Studying Social Inequality with Data Science

INFO 3370 / 5371
Spring 2023

Causal Estimators:

Nonparametric estimation with subgroup means

Learning goals for today

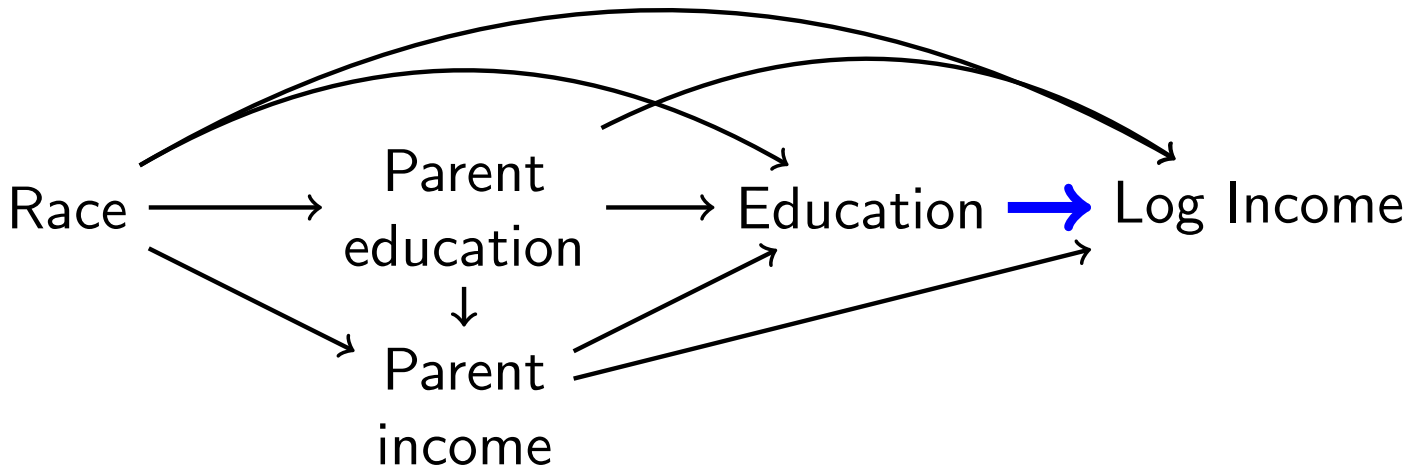
By the end of class, you will be able to

- Understand statistical adjustment as subgroup analysis
- Estimate causal effects nonparametrically

Data

We will examine the effect of education on log income using the data from the PSID Prediction Challenge

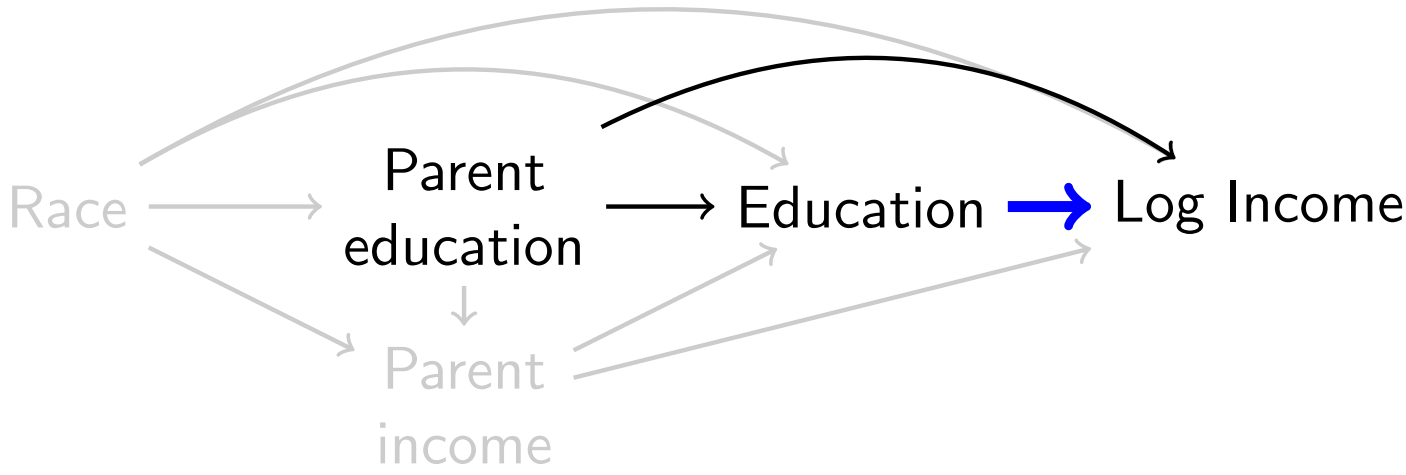
Assume this causal DAG:



Ultimately, we will adjust for race, parent education, and parent income

Nonparametric adjustment

First, pretend that parent education is a sufficient adjustment set (ignore everything that is gray)

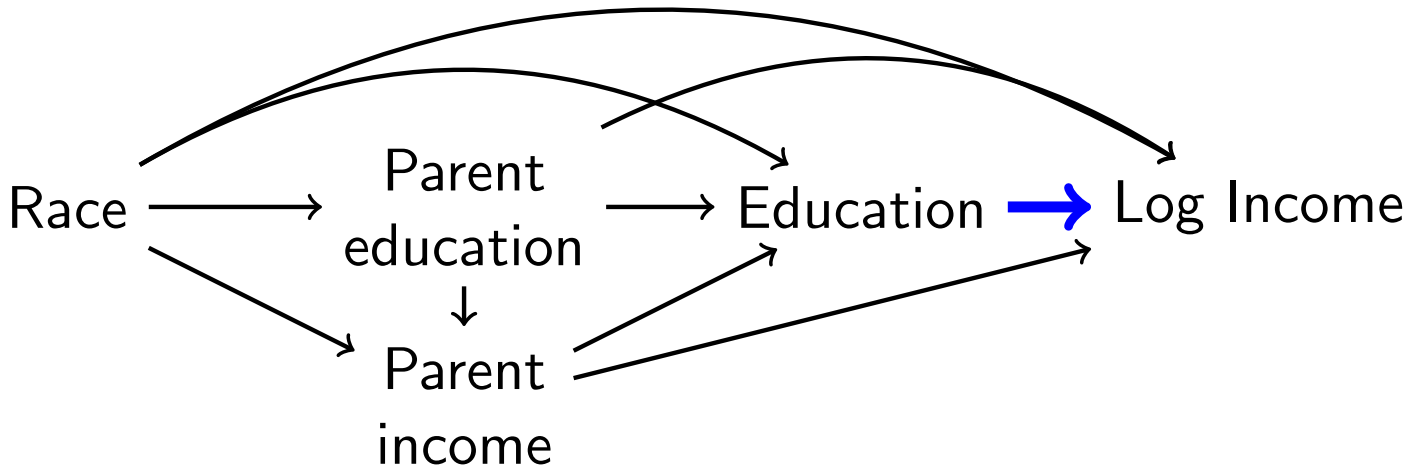


How to adjust nonparametrically:

1. take subgroups by confounders and treatment — `group_by`
2. take means within subgroups — `summarize`
3. difference over treatments — `pivot_wider` and `mutate`
4. (optional) average over confounder subgroups weighted by size

Preview for Monday

Now consider all variables. Is nonparametric adjustment possible?



Learning goals for today

By the end of class, you will be able to

- Understand statistical adjustment as subgroup analysis
- Estimate causal effects nonparametrically