Jeffrey Hou

5/6/19

Data Science 110

Does Socioeconomic factors affect test scores?

RPUBS-<http://rpubs.com/Jeffo003/494490>

For this final project, I decided to focus on a data set that included student’s test scores in 3 different subjects (Math, Reading, and Writing). The dataset included 8 different variables with 5 factors and 3 integers. There were 1000 students that are in the sample. The 5 factors were gender, race/ethnicity, parental level of education, lunch, and test preparation. The integers were the 3 different scores which were math score, reading score, and writing score. For the 5 factors, gender was male or female, race had 5 different choices (A-E), parental ranged from some high school to master’s degree, lunch was either free or standard, and test preparation was completed or not.

This specific data set came from Kaggle and not only is this an important data set, its also incredibly beneficial to us as students. This dataset came pretty clean, as there was not any missing info. The only minor tweak I had to make was to reorder the parental level of education from some high school to master’s degree. Using the dplyr package, I was able to use the gather function to make a whole new data set which included each score for each subject as a separate entity. This expanded the data set from 1000 rows to 3000 rows. I also utilize the cor.test to find correlation between different subjects. I found out there was a strong correlation between writing and reading.

I chose this data set because the relationship between social economic status and education is quite important in our society nowadays. The scholastic gap as we so call it can be seen comparing a test score from Baltimore vs a test score in North Potomac. Students in poorer areas have a hard time getting the help they need, while the students with wealthier backgrounds may have tutors. As long as the income gap in the US exists, these students will always struggle with education due to the fact that many of their families have never gone to college and can’t help the students with their work. The main factor in these gaps is the school environment. For example, the high school I attended was Wootton, and the students there are all super studious, and many go on to attend prestigious colleges. This would not be the same for students who only see their upperclassmen fail out of school. Only the most determined students would be able to escape that type of environment and excel. According to article written by Greg J. Duncan and Katherine A. Magnuson, “Racial and ethnic differences are also apparent in the total number of hardships that children face. The vast majority of black and Hispanic children suffer at least one hardship, compared with just over half of white children.” Students who may have to work to support their family will obviously have less time to study, as they have to raise their siblings and possibly their parents as well. Another important factor is family background, as a student with parents with college degrees are very likely to also go to college, as that is considered the norm in the family. A student who may be the first in their family to go to college may not have that added pressure from their family background. Children with highly educated parents routinely score higher on cognitive and academic achievement tests than do children of parents with less education. According to Sue Thomson, parents with higher education are also “more likely to provide a more stimulating home environment to promote cognitive development. “ Family structure also is an important piece of the puzzle when it comes to this income gap problem that exists in our society. Many low-income families are often broken up, either single dad or single mom. The children that are raised in that type of family lose a key role model that could have been beneficial to their growth. Life is very different for a family with a single parent compared to one with one parent working and one staying at home as a caregiver. Many single parent children grow up to have a distaste towards family. One family is most likely going to have much more hardships due to the fact that many issues arise which can be hard for just one person to solve.

For this project, I decided to analyze how these specific factors such as lunch or test preparation affected the student’s ability to excel in their subjects. An interesting pattern arose as students with reduced lunch typically did worse than students with standard lunch. Also students who have test preparation also did better than students without. While these students may go to the same school, it is very likely that their social economic status is very different. Many families may have all their income go towards their child’s education compared to students who may be wealthy enough that money is not an issue. An interesting observation I made was that when it came to parents with master’s degrees, Students in the Race A actually managed to have the lowest average math score. For all other races, the master’s degree was typically top 1 or 2. Group A was the one group where college degrees didn’t necessarily mean higher test grades. It is very likely that Group A had a lot of students who were super motivated even though they might not have come from a wealthy family. A lot of the factors in this dataset were socioeconomic based such as family education and reduced lunch. The students with families with low education backgrounds as expected did worse than students with better backgrounds. The other interesting point I found was the fact that usually for all parental backgrounds, the higher the background the higher the test scores. This for some reason was not the case for High school and some high school, as some high school typically did better than students with parents who got their high school diploma.

A couple things I would have liked to have worked was to figure out which ethnicity and race each group was and analyze the socioeconomic factors for each group. I also would have liked to have extra factors possibly such as school location in order to build a geospatial map to see which schools or areas were hot zones for high test grades. I also could have been able to locate areas where specific racial groups were prominent in and see if neighborhood environment affected their school work.

Bibliography

Duncan, Greg J, and Katherine a Magnuson. “Can Family Socioeconomic Resources Account for Racial and Ethnic Test Score Gaps?” *Futureofchildren*, 2005, Spring , files.eric.ed.gov/fulltext/EJ795842.pdf.

Thomson, Sue. “Achievement at School and Socioeconomic Background-an Educational Perspective.” *Nature News*, Nature Publishing Group, 23 Mar. 2018, www.nature.com/articles/s41539-018-0022-0.