Exploratory Data Analysis

DSC530-T302

Joshua Greenert

8/11/2022

**What is the outcome of the EDA?**

The outcome of the exploratory data analysis failed to answer the question, “Are men better test-takers than women'? Using the data to confirm the inverse proved ineffective as well. There were points where the p-values suggested a statistical significance when comparing the parent’s level of education with each respective gender’s math scores, but this was nullified by further data comparisons along with multiple regression. It is possible that more data to compare would shed light on the question for future analysis along with other implementations of numeric data points.

**What do you feel was missed during the analysis?**

During the analysis, after cleaning the data, it was apparent there was a lack of viable information. Further estimates might have proven more useful given a larger amount of data points to contrast and compare with.

**Were there any variables you felt could have helped in the analysis?**

If the students were to take a psychology test prior to their exams, the data could carry more points for clarification on anxiety, mood, or overall well-being. These data could integrate within the analysis to pinpoint additional attributes that affect the overall testing experience. Likewise, these data may carry statistical significance towards the reasoning for some students scoring less than fifty percent, while also explaining while some students scored nearly one hundred percent.

**Were there any assumptions made you felt were incorrect?**

The only assumption made during this exploratory analysis was that a student’s test score may be dependent on their parent’s respective academic success. Beyond this assumption, all other points were verified through testing instead and negated themselves from the conclusion.

**What challenges did you face, what did you not fully understand?**

The thinkstats2 library doesn’t appear to have much documentation or users at the current moment. This lack of information greatly impacted my ability to use the featured libraries to aid my research. Ultimately, I used more data within the numpy library to produce the same results so that I could understand what the code was doing and how to replicate each respective test.