ORIE 5741 Project Proposal

To: World Health Organization
From: Data Scientist Team from Cornell University (Zhuo Zhou, Tianyi Nie)
Date: Oct 3rd, 2021
Subject: Statistical Analysis on Various Factors of Expected Life Expectancy

1 Introduction

Typically, when we look at the factors that contribute to life expectancy, we take into account factors such as population, income, and mortality. However, few predictions and analyses have been conducted for each country; Instead, they were implemented throughout the whole world. Also, there are many other factors such as vaccination that could be a very important factor.

2 Objective

We would like to conduct multiple linear or non-linear analyses of a large number of factors for different countries to identify factors that contribute to low expected life expectancy. Since the situation for countries might be significantly different, this will help to suggest which areas a specific country should focus on in order to effectively increase the life expectancy of its population. In this project, we will use a combined dataset to explore various factors including the health system, population, social effect, macroeconomic index and etc.

3 Dataset

The dataset source is from Kaggle:
https://www.kaggle.com/kumarajarshi/life-expectancy-who

This dataset contains 20 potential factors that might influence the expected life expectancy of 193 countries from 2000 to 2015. It is provided by WHO and collected on Kaggle.

Based on the dataset above, we want to specifically answer the following questions after our analysis:

1. What are the significant features that actually affecting life expectancy?

2. How much the health system for a country affects its expected life expectancy?

3. What other actions should a government take to improve the expected life expectancy?