Pre-Class Response for Lecture 10:

We're going to begin the "Returns to Education" unit! The goal is to understand the effect of education (that is, an extra year of high school; or going to college; etc) on outcomes further down the road (for example, earnings at age 30).

Here are two first attempts to get at this effect, given that a randomized experiment is difficult/unethical.

Attempt 1: Look at data from the US census, and run a multivariate regression, where:

- The independent variable is X = years of education,
- The dependent variable is Y = log(earnings).

Imagine that the regression controls for a bunch of possible confounders: Parents' income, parents' education, race, gender, age, IQ tests taken as a child, ...

Interpret the coefficient $\hat{\beta}$ on X as the "earnings returns to education." That is, if someone gets one more year of education, you would expect that their log(earnings) would increase by $\hat{\beta}$, or alternatively that their earnings by would increase by a multiplicative factor of $e^{\hat{\beta}}$.

Attempt 2: Look at data from a few hundred pairs of identical twins. Each pair of twins have the same parents, race, gender, age, and so on. Run a regression where, for each pair of twins:

- The dependent variable is Y = log(earnings of twin 1) - log(earnings of twin 2)
- The independent variable is X = (years of education for twin 1) - (years of education for twin 2)

Interpret the coefficient on X as the "earnings returns to education" (in the same sense as in Attempt 1).

For both attempts, don't worry for now about why we are measuring log(earnings) instead of earnings – we'll talk about that in class!

Question: Which of these two attempts would you find more compelling and why? Also, what additional information about the two attempts would be most helpful in answering that question?