**Proposal and Data Assembly**

**Ian Musumba**

1. With the assistance of Deeksha Russell and Duan Wang, the dataset was acquired from the repository databases of WHO and United Nations website. The data consists of health factors for 193 nations and its equivalent economic data was obtained from the United Nations website. Only the most representative critical factors from each category of health-related factors were selected. It has been noted that there has been significant progress in the health sector over the previous 15 years, which has improved human mortality rates. In order to do further analysis for this project, data for 193 nations from the year 2000 to 2015 were taken into account. 22 columns and 2938 rows make up the final dataset, which means 20 predicting variables.
2. Relative research questions
   1. Does a country’s status have an effect on life expectancy?
   2. Which variables are the best predictor for life expectancy?

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable Name** | **Original Definition** | **Units** | **Range or Levels** | **Possible Recording** | **Rationale** |
| Status | Developing or developed |  |  |  | Explanatory variable |
| Life expectancy | Life expectancy in age | age | 0-100 age |  | Response variable |
| Schooling | Number of years of Schooling(years) |  |  |  | Explanatory variable |
| Adult Mortality | Adult Mortality Rates of both sexes (probability of dying between 15 and 60 years per 1000 population) |  | 15-60 age |  |  |
| Infant deaths | Number of infants deaths |  |  |  | Explanatory variable |
| GDP | Gross Domestic Products |  |  |  | Explanatory variable |
| HIV/AIDS | Human immunodeficiency virus |  |  |  | Explanatory variable |
| Hepatitis B | Hepatitis B vaccine |  |  |  | Explanatory variable |
| BMI | Body Mass Index |  |  |  | Explanatory variable |
| Income composition of resources | Aggregated total income of that group/area |  |  |  | Explanatory variable |

1. Relevant Sources
   1. <https://www.cdc.gov/vaccines/parents/by-age/newborn-birth.html>

Summary: All babies should get the first shot of Hepatitis B vaccine within 24hours of birth.

* 1. <https://genus.springeropen.com/articles/10.1186/s41118-019-0071-0>

Summary: The findings demonstrate that population health and socioeconomic development in a nation have a significant impact on life expectancy at birth; in other words, when population health and socioeconomic development in a nation are improving, infant mortality rates decline, and as a result, life expectancy at birth appears to have increased. By fostering more economic growth and development in a nation, GDP per capita raises the life expectancy at birth and, in turn, promotes longevity.

* 1. <https://www.disabled-world.com/fitness/longevity/>

Summary: Life expectancy at birth has risen rapidly during the last century due to many factors, including reductions in infant mortality, rising living standards, improved lifestyles and better education, as well as advances in healthcare and medicine.

1. We plan to run multiple linear regression models with Life.expectancy as response and Status, Adult.Mortality, infant.deaths, Hepatitis.B, BMI, GDP, Population, Income.composition.of.resources, Schooling as explanatory variables, possibly examining interactions between all. We will also perform any necessary transformation to better my final model findings.