Jonothan Meyer
Project 1: Exercise 3
Linear Regression
1/29/21

Problem Title: The Opioid Crisis - 2019

1) What data collection strategies for the problem would you consider. Would you used a controlled experiment or an observational study approach?

I would use an observational study approach. There is a wealth of data available regarding doctor opiate prescriptions, incarcerations due to opiates, hospitalizations as a result of overdose, ext. I would draw from all that data to try to find interactions between opiate abuse and economic impact. I think an important first step that might be easily overlooked, is looking at the interaction between the flood of opium to China and the most significant impact it had on their economy between 1820 and 1840, prior to the Opium Wars. Obviously based on the time and place America's Crisis is not directly comparable to this event. However, I believe you would be able to derive useful interactions from that event that effectively model the same economic impact today, here in the states.

2) Identify several possible predictor variables of interest and a response variable of interest; how might you reduce this set of predictor variables for an initial analysis of a simple version of the model?

Predictor: Opiate related overdoses per capita
Response: Unemployment Rate per capita

Predictor: Opiate related arrests per capita
Response: College Level Academic Enrollment per capita

Predictor: Amount of RX's written for opiates
Response: Work related injuries or deaths

To reduce the number of predictor variables for analysis I would first look at which of these interactions can best be represented using a mathematical model such as linear regression. Some of the predictor variables could possibly be grouped together to reduced the set of predictors, as well as increase the overall sample size. An example of this would be to combine the amount of synthetic opiate drug busts with the amount of doctor opiate prescriptions. This gives a good general total of illicit and legal opiates out on the streets to compare with something like college enrollment.

3) What variable reductions should be considered to simplify the model?
Ideally I would want to find one or two response variables that had a strong correlation with opiate addiction and served as a reliable indicator of negative economic impact. Age, region, demographic, lifestyle, sex, and type of opiate preferred (heroin vs. prescription opiates) might all be reduced or combined to reduce variance in the analysis and provide simplicity. The goal, ultimately, is to find one predictor variable related to the opiate epidemic that strongly correlates with a single response variable related to negative economic impact. One possible example of this would be the number of opiate overdose hospitalizations vs. the increase in costs for insurance premiums.

4) What data or predictions would you use to validate your model?

As discussed previously, I think it would be insightful to compare The United States economic recovery as opiate availability diminishes vs. China's economic recovery after The Opium Wars. In addition, I think it would be valuable to use a region in America that has very low opiate use and use it's economic performance as a baseline to compare with my response variable. If several control regions in The US have comparable economic performance to the study despite having low levels of opiate use then I will know that the model I constructed is invalidated, and needs to be refined.