

ORIE 4741: Project Proposal

Anna Halldorsdottir, Stephanie Hellman

September 2020

Problem

Our goal is to identify risk factors for drinking in teenagers. To do this, we will examine the student's age, gender, whether they live in an urban or rural area, and their success in school, among other factors.

Data Set

We have a data set of about 700 students from Portugal, all of whom took a survey in their math or Portuguese class in the 2005-2006 school year. For each student, we have data regarding their family life, social life, education habits, grades, and whether or not they drink. We downloaded the data set from Kaggle.com, <https://www.kaggle.com/uciml/student-alcohol-consumption>.

We plan to use combinations of the variables above, trying to avoid confounding variables, to predict whether or not a student will drink.

Value

This is an important question to consider as teenage drinking is a concern in high schools, as it can lead to brain damage, and is also illegal. The results of this analysis could be applied in school counseling offices to determine which students may need extra attention. Parents could also use this information to see if their family situation may cause risk for underage drinking, and perhaps change their habits.

We believe we can answer this question with our data set, because of how much information the data set contains. Not only does it contain solely education habits, social habits, and family habits: it combines all three.