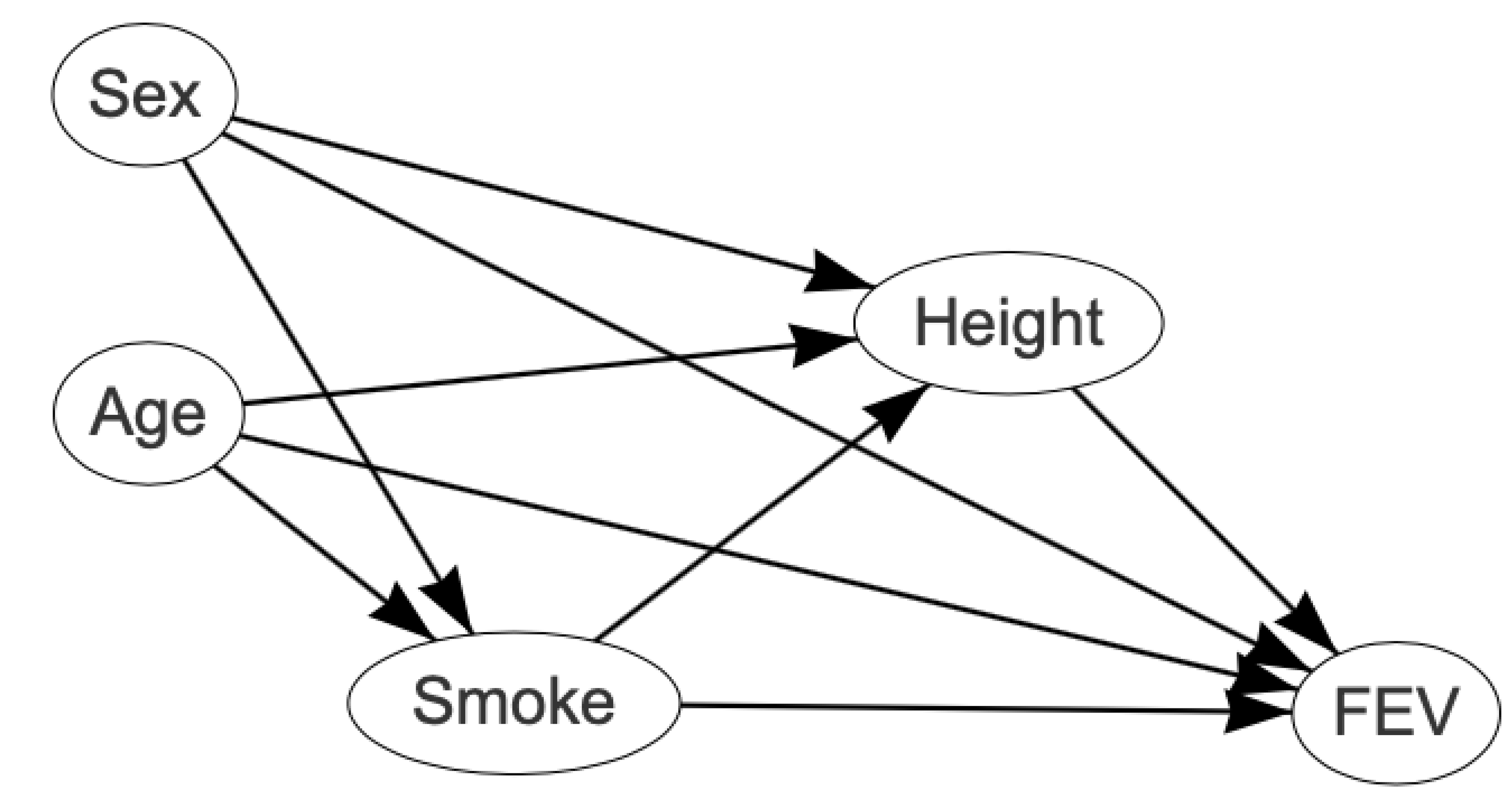

Figure: Causal diagram depicting causal relationships between variables in the youth smoking study.

Resources and References

1. Carver, Robert, et al. “Guidelines for assessment and instruction in statistics education (GAISE) college report 2016." (2016).
2. Pearl, J. . “Causal diagrams for empirical research." Biometrika, 82(4):669âĂŞ688, (1995).
3. Pearl, Judea, and Dana Mackenzie. *The book of why: the new science of cause and effect.* Basic Books, (2018).
4. Kahn, Michael. “An exhalent problem for teaching statistics." Journal of Statistics Education 13.2 (2005).
5. Hernan, M. and Robins, J. *Causal Inference.* Boca Raton: Chapman and Hall/CRC, (2018).
6. Online course by Miguel Hernan. “Causal diagrams: Draw your assumptions before your conclusions. Available at `www.edx.org`.