Context

Although there have been lot of studies undertaken in the past on factors affecting life expectancy considering demographic variables, income composition and mortality rates. It was found that affect of immunization and human development index was not taken into account in the past. Also, some of the past research was done considering multiple linear regression based on data set of one year for all the countries. Hence, this gives motivation to resolve both the factors stated previously by formulating a regression model based on mixed effects model and multiple linear regression while considering data from a period of 2000 to 2015 for all the countries. Important immunization like Hepatitis B, Polio and Diphtheria will also be considered. In a nutshell, this study will focus on immunization factors, mortality factors, economic factors, social factors and other health related factors as well. Since the observations this dataset are based on different countries, it will be easier for a country to determine the predicting factor which is contributing to lower value of life expectancy. This will help in suggesting a country which area should be given importance in order to efficiently improve the life expectancy of its population.

Content

The project relies on accuracy of data. The Global Health Observatory (GHO) data repository under World Health Organization (WHO) keeps track of the health status as well as many other related factors for all countries The data-sets are made available to public for the purpose of health data analysis. The data-set related to life expectancy, health factors for 193 countries has been collected from the same WHO data repository website and its corresponding economic data was collected from United Nation website. Among all categories of health-related factors only those critical factors were chosen which are more representative. It has been observed that in the past 15 years , there has been a huge development in health sector resulting in improvement of human mortality rates especially in the developing nations in comparison to the past 30 years. Therefore, in this project we have considered data from year 2000-2015 for 193 countries for further analysis. The individual data files have been merged together into a single data-set. On initial visual inspection of the data showed some missing values. As the data-sets were from WHO, we found no evident errors. Missing data was handled in R software by using Missmap command. The result indicated that most of the missing data was for population, Hepatitis B and GDP. The missing data were from less known countries like Vanuatu, Tonga, Togo, Cabo Verde etc. Finding all data for these countries was difficult and hence, it was decided that we exclude these countries from the final model data-set. The final merged file(final dataset) consists of 22 Columns and 2938 rows which meant 20 predicting variables. All predicting variables was then divided into several broad categories:​Immunization related factors, Mortality factors, Economical factors and Social factors.

Acknowledgements

The data was collected from WHO and United Nations website with the help of Deeksha Russell and Duan Wang.

Inspiration

The data-set aims to answer the following key questions:

1. Does various predicting factors which has been chosen initially really affect the Life expectancy? What are the predicting variables actually affecting the life expectancy?
2. Should a country having a lower life expectancy value(<65) increase its healthcare expenditure in order to improve its average lifespan?
3. How does Infant and Adult mortality rates affect life expectancy?
4. Does Life Expectancy has positive or negative correlation with eating habits, lifestyle, exercise, smoking, drinking alcohol etc.
5. What is the impact of schooling on the lifespan of humans?
6. Does Life Expectancy have positive or negative relationship with drinking alcohol?
7. Do densely populated countries tend to have lower life expectancy?
8. What is the impact of Immunization coverage on life Expectancy?

Columns :

Country

Country

Year

Year

Status

Developed or Developing status

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

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 1 000 live births HIV/AIDS (0-4 years)

GDP

Gross Domestic Product per capita (in USD)

Population

Population of the country

thinness 1-19 years

Prevalence of thinness among children and adolescents for Age 10 to 19 (% )

thinness 5-9 years

Prevalence of thinness among children for Age 5 to 9(%)

Income composition of resources

Human Development Index in terms of income composition of resources (index ranging from 0 to 1)

Schooling

Number of years of Schooling(years)