

Asking a Causal Question

STSCI / INFO / ILRST 3900

9 Sep 2026

Course Project

- ▶ Opportunity to engage with the course content via a real-world example
- ▶ Students in class come in with different backgrounds and skills
- ▶ The project will require various skill sets
 - ▶ Asking a causal question
 - ▶ Reasoning about assumptions
 - ▶ Data Analysis
 - ▶ Communicating Results

Overview

Pieces of the project will be completed throughout the semester

Individual Tasks

- ▶ Define a causal question
- ▶ Draw and reason about a causal graph

Group Tasks

- ▶ Gather and process relevant data
- ▶ Select an identification strategy and analyze the data
- ▶ Communicate the results with short report

More details for each task will be posted on the course website

Asking a Causal Question

- ▶ The first step in answering a causal question is to be precise about the question you want to answer
- ▶ To answer the question, we must ensure appropriate data is available
- ▶ Two options for data sources:
 1. Use the Add Health Data set and formulate a question using its variables
 2. Come up with a question independently and locate an appropriate data source

Add Health Data

- ▶ [National Longitudinal Study of Adolescent to Adult Health](#)
- ▶ Survey of 20,000 high-school aged participants in 1994-1995
- ▶ Followed up at 4 subsequent time points (as recently as 2018, ages mid-20s to early 30s)
- ▶ **Topics covered:**
 - ▶ Crime/Delinquency and Victimization, Demographic Characteristics, Education, Family, Medication and Substance Use and Abuse, Psychological Well-being and Cognition, Reproductive Health, SES, Labor Market and Occupation
- ▶ Find a more comprehensive list of topics and specific survey questions [here](#)
- ▶ To get an idea of the types of questions people have used the data to answer, take a look at journal articles citing Add Health [here](#)

In-Class Assignment

Today we will practice:

- ▶ Asking a causal question
- ▶ Defining the treatment and outcome variables

This is NOT part of the project

- ▶ You will pick a causal question for the project later

Defining Variables

- ▶ A **treatment variable** represents the variable used to indicate which treatment an individual receives
- ▶ A **treatment** is a specific category or value of the treatment variable that is assigned to an individual
- ▶ If your treatment is a variable that can take many different values, you could consider making it binary

Example

If the **treatment variable** is the number of hours spent studying each day, the **treatment** would be the specific number, such as 0, 1, 2, 3, or 4 hours of studying per day. You can simplify by dichotomizing the treatment in the following way:

$$A_i = \begin{cases} 0 & \text{if Hours} \leq 2 \\ 1 & \text{if Hours} > 2 \end{cases}$$

Then the **treatment** would be studying more than 2 hours per day.

Action Items

Question 1:

- ▶ Ask a causal question so someone outside this class would understand
- ▶ Why are you interested in this question?
- ▶ How could answering this allow better decision-making?

Question 2:

- ▶ Identify the treatment and outcome variables
- ▶ Write out potential outcomes using class notation

Question 3:

- ▶ How does the fundamental problem of causal inference apply directly to your question?

Make sure to hand in the assignment before leaving!