Title: Life Expectancy and Its Factors

Project Members:

Shachi Hardi :- SY BSc ASA, SAP ID: 75252019060

Keisha Mehta:- SY BSc ASA, SAP ID: 75252019051

Vedant Wakankar:- SY BSc ASA, SAP ID: 75252019033

Sanjana Waghray:- SY BSc ASA, SAP ID: 75252019137

Malhar Khatu :- SY BSc ASA, SAP ID: 75252019118

Project Supervisor:

Dr. K. S. Madhava Rao.

Professor, Statistics at Sunandan Divatia School Of Science, NMIMS

Glossary:

Motivation: Here you should bring out that aspect of your project which motivated you. Justify why a study on life expectancy is indeed needed and so on.

To determine if there is a relationship between life expectancy and its various factors, as well as to look into a dependency amongst the internal factors of life expectancy. In addition, to determine whether there is a change in life expectancy over time.

You can delete the highlighted text as it does not relate to motivation!

Objectives:

To find out if there is a correlation between GDP of a country and its life expectancy.

To test the hypothesis that the average life expectancy in 2015 is 60. (You need to justify why this gold standard value is assumed?)

To find out if there is relation between Reported cases of measles per 1000 population and thinness % of population (10-19 years) in Asian countries in 2015.

To find the relation between life expectancy and population of countries around the world in 2015.

To find out the correlation between percentage expenditure and under-five deaths in Asian countries for the year 2014.

To find the correlation between Adult Mortality and Alcohol Consumption.

Literature Review:

Source of Data:

Data has been collected from Kaggle using dataset published by Kumar Rajarshi, who has in turn collected the data from WHO and United Nations websites.

Data cleaning has been carried out on Microsoft Excel, with all null values being removed from the dataset.

Description of dataset:

There are 22 columns and 2938 rows of data.

Country:- Country

Year:- Year

Status:- Developing or Developed

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 1000 live births HIV/AIDS (0:-4 years)

GDP:- Gross Domestic Product per capita (in USD)

Population:- Population of the country

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

Thinness (5:-9) – Prevalence of thinness among children for Age 5 to 9(%)

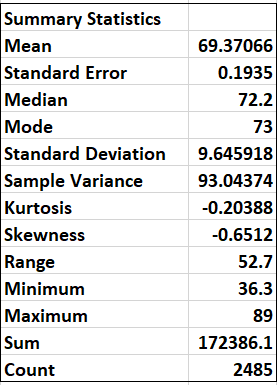
Income composition:- Human Development Index in terms of income composition of resources (index ranging from 0 to 1)

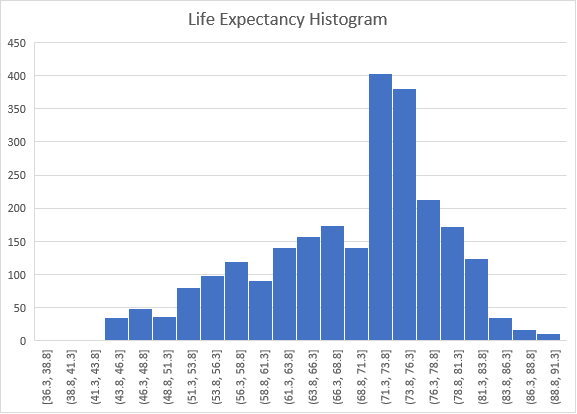
Schooling :- Number of years of Schooling (years)

Exploratory Data Analysis:

Exploratory Data Analysis was carried out, keeping the objectives in mind. All null values were removed from the data. Zero values were allowed to remain.

Exploratory Data analysis was first carried out on all data points of Life Expectancy, barring the missing values. Shown below is the frequency distribution and the summary statistics for Life Expectancy of all countries included, from the year 2000 to 2015. The data analysis package in Microsoft Excel was used to find the summary statistics.



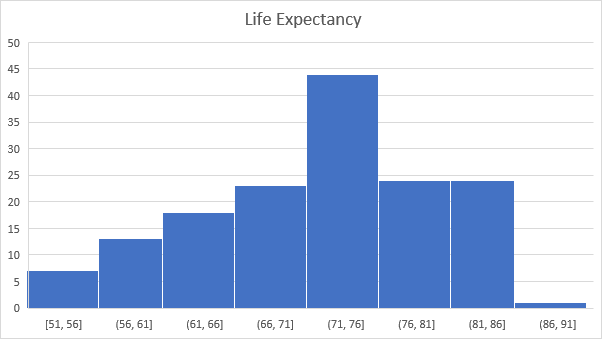


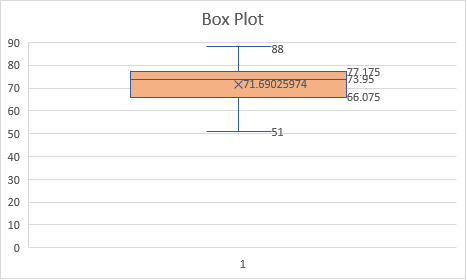
This seems to resemble a slightly negatively skewed ~~normal~~ distribution.

As the data points were deemed too sizeable, it was decided that EDA should be conducted on a more microscopic level, hence we then looked at data only from the year 2015.

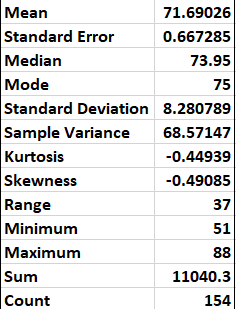
Visualisation of life expectancy according to country:

Visualisation of distribution of life expectancy in the year 2015:





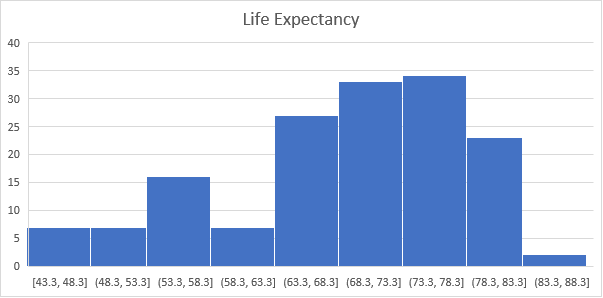
Summary Statistics for Life Expectancy for the year 2015:

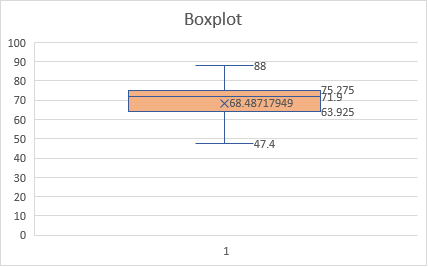


Similar analysis was performed for the year 2005:

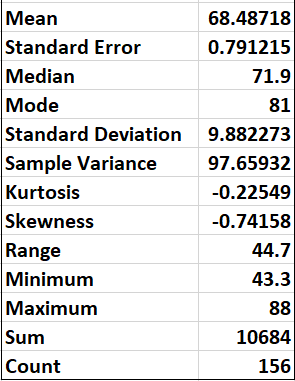
Visualisation of life expectancy according to country

Visualisation of distribution of life expectancy in the year 2005:

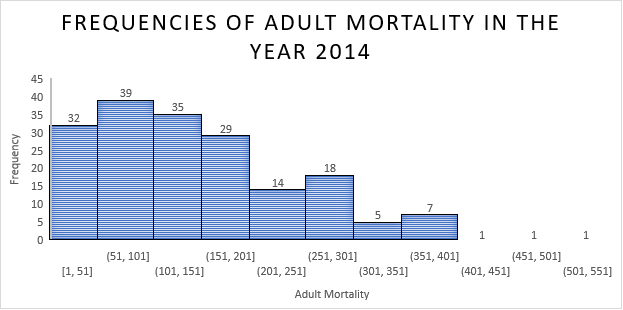


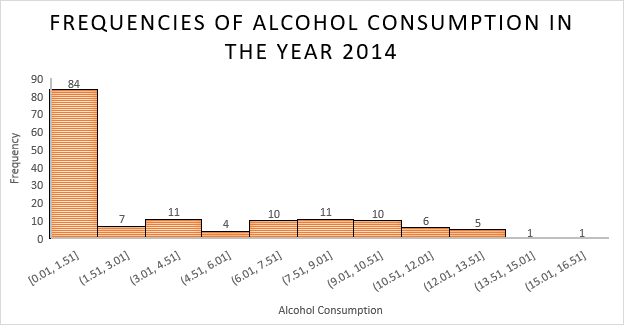


Summary Statistics for Life Expectancy for the year 2005:



Next, to visualise the frequencies of Adult Mortality and Alcohol Consumption, we use histograms. To get a detailed view, we specifically consider data of different countries only for the year 2014.

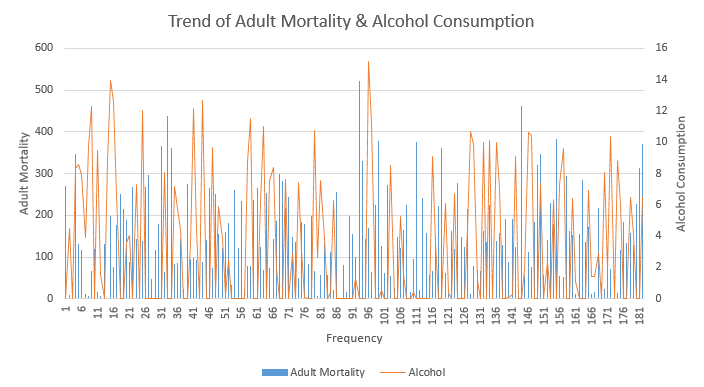




We can see that from the first diagram, that lower Adult Mortality has higher frequency.

As expected, we see in the second diagram that lower Alcohol Consumption has higher frequency.

In order to gain understanding of both the parameters together, we have used the visual aid given below:



**Moving on, we visualize the data pertaining to the number of reported cases of measles per thousand population and the percentage of thinness in the population who’s age ranges from 10 years to 19 years. As the data points were deemed too sizeable, only Asian countries have been considered for the year 2015.**

**Visualization of reported cases for Measles per 1000 population according to country**

**Visualization of percentage of thinness in population aged 10-19 years according to country**

**Summary statistics for Cases for Measles per thousand and Percentage of thinness (10-19 years)**

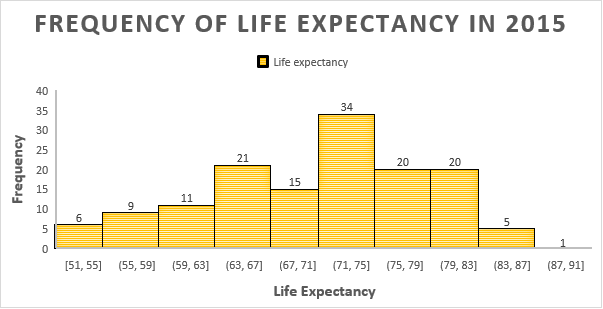
|  |  |
| --- | --- |
| ***Measles*** |  |
|  |  |
| **Mean** | 2420.555177 |
| **Standard Error** | 211.7032489 |
| **Median** | 16.5 |
| **Mode** | 0 |
| **Standard Deviation** | 11471.11275 |
| **Sample Variance** | 131586427.7 |
| **Kurtosis** | 114.7793474 |
| **Skewness** | 9.438034149 |
| **Range** | 212183 |
| **Minimum** | 0 |
| **Maximum** | 212183 |
| **Sum** | 7106750 |
| **Count** | 2936 |

|  |  |
| --- | --- |
| ***Thinness*** |  |
|  |  |
| **Mean** | 4.768813441 |
| **Standard Error** | 0.08227789 |
| **Median** | 3.3 |
| **Mode** | 1 |
| **Standard Deviation** | 4.397828249 |
| **Sample Variance** | 19.3408933 |
| **Kurtosis** | 4.246805581 |
| **Skewness** | 1.766439704 |
| **Range** | 27.6 |
| **Minimum** | 0.1 |
| **Maximum** | 27.7 |
| **Sum** | 13624.5 |
| **Count** | 2857 |

**Visualization of reported cases of measles per thousand population vs thinness in the population % aged (10-19 years)**

**As we see in the graph above, we can speculate that whenever thinness has peaked, measles has peaked too, however only in some places. Exceptions exist.**

Next, to visualize the relation between life expectancy and population, we use histograms and line diagrams. To get a detailed view, we specifically consider data of different countries only for the year 2015.



The above frequency histogram depicts the life expectancies of 142 countries around the world for the year 2015. The life expectancy is grouped in bins of 4 to aid visualization.

The above line diagram represents the variation in population across the countries for the year 2015.

The above line chart depicts the variations in life expectancies of different countries for the year 2015. As we can see in the chart, the life expectancy is concentrated between 60-80 for all countries considered, in the year 2015.

|  |  |
| --- | --- |
| **Population** | |
|  |  |
| Mean | 12788046 |
| Standard Error | 1284434 |
| Median | 1376133 |
| Mode | 444 |
| Standard Deviation | 61209729 |
| Sample Variance | 3.75E+15 |
| Kurtosis | 296.0741 |
| Skewness | 15.86513 |
| Range | 1.29E+09 |
| Minimum | 34 |
| Maximum | 1.29E+09 |
| Sum | 2.9E+10 |
| Count | 2271 |

**Summary Statistics for Life expectancy and Population of different countries for 2015**

|  |  |
| --- | --- |
| **Life Expectancy** | |
|  |  |
| Mean | 69.27284 |
| Standard Error | 0.176361 |
| Median | 72.05 |
| Mode | 73 |
| Standard Deviation | 9.451376 |
| Sample Variance | 89.32851 |
| Kurtosis | -0.23368 |
| Skewness | -0.62766 |
| Range | 52.7 |
| Minimum | 36.3 |
| Maximum | 89 |
| Sum | 198951.6 |
| Count | 2872 |

**Visualization of Life Expectancy Vs Population of different countries for the year 2015.**

As we can see in the above scatter plot, the majority of the population of all countries for the year 2015, have a life expectancy ranging between 60 and 80.

Next, we are trying to visualize the data between the percentage expenditure and the number of under-five deaths per 1000 population. We are considering the values only for the year 2014 in Asian countries as the data is sizeable.

***Visualization of percentage expenditure according to the countries.***

***Visualization of Under five deaths per 1000 according to the countries.***

***Summary statistics for percentage expenditure and under-five deaths.***

|  |  |
| --- | --- |
| ***Percentage Expenditure*** | |
| *Mean* | 749.6198 |
| *Standard Error* | 37.2337 |
| *Median* | 67.73374 |
| *Mode* | 0 |
| *Standard Deviation* | 2001.982 |
| *Sample Variance* | 4007934 |
| *Kurtosis* | 26.11398 |
| *Skewness* | 4.612456 |
| *Range* | 19479.91 |
| *Minimum* | 0 |
| *Maximum* | 19479.91 |
| *Sum* | 2167151 |
| *Count* | 2891 |

|  |  |
| --- | --- |
| ***Under-five Deaths*** | |
| *Mean* | 41.92148 |
| *Standard Error* | 3.00805 |
| *Median* | 4 |
| *Mode* | 0 |
| *Standard Deviation* | 161.7369 |
| *Sample Variance* | 26158.82 |
| *Kurtosis* | 108.0008 |
| *Skewness* | 9.422405 |
| *Range* | 2500 |
| *Minimum* | 0 |
| *Maximum* | 2500 |
| *Sum* | 121195 |
| *Count* | 2891 |

***Visualization of Under-five deaths VS percentage expenditure.***

From the above line chart we can see the pattern of percentage expenditure and Under-five deaths.