The usage of Substances such as Alcohol and Marijuana has been shown to have negative effect on the academic performance of American college students, yet it is common for students to consume these substances on a regular basis, particularly marijuana and alcohol, with alcohol often being consumed in excessively. It has been further shown that the consumption of both alcohol and marijuana having separately tend to alter the brain structure and function respectively and the combined consumption for adolescent leads to the development of abnormal neural tissue, that adolescent consuming alcohol and the smoking of tobacco cigarettes or marijuana are more prone to develop substance usage disorders as younger adults in addition to the consumption of alcohol and marijuana, individually or combined, often being associated with negative effects, such as impaired decision making, poorer attention or a drops IQ, among others, with earlier onset of substance usage generally having stronger negative effects.

To examine the effect of the combined usage of alcohol and marijuana on the academic performance of college students, the authors of [**?**] 2016 conducted an explorative longitudinal study in which they assigned the students into based on their substance usage into different cluster and then evaluated the study findings through multiple statistical models. The study was of interest to the [**?**] due to the academic performance having a potentially direct and strong affect on the long term academic success of students, as well as potential larger societal costs due to substance abuse. The authors were also motivated by a perceived lack of research into the combined substance abuse at the time of the study.

In this undergraduate thesis I'm trying to computationally recreate the statistical analysis of [**?**] with the published data from their conducted study. Due to often a lack of details and succinct explanations, I'm also providing my providing my replication with a more extensive descriptive illustration of the data, increased statistical context and interpretation of the methods used in the study, as well as the findings and performance of the used models. In section 1 will contain the most important descriptive illustrations and replication of [**?**]

—— Overview of the results here later ——

In section 1 I will provide a recreation of the used models in the paper, as well as their results and performance, while in section 3 I will provide several robustness checks. Section 4 I will discuss and summaries the results of the previous sections.

Introduction

- What is the subject of the study? Describe the economic/econometric problem.

- What is the purpose of the study (working hypothesis)?

- What do we already know about the subject (literature review)? Use citations. For example, see [**?**].

- What is the innovation of the study?

- Provide an overview of your results.

- Outline of the paper:
  The paper is organized as follows. The next section describes the model under investigation. Section "[Data]" describes the data set and Section "[Results]" presents the results. Finally, Section "[Conclusion]" concludes.

- The introduction should not be longer than 4 pages.