Asuncion Los Banos

DS-670 Capstone: Big Data & Business Analytics

Assignment 11

**Discussion**

My competitor’s research correlated family income with test scores. They supported their hypothesis, linking test scores with household income, by using data based on five repeated measures of cognitive test scores per child. Using data derived from the U.S. government’s Earned Income Tax Credit (EITC) records, they identified nonlinear changes that indicated upward shifts, or expansions, in family income over twelve years (1987-1999), a twenty-percent increase of around $2,100, between 1993 and 1997. They applied an “instrumental variable strategy” that established a causal relationship between measurable expansions in family income and corresponding increases in children’s math and reading scores. Testing to a variety of independent variables, they found short term score improvements of six percent, with one standard deviation for every $1,000 increase in income, for low-income families. The conclusions reached by recent studies suggest that unobserved heterogeneity and endogenous income shocks are important concerns. Furthermore, they suggest that income effects may be greatest among economically disadvantaged families. My competitors, Dahl and Lochner, outlined an instrumental variables strategy that eliminates omitted variable biases due to both permanent and temporary shocks correlated with family income and alleviates bias due to measurement error in income.

Dahl and Lochner completed a correlation on their study to find a cause and effect in family income and test scores. My approach was a regression where I used independent variables to find influential factors to a student’s test score. Dahl and Lochner’s results implied that a $1,000 increase in income raises combined math and reading test scores by 6 percent of a standard deviation in the short run. The regression I completed for the different types of average income by household type showed a $10,000 increase in in average income translate into an increase in the average test score of the county for 2014, 2010, and 2007.

The competitor completed a correlation versus I ran a regression on my data variables. They simply computed a correlation coefficient that tells how much one variable tends to change when the other one does. Whereas linear regression finds the best line that predicts Y from X, correlation does not fit a line. With linear regression, the X values can be measured or can be a variable that I could control. Finding the cause and effect is quite hard to find in exactly how correlated income can be to test scores because other events might be going on in the different types of household students live in. Also, while the competitor’s study reveal the correlations between income and child outcomes, they do not necessarily estimate a causal relationship. Children living in poor families may have a worse home environment or other characteristics that the researcher does not observe. These omitted variables may be part of the reason for substandard achievement and may continue to affect children’s development even if family income were to rise. Therefore, it is really hard to say if income really correlates to lower or higher test scores because there is inner working inside a household that cannot be seen or weighted.

**Conclusion**

The influential factor for a student’s test score is important to understand. Although a student cannot decide which race they are born into, understanding the powerful factor of household income is an important question. Regardless of household type, each one had a prominent factor to test scores. The results for average income by household and race shows positive influential factor between the variables and the average scores. From 2007 to 2010 the coefficients for all household types rose and then fell by a couple of points in 2014. The results from the simple linear regression for household, family, and married couple were significant at 0.001. The non family household were significant at 0.001 in 2014 and 2010, but was significant at 0.01 in 2007. The results from all years showed significant where a $10,000 increase in average income for any household type translated into an increased in the average test score in the county.

The coefficient for African American as a race for all three years were always statistically significant than the other race. African American was significant at 0.01 in 2014 and 2010 and significant at 0.001 in 2007. Hispanic or Latino was significant at 0.05 for all three years. African American and Hispanic students fall in the “achievement gap” where socio-economic factors including income levels, educational attainment, employment rates, housing options, neighborhood crime rates, and resources available to schools are worse for African Americans and Hispanics, on average, than for Whites (Patterson, C. J., et. al., 1990). The “achievement gap” often lead to fewer opportunities for African American and Hispanic children to access a wide range of activities and experience an enriched educational environment. Caucasian was significant at 0.05 in 2014 and 0.01 for 2010 and 2007. As for Asians, we cannot conclude that a significant difference exists because the p-value for all three years were over 0.05. However, because a regression is looking for the most influential factor it can mean that because Asian was not statistically significant it can mean that Asian American students excel in school because of culture. A study published in the journal PNAS, stated “Asian and Asian American youth are harder working because of cultural beliefs that emphasize the strong connection between effort and achievement. Studies show that Asian and Asian American students tend to view cognitive abilities as qualities that can be developed through effort, whereas white Americans tend to view cognitive abilities as qualities that are inborn” (Hsin, A., & Xie, Y., 2014).

In all, average income and race has an effective factor on standardized test scores, yet it may not be the main influencing factor. Students living in poor households may have a worse home environment or other characteristics that cannot be observed. Although, the regression shows income has an influential factor to test scores the omitted variables in the student’s home environment may be part of the reason for substandard and may continue to affect student’s development even if family income were to rise.

**Reference**:

Hsin, A., & Xie, Y. (2014). Explaining Asian Americans’ academic advantage over whites. *Proceedings of the National Academy of Sciences*, *111*(23), 8416-8421.