STAT-S670 Final Project Proposal

Group 19 :- Ruchik Dama | Shashank Kumar | Sumanth Gopalkrishna | Aniket Kale

---

Project Description:

The aim of the project is to understand how life expectancy is dependent upon different factors like economical, immunization rate, diseases, lifestyle and health care expenditure in a country. The data is collected from 2000 to 2015 for 193 countries.

Data Description:

The dataset is available at **https://data.worldbank.org/indicator/SP.DYN.LE00.IN.**

The data-set related to life expectancy and health factors for 193 countries has been collected from the same WHO data repository website and its corresponding economic data was collected from the United Nation website. In total, there are 20 columns and 3000 rows in the final merged file.

An initial look at the data revealed some missing values. Since the data sets come from WHO, we found no obvious errors. Based on the results, most of the missing data were related to population, Hepatitis B, and GDP. The variable descriptions:

The variable description is as follows:

- **Country :** Country Name
- **Year :** Year of the observation
- **Status:** Developed or Developing status of the Country
- **Life expectancy:** Life Expectancy in age
- **Adult Mortality:** Adult Mortality Rates of both sexes (probability of dying between 15 and 60 years per 1000 population)
- **Infant deaths:** Number of Infant Deaths per 1000 population
- **Alcohol:** Alcohol, recorded per capita (15+) consumption (in litres of pure alcohol)
- **Percentage expenditure:** Expenditure on health as a percentage of Gross Domestic Product per capita(%)
- **Hepatitis B:** Hepatitis B (HepB) immunization coverage among 1-year-olds (%)
- **Measles :** Number of reported cases per 1000 population
- **BMI:** Average Body Mass Index of entire population
- **Under-five deaths:** Number of under-five deaths per 1000 population
- **Polio:** Polio (Pol3) immunization coverage among 1-year-olds (%)
- **Total expenditure:** General government expenditure on health as a percentage of total government expenditure (%)
- **Diphtheria:** Diphtheria tetanus toxoid and pertussis (DTP3) immunization coverage among 1-year-olds (%)
- **HIV/AIDS:** Deaths per 1000 live births HIV/AIDS (0-4 years)
- **GDP:** Gross Domestic Product per capita (in USD)
- **Population:** Population of the country
- **Income composition:** Human Development Index in terms of income composition of resources (index ranging from 0 to 1)
- **Schooling:** Number of years of Schooling(years)

Research Questions:

To answer how life expectancy is dependent upon different factors, we need to explore our data and come up with the hypothesis which will help us answer it. We have divided our research question into four major segments. Below are the questions

Trends:

1. Did life expectancy grow over the years for developing and developed countries? Is the rate of growth constant among these countries?
2. How is adult mortality and infant mortality changing over time? Is the change positively or negatively correlated with life expectancy?

Health and Lifestyle metrics:

1. Dependency of life expectation with the immunization taken for different diseases. Does taking a vaccine increase life expectancy?
2. Number of deaths and cases of communicable diseases like measles and HIV are related to life expectancy?
3. People with high alcohol consumption and BMI ratio live shorter lives?

Economy metrics:

1. Does a country spending more on health care have a higher life expectancy?
2. Does population and population density influence life span?
3. Does education play a role in life expectancy?

Mortality rates:

1. Relationship between adult mortalities and infant mortalities with life expectancy. We want to check the hypothesis that if the infant mortalities rate are high the average life expectancy would be shifted to lower range.

Dataset Citations:

Alcohol data: *https://ourworldindata.org/alcohol-consumption*
Population data: *https://www.census.gov/popclock/world*
GDP data: *https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG?most_recent_value_desc=false*
Healthcare expenditure: *https://data.oecd.org/healthres/health-spending.htm*
Polio : *https://data.oecd.org/healthres/health-spending.htm*