UNITED STATES MILITARY ACADEMY
WEST POINT®

Multivariable thinking, confounding, and causal inference in intro stats

Kevin Cummiskey

Department of Mathematical Sciences, West Point

Joint Statistical Meetings
August 3, 2020

Views expressed in this presentation are mine and do not represent the official position of the U.S. Department of Defense, the Army, or West Point.

Multivariable thinking

"We must prepare our students to answer challenging questions that require them to investigate and explore relationships among many variables." (revised GAISE college report)

"All students also should receive experience with multivariable thinking in the introductory course." (Wood et al., 2018)

"The lack of appreciation for simple multivariable methods is a major limitation in too many of our courses." (Horton, 2015)

Multivariable thinkers

Consider a multivariable thinker in your life.

- *What makes this person a multivariable thinker?*
- *How do we make our students think more like this person?*

Lieutenant General Tom Vandal, a multivariable thinker in my life, briefing Ms. Kang Kyung-hwa, the Foreign Minister of South Korea.

Multivariable thinkers

How do they think?

- Appreciate complex relationships may exist among several variables.
- Have an intuitive understanding of association, causation, interactions, and confounding.
- Recognize the role of study design in making appropriate conclusions.
- Develop multiple “causal stories” consistent with observed data.
- Determine additional information required to refine their “causal stories”.

Multivariable thinkers

What can they do?

- Interpret multivariable visualizations
- Recognize when stratification would be helpful
- Compare summary statistics of appropriate conditional distributions of data

The distinction between multivariable thinking and multivariable statistical methods is important.

Multivariable thinkers

How do we develop them?

- Base of knowledge
- Add additional variables (and do so right away!)
- Stratification
- Multivariable methods (multiple regression)
- Causal diagrams
- Practice!

"Stratification requires no advanced methods, nor even any inference, though some instructors may incorporate other related concepts and approaches such as multiple regression." (revised GAISE college report)

Causal Diagrams

Key Idea: Causal diagrams contain all common causes of the treatment and outcome.

Figure 1. A is a cause of Y. They have no common causes.

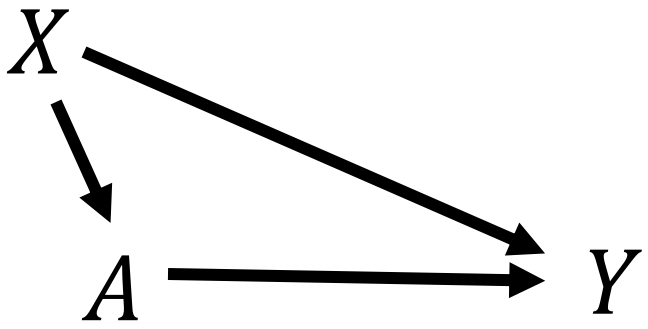


Figure 2. Confounding: X is a confounder of the effect of A on Y.

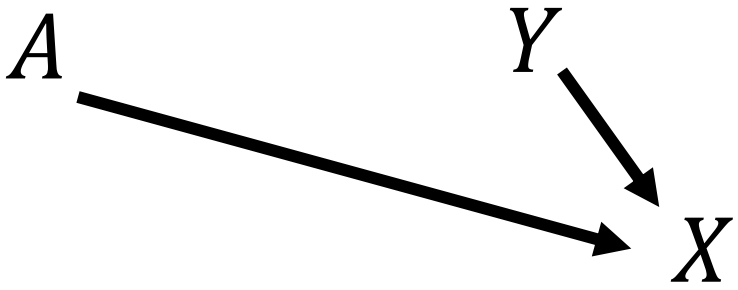


Figure 3. Collider - A and Y are common causes of collider X.

Figure 4. X is a mediator of the effect of a A on Y.

A "small teaching" example

Class Exercise

Time Required: 15-20 minutes

Instructor selects an advertisement that makes a causal claim based on statistical results.

Students:

1. Identify the causal claim in the advertisement.
2. Discuss and draw causal diagrams relating variables relevant to the claim.
3. Write short paragraphs describing “causal stories" consistent with the reported result.
4. Identify additional information required to assess the claim.

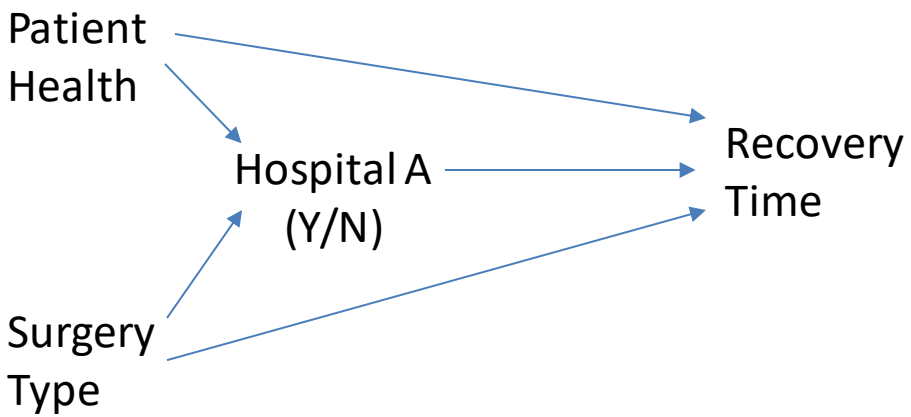
Repeat this exercise 3-4 times over the semester with different contexts.

A "small teaching" example

An advertisement for a Hospital A in Westchester County states they have the shortest recovery times from surgery in New York.

1. Causal claim: If I have surgery at this hospital, my recovery time will likely be shorter than if I have surgery at another hospital.

2. Causal diagram:



A "small teaching" example

3. <u>The following causal stories are consistent with the advertisement</u>:
   i. In the long run, there is no difference in average recovery times between Hospital A and other hospitals. The observed difference is due to sampling variability. ***(random chance)***
   ii. Hospital A has more highly trained personnel, better procedures, or more modern equipment resulting in shorter recovery times. ***(causation)***
   iii. The patients at Hospital A are healthier in general or get different types of surgeries than other hospitals resulting in shorter recovery times. ***(association)***

4. <u>Additional information required</u>:
   – Health characteristics
   – Type of surgery

See Lock Morgan 2019/2020 for a framework for causality in introductory statistics.

UNITED STATES MILITARY ACADEMY
WEST POINT

An example with data

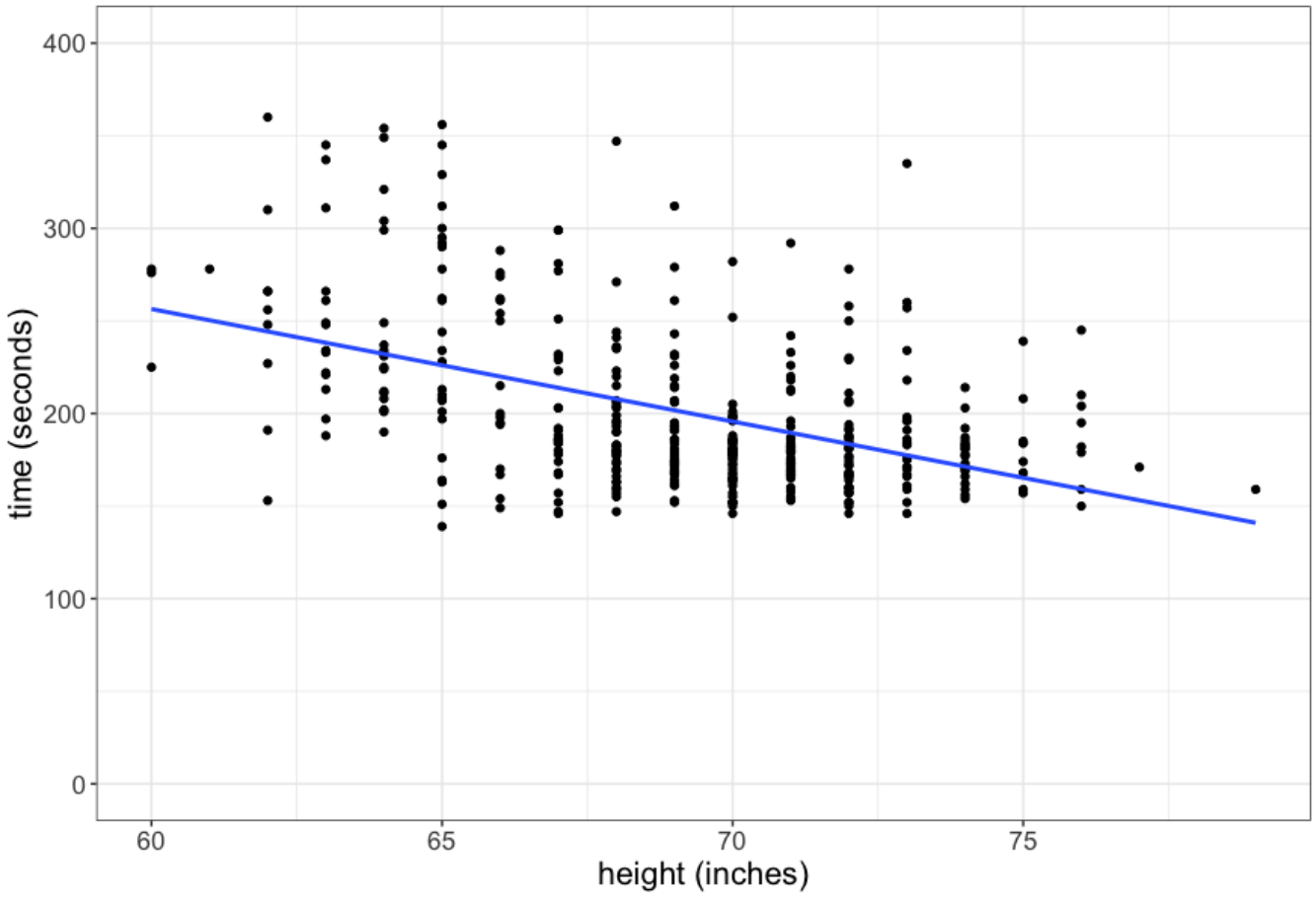
The West Point Indoor Obstacle Course (video here)

Research Question: *Is being taller an advantage*?





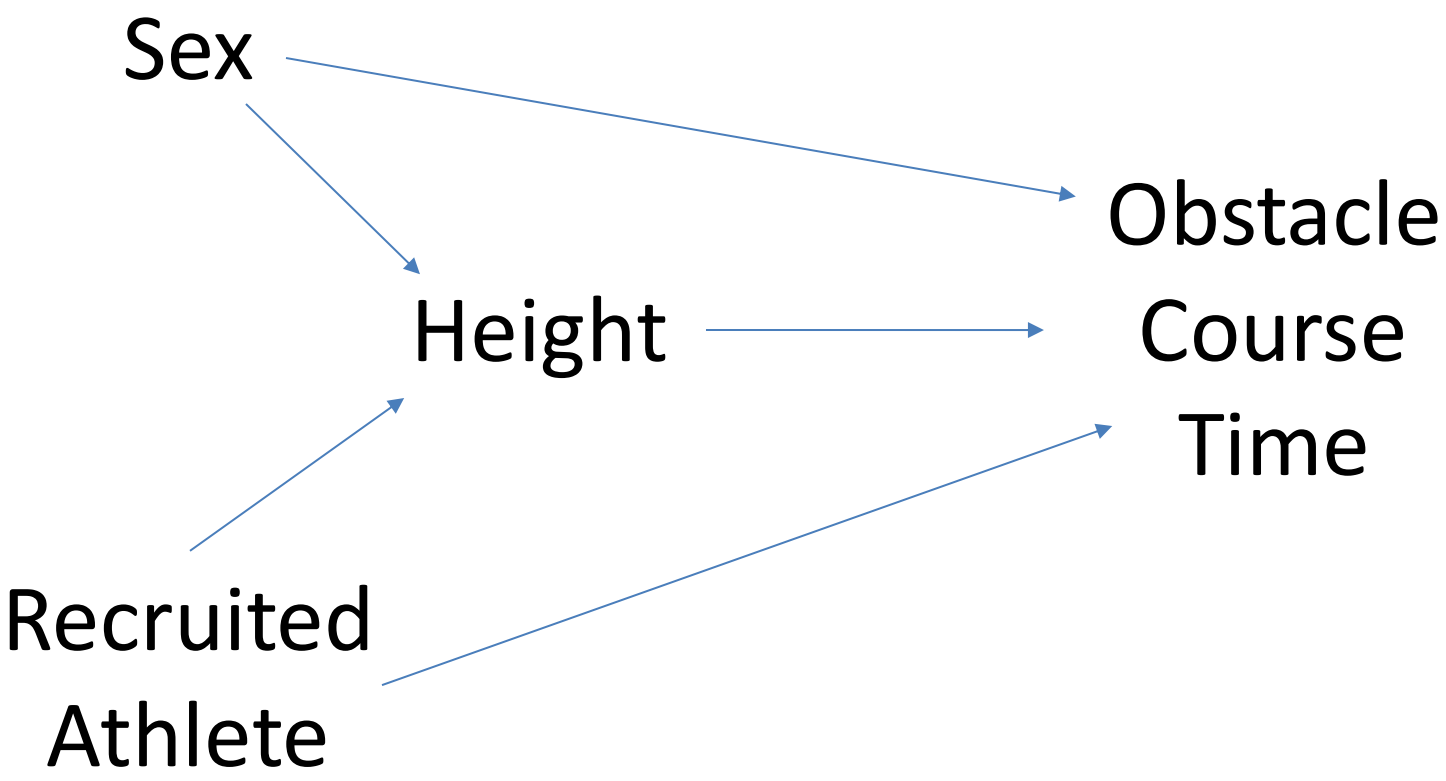

Is being taller an advantage?



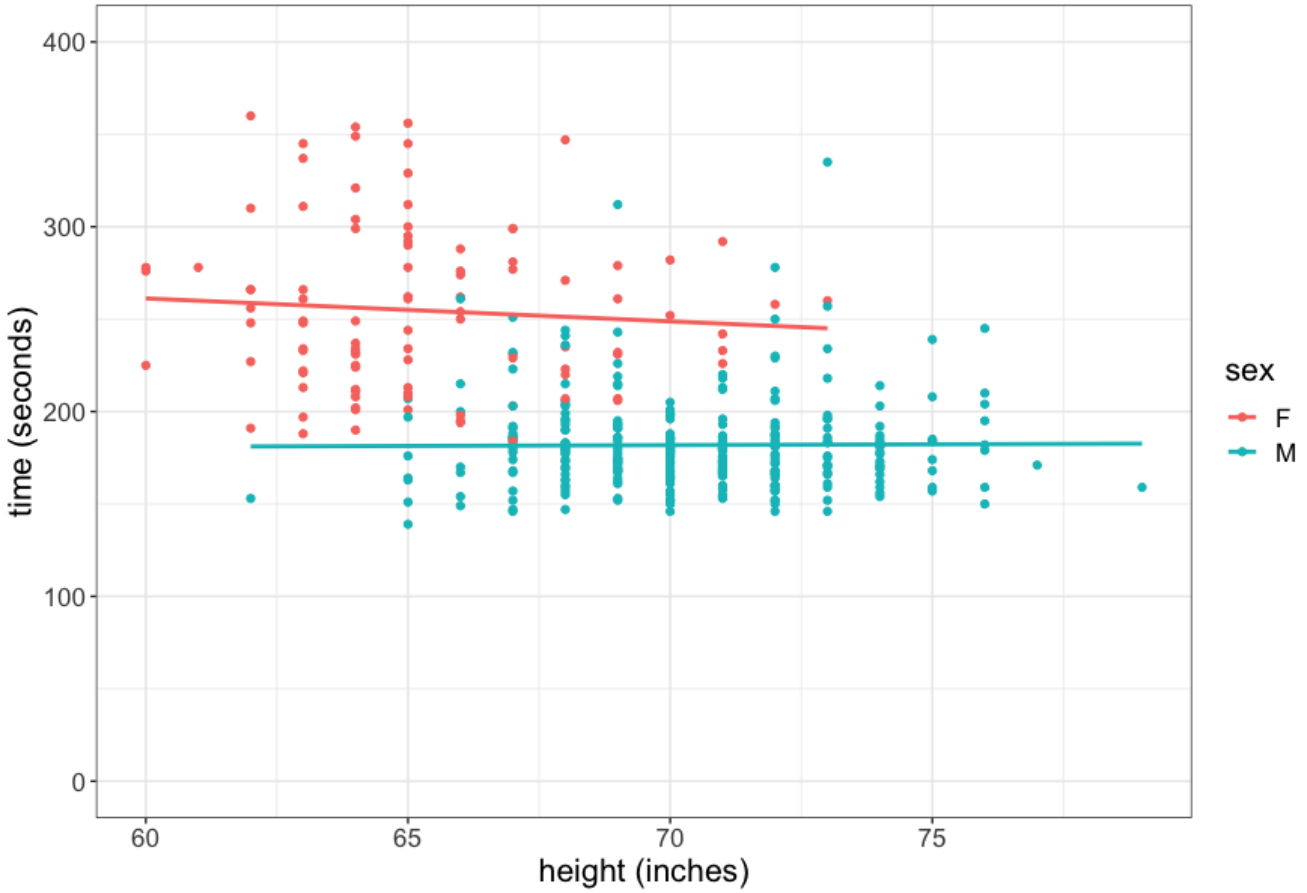
- An inch of height is associated with a 7.0 second (95% C.I.: 5.6, 8.4) faster time.
- This is an importance difference – about a half a letter grade per inch of height.
- *Why is being taller an advantage?*

Credit: Bryan Adams

Causal diagram



Is being taller an advantage?



- Being taller is not an advantage once we look within strata of sex.

"It's important to have students look for possible confounding factors when the relationship isn't what they expect, *but it is also important when the relationship is what is expected*." (revised GAISE college)

Another example – gender and pay

Data: 2016 Current Population Survey (Bureau of Labor Statistics)

Applet: https://army-math-west-point.shinyapps.io/CEPRVisualizer/

Student Activity:

- Read scientific article on the gender wage gap.
- Investigate earnings data for 180,000 individuals.
- Consider relationship between gender and earnings.
- Materials currently under development.

Credit: Rob Lasater and Anny-Claude Joseph

Key Ideas

- Multivariable thinking is important to develop in introductory statistics.
- Multivariable thinking and multivariable methods are different.
- We can make small changes to our existing courses to develop multivariable thinking.
- Causal diagrams are useful for structuring multivariable thinking.

References

1. Carver, Robert, et al. "Guidelines for assessment and instruction in statistics education (GAISE) college report 2016." (2016).
2. Cummiskey, Kevin, Bryan Adams, James Pleuss, Dusty Turner, Nicholas Clark, and Krista Watts. "Causal Inference in Introductory Statistics Courses." *Journal of Statistics Education* 28, no. 1 (2020): 2-8.
3. Cummiskey, Kevin and Bryan Adams. "Causal Inference in Introductory Statistics: Why, What, and How?" CAUSEweb. https://www.causeweb.org/cause/webinar/teaching/2019-10
4. Greenland, Sander, Judea Pearl, and James M. Robins. "Causal diagrams for epidemiologic research." *Epidemiology* 10 (1999): 37-48.
5. Hernan, Miguel, and J. M. Robins. "Causal inference book" (2017) https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/
6. Hernan, Miguel. "Causal Diagrams: Draw your assumptions before your conclusions" Online course available at https://www.edx.org/course/causal-diagrams-draw-your-assumptions-before-your-conclusions.
7. Horton, Nicholas J. "Challenges and opportunities for statistics and statistical education: looking back, looking forward." *The American Statistician* 69, no. 2 (2015): 138-145.
8. Lock Morgan, Kari. "Teaching introductory students how to evaluate evidence." USCOTS 2019 (https://www.causeweb.org/cause/uscots/uscots19/keynote/4)
9. Lock Morgan, Kari. "Two questions to ask before making causal conclusions." Guest post on Allan Rossman's *Ask Good Questions.* (https://askgoodquestions.blog/2020/07/27/56-questioning-causal-evidence/)
10. Pearl, Judea. "Causal diagrams for empirical research." *Biometrika* 82.4 (1995)
11. Tintle, Nathan, et al. *Introduction to statistical investigations*. John Wiley & Sons, 2015.
12. Wood, Beverly L., Megan Mocko, Michelle Everson, Nicholas J. Horton, and Paul Velleman. "Updated guidelines, updated curriculum: the GAISE college report and introductory statistics for the modern student." *Chance* 31, no. 2 (2018): 53-59.

Questions?

"If students do not have exposure to simple tools for disentangling complex relationships, they may dismiss statistics as an old-school discipline only suitable for small sample inference of randomized studies." (revised GAISE college report)