BOSP 350
Final Project Proposal

Group members: Ebubekir Karamustafa, İlkin Umut Melanlıoğlu

Type of project: Empirical Project

Dataset: Oregon Health Insurance Experiment, provided by Isaac Sorkin.

Possible Techniques: ATE/LATE calculation, linear regression analysis

Questions:

- Does having health insurance either increase or decrease the rate of smoking?
- How do the following insurance-related variables relate to the change in smoking rates:
    - Asthma diagnosis
    - Depression diagnosis
    - Hypertension diagnosis
    - High cholesterol diagnosis
    - Availability of dental care

Initially, we were planning to come up with a final project that investigated the causal effect of having health insurance on obesity. The inspiration for this particular idea came from a series of academic publications that argued that people with health insurance are more likely to become obese since their health insurance would significantly decrease the additional financial strain caused by obesity-related medical expenses, giving them an economic incentive to gain weight. The Oregon Health Insurance Experiment dataset was given to us by our BOSP 350 instructors as a reliable and comprehensive source of data, which we further examined in the following weeks. The population of the RCT-based experiment consisted of low-income Oregon-based adults. Then, various health-related effects of having health insurance were investigated by essentially giving a randomized subset of the population the ability to apply for Medicaid (most of the applicants received the insurance). Upon having a look at the dataset, which contained more than 90 thousand data points, it became clear to us that it would be a perfect building group for an empirical project, enabling us to use multiple data analysis techniques that we had learned in the previous lectures.

However, this dataset was unfortunately unusable for our obesity-related investigation because it did not contain height, weight, or BMI readings of the experimental population at the initial stage of the experiment. This unavailability of initial obesity-related data would significantly hinder our ability to come up with a viable assessment of the treatment effect on obesity. Nevertheless, we realized that we could adjust the topic of our project to one that would be thoroughly complemented by this ‘goldmine’ of a dataset. Thus, after further deliberating on the possible medical outcomes of having health insurance, we decided to investigate the effect of having health insurance on (cigarette) smoking rates.

Understanding the effect of modern-day health insurance on smoking rates is particularly important because although the smoking rates have been dropping throughout the past decades, approximately 13.7% of the adult US population is still engaged in smoking habits. While these smoking habits are well known to have dire long-term medical consequences – such as increased vulnerability to cardiovascular diseases, various types of cancer, and immunodeficiencies, exploring the effect of seemingly unrelated health insurance on the smoking rates could provide valuable insights into the complex relationship between healthcare access and smoking behaviors.

We are expecting one of the two following scenarios:

1) People who obtained Medicaid health insurance (treatment group) are more likely to stop/decrease their smoking compared to those in the control group due to the increased availability of medical screening/tools that will uncover the pre-existing health deficiencies, leading treatment group people to lessen their degree of smoking.
2) As it was the case with obesity, smoking people with health insurance, which will reimburse the increased medical costs induced by smoking, will have more to gain from having health insurance, compared to those who do not. Hence, people in the treatment group might start to inadvertently increase their level of smoking. Then, we would be able to establish a negative treatment effect on smoking.

To this end, we have first outlined possible confounding variables (such as asthma and hypertension diagnosis rates). We are going to build a linear regression model to evaluate the impacts of such diagnoses, coupled with the availability of dental care, on smoking rates. Afterward, we are going to validate our findings by computing average treatment effects (ATEs) and local average treatment effects (LATEs), using basic formulations familiar from our initial classes. Consequently, we will be able to deduce both the individual and joint impacts of insurance-related variables on smoking rates and perhaps establish strong causal relationships. Moreover, the RCT nature of the experiment enables us to establish causal inferences without worrying about the selection effects/confounding variables.