Data 621 – CUNY SPS

Fall 2019 – Group 2

"Omar Pineda, Jeff Littlejohn, Sergio Ortega Cruz, Chester Poon, Simon Ustoyev"

Final Project Proposal

Despite the national narrative that America affords its inhabitants an unprecedented land of opportunity, intergenerational social mobility, defined as the likelihood that a child born to parents in the bottom fifth of the income distribution reaches the top fifth, is higher in many other advanced countries. Fewer than eight percent of Americans born in the bottom 20% of the income distribution reach the top 20%, whereas more than 13 percent of Canadians do. However, as Harvard Economist Raj Chetty has demonstrated, significant differences in upward mobility rates exist across the United States. “In this country, of all countries, a person’s zip code shouldn’t decide their destiny,” President Barack Obama said in 2015.

We wish to understand which demographic and geographic factors are associated with social mobility. We will use the CollegeDistance data set that is included in the AER R package. It contains survey data from 1100 USA residents who were high school seniors in 1980 with follow-up data regarding education attainment in 1986. While educational attainment is not a perfect proxy for income, we will use it as our outcome variable, our metric of interest. We will start the construction of the regression model with the 13 other variables, including gender, ethnicity, a composite test score, school and community demographic factors, and family income, among others. After exploratory data analysis that includes checking multicollinearity, we will endeavor to construct a multiple linear or multiple logistic regression model that predicts educational attainment based on the independent variables.

While our data captures educational attainment from more than 30 years ago, we think the construction of a model that predicts such a metric is nonetheless a worthwhile endeavor. An interesting follow-up study may be to repeat such a survey in the present day, construct a model, and then analyze differences.

References (to be converted to APA format for project):

Relative national intergenerational social mobility: <http://www.ecineq.org/ecineq_nyc17/FILESx2017/CR2/p256.pdf>

Chetty paper on geography of intergenerational mobility: <http://www.equality-of-opportunity.org/assets/documents/mobility_geo.pdf>

Obama quote: <https://talkpoverty.org/2015/12/17/american-dream-zip-codes-affordable-housing/>

CollegeDistance data info: <http://rdocumentation.org/packages/AER/versions/1.2-7/topics/CollegeDistance>

<http://wps.pearsoned.co.uk/wps/media/objects/12401/12699039/empirical/empex_tb/CollegeDistance_DataDescription.pdf>