Models for causal inference

Here are slides on outcome modeling and treatment modeling.

Models are useful when we need subgroup summaries but we do not observe very many units in each subgroup. This situation is common in causal inference: we assume that $\vec{X}$ is a sufficient adjustment set so that conditional exchangeability holds, and this allows us to identify the causal quantity $\text{E}(Y^a \mid \vec{X} = \vec{x})$ by the statistical quantity $\text{E}(Y \mid A = a, \vec{X} = \vec{x})$. But that empirical quantity—the subgroup mean among those with treatment value a and adjustment set value $\vec{x}$—may be the mean of a subgroup that is unpopulated. This is especially true in practice because the adjustment set $\vec{X}$ is often most plausible when it includes many variables, leading to a curse of dimensionality and small subgroup sample sizes. For this reason, causal inference approaches that adjust for measured variables often require us to estimate the means in many subgroups that are sparsely populated.

This page introduces outcome models for causal inference. To run the code on this page, you will need the tidyverse.

```
library(tidyverse)
```

Motivating example

To what extent does completing a four-year college degree by age 25 increase the probability of having a spouse or residential partner with a four-year college degree at age 35, among the population of U.S. residents who were ages 12–16 at the end of 1996?

We used this example on the Why Model? page and will continue with it here. For those jumping in on this page, here is a refresher.

This causal question draws on questions in sociology and demography about assortative mating: the tendency of people with high education, income, or status to form households together[1]. One reason to care about assortative mating is that it can contribute to inequality

[1]For reviews, see Mare 1991 and Schwartz 2013.