Abstract

This project examines which demographic and geographic factors are associated with educational and social mobility. Using data included in the AER R package from a 1980 survey of 1100 high school seniors that included a follow-up six years later to gauge the level of education that had been obtained, we constructed a regression model that predicts educational attainment based on a selection of attributes included in the dataset. To start, we performed exploratory data analysis to understand the data and detect issues problematic with regression such as multicollinearity. Next, we constructed a simple multiple linear regression model using an achievement test score, average tuition for a four-year college in the student’s state, a student’s home distance from a four-year college, and whether their high school was in an urban area. We checked the model’s assumptions, including testing for influential outliers, before reviewing model diagnostics. Next, we created a new model that included an interaction term between the distance and urban variables. This model yielded similar results. In the end, we found that an achievement test score had the highest positive correlation with educational attainment, while average state school tuition had the largest negative association. Distance or whether a high school was located in an urban area displayed little correlation.

Introduction (Revised)

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 (Chetty el al. 2016). “In this country, of all countries, a person’s zip code shouldn’t decide their destiny,” President Barack Obama said in 2015 (Kaufman).

We wish to understand which demographic and geographic factors are associated with social mobility. While educational attainment is not a perfect proxy for income, we will use it as our outcome variable, our metric of interest. Although our data set captures a survey from more than 30 years ago, we think the construction of a model that predicts such a metric is nonetheless a worthwhile endeavor.

Additional References:

Kaufman, Greg (2015, December 17). Why Achieving the American Dream Depends on Your Zip Code. *Talk Poverty.* Retrieved from <https://talkpoverty.org/2015/12/17/american-dream-zip-codes-affordable-housing/>

Kleiber C, Zeileis A (2008). Applied Econometrics with R. Springer-Verlag, New York. ISBN 978-0-387-77316-2, [https://CRAN.R-project.org/package=AER](https://cran.r-project.org/package=AER).