Additional materials for the Sabina's stats series lecture "Who counts? Denominator, bias and illusion of equity"

Lecture recording (April 26, 2026): https://www.youtube.com/watch?v=zjakvEWA6WY

Summary:

During the lecture we talked about Confounding, Causal inference, Reuben Causal model, Fundamental Problem of Causal Inference.

This document includes brief description of key elements used in the lecture to help you understand it better.

Data dictionary:

- *What is counterfactual?* It is a "what-if" statement about a scenario that did not actually occur, means it is literally "contrary to the facts".
  In causal inference, it refers to the potential outcome for an individual under a version of the world that we did not observe.
- *What is observed?* It refers to the actual, realized data that is physically recorded or measured in the real world.
- *What is true effect?* It is simply the difference between the observed outcome and the counterfactual outcome for the same individual
- *What is confounding?* It is a form of systematic bias that occurs when an extraneous variable (the confounder) is associated with both the treatment and the outcome, making it look like the treatment caused the outcome when it was the third variable doing the work.

Rubin Causal Model (RCM)

The core principals of this theorem included in the lecture:

- The Counterfactual Question
- Fundamental Problem of Causal Inference
- True effect
- Missing Data Perspective
- Shift to Study Design

Link to original paper:
https://www.academia.edu/93080536/Estimating_causal_effects_of_treatments_in_randomized_and_nonrandomized_studies

Donald Rubin's landmark 1974 paper didn't just provide a new formula; it provided a new way of seeing. It shifted the focus of research from describing what happens to investigating what might have been.

At its core, Rubin's framework asks a deceptively simple question:

What would have happened to the same individual if, under identical conditions, they had both received and not received the intervention?

Of course, we can never observe both realities simultaneously. One outcome is observed, while the other remains counterfactual, the unobserved potential outcome.

1. *The Counterfactual Question: A Tale of Two Realities*

   At its heart, the framework relies on a concept called the Counterfactual. Imagine an individual at a crossroads. To understand the "True Effect" of a choice, we need to see two parallel versions of that person:

   - The version who received the intervention.
   - The version who, under identical conditions, did not.

   The True Effect is simply the difference between these two potential outcomes. It is the pure impact of the intervention on that specific person.

   You can look at it as seeing yourself in the "mirror universe", a version of yourself that reflects every characteristic, history, and biological trait, but exists in an alternate reality where the treatment was withheld.

This transformed causal inference from philosophical discussion into a statistical problem of missing data, usually called the ***Fundamental Problem of Causal Inference***.

Instead of asking:
"Are groups different?"

We ask:
"What would the outcome have been under an alternative exposure state?".

But this information is missing, we only live in one timeline. We can observe the "factual" outcome, but the "counterfactual" remains forever hidden. This realization transformed causal inference from a philosophical debate into a Missing Data Perspective. Statistics, under this framework, is essentially an exercise in "filling/estimating in the blanks" for the reality we cannot see.

In statistics, one of the ways to make the "counterfactual" reality more comparable is to adjust to relevant and available characteristics.

In the lecture I gave example of tutoring program and student performance.

In this example we need to distinguish between the

- The Program Effect: The actual knowledge the tutor provides.
- The Selection Effect: The natural advantage motivated students already have (they study more, attend more, etc.).

The question is how we can measure the confounding "motivation". We can look at the Prior Attendance Rates, Past Academic Performance (GPA), Early Submission Habits as possible proxies for the motivation.

Rubin's framework fundamentally reorients the researcher's focus. It moves the priority away from the mathematical mechanics of the statistical model and toward the logic of study design and the assignment mechanism.

Consider a study on a Tutor Program. A simple regression might show that students with tutors get higher grades. But a regression model alone does not create causality. The framework forces us to step back and ask the Causal Question first: Would these specific students have performed differently if they hadn't had the tutor?

We need to look at multiple things at once.

- Treatment Assignment: Why did some students get a tutor? If "highly motivated" students are the ones seeking help, then Motivation is the engine assigning the treatment.

- Exchangeability: Are the groups "swappable"? If we swapped the students, would the "No Tutor" group have performed just as well as the "Tutor" group? If the answer is no (because they differ in motivation or baseline ability), they are not exchangeable.

- Confounding: Motivation acts as a "mixing" factor. It influences both the choice to get a tutor and the final test score. Without addressing this, we can't tell where the student's drive ends, and the tutor's help begins.

- Construction of Valid Counterfactuals: We use Adjustment to find a "mirror" for our treated students. By adjusting for pure covariates (like age or baseline IQ) and motivation proxies (like prior attendance), we attempt to "fill in the blanks" for the missing counterfactual. This is why the same regression equation can be predictive in one study, descriptive in another, and causal in a third. It depends entirely on whether your design assumptions have successfully isolated the "True Effect" .

- In this framework, the "True Effect" is the "Ground Truth" or the "Gold Standard" we are trying to hit. It exists in theory, but because of the Fundamental Problem of Causal Inference, it is physically impossible to observe

- We usually settle for estimating the "Average Treatment Effect" (ATE). Because we only observe one potential outcome for everyone, the individual-level true causal effect is fundamentally unobservable and is the missing counterfactual. To address this, we compare groups that are exchangeable, meaning they are comparable in all relevant ways except for the treatment itself. By using the observed outcomes in one group to approximate the missing potential outcomes in the other, we can estimate the Average Treatment Effect (ATE).

- For researchers in healthcare and social sciences, this framework remains foundational because it forces the ultimate question: Does our design truly support the causal claim, or are we just measuring the "Selection Effect" of who walked through the door?

How this related to equity/fairness in the research?

The Rubin Causal Model is closely connected to equity and fairness research because it forces us to ask whether the groups being compared are truly comparable. Observed differences between populations may reflect confounding, unequal access to care, referral patterns, measurement differences, or structural bias rather than a true causal effect. The framework reminds researchers that fairness is not simply achieved through statistical adjustment or mathematical models. It depends on valid study design, appropriate denominator definition, exchangeability between groups, and careful construction of meaningful counterfactual comparisons. In this sense, equity research is not only about measuring differences, but about understanding whether those differences reflect true effects or the underlying structures that shape who become visible in the data.