How to Improve Lifespan Through National Policy

Group 2

Deep Karki, James Hawkins, Tomiwa Omotesho, Austin Hadamuscin

1.1 Introduction

There are many aspects when it comes to having a healthy life. Measuring a healthy life and determining the factors that lead to a healthy life is a tricky task, but if we were to take a broad view of health, one could simply quantify this by measuring similar individuals' lifespans. An easy and convenient way to group similar individuals is by country. Although most countries contain residents from many different ethnicities, social statuses, religions, etc., they share the same government policies and a general geographical area.

Our research topic is related to the World Health Organization's (WHO) sustainable development goal of "good health and well-being." We describe individuals with good health and well-being as those who have a longer than average lifespan. We feel that the indicators below are a good resource to model an individual's average longevity. There have been numerous studies linking BMI, education level, alcohol consumption, and disease prevalence along with other factors to lifespan. We plan to use this data to find solutions for countries with shorter life expectancies.

1.2 Objectives

1. To create easily understandable data visualizations that effectively show our results.

2. To determine the most important variables for improving countries' average lifespan.

3. To determine whether there are better ways to group countries other than the developed and undeveloped for the purpose of modeling life expectancy.

1.3 Data Mining Techniques

We will conduct clustering to account for cultural and geographical differences and use a cross-validation approach to lasso or ridge regression to model life expectancy for each cluster. If clustering does not improve our ability to model life expectancy, we will use our original data set to fit our model.

2.1 Data

Our data comes from the Kaggle data set titled Life Expectancy (WHO) and contains 21 variables. We will be using the average life expectancy by country as our response variable and the remaining variables will be our potential regressors. Our regressors include both qualitative and quantitative data. The quantitative regressors are:

1. Country – name of the county
2. Development status – differentiates developed and undeveloped countries.

The quantitative regressors are:

1. Life Expectancy – life expectancy in age
2. Adult Mortality – mortality rate per 1000 people aged 15-60
3. Infant Deaths – number of infant deaths per 1000 population
4. Alcohol – alcohol consumption (ages 15+) per capita in liters of pure alcohol
5. Percentage Expenditure – percent of GDP spent on health expenditures
6. Hepatitis B – hepatitis b (HepB) immunization coverage among 1-year-olds (%)
7. Measles – number of reported cases per 1000 population
8. BMI – average body mass index of population
9. Under-Five Deaths – number of under-five deaths per 1000 population
10. Polio – polio (Pol3) immunization coverage among 1-year-olds (%)

11. Total expenditure – general government expenditure on health as a percentage of total government expenditure (%)

12. Diphtheria – diphtheria tetanus toxoid and pertussis (DTP3) immunization coverage among 1-year-olds (%)

13. HIV/AIDS – deaths per 1000 live births with HIV/AIDS (0-4 years)

14. GDP – gross domestic product

15. Population – population of the country

16. thinness 1-19 years – prevalence of thinness among children and adolescents for age 10 to 19 (%)

17. thinness 5-9 years – prevalence of thinness among children for age 5 to 9 (%)

18. Income composition of resources - human development index in terms of income composition of resources (index ranging from 0 to 1)

19. Schooling – number of years of schooling (years)