IB Mathematics A&A (SL) Internal Assessment

To what extent do the *Life Expectancy* rates of developing and developed countries depend on their respective *Healthcare Expenditure Per Capita*?

Topic: Statistical Analysis of "Life Expectancy at Birth"
Candidate Code: jdk323
Session: May 2021
Page Count: 20 (excluding bibliography, appendix and cover page)

INTRODUCTION

Around this time last year, when we were studying the HDI Index in geography class, I got surprised by some of the extreme average life expectancies listed in the book, such as 53 for the Central African Republic and 84 for Japan, or 54 for Nigeria and 82-83 for most of the European countries. Then I wondered what were the main factors creating such huge gaps between the life expectancy levels of developing and developed countries. After doing some research on the World Bank database, I found that the developed countries on average spend \$2 000-\$10 000 per capita on health, whereas developing countries spend \$30-\$700 with an exception of some more robust economies spending up to \$1 000. Just by looking at the values, I noticed that there was a clear correlation between the health expenditure per capita and the life expectancy of countries. Therefore, for this Internal Assessment, I wanted to investigate the correlation of these variables in exact values and try to come up with a formula that can be applied to estimate the life expectancy of countries with their given health expenditure per capita values.

To carry out my investigation, I will use univariate and bivariate statistic data analysis in multiple steps:

Initially, for the univariate data analysis, I will explore the distribution of my independent (Health Expenditure per Capita, \$US) and dependent (Life Expectancy at birth, years) variables one by one first for developing and then developed countries using box-whisker plots, values from five-number summaries, and test for outliers. In case there are outliers, I will identify them and try to uncover the reasons by doing online research. Lastly, I will create parallel box-whisker plots displaying the data from both developing and developed countries collectively on a common scale, which I expect will give a clear understanding of the gap between them, if there is.

Next, for the bivariate data analysis, I will investigate the correlation between the two variables for developing and developed countries individually, using Pearson's Correlation Coefficient test, scatter diagrams, and least-squares regression method to come up with a formula to estimate the life expectancy of a country with its given health expenditure per capita. If I succeed in generating the formulas, I will test their validity by applying them to some countries and estimating their life expectancy values with their given health expenditure per capita values. I will then compare the obtained results with the real values from the World Bank Data, and calculate their percentage errors to check the extent to which the estimations were correct. Hence, the percentage error tests and the formulas to be created along with the results from Pearson's correlation test will help me to come up with an exact value for the correlation of two variables both for developing and developed countries and ultimately I will be able to answer my research question **"To what extent do the Life Expectancy rates of developing and developed countries depend on their respective Health Expenditure per Capita?"**.

Key Term Definitions:

Below are the definitions for key terms used throughout the whole investigation. Note that all these definitions were taken from reliable sources such as WHO, OECD Data, and World Bank:

★ **Life expectancy at birth** - the average number of years a newborn can expect to live if current death rates do not change. (WHO)

★ **Current health expenditure per capita** - the total amount of money each country spends on healthcare divided by the population number, in current US$. (OECD Data)

★ **Developing countries** - a country-classification terminology referred to low and middle-income countries, whose GNI per capita is less than $12,535 as of the 2021 fiscal year. (World Bank)

★ **Developed countries** - a country-classification terminology referred to high-income countries only, whose GNI per capita is more than $12,535 as of 2021 fiscal year. (World Bank)

★ **Gross National Income per capita, GNI/capita** - the country's annual total income in current US$ divided by its midyear population. (World Bank Data)

Data Obtaining Method:

To carry out the investigation within the IA scope 50 samples were to be chosen in total out of 218 provided by the World Bank Data as representatives from both developing and developed countries. Therefore, to remove the bias while choosing and to give every country an equal chance of being picked, www.random.org/lists - an online random sampling generator has been used to randomly choose 25 countries from each category. To do this, all 135 developing countries provided were placed on the online generator www.random.org/lists which then assorted them in random order, and then the first 25 were selected to be the investigation representatives. The same method was repeated to randomly choose 25 developed countries out of 83, provided by the World Bank Data.

Therefore, this method enabled me to investigate nearly ¼ of all the world countries provided by the World Bank Data to study the correlation between the "Life Expectancy at Birth" of a country and its "Health Expenditure per Capita" both for developing and developed countries.

UNIVARIATE DATA ANALYSIS

Given below are the randomly selected countries for investigation, 25 from each category:

Table 1. Randomly selected 25 developing and 25 developed countries

Developing Countries	Developed Countries
1. Thailand	1. New Zealand
2. Sudan	2. Germany
3. Philippines	3. United Kingdom
4. Georgia	4. Belgium
5. Burkina Faso	5. Italy
6. Belarus	6. Denmark
7. Tajikistan	7. Israel
8. Algeria	8. Canada
9. Egypt	9. Singapore
10. Paraguay	10. Australia
11. Uzbekistan	11. United States
12. Madagascar	12. Luxemburg
13. Timor-Leste	13. Japan
14. Azerbaijan	14. Sweden
15. Peru	15. Norway
16. Liberia	16. Iceland
17. Cameroon	17. Ireland
18. El Salvador	18. Finland
19. Kyrgyzstan	19. Greece
20. Cuba	20. Saudi Arabia
21. Venezuela	21. Netherlands
22. Turkmenistan	22. Switzerland
23. Kenya	23. Cyprus
24. Tanzania	24. Spain
25. Haiti	25. Austria

After generating the random country samples, data values of Life Expectancy at Birth (total) and Health Expenditure per Capita (in current $US) were extracted from the World Bank Data using 2018 statistics for both variables:

Table 2. Current Health Expenditure per Capita and Life Expectancy at Birth values, developing countries

Developing Countries	Health Expenditure per capita (current US$)	Life expectancy at birth, total (years)
1. Thailand	275.92	77
2. Sudan	60.17	65
3. Philippines	136.54	71
4. Georgia	312.75	74
5. Burkina Faso	40.25	61
6. Belarus	356.25	74
7. Tajikistan	59.84	71
8. Algeria	255.87	77
9. Egypt	125.55	72
10. Paraguay	400.39	74
11. Uzbekistan	82.27	72
12. Madagascar	22.05	67
13. Timor-Leste	93.69	69
14. Azerbaijan	165.77	73
15. Peru	369.08	77
16. Liberia	45.42	64
17. Cameroon	54.14	59
18. El Salvador	288.52	73
19. Kyrgyzstan	85.74	71
20. Cuba	986.94	79
21. Venezuela	256.95	72
22. Turkmenistan	460.18	68
23. Kenya	88.39	66
24. Tanzania	36.82	65
25. Haiti	64.25	64

Table 3. Current Health Expenditure per capita and Life Expectancy at Birth values, developed countries

Developed Countries	Current Health Expenditure per capita (current US$)	Life expectancy at birth, total (years)
1. New Zealand	4037.46	82
2. Germany	5472.2	81
3. United Kingdom	4315.43	81
4. Belgium	4912.7	82
5. Italy	2989	83
6. Denmark	6216.77	81
7. Israel	3323.65	83
8. Canada	4994.9	82
9. Singapore	2823.64	83
10. Australia	5425.34	83
11. United States	10623.85	79
12. Luxemburg	6227.08	82
13. Japan	4266.59	84
14. Sweden	5981.71	83
15. Norway	8239.1	83
16. Iceland	6530.93	83
17. Ireland	5489.07	82
18. Finland	4515.68	82
19. Greece	1566.9	82
20. Saudi Arabia	1484.59	75
21. Netherlands	5306.53	82
22. Switzerland	9870.66	84
23. Cyprus	1954.41	81
24. Spain	2736.32	83
25. Austria	5326.44	82

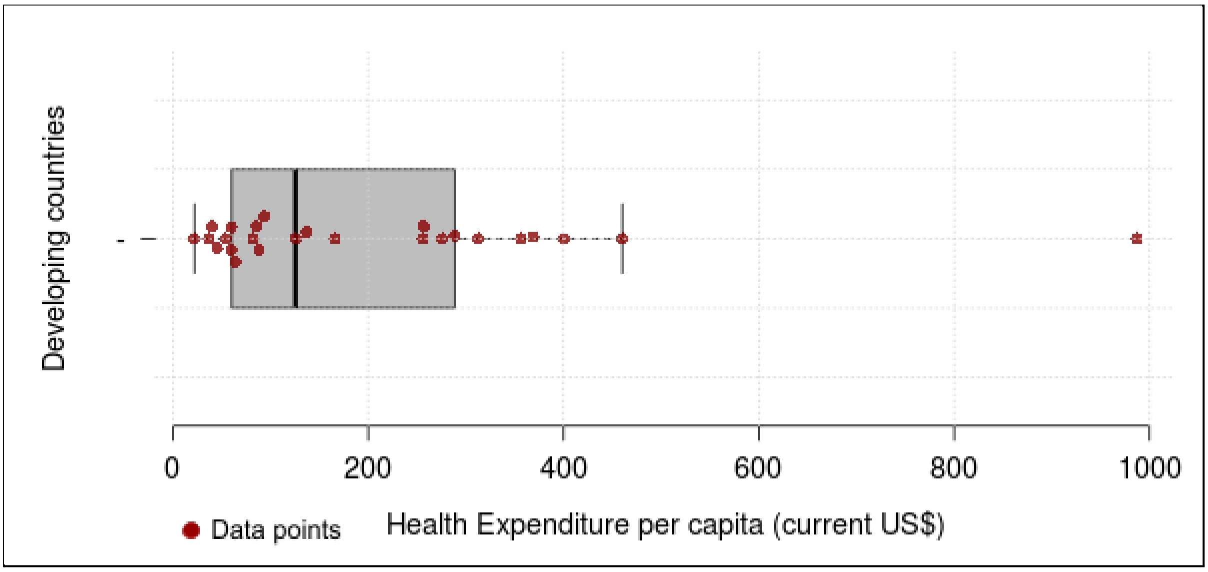
Step 1:

After extracting the data values of both variables for randomly selected developing and developed countries, the first step of my investigation was to look into the distribution of the independent variable - Current Health Expenditure per Capita (current US$). To do this, the TI-84 Plus CE graphic calculator, Google Spreadsheets, and BoxPlotR online boxplot generator at http://shiny.chemgrid.org/boxplotr/ were used to individually graph the Health Expenditure per Capita values of developing and developed countries in the form of a box-whisker plot, which allowed me to visually analyze the distribution of all the data. At the same time, it allowed me to obtain the five-number summaries, which consist of five values; maximum and minimum, upper and lower quartiles, and median. Next, the test for outliers was carried out to check if there is an outlier country from the list of given 25 countries. Hence all this information gave me a better understanding of health expenditure per capita distributions for each data set, developing and developed world:

Table 4. Five-number summaries of Health Expenditure per capita in current US$

Five-Number Summaries of Health Expenditure per capita in Current US$		
	Developing Countries	Developed Countries
Minimum	22.05	1484.59
Lower Quartile	60.17	3323.65
Median	125.55	4994.9
Upper Quartile	288.52	5981.71
Maximum	986.94	10623.85

Graph 1. Box-Whisker Plot for Developing Countries' Health Expenditure per capita in current US$



$$\begin{aligned} \text{Range} &= \text{max - min} \\ &= 986.94 - 22.05 \\ &= \$964.89 \end{aligned} \qquad \begin{aligned} \text{IQR} &= Q_3 \text{ - } Q_1 \\ &= 288.52 - 60.17 \\ &= \$228.35 \end{aligned}$$

Test for Outliers:

The upper limit = $Q_3 + 1.5 \times IQR$
= $288.52 + 1.5 \times 228.35$
= \$631.045

★ There is only one value above the upper limit, which is \$986.94 belonging to Cuba. So Cuba is the outlier.

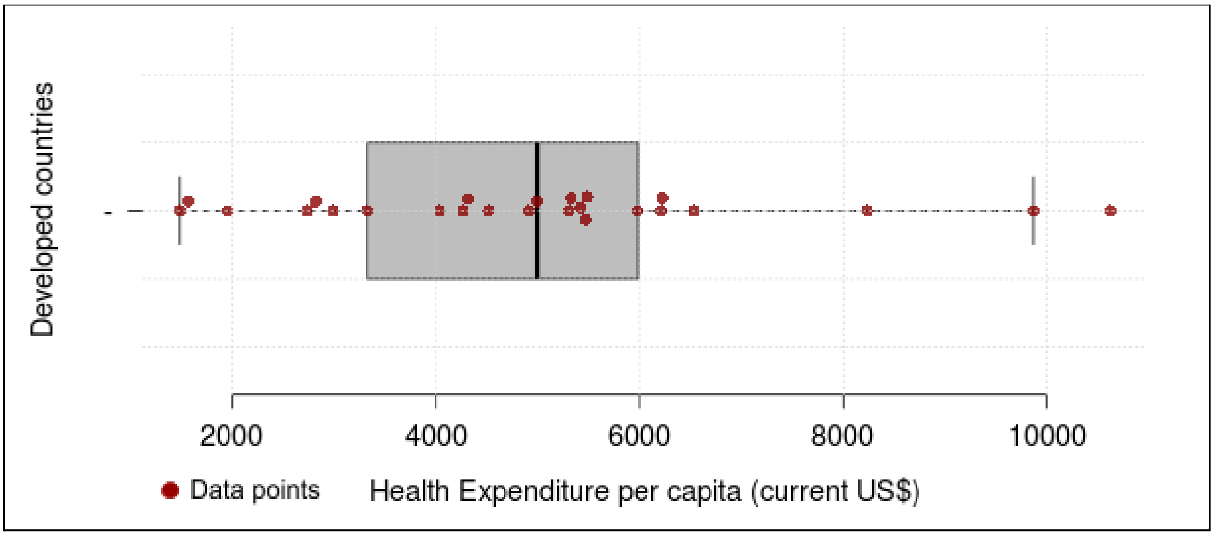
The lower limit = $Q_1 - 1.5 \times IQR$
= $60.17 - 1.5 \times 228.35$
= \$ -282.355 US

★ There is no value less than the lower limit, so there is no outlier.

The box-whisker plot above is in a positively skewed distribution, meaning that most of the data fall around the low values as can be seen from the spread of red data points. Additionally, this distribution has a median value of \$125.55, a range of \$964.89, and an IQR of \$288.35.

There appears to be only one outlier or an extreme value, \$986.94, which belongs to Cuba exceeding the upper limit of \$631.045. I did some research on this matter and found out that Cuba's healthcare system is indeed considered one of the best in the world (Warner), which the UN secretary-general Ban Ki-moon says "a model for many countries" (United Nations). The country has a GDP per capita of \$8,821.8 for its just over 11 million population. In addition to that, the country runs a current account surplus of \$985.4 million, which implies that the country exports more than it imports indicating high economic productivity. This high economic productivity, in turn, secures higher incomes and more job opportunities for people in the country. Therefore, the government and people have more money compared to most of the developing countries to spend on healthcare, explaining why it appeared as an outlier in our investigation.

Graph 2. Box-Whisker Plot for Developed Countries' Health Expenditure per capita in current US$



$$\text{Range} = \text{max - min} = 10\,623.85 - 1\,484.59 = \$9\,139.26$$

$$\text{IQR} = Q_3 - Q_1 = 5\,981.71 - 3\,323.65 = \$2658.06$$

Test for Outliers:

The upper limit = $Q_3 + 1.5 \times \text{IQR}$
= $5\,981.71 + 1.5 \times 2\,658.06$
= $9 968.8

★ There is only one outlier, $10 623.85, which belongs to the United States.

The lower limit = $Q_1 - 1.5 \times \text{IQR}$
= $3\,323.65 - 1.5 \times 2\,658.06$
= $ -663.44

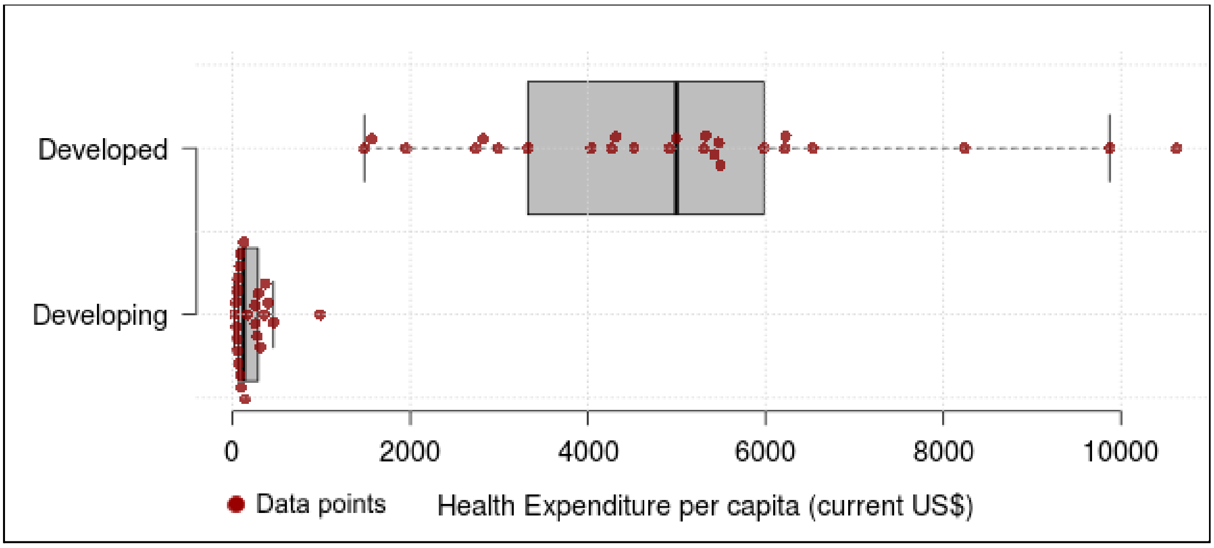
★ There is no value less than the lower limit, so there is no outlier.

The box-whisker plot in Graph 2 has a negatively skewed distribution with a median of $4 994.9, a range of $9 139.26, and an IQR of $2 658.06. As can be seen in the box-whisker plot, most of the data fall around the median, with exceptional extreme values and one outlier.

The outlier identified belongs to the United States, $10 623.85, exceeding the generated upper limit. After some online research, I found out that the main reasons for this are relatively expensive pharmaceuticals, medical care, expensive tests, and administrative costs. For example, insulin for diabetes in the US costs $186 monthly, whereas it costs only a third of that in Canada. Another fact from the same source, people in the US pay $86 for cholesterol-lowering medication, whereas people in Germany pay less than half of this price (Dr. Hector Florimon). Hence, these higher prices add up to the higher expenditure value, explaining why it appeared as an outlier in our investigation.

Individually generated box-whisker plots above have different scales, which limits our understanding of their distribution comparing the two. Therefore, a parallel box-whisker plot containing both data sets was generated enabling us to visually compare their respective distribution in relation to each other:

Graph 3. Parallel Box-Whisker Plot of Health Expenditure per Capita in current US$, developing and developed countries



The parallel box-whisker plot above shows the difference in Health Expenditure per Capita of developing (bottom one) and developed (top one) countries on a common scale, enabling us to visually compare the gap between the two. The median for developing countries is $125.55, whereas developed countries have a median of $4994.9, nearly 40 times bigger. Therefore, it can be concluded that developed countries, on average, spend more than developing countries on healthcare. The reasons can be higher incomes, expensive pharmaceuticals and medical care, higher government spendings on healthcare, and so on.

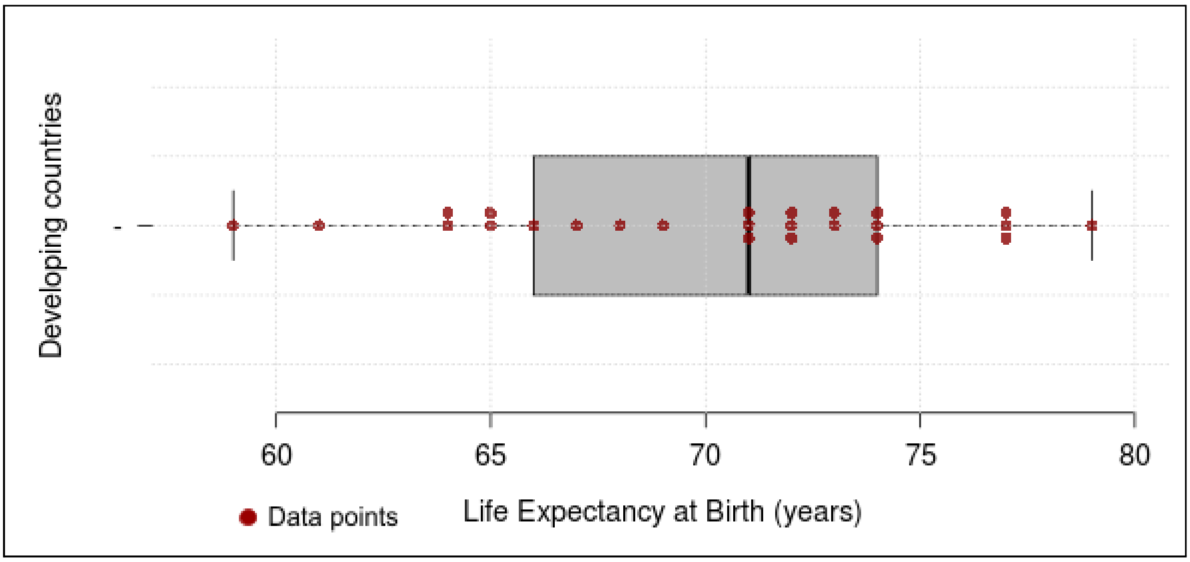
Step 2:

From the first step of the investigation, we got a clear understanding of Health Expenditure per capita distribution for developing and developed countries. Therefore, the second step of my investigation was to look at the distribution of the Life Expectancy at Birth rates for developing and developed countries following the same method, first plotting individually and afterward, in a parallel box-whisker plot to better analyze their distribution in relation to each other. To do this, the TI-84 Plus CE graphing calculator, Google Spreadsheets, and an online box-whisker plot generating tool, BoxPlotR, were used. To display the data, again box-whisker plot was chosen, enabling me to visually analyze the distribution of data at the same time as helping me to obtain the five-number summaries using the values from Table 2 and Table 3 above:

Table 5. Five-number summaries of Life Expectancy at Birth

Five-Number Summaries of Life Expectancy at Birth, total (years)		
	Developing Countries	Developed Countries
Minimum (=min)	59	75
Lower Quartile (= Q_1)	66	82
Median (=med)	71	82
Upper Quartile (= Q_3)	74	83
Maximum (=max)	79	84

Graph 4. Box-Whisker Plot for Developing Countries' Life Expectancy at Birth, total



$$\text{Range} = \text{max - min} = 79 - 59 = 20 \text{ years}$$

$$\text{IQR} = Q_3 - Q_1 = 74 - 66 = 8 \text{ years}$$

Test for Outliers:

$$\text{The upper limit} = Q_3 + 1.5 \times \text{IQR} = 74 + 1.5 \times 8 = 86 \text{ years}$$

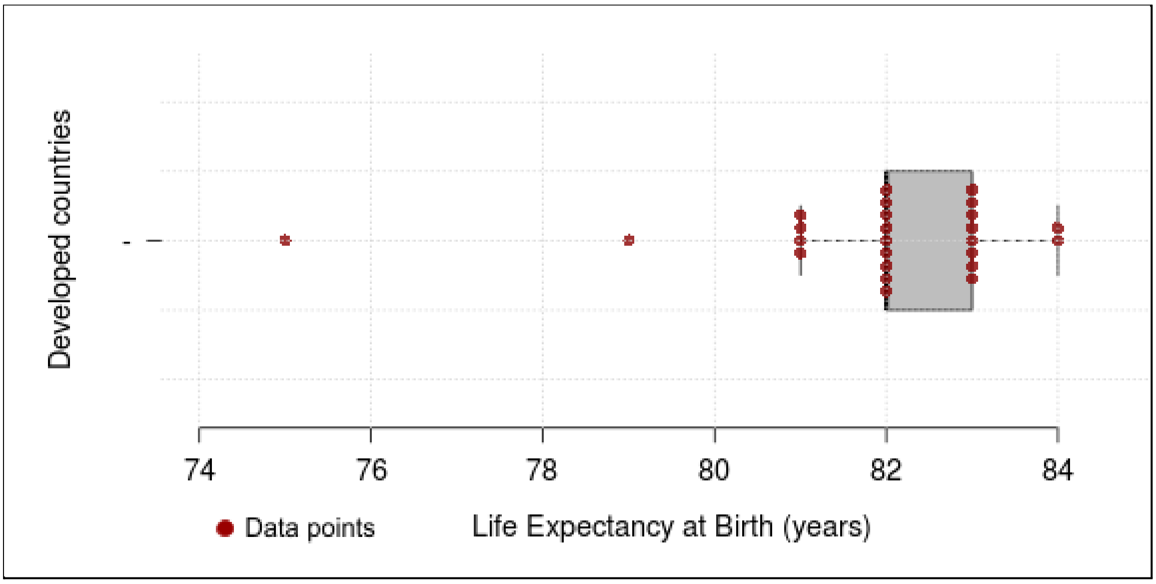
★ There is no value more than the upper limit, so there is no outlier.

$$\text{The lower limit} = Q_1 - 1.5 \times \text{IQR} = 66 - 1.5 \times 8 = 54 \text{ years}$$

★ There is no value less than the lower limit, so there is no outlier.

The box-whisker plot in Graph 4 is in a negatively skewed distribution, meaning that most of the data fall around the upper values or on the right side of the median, with some extreme values in the first and fourth quarters. This can easily be noticed with the distribution of red data points. Additionally, this distribution has a median value of 71 years, a range of 20 years, and an IQR of 8 years. The lower limit is 54 years whereas the upper limit is 86 years, and since there is no developing country whose Life Expectancy at Birth value exceeds the limits, there was no outlier noticed.

Graph 5. Box-Whisker Plot for Developed Countries' Life Expectancy at birth, total



$$\text{Range} = \text{max - min} = 84 - 75 = 9 \text{ years}$$

$$\text{IQR} = Q_3 - Q_1 = 83 - 82 = 1 \text{ year}$$

Test for Outliers:

$$\textbf{The upper limit} = Q_3 + 1.5 \times \text{IQR} = 83 + 1.5 \times 1 = 84.5 \text{ years}$$

★ There is no value more than the upper limit, so there is no outlier on the right side.

$$\textbf{The lower limit} = Q_1 - 1.5 \times \text{IQR} = 82 - 1.5 \times 1 = 80.5 \text{ years}$$

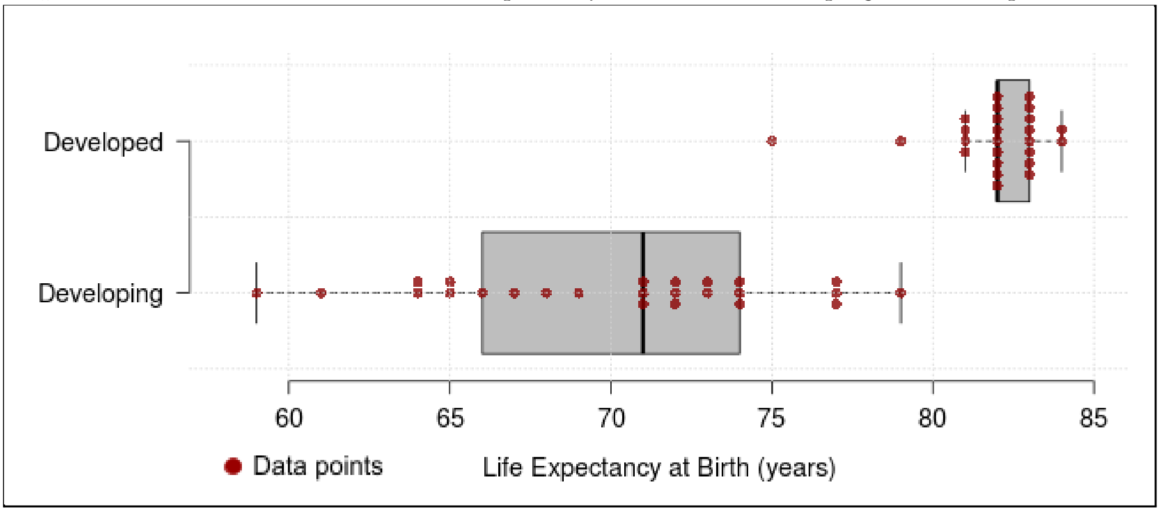
★ There are two outliers with values below the lower limit. These belong to Saudi Arabia with 75 years, and the US with 79 years.

The box-whisker plot in Graph 5 has a very strong positively skewed distribution, meaning that most of the data fall in the first half, which also can be seen by the data points on the graph shown in red colors. The box-whisker plot, surprisingly, has a mean with the same value as the Q_1, 82. Besides, it has a range of 9 years and an IQR of 1 year.

There are two outliers in the distribution in total, belonging to Saudi Arabia (75 years) and the US (79 years), exceeding the lower limit of the box-whisker plot. Further research was carried out to find the reasons for this for both countries. Firstly, for Saudi Arabia, it was found that the country has newly joined the team of developed countries due to its high income from oil exports. Therefore, the healthcare system of the country is still in the process of improvement. Besides, the climate of the country is marked by really high temperatures, following a pattern of a desert climate (Weather Online). Hence this extreme climate links to one of the main reasons for Saudi Arabia's death rates - heat stroke (Tyrovolas et al.). Therefore, the level of Saudi Arabia's life expectancy hasn't reached the level of other rich countries yet. Secondly, for the US, it was found that the main reasons for shorter life expectancy are high levels of smoking, homicides, opioid overdoses, and suicides (Roser). However, for the case of the US, it is interesting that it also appeared as an outlier in the Health Expenditure per Capita investigation for spending more on healthcare than other rich countries. This means that even though people spend much more on healthcare in the US, the life expectancy is yet to reach the level of other developed countries. Therefore, for the reasons mentioned above, it is no surprise that Saudi Arabia and the US appeared as outliers in my investigation.

Again, individually generated box-whisker plots above have different scales, limiting our understanding of the real distribution of data sets to some extent. Therefore, a parallel box-whisker plot containing both data sets was generated to visually compare their respective distribution in relation to each other:

Graph 6. Parallel Box-Whisker Plot of Life Expectancy At Birth for Developing and Developed countries

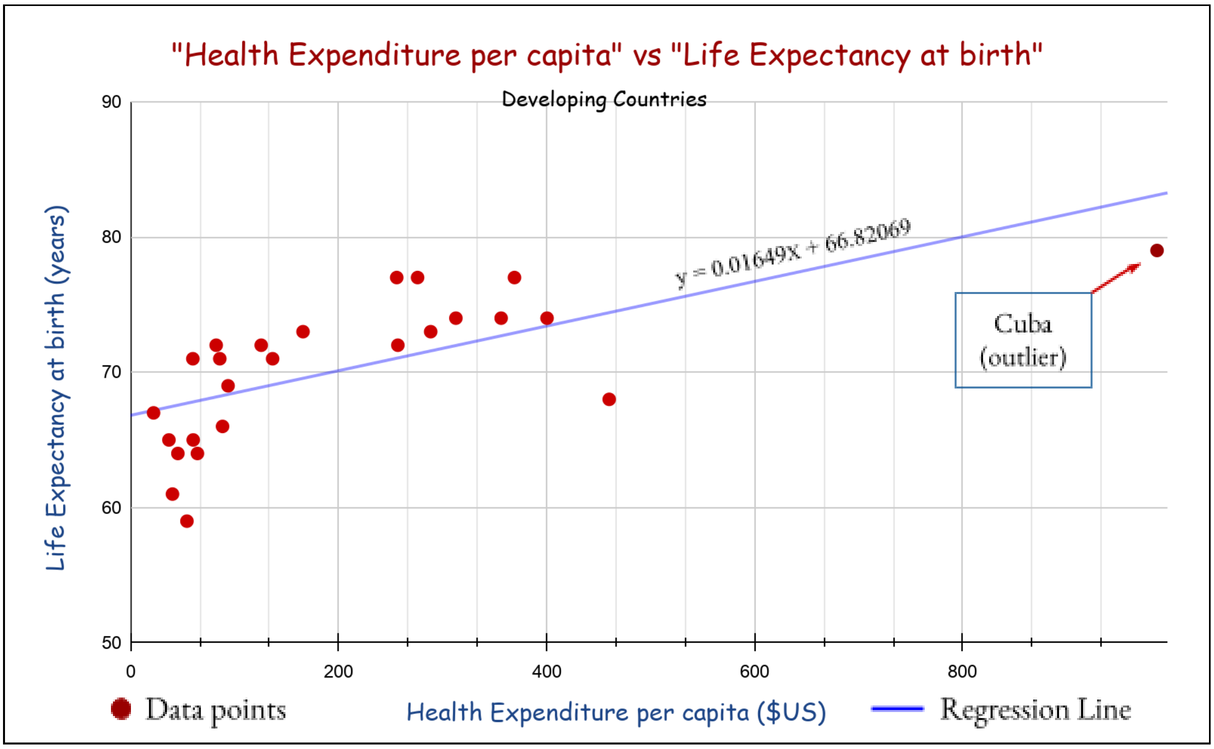


The parallel box-whisker plot above shows the difference in Life Expectancy at Birth for the total population of developing (bottom one) and developed (top one) countries on a common scale, which enables us to visually compare the gap between the two. From the graph, it can be seen that the median for developing countries is 71 years with a minimum value of 59 and a maximum value of 79, whereas the median for developed countries is 82 years with a minimum value of 75 and a maximum value of 84. Hence, it can be concluded that people in developed countries, on average, live longer than people do in developing countries.

BIVARIATE DATA ANALYSIS

Step 3 (a):
After exploring the distribution of Health Expenditure per capita and Life Expectancy at birth in developing and developed countries individually and together, the third step of my investigation was to analyze the correlation between the two variables. To do this, I generated their respective scatter diagrams using Google Spreadsheets, and then used Pearson's Correlation Coefficient test to check the nature of the correlation.

Graph 7. Two-Variable Scatter Diagram for Developing Countries



Pearson's Correlation Coefficient formula:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

where;

r = correlation coefficient
x_i = x-variable values of the sample
$\bar{x}$ = mean of the x-variable values
y_i = y-variable values of the sample
$\bar{y}$ = mean of the y-variable values

The results are obtained by using TI-84 Plus CE and online Pearson Correlation Coefficient Calculator at www.socscistatistics.com:

Correlation coefficient, r (25) = 0.6599

The correlation coefficient obtained shows that there is a strong positive correlation between the Health Expenditure per Capita and Life Expectancy at Birth values of developing countries. The positiveness of the correlation means that the more the healthcare expenditure, the higher are the life expectancy levels.

***P*-value, $p = 0.000332$, significant at $p < 0.01$**

The probability of occurrence by chance is really small, 0.03%. Therefore, with 99.97% confidence, we can say that there is a correlation between the two variables.

Coefficient of determination, $r^2 = 0.4355$

The coefficient of determination value, r^2, indicates that in developing countries 43.55% of the variation in Life Expectancy At Birth can be explained by the variation in Health Expenditure per Capita, leaving out the remaining 56.45%, to link to other factors.

Least squares regression line, $y = 0.01649x + 66.82069$

The line of best fit, shown in Graph 7, was found using the TI-84 Plus CE, and then tested using the online Linear Regression Calculator on www.socscistatistics.com/tests/regression/. The equation can be applied as follows:

Life Expectancy At Birth (years) = 0.01649 × *Health Expenditure Per Capita($US)* + 66.82069

y (\$0) = 66.82069 years		
	1.649 years	Differences
y (\$100) = 68.46969 years		
	1.649 years	
y (\$200) = 70.11869 years		
	1.649 years	
y (\$300) = 71.76769 years		

Therefore, from the calculations above using the obtained equation of the best fit line, it can be summarized that, in developing countries, for every additional \$100 health expenditure the life expectancy increases by 1.649 years.

To test its validity, the generated formula was applied to four randomly chosen developing countries, two from the list in this investigation, Tanzania and Madagascar, and two that are not being explored in this investigation, Zambia and Bulgaria. Note that the values for Current Health Expenditure per capita and Life Expectancy at birth of Zambia and Bulgaria were extracted from the same source as other countries, World Bank Data.

Tanzania → $y(\$36.82) = 0.01649 \times 36.82 + 66.82069 = 67.428$ years

Madagascar → $y(\$22.05) = 0.01649 \times 22.05 + 66.82069 = 67.184$ years

Zambia → $y(\$75.99) = 0.01649 \times 75.99 + 66.82069 = 68.074$ years

Bulgaria → $y(\$689.91) = 0.01649 \times 689.91 + 66.82069 = 78.197$ years

Table 6. Data from World Bank vs generated values using the formula

Country	Current Health Expenditure per capita (US$), World Bank	Life Expectancy at birth (years), World Bank	Life Expectancy at birth (years), generated equation
Tanzania	36.82	65	67.428
Madagascar	22.05	67	67.184
Zambia	75.99	64	68.074
Bulgaria	689.91	75	78.197

Using the findings from the generated formula and the data from the World Bank, the percentage error values were estimated with the formula below:

$$\text{Percentage error} = \left| \frac{\text{estimated value - real value}}{\text{real value}} \right| \times 100\%$$

$$\text{Tanzania} \rightarrow \left| \frac{67.428 - 65}{65} \right| \times 100\% = 3.735\%$$

$$\text{Madagascar} \rightarrow \left| \frac{67.184 - 67}{67} \right| \times 100\% = 0.274\%$$

$$\text{Zambia} \rightarrow \left| \frac{68.074 - 64}{64} \right| \times 100\% = 6.366\%$$

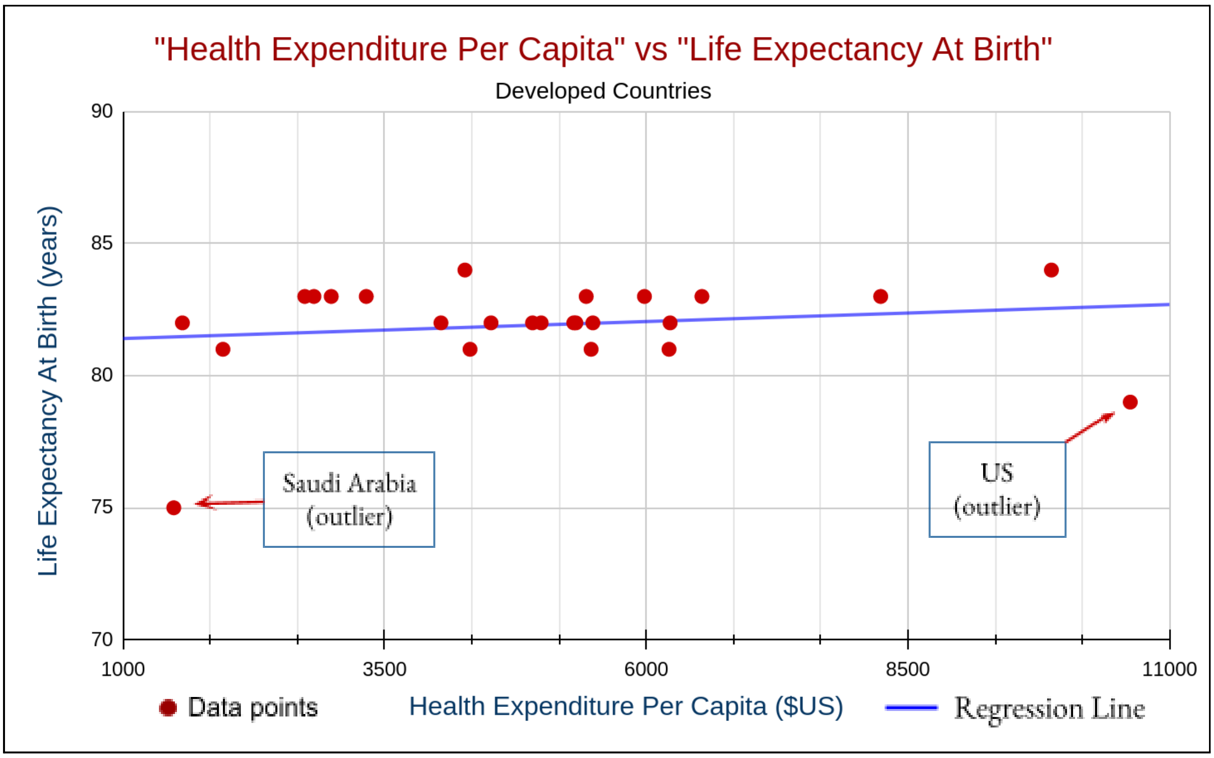
$$\text{Bulgaria} \rightarrow \left| \frac{78.197 - 75}{75} \right| \times 100\% = 4.263\%$$

The generated Life Expectancy at Birth value of Tanzania, using the formula, was 67.428 years compared to the 65 years given by World Bank Data. Therefore, the estimated percentage error was found to be just 3.735%. Using the same formula, the estimated percentage errors for the other three countries are all below 7%, indicating that the least-squares regression line has been an appropriate tool to predict the Life Expectancy At Birth values of developing countries with the given Healthcare Expenditure per capita values.

Step 3 (b):

Below is the scatter diagram displaying the Health Expenditure per capita values in relation to Life Expectancy at birth values of **developed** countries:

Graph 8. Two-Variable Scatter Diagram for Developed Countries



The TI-84 Plus CE graphing calculator and online tool Pearson Correlation Coefficient Calculator at www.socscistatistics.com were used to obtain all the results below:

Correlation coefficient, *r* (25) = 0.1628

The Pearson's correlation coefficient obtained shows that there is a weak positive correlation between the Health Expenditure Per Capita and Life Expectancy At Birth values of 25 given developed countries.

***P*-value, *p* = 0.436843**

The probability of occurrence by chance is really big, 43.68% in percentage. Therefore, the correlation results cannot be considered significant.

Coefficient of determination, r^2 = 0.0265

The coefficient of determination value, r^2, indicates that in developed countries only 2.65% of the variation in Life Expectancy At Birth can be explained by the variation in Health Expenditure Per Capita, which also signifies that the remaining 97.35% of variation links to other factors.

Least squares regression line, $y = 0.00013x + 81.28047$

The line of best fit, also shown in Graph 8, was found using the TI-84 Plus CE, and later tested using the online Linear Regression Calculator on www.socscistatistics.com/tests/regression/ . The equation can be applied as follows:

$$\textit{Life Expectancy At Birth (years)} = 0.00013 \times \textit{Health Expenditure Per Capita(\$US)} + 81.28047$$

y (\$0) = 81.28047 years
y (\$100) = 81.29347 years
y (\$200) = 81.30647 years
y (\$300) = 81.31947 years

0.013 years
0.013 years
0.013 years
Differences

The calculations above are estimated Life Expectancies of the general population for developed countries with their differences for each additional \$100 using the line regression equation obtained earlier.

The results signify that for every additional \$100 expenditure on health, the Life Expectancy at Birth values increase by just 0.013 years.

Once again, to test its validity, the generated formula was applied to four randomly chosen developed countries, two from the list in this investigation, Ireland and Denmark, and two that are not being explored in this investigation, UAE and Portugal. Note that the values for current Health Expenditure per capita and Life Expectancy at Birth of UAE and Portugal were extracted from the same source as other countries, World Bank Data.

Ireland → y (\$5489.07) = 0.00013×5489.07 + 81.28047
= 81.994 years

Denmark → y (\$6216.77) = 0.00013×6216.77 + 81.28047
= 82.089 years

UAE → y (\$1817.35) = 0.00013×1817.35 + 81.28047
= 81.517 years

Portugal → y (\$2215.17) = 0.00013×2215.17 + 81.28047
= 81.568 years

Table 7. Data from World Bank vs generated values using the formula

Country	**Health Expenditure per capita (current US$), World Bank**	**Life Expectancy at birth (years), World Bank**	**Life Expectancy at birth, generated equation**
Ireland	5489.07	82	81.994
Denmark	6216.77	81	82.089
UAE	1817.35	78	81.517
Portugal	2215.17	81	81.568

Afterward, using the findings from the generated formula and the data from the World Bank, the percentage error values were estimated with the same formula used earlier:

$$\text{Ireland} \rightarrow \left|\frac{81.994 - 82}{82}\right| \times 100\% = 0.007\%$$

$$\text{Denmark} \rightarrow \left|\frac{82.089 - 81}{81}\right| \times 100\% = 1.345\%$$

$$\text{UAE} \rightarrow \left|\frac{81.517 - 78}{78}\right| \times 100\% = 4.509\%$$

$$\text{Portugal} \rightarrow \left|\frac{81.568 - 81}{81}\right| \times 100\% = 0.701\%$$

The calculated Life Expectancy at Birth values of all four randomly chosen countries were really close to the real values extracted from World Bank Data. Therefore, the percentage error values were all below 5%, indicating that the least-squares regression line equation has been an appropriate tool to predict the Life Expectancy at Birth values of developed countries with their given Healthcare Expenditure per capita values, even though the correlation coefficient was really small

CONCLUSION

Throughout the investigation, it was found that Life Expectancy at Birth values of countries, to a certain extent, depend on their Health Expenditure per Capita. To explore this, the investigation was carried out in 3 steps:

The First Step was dedicated to looking at the distribution of the independent variable, Health Expenditure per Capita ($US), for developing and developed countries individually. For developing countries, it was found that the data have a positively skewed distribution, indicating that most of the values fall between the first quartile ($60.17) and the median ($125.55). The box-whisker plot used to display these values also showed that the distribution has a range of $964.89 and IQR of $228.3. In the test for outliers, only Cuba's value ($986.94) appeared to exceed the calculated upper limit ($631.045), being considered as the only outlier of the 25 randomly chosen developing countries. This was explained by the economic productivity of the country and the high GDP per capita ($8,821.8) for its small population. For developed countries, on the other hand, the box-whisker showed a negatively skewed distribution, with a median of $4994.9, a range of $9 139.26, and an IQR of $2 658.06. With the test, the United States was found to be the only outlier out of the 25 with the value of $10 623.85, which was then explained by expensive pharmaceuticals and medical care. Lastly, to gain a better understanding of the distribution a parallel box-whisker plot was formed displaying all the data on a single scale.

In the Second Step, the same method was used to analyze the distribution of Life Expectancy at Birth values for developing and developed countries via box-whisker plots, five-number summaries, and tests for outliers. For developing countries, it was found that the distribution had a negatively skewed shape with a median of 71 years, range of 20 years, and IQR of 8 years. There were no outliers identified for this distribution. For the developed countries a very strong positively skewed distribution was observed with the median of the same value as the first quartile, 82 years. Additionally, a range of 9 years and an IQR of just 1 year were observed with the US and Saudi Arabia being the outliers. Different statistics were given to explain these cases by doing some research. Lastly, again to better understand the distribution of the data in general, a parallel box-whisker plot was created displaying all 50 values of Life Expectancy at Birth on a single scale.

In the Third Step, I carried out a bivariate data analysis investigating the correlation of two variables for both developing and developed countries. To do this, I used scatter plot diagrams, Pearson's Correlation Coefficient test, and the least-squares regression line equation. For developing countries, the results showed a strong positive correlation, r=0.6599, with a highly significant p-value of 0.000332 and an r^2 value of 0.4355, indicating that 43.55% of the variation in Life Expectancy can be explained by variation in Health Expenditure per Capita. Afterward, using the

TI-84 Plus CE and online tools, a formula with the equation of $y = 0.01649x + 66.82069$ was generated to estimate the life expectancy of any developing country with its given health expenditure per capita value. This formula then was tested by estimating the life expectancy values of four developing countries, whose percentage errors were found to be just below 7% indicating that the formula generated was a valid tool to use. In the second section of this Step, the same method was used to find the correlation of variables for developed countries. It was then found that there is a weak correlation between the two variables, with an r-value of 0.1628, p-value of 0.436843, and determination coefficient (r^2) of 0.0265. The created formula for the least-squares regression line for this was $y = 0.00013x + 81.28047$, which was then tested for estimating the life expectancy values of 4 developed countries with their given health expenditure per capita values.

With all these findings now I am able to answer my initial research question, "To what extent do the Life Expectancy rates of developing and developed countries depend on their respective Health Expenditure per Capita?". The correlations highly vary between developing and developed countries. For developing countries, for every additional $100 expenditure on health, the life expectancy goes up by 1.649 years, whereas this value is 0.013 years for developed countries. which were obtained with their respective regression line equations. This tells me that for developing countries there is a strong correlation between these two variables whereas for developed countries it is really low.

BIBLIOGRAPHY

Dr. Hector Florimon. "Why US Health Care Costs More, but Isn't Better than Other Countries: Study." *ABC News*, 13 Sept. 2019, abcnews.go.com/Health/us-spends-health-care-countries-fare-study/story?id=53710650. Accessed 5 Apr. 2021.

Haahr, Mads. "RANDOM.ORG - List Randomizer." *Random.org*, www.random.org/lists/. Accessed 5 Apr. 2021.

IMathAs. "Boxplot Grapher." *Www.imathas.com*, www.imathas.com/stattools/boxplot.html. Accessed 4 Apr. 2021.

OECD Data. "Health Resources - Health Spending - OECD Data." *Data.oecd.org*, 2019, data.oecd.org/healthres/health-spending.htm. Accessed 4 Apr. 2021.

Roser, Max. "Why Is Life Expectancy in the US Lower than in Other Rich Countries?" *Our World in Data*, 29 Oct. 2020, ourworldindata.org/us-life-expectancy-low. Accessed 6 Apr. 2021.

Stangroom, Jeremy. "Pearson Correlation Coefficient Calculator." *Socscistatistics.com*, www.socscistatistics.com/tests/pearson/Default2.aspx. Accessed 5 Apr. 2021.

---. "Quick Linear Regression Calculator." *Www.socscistatistics.com*, www.socscistatistics.com/tests/regression/. Accessed 5 Apr. 2021.

---. "Quick P Value from Pearson (R) Score Calculator." *Www.socscistatistics.com*, www.socscistatistics.com/pvalues/pearsondistribution.aspx. Accessed 5 Apr. 2021.

The World Bank. "Life Expectancy at Birth, Total (Years) | Data." *Data.worldbank.org*, 2019, data.worldbank.org/indicator/SP.DYN.LE00.IN. Accessed 4 Apr. 2021.

---. "World Bank Country and Lending Groups – World Bank Data Help Desk." *Www.worldbank.org*, 2021, datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups. Accessed 4 Apr. 2021.

Tyrovolas, Stefanos, et al. "The Burden of Disease in Saudi Arabia 1990–2017: Results from the Global Burden of Disease Study 2017." *The Lancet Planetary Health*, vol. 4, no. 5, 1 May 2020, pp. e195–e208. *The Lancet*, www.thelancet.com/journals/lanplh/article/PIIS2542-5196(20)30075-9/fulltext, 10.1016/S2542-5196(20)30075-9. Accessed 6 Apr. 2021.

United Nations. "Secretary-General Hails Cuba for Training Medical 'Miracle Workers', Being on Frontlines of Global Health | Meetings Coverage and Press Releases." *Www.un.org*, 28 Jan. 2014, www.un.org/press/en/2014/sgsm15619.doc.htm. Accessed 6 Apr. 2021.

Warner, Rich. "Is the Cuban Healthcare System Really as Great as People Claim?" *The Conversation*, 30 Nov. 2016, theconversation.com/is-the-cuban-healthcare-system-really-as-great-as-people-claim-69526. Accessed 5 Apr. 2021.

Weather Online. "Climate of the World: Saudi-Arabia | Weatheronline.co.uk." *Weatheronline.co.uk*, www.weatheronline.co.uk/reports/climate/Saudi-Arabia.htm. Accessed 5 Apr. 2021.

WHO. "Life Expectancy at Birth." World Health Organization.

World Bank Data. "Current Health Expenditure per Capita (Current US$) | Data." *Data.worldbank.org*, 2018, data.worldbank.org/indicator/SH.XPD.CHEX.PC.CD. Accessed 4 Apr. 2021.

---. "Glossary | DataBank." *Databank.worldbank.org*, databank.worldbank.org/metadataglossary/health-nutrition-and-population-statistics/series/NY.GNP.PCAP.CD. Accessed 4 Apr. 2021.

APPENDIX

Appendix 1. Pearson's Correlation Coefficient, r, for developing countries from www.socscistatistics.com

Result Details & Calculation

X Values
$\sum = 5123.74$
Mean = 204.95
$\sum(X - M_x)^2 = SS_x = 1060216.338$

Y Values
$\sum = 1755$
Mean = 70.2
$\sum(Y - M_y)^2 = SS_y = 662$

X and Y Combined
$N = 25$
$\sum(X - M_x)(Y - M_y) = 17481.372$

R Calculation
$r = \sum((X - M_y)(Y - M_x)) / \sqrt{((SS_x)(SS_y))}$

$r = 17481.372 / \sqrt{((1060216.338)(662))} = 0.6599$

Meta Numerics (cross-check)
r = 0.6599

Key

X: X Values
Y: Y Values
M_x: Mean of X Values
M_y: Mean of Y Values
$X - M_x$ & $Y - M_y$: Deviation scores
$(X - M_x)^2$ & $(Y - M_y)^2$: Deviation Squared
$(X - M_x)(Y - M_y)$: Product of Deviation Scores

Appendix 2. Least Squares Regression Line for developing countries from www.socscistatistics.com

Calculation Summary

Sum of X = 5123.74
Sum of Y = 1755
Mean X = 204.9496
Mean Y = 70.2
Sum of squares (SS_X) = 1060216.3383
Sum of products (SP) = 17481.372

Regression Equation = $\hat{y} = bX + a$

$b = SP/SS_X$ = 17481.37/1060216.34 = 0.01649

$a = M_Y - bM_X$ = 70.2 - (0.02*204.95) = 66.82069

$\hat{y} = 0.01649X + 66.82069$

Appendix 3. Pearson's Correlation Coefficient, r, for developed countries from www.socscistatistics.com

Result Details & Calculation

X Values
$\sum$ = 124630.95
Mean = 4985.238
$\sum(X - M_x)^2 = SS_x$ = 125284879.003

Y Values
$\sum$ = 2048
Mean = 81.92
$\sum(Y - M_y)^2 = SS_y$ = 77.84

X and Y Combined
N = 25
$\sum(X - M_x)(Y - M_y)$ = 16072.176

R Calculation
$r = \sum((X - M_y)(Y - M_x)) / \sqrt{((SS_x)(SS_y))}$

r = 16072.176 / √((125284879.003)(77.84))
= 0.1628

Meta Numerics (cross-check)
r = 0.1628

Appendix 4. Least Squares Regression Line for developed countries from www.socscistatistics.com

Calculation Summary

Sum of X = 124630.95
Sum of Y = 2048
Mean X = 4985.238
Mean Y = 81.92
Sum of squares (SS_X) = 125284879.003
Sum of products (SP) = 16072.176

Regression Equation = $\hat{y} = bX + a$

$b = SP/SS_X$ = 16072.18/125284879 = 0.00013

$a = M_Y - bM_X$ = 81.92 - (0*4985.24) = 81.28047

$\hat{y} = 0.00013X + 81.28047$