DRP Final Report

This quarter, broadly speaking, I learned an introduction to casual inference and techniques used in casual inferencing to remove the effects of confounders. I began learning about the foundational problem of casual inference, which is the reconciliation of the fact that you can't observe what would have been the response of a unit had it been treated and untreated at the same time. We live in the factual world where something may have been treated or not but in the counterfactual world, we don't know what would've happened if the effect was the opposite.

In assessing for a conclusion on whether a treatment has some sort of effect on another variable, I learned about potential outcomes. Where A is denoted to be the treatment on whether or not something is treated, and Y is the potential outcome. Y1 being the observed outcome of a treated unit and Y0 being the observed outcome of an untreated unit. Finding the average treatment effect, which is the difference between the means of Y1 and Y0, provides a statistical estimate of the magnitude of the difference the treatment caused. The SUTVA (stable unit treatment value assumption) states that the treatment should be well defined and lead to know interfering of treated units effecting untreated units' outcomes.

The statistical estimate may be distorted by confounders, which are variables that affect both the treatment variable and potential outcome variable. To control for confounders in experimental studies, one can assign an assignment probability wherein each unit has an equal likelihood of being treated or not. This creates exchangeability, where both the control and treated groups are similar enough that if they were switched the results would be roughly the same. When the confounder is known through stratified randomization by randomizing within groups, the use of standardization and inverse probability weighting can be used to provide an average treatment effect with the confounder removed.

Since observational studies have characteristics that prevent the use of assigning assignment probabilities and thus randomization, I learned techniques that help in simulating a randomized condition of the data. This includes inverse probability weighting through logistically regressing each unit to find a propensity score and standardization through linear regression being applied to a treated and control group. To find a standard error, I learned about bootstrapping to sample the dataset I would be analyzing multiple times and find a variance of means that can be used as the standard error. To apply these learnings, I used a dataset called Lalonde from the R package library "Matching" that was about a job training program's effect on real earnings in 1978.