HW3 Proposal

RJ Cass

1. Understanding the Problem

Background:

We have a dataset of various factors of conditions in a school, and the corresponding test scores of that school. We want to use these data to understand the how the different factors influence test scores, as well as be able to predict test scores.

Goals:

In this analysis, we want to address several key questions:

i. Is there evidence of diminishing returns on extracurricular activities in terms of student learning?

- To answer this we will examine the relationship between Income (the factor that measure spending on extracurricular activities) and Score. If there are diminishing returns of the effect of Income on Score, we would expect the graph to show a relationship between the variables where the slope decreases as Income increases.

ii. Is English as a second language a barrier to student learning?

- To answer this, we will develop a model indicating the significance of each factor. If the English factor is indicated as being significant, then it will show that English as a secone language is a barrier to student learning.

iii. In your opinion and based on the data, what can be done to increase student learning?

- To answer this, we will develop a model that can identify important factors in relation to score. From this we will be able to indicate what actions would most impact scores (ie. student learning).

iv. How well does the model predict compared to alternatives?

- To answer this, we will examine 2 different models in our analysis and determine their predictive capabilities.

2. Exploratory Data Analysis

3. Desired Attributes

Model attributes required from the research questions

The research goals require an model with the ability to identify important factors, and that can predict.

Model attributes required from the data

The data show the selected model will need to address collinearity amongst the explaining factors. The model also needs to work for non-linear relationships with the factors.

Any other anticipated problems

What goes wrong if the above are not accounted for?

4. Proposed Method

Appropriate models

Specific model proposal

Method strengts/weakness

i. How this method accounts for issues in the dataset

ii. How this method accomplishes research/analysis goals and yields appropriate estimators

iii. How this method will answer the research questions

iv. What assumptions are needed to use the model adequately? Are they reasonable to assume and explained well?