STAT420 Final Project Proposal

1. Team Members: Jingqi Kou(jingqik2); Zhan Shi(zhans5); Lang Zhou(langz2);

Yiyang Li(yiyang18)

1. The Tentative Title: What Influences Life Expectancy?
2. Description of Dataset:

The data we got for this project is from Kaggle and the dataset is mainly about the life expectancy and distinct health-related factors in 193 countries from 2000 to 2015. It contains 2938 observations and 20 predicting variables which can be divided into 4 categories: immunization related factors, mortality factors, economic factors, and social factors.

The variables may strongly influence the life expectancy:

* **Status**: “Developed” or “Developing” status of a country
* **Adult Mortality** (%): the adult mortality rate per 1000 population (i.e. number of people dying between 15 and 60 years per 1000 population)
* **Infant Mortality** (%): number of infant deaths per 1000 population
* **GDP** (%): Gross Domestic Product per capita in USD
* **Total expenditure** (%): government expenditure on health as a percentage of total government expenditure
* **Income composition of resources**: Human Development Index in terms of income composition of resources (index ranging from 0 to 1)
* **Hepatitis b** (%): number of 1 year old with Hepatitis B immunization over all 1 year old in population
* **Polio** (%): number of 1 year old with Polio immunization over all 1 year old in population
* **Diphtheria** (%): Diphtheria tetanus toxoid and pertussis (DTP3) immunization rate of 1 year old

1. Background Information:

The data set is merged by Kumar Rajarshi from Kaggle. The concrete data was collected from WHO and United Nations websites with the help of Deeksha Russell and Duan Wang. Particularly, Kumar took the immunization factors and human development index into consideration in the dataset unlike other previous data analysis considering mortality factors and income factors. There were missing data of some small countries and Kumar decided to exclude those countries from the final dataset since it was hard to find the relevant information.

Credits:

Kumar Rajarshi <https://www.kaggle.com/kumarajarshi/life-expectancy-who>

1. Points of Interest:

As stated above, we somewhat conclude that the life expectancy has correlation with mortality factors and income factors from many studies. It will be interesting to explore whether there is some certain relationship between the immunization factors and life expectancy. We want to know how much the expected variables that affect life expectancy have their respective effects on life expectancy. How does Infant and Adult mortality rates affect life expectancy? How does Infant and Adult mortality rates affect life expectancy?

Does Life Expectancy have positive or negative correlation with eating habits, lifestyle, exercise, smoking, drinking alcohol etc.

1. Evidence

