STAT 151a Final Project Proposal

Team Members: Brandon Rodriguez, Gavin Yang, Xuejing Yan

Cancer is a leading cause of mortality worldwide, and understanding the factors that affect cancer mortality can help state and federal governments effectively allocate capital and resources in the correct areas. In our final project, we aim to use regression modeling techniques to analyze the factors affecting cancer mortality using an organized dataset sourced from cancer.gov, the American Community Survey (census.gov), and clinicaltrials.gov. Our research question is: What factors most affect cancer mortality rates, and what, if any, are the interactions between these factors.

We will use exploratory data analysis to select factors of interest such as income, age, gender, socioeconomic status, geographic location, and more. Next, we will use regression modeling techniques to analyze the relationships between these factors and Mean *per capita* (100,000) cancer mortalities in each area, which we use as a proxy for cancer mortality rates. We will also examine the interactions between different factors to examine possible collinearity issues and interactions between them.

As listed in our research question, the outcomes of this study will include identifying the most significant factors that affect cancer mortality rates and quantifying their impact, as well as the interactions between them. The data used comes entirely from locations in the United States, and was collected from 2010-2016, and thus we limit our conclusions to the areas of the United States. These results can inform public health policies and interventions aimed at reducing cancer deaths, and hopefully lead to more efficient uses of federal aid to decrease cancer mortality rates nationwide.

Data citation:

N.Rippner. Cancer Trials, 2017. Retrieved from http://data.world/exercises/ linear-regression-exercise-1/workspace/file?filename=cancer_reg.csv.