

Quantitative Reasoning – LAB 413

Causation, Experimental Design & Sampling

Week 5 Recap ■ Lab 4

Instructor: **Subin Na**

September 24, 2025

Today's Plan

1. Quick Recap

- Sources of Data
- Causation vs. Correlation
- What is Causality?
- RCTs

- Problems with Experiments
- Sampling
- Survey Cautions

2. RStudio Practice

01 Sources of Data: Overview

- **Anecdotal data:** individual striking cases (often outliers, not representative).
- **Observational data:** existing data collected for other purposes.
- **Sample surveys:** collecting opinions from representative samples.
- **Experiments:** deliberately imposing treatments to measure causal effects.

02 Causation vs. Correlation

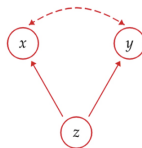
Key Problem

Association $\neq$ Causation. Three explanations for an X - Y correlation:

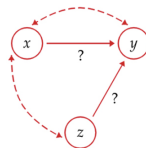
- **Direct causation:** X causes Y .
- **Common response:** Z causes both X and Y .
- **Confounding:** X causes Y , but Z also affects Y and correlates with X .



Causation



Common response



Confounding

Solution: well-designed experiments can avoid confounding.

03 What is Causality?

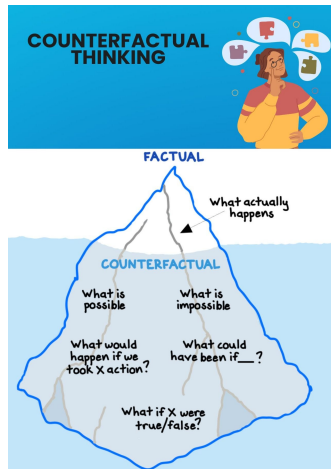
Fundamental Problem

We can never observe counterfactuals.

- **Factual:** what actually happened.
- **Counterfactual:** what would have happened otherwise.

$$\text{Causal Effect} = Y(1) - Y(0)$$

- $Y(1)$: outcome if treated.
- $Y(0)$: outcome if not treated.



04 Randomized Experiments (RCTs)

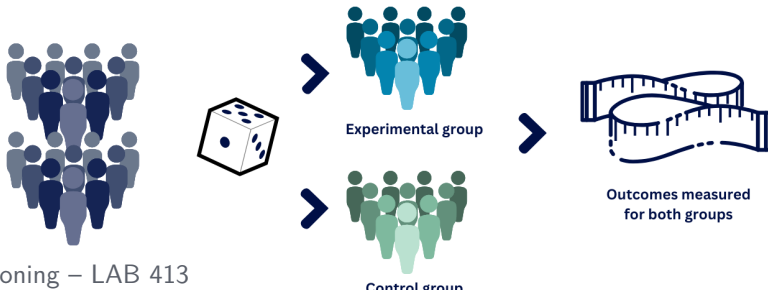
Key Components

- **Treatment:** condition applied to experimental units.
- **Control group:** valid counterfactual.
- **Random assignment:** treatment determined by chance.

Why randomize?

- Creates balance between treatment / control groups.
- Groups identical “on average” (observable + unobservable characteristics).
- Only difference is the treatment effect.

Randomised controlled trial



05 Problems with Experiments

- **Selection bias:** voluntary participation creates unrepresentative samples.
- **Placebo effects:** responding to attention rather than treatment.
- **Hawthorne effects:** acting differently when observed.
- **External validity:** may not reflect real-world conditions.

Remedies

- Double-blind experiments.
- Balance checking on pre-treatment variables.

06 Sampling (1)

Key Concepts

Making inferences from a small group to a larger population.

- **Population:** complete set we want to learn about.
- **Sample:** part we actually collect data from.

Good sampling

- **Simple Random Sampling (SRS):** everyone has an equal chance of being selected.
- **Stratified Sampling:** divide the population into groups, then randomly sample within each group.
- **Multistage Sampling:** sample in stages (e.g., cities → neighborhoods → households).



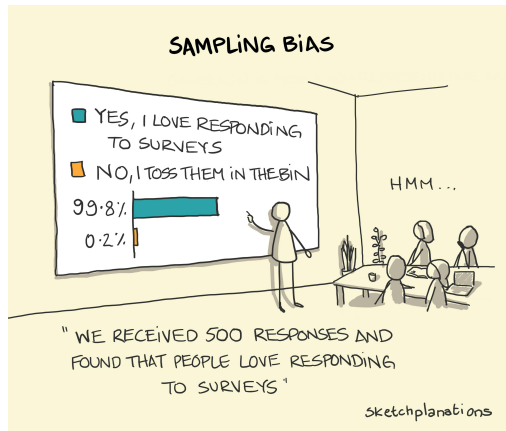
06 Sampling (2)

Bad sampling

- **Convenience:** just picking whoever's easy to reach.
- **Voluntary response:** letting people self-select into your study.

Key Principle

The sample needs to be representative of the target population!



07 Survey Cautions

- **Undercoverage:** missing population groups.
- **Nonresponse:** people don't participate.
- **Response bias:** systematic incorrect responses.
- **Question wording:** the most important influence on responses.

“Should we reduce government spending?” vs. “Should we cut funding for education and healthcare?”

Always pilot test surveys!

RStudio Practice

School Data: Summaries, “Top” Schools & Regression

See `Lab4.Rmd` / `Lab4_0924.R`