Stats506 F19 – Project Proposal, Group 9

Daxuan Deng, Ming-Chen Lu, Ningyuan Wang

November 27, 2019

Question: How are drinking habits related to general health condition for adults in US?

Topic: Semiparametric bootstrap - resampling residuals

Details

Drinking alcohol is a common behavior among adults. People always think drinking alcohol is good for mood but bad for health. However, we are curious about the drinking habits of people with general "good" health condition in US. Since there is a higher risk of response bias towards drinking question for young people, we decide to limit the age of analysis to the US legal drinking age of 21. Additionally, the histogram of number of drinking times shows that there is an "excessive" number of individuals with a count of 0. This zero-inflated problem motivates the use of semiparametric bootstrap in our project. We will first estimate a linear regression model, and then simulate by resampling residuals to that estimate and adding them back to the fitted values. The work will be done through the programming languages of Matlab, Python, and R.

Dataset Descriptions

For the entire analysis, we will be using datasets from NHANES 2005-2006.

The response variable would be *ALQ120Q* (how often drink alcohol over past 12 months) from Alcohol Use. The predictors include *HSD010* (general health condition with levels of good, very good, and excellent) from Current Health Status and *SEQN*, *RIDAGEYR* (age), *RIAGENDR* (gender), *DMDEDUC2* (education), *INDFMPIR* (poverty income ratio) from Demographics.

Responsibility

- Daxuan Deng - Matlab
- Ming-Chen Lu - Python
- Ningyuan Wang - R