Contents

[Special Thanks: 2](#_Toc102975017)

[Section 1: Introduction 3](#_Toc102975018)

[1.1: Life Expectancy 3](#_Toc102975019)

[1.2: Report 3](#_Toc102975020)

[1.3: Data Sources and Software 3](#_Toc102975021)

[SECTION 2: DESCRIPTIVE STATISTICS 4](#_Toc102975022)

[SECTION 2.1: LIFE EXPECTANCY AND GDP PER CAPITA 4](#_Toc102975023)

[SECTION 2.2: LIFE EXPECTANCY AND HEALTHCARE EXPENDITURE 6](#_Toc102975024)

[SECTION 2.3: LIFE EXPECTANCY RELATED TO GENDER 8](#_Toc102975025)

[2.4 Life Expectancy and Countries in Different Continents. 11](#_Toc102975026)

[Section 2.5: Life Expectancy of Developed Countries vs Undeveloped Countries 13](#_Toc102975027)

[Section 3: Regression Analysis 17](#_Toc102975028)

[3.1.1 Regression analysis of GDP per capita and Life expectancy 17](#_Toc102975029)

[3.1.2 Regression Interpretation 18](#_Toc102975030)

[3.1.3 Regression prediction 18](#_Toc102975031)

[3.1.4 Conclusion 18](#_Toc102975032)

[3.2.1 Regression analysis of Health care expenditure and Life expectancy 19](#_Toc102975033)

[3.2.2 Regression Interpretation 20](#_Toc102975034)

[3.2.3 Regression prediction 20](#_Toc102975035)

[3.2.4 Conclusion 20](#_Toc102975036)

LIFE EXPECTANCY

PROJECT – INTRODUCTION TO STATISTICS - JAVED IQBAL

HASSAAN BOKHARI

KUMAIL ALIM

IRTIZA ZAIDI

# Special Thanks:

We would like to extend our gratitude to Dr. Javed Iqbal for actively helping us draft this report, and guiding us through the necessary steps required.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Name | ERP ID | Class ERP | Percentage Contribution |
| 1 | SYED MUHAMMAD IRTIZA ZAIDI | 23976 | 60616 | 33.30% |
| 2 | KUMAIL ALIM | | 60616 | 33.30% |
| 3 | SYED HASSAN BOKHARI | | 60616 | 33.30% |

Section 1: Introduction

## 1.1: Life Expectancy

The number of years a person can expect to live is referred to as "life expectancy." By definition, life expectancy is based on a prediction of the average age at which members of a population group would die. Life expectancy is an important measure of the judgment of a country. If life expectancy in a particular country is high, it directly relates to the well-being of that country, and the higher preference of people to migrate to that country.

## 1.2: Report

This report has been drafted courtesy of our professor, Dr. Javed Iqbal to explain indicators and variables for life expectancy. We have divided our report into three sections, with the next two sections named: Comparative Analysis, and Regression Analysis.

**Comparative Analysis (Descriptive Statistics):** This section has been divided into **five** subsections, comparing and explaining the differences and reasons for differences between varying life expectancies. The section includes descriptive statistics as a model to compare data and makes use of **boxplots, line graphs, histograms, and scatterplots** to elaborate on the differences.

**Regression Analysis**: The third section of this report was focused on providing a detailed regression analysis on two topics. We analysed how expenditure on **health care per capita**, and **GDP per capita** were two variables that brought changes to the life expectancy.

## 1.3: Data Sources and Software

This report has been created on Microsoft Word, using research data from the world-bank site. Microsoft Excel was then used to sort data, and plot different graphs using this data.

## SECTION 2: DESCRIPTIVE STATISTICS

This section of our report includes several distinct descriptive statistics used to compare the life expectancy of different groups of individuals, based on their gender, country of residence, their economic well-being, the health care they are provided with, and the crime-rate of the country they live in. The report consists of histograms, scatterplots, and box plots for differentiating purposes.

### SECTION 2.1: LIFE EXPECTANCY AND GDP PER CAPITA

We believe that the difference in life expectancy observed throughout the world, is largely because of the difference in Gross Domestic Product (GDP) Per Capita. GDP Per Capita is a value that allows us to measure the economic well being of the country with respect to the population size.

Table : Data for GDP

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *High Income* | *Upper Middle Income* | *Middle Income* | *Lower Middle Income* | *Low Income* |
| |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | |  | | --- | | Mean | | Standard Error | | Median | | First Quartile | | Third Quartile | | Interquartile | |  | | Standard Deviation | | Sample Variance | | Kurtosis | | Skewness | | Range | | Minimum | | Maximum | | Sum | | Count | | |  | | |  | | --- | | 76.63046 | | 0.428588 | | 76.65693 | | 74.2565 | | 79.36598 | | 5.10948 | |  | | 3.000117 | | 9.000703 | | -1.21632 | | -0.15604 | | 9.840107 | | 71.13795 | | 80.97805 | | 3754.892 | | 49 | | |  | | --- | | 69.8604 | | 0.541069 | | 69.55136 | | 67.39644 | | 72.86479 | | 5.468345 | |  | | 3.78748 | | 14.345 | | -0.76927 | | -0.14236 | | 13.99209 | | 61.98537 | | 75.97746 | | 3423.159 | | 49 | | |  | | --- | | 65.35882 | | 0.614134 | | 65.2834 | | 62.23821 | | 68.85936 | | 6.621144 | |  | | 4.298935 | | 18.48084 | | -0.94044 | | -0.1334 | | 15.25678 | | 56.91103 | | 72.16781 | | 3202.582 | | 49 | | |  | | --- | | 61.13062 | | 0.754456 | | 61.35444 | | 57.00265 | | 65.53706 | | 8.534413 | |  | | 5.281193 | | 27.891 | | -1.07124 | | -0.15333 | | 18.01058 | | 51.27224 | | 69.28282 | | 2995.401 | | 49 | | |  | | --- | | 53.59577 | | 0.813868 | | 51.80154 | | 49.40719 | | 58.49542 | | 9.088231 | |  | | 5.697076 | | 32.45667 | | -1.06622 | | 0.439617 | | 19.01337 | | 45.04039 | | 64.05376 | | 2626.193 | | 49 | |

This data was used to create a boxplots for better understanding, as shown in figure 1.

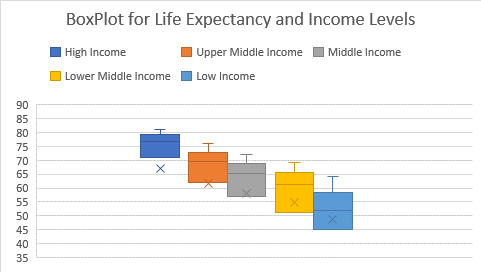


Figure 1: Boxplots for income levels

As is seems evident, the high income levels have the highest life expectancy, with lowest variation and range, whereas the low income levels have the lowest life expectancy, touching almost a minimum of 45 years of life!   
  
We then created another line graph, plotting life expectancy on the y axis and years on the x axis. This graph was used to display the average life expectancy of all income levels from 1970 to 2020. It shows how there Is a significant difference in life expectancy of lower middle income level and lower income, which is not visible in upper middle income level and high income! The graph shows a trend of how high income always generates the highest life expectancy, but also shows that high income’s life expectancy was most stable and changed slightly as opposed to lower income levels, where there is a steep increase in life expectancy.

Figure 2: Line Graph for income levels

### SECTION 2.2: LIFE EXPECTANCY AND HEALTHCARE EXPENDITURE

Healthcare is one of the most significant inputs to good health. We investigate evidence of a relationship between aggregate healthcare consumption and production and health outcomes across countries.

Current health expenditure is a typical means of evaluating national healthcare consumption and production. We created a life expectancy line graph, as well as a scatterplot. The line graph shows how

Table : Data for health care expenditure

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *High Income* | | *Middle income* | | *Low income* | |
| Mean | 4327.803 | Mean | 177.2547 | Mean | 31.6207 |
| Standard Error | 231.0208 | Standard Error | 18.77558 | Standard Error | 2.47978 |
| Median | 4563.613 | Median | 180.1602 | Median | 34.97951 |
| Mode | #N/A | Mode | #N/A | Mode | #N/A |
| Standard Deviation | 1033.156 | Standard Deviation | 83.96693 | Standard Deviation | 11.08991 |
| Sample Variance | 1067412 | Sample Variance | 7050.446 | Sample Variance | 122.9862 |
| Kurtosis | -0.92323 | Kurtosis | -1.64858 | Kurtosis | -1.32187 |
| Skewness | -0.55409 | Skewness | -0.13893 | Skewness | -0.50116 |
| Range | 3279.522 | Range | 230.0617 | Range | 30.89537 |
| Minimum | 2456.442 | Minimum | 63.87131 | Minimum | 13.6966 |
| Maximum | 5735.964 | Maximum | 293.933 | Maximum | 44.59197 |
| Sum | 86556.06 | Sum | 3545.094 | Sum | 632.4139 |
| Count | 20 | Count | 20 | Count | 20 |
|  | 0 |  | 0 |  | 0 |

As it can be seen by the table above, there is a vast difference between the expenditure levels between high income groups, and medium and low income groups. This hence, has a direct relationship between the life expectancy of population living in those countries.

For example, a country with a higher budget for health expenditure tends to show a greater life expectancy than those with lower budgets.

Figure 3: Line Graph for Health Care

We plotted the graph above to show the relationship between expenditure on health and life expectancy. The above graph shows how there is a positive relationship in most instances between health expenditure per capita and life expectancy.

Further, we created a scatterplot below to show how life expectancy has varied with an increase in health expenditure. This time, we plotted health expenditure as a percentage of GDP against life expectancy at birth.

Figure 4: Scatterplot for Health Care

### SECTION 2.3: LIFE EXPECTANCY RELATED TO GENDER

Life expectancy can also be compared on the basis of gender. The life expectancy between males and females at birth has shown variation, which can be shown in the table below.

It is evident after studying the data, that women have a greater life expectancy than men, a tradition that has been carrying out since ages. This is mostly due to the fact that men on average, are more involved in more risky jobs such as construction workers, soldiers for war, and also are more involved in practices like over speeding, drink and drive etc.

Table : Data of life expectancy wrt genders.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Life expectancy at birth, male (years)* | |  | *Life expectancy at birth, Female (years)* | |
|  |  |  |  |  |
| Mean | 70.57442 |  | Mean | 75.46521 |
| Standard Error | 0.508699 |  | Standard Error | 0.538261 |
| Median | 71.168 |  | Median | 77.2 |
| Mode | 78.6 |  | Mode | 83.6 |
| Standard Deviation | 7.176087 |  | Standard Deviation | 7.593102 |
| Sample Variance | 51.49622 |  | Sample Variance | 57.65519 |
| Kurtosis | -0.44244 |  | Kurtosis | -0.29066 |
| Skewness | -0.45529 |  | Skewness | -0.6833 |
| Range | 31.446 |  | Range | 32.125 |
| Minimum | 51.454 |  | Minimum | 55.875 |
| Maximum | 82.9 |  | Maximum | 88 |
| Sum | 14044.31 |  | Sum | 15017.58 |
| Count | 199 |  | Count | 199 |

Further we can note that both data are skewed negatively, and females have had a higher median, range, significantly higher maximum life (88 vs 82.9), and surprisingly they have a greater standard deviation as well as compared to males.

To further elaborate on this comparison, we have created a boxplot as well to visualise the change.

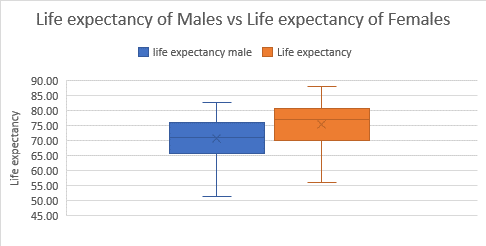


Figure 5: BoxPlot for Genders and LE

The boxplot clearly shows the huge difference in life expectancy between males and females. It shows how the mean of females is significantly larger than that of males, but also shows how the life expectancy is stretched more than that of males.

Further we created a scatterplot of males’ life expectancy sketched against females’ to note that females’ life expectancy has always been plotted above males.

Figure 6: Line Graph between Male and Females' Life Expectancy

We also plotted life expectancy of both the genders across the country against a time frame of 1972 to 2020, with time on the x-axis and life expectancy on the y-axis.

We were successful in finding one interesting fact, that throughout these 50 years, there was no year that the life expectancy of females would drop lower than men! This fact was very interesting for our group.

Figure 7: Line Graph for Average World Life Expectancy

### 2.4 Life Expectancy and Countries in Different Continents.

It is no surprise that life expectancies hugely vary from continent to continent. Some continents, like Europe consist of many developed countries like France, Italy etc. whereas some continents like Asia are majorly dominated by struggling countries. The data below provides information regarding life expectancy in different continents.

The reasons behind such differences in life expectancies is mainly due to limited and constrained health budgets, and staggering GDP per capita, two factors already discussed earlier that can highly change life expectancies.

Table : Life Expectancies for different continents

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | *Africa* | *North America* | *South America* | *Asia* | *Australia* | *Europe* |
|  |  |  |  |  |  |  |
| Mean | 52.16375 | 76.18499369 | 75.9605774 | 60.65518 | 78.137232 | 77.2813 |
| Standard Error | 0.666164 | 0.319593605 | 0.310454895 | 0.882979 | 0.496015 | 0.47649 |
| Median | 50.21457 | 76.24010485 | 76.02682927 | 61.06099 | 78.078049 | 77.2184 |
| First Quartile | 49.62588 | 74.72538912 | 74.56341463 | 55.62693 | 75.387805 | 74.5873 |
| Third Quartile | 45.35143 | 71.32627262 | 71.15609756 | 49.20068 | 71.457561 | 71.6054 |
| IQR | 4.274448 | 3.3991165 | 3.407317073 | 6.426253 | 3.9302439 | 2.98194 |
| Standard Deviation | 4.663145 | 2.237155237 | 2.173184267 | 6.180856 | 3.4721052 | 3.3354 |
| Sample Variance | 21.74492 | 5.004863554 | 4.72272986 | 38.20298 | 12.055514 | 11.1249 |
| Kurtosis | -0.47669 | -0.820094316 | -0.769317361 | -1.17682 | -1.142489 | -1.2489 |
| Skewness | 0.815888 | -0.364167402 | -0.378884249 | -0.16994 | -0.256108 | -0.0798 |
| Range | 16.59968 | 7.811810295 | 7.685365854 | 20.66717 | 11.742439 | 10.7208 |
| Minimum | 45.35143 | 71.32627262 | 71.15609756 | 49.20068 | 71.457561 | 71.6054 |
| Maximum | 61.95111 | 79.13808291 | 78.84146341 | 69.86785 | 83.2 | 82.3262 |
| Sum | 2556.024 | 3733.064691 | 3722.068293 | 2972.104 | 3828.7244 | 3786.78 |
| Count | 49 | 49 | 49 | 49 | 49 | 49 |

We plotted a box plot, as shown below for different continents and their respective life expectancies, and it was no surprise that Africa and Asia were among the two bottom countries with the lowest life expectancies. However, we realised a high variation in Asia, probably because of the presence of both, developed countries such as **China**, and undeveloped countries like **Afghanistan.**

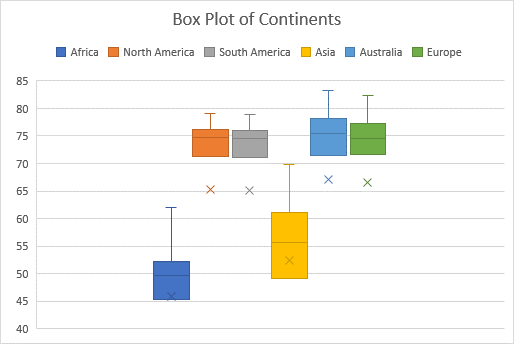


Figure 8: Boxplot for different continents.

Next, we plotted line graph of continents against time on the x axis, and we realised that there were only two continents, namely Asia and Africa that had a staggering life expectancy. However, throughout the time, there has been significant improvement in terms of life expectancy for Africa and Asia as compared to other continents.

Figure 9: Line Graph for different continents

### Section 2.5: Life Expectancy of Developed Countries vs Undeveloped Countries

To begin with our work, we collected the data on life expectancies of 15 developed and developing countries for the year 2020.

**DEVELOPING COUNTRIES DEVELOPED COUNTRIES**

Table : Tables for developed and undeveloped countries

|  |  |
| --- | --- |
| ***LIFE EXPECTANCY*** | |
|  |  |
| **Mean** | 71.4 |
| **Standard Error** | 1.433610162 |
| **Median** | 71 |
| **Mode** | 77 |
| **Standard Deviation** | 5.552348281 |
| **Sample Variance** | 30.82857143 |
| **Kurtosis** | -1.383575883 |
| **Skewness** | -0.284504448 |
| **Range** | 16 |
| **Minimum** | 62 |
| **Maximum** | 78 |
| **Sum** | 1071 |
| **Count** | 15 |
| **Largest(1)** | 78 |
| **Smallest(1)** | 62 |

|  |  |
| --- | --- |
| ***LIFE EXPECTANCY*** | |
|  |  |
| **Mean** | 82.33333333 |
| **Standard Error** | 0.303419663 |
| **Median** | 82 |
| **Mode** | 82 |
| **Standard Deviation** | 1.175139303 |
| **Sample Variance** | 1.380952381 |
| **Kurtosis** | 0.366870941 |
| **Skewness** | 0.767446686 |
| **Range** | 4 |
| **Minimum** | 81 |
| **Maximum** | 85 |
| **Sum** | 1235 |
| **Count** | 15 |
| **Largest(1)** | 85 |
| **Smallest(1)** | 81 |

As can be observed in table, the mean life expectancy for developed countries (**82.3**) in 2020 exceeded that of developing countries (**71.4**). In general, wealthy nations have a greater average life expectancy than poorer countries, which may be attributed to higher living standards, more effective health systems, and more money invested in health determinants (e.g. sanitation, housing, education).

Another thing to notice is that the standard deviation of life expectancies for developing countries is greater than that of developed countries. This means that the life expectancy of developing countries deviates more from the mean. It is also notable that the sample variance for the developing countries is higher than that of the developed countries.

In the following box plot, the blue box represents the developed countries whereas the orange one represents the developing countries. It can be seen clearly that even the minimum value of the developed countries is even higher than the maximum value of developing countries. Another observation is that the difference in the quartiles is greater in the orange box plot.

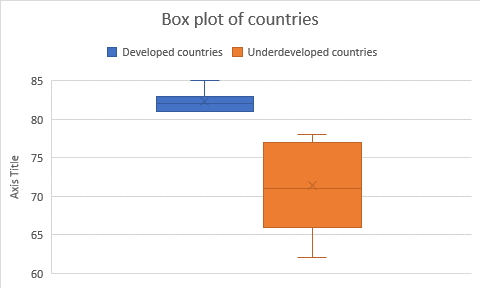


Figure 10:Boxplots of different countries

Our final goal was to determine the general historical pattern of life expectancy in both groups of countries. We used World Bank statistics from 2000 to 2020 to do this. We were able to generate a time-series graph, as shown in Figure 3, that illustrates the average worldwide life expectancy for developed and developing countries. We discovered that the average worldwide life expectancy for both types of countries increased in a comparable way across time.

Figure 11: Line Graph of Developed Countries

Figure 12: Line Graph of undeveloped countries

The two graphs plotted above show life expectancy between developing countries and developed countries. As we can notice, there is a vast difference between the life expectancies of these two type of countries, with developed countries having a significantly higher life expectancy than underdeveloped countries.

## Section 3: Regression Analysis

Regression is a statistical method that attempts to determine the strength and character of the relationship between one dependent variable (usually denoted by Y) and a series of other variables (known as independent variables). We will use simple regression in this case. Regression takes a group of random variables, thought to be predicting Y, and tries to find a mathematical relationship between them. This relationship is typically in the form of a straight line (linear regression) that best approximates all the individual data points.

### 3.1.1 Regression analysis of GDP per capita and Life expectancy

Hence, we used the GDP per capita is the independent variable, while life expectancy at birth is the dependent variable in our analysis. The sample size was 174 countries and the year used for the analysis was 2019.Relevant data was collected From <https://www.worldbank.org/en/home> and we used the excel to arrange the data and then find the regression equation , graph and summary for the relationship between GDP per capita and Life expectancy .

Each dot represents a single country

Figure : Regression between Life expectancy and GDP per capita($US)

The regression equation for this analysis is

**y = 0.0002x + 69.036**

Where y=Life expectancy

and x = GDP per capita($US)

### 3.1.2 Regression Interpretation

The regression eq represents 2 main aspects:

1)Interpretation of slope

According to our regression equation, for every $1 rise in a country's GDP per capita, its average life expectancy improves by 0.0002 years.

2)Interpretation of y-intercept

Furthermore, the y intercept shows that the average life expectancy of a country whose GDP per capita is $0 is 69.036 years on average.

However ,Regression model should be used of within sample prediction (interpolation) but not for outside the sample range prediction (extrapolation). because the minimum value of GDP per capita in the dataset is $ 228.2136, this is an unrealistic conclusion to make, as the value $0 falls outside the range.

### 3.1.3 Regression prediction

Prediction table

Table : Perdition table for Regression equation of life expectancy and GDP per capita ($US)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Countries | GDP per capita | Actual Life expectancy | predicted Life expectancy | Difference |
| Pakistan | 1288.56 | 67.27 | 69.29 | -2.02 |
| United States | 65279.53 | 78.79 | 82.09 | -3.30 |
| Japan | 40777.61 | 87.26 | 77.19 | 10.06 |
| Australia | 54875.29 | 82.90 | 80.01 | 2.89 |
| Italy | 33641.63 | 83.50 | 75.76 | 7.73 |

### 3.1.4 Conclusion

Our analysis also found that the value of R^2 is 0.433, correct to three significant figures. This means that 43.3% of the variation in average life expectancy is explained by the GDP per capita of the countries

Using our value for R^2 , we also calculated the correlation coefficient (r) as 0.658 .This means that there is a strong positive correlation between GDP per capita and life expectancy at birth.

### 3.2.1 Regression analysis of Health care expenditure and Life expectancy

In our analysis, we used health care expenditure as the independent variable and life expectancy at birth as the dependent variable. The representative sample comprised of 174 countries, with 2019 as the year of analysis. Relevant data was gathered from https://www.worldbank.org/en/home, and we used Excel to organise the information before calculating the regression equation, graph, and summary for the association between healthcare expenditures and life expectancy.

Table : Regression between Life expectancy and health care expenditure

Regression equation is

**y = 0.0024x + 69.86**

Where y=Life expectancy

And x =Health care expenditure

### 3.2.2 Regression Interpretation

The regression eq represents 2 main aspects:

1)Interpretation of slope

According to our regression equation, for every $1 rise in a country Health care expenditure, its average life expectancy improves by 0.0024 years.

2)Interpretation of intercept

The y intercept shows that the average life expectancy of a country whose Health care expenditure is $0 is 69.864 years on average.

However, regression models should be used for within-sample prediction (interpolation), not for prediction outside the sample range (extrapolation). This is an unrealistic conclusion to draw because the minimum value of Health care expenditure is $19.89, and the value $0 is outside the range.

### 3.2.3 Regression prediction

Prediction table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Countries | Health care expenditure | Actual Life expectancy | predicted Life expectancy | Difference |
| Pakistan | 39.499 | 67.27 | 69.96 | -2.69 |
| United States | 10921.01 | 78.79 | 96.07 | -17.29 |
| Japan | 4360.474 | 87.26 | 80.33 | 6.93 |
| Australia | 5242.183 | 82.90 | 82.45 | 0.45 |
| Italy | 2905.5 | 83.50 | 76.84 | 6.66 |

Table : Perdition table for Regression equation of life expectancy and Health care expenditure

### 3.2.4 Conclusion

Our analysis also found that the value of R^2 is 0.372, correct to three significant figures. This means that 37.2% of the variation in average life expectancy is explained by the health care expenditure the of the countries

Using our value for R^2 , we also calculated the correlation coefficient (r) as 0.61 .This means that there is a strong positive correlation between GDP per capita and Health care expenditure