Mathematics Exploration
Applications and Interpretation HL

To predict a model function on Life Expectancy for any Country.

Session: May 2021

Exploration Page Count - 23 Pages

Table of Contents

Introduction

This exploration aims to predict an appropriate model that can ascertain the Life Expectancy for any country. Perusing countless number of articles pertaining to the subject of the higher death rates in India as compared to other nations due to alarming factors, drove me to further explore this topic. In practically all the news articles, the high degree of pollution and sanitation were the primary causes of these deaths. This caused me to consider about whether there were other such determinants that could likewise influence the span of life. Consequently, I looked through various reports and projects on the net, collected and finally collated data for **three independent variables** – **Homicide Rate, Total Expenditure on Health per Capita, and GDP per Capita** from a scope of different variables.

I have utilised the numerical concept of Multivariable Regression in this investigation to anticipate the model equation. The three independent variables had been chosen that could influence the principle **dependent variable**, which is **Life Expectancy**. The model will initially be done manually by hand and the concept of summation, and then will be followed by utilising statistical software to get a more precise form of the desired equation. A significant fact is that, all the information and data has been gathered taking India in consideration. Thus, the model will be predicted on the basis of India's respective statistical data points. The investigation tells us which factor majorly influences life expectancy in a negative way and subsequently should be controlled to bring about a change in the life expectancy of an individual.

Life Expectancy

Life expectancy is the most commonly used measure to portray the health and wellbeing of a population and reflects the general mortality level of a population. Life expectancy measures how long, on average, an individual is expected to live based on current age and gender-specific death rates.[1] This is a measure that can be determined for a specific range of ages, a specific sex and even for various metropolitan and rural areas.

Life expectancy is said to change over time as circumstances can turn out to be better or more terrible. Subsequently, there are several factors that can impact / influence this measure. These range from the most broadly known indicators such as GDP per Capita of a nation, Income of the individual, access to healthcare services, crime rate etc to tiny factors such as the degree of violence against a particular densely populated group, and a measure of individuals suffering due to certain diseases. Hence, only the most significant variables are contemplated for this exploration. The three key chosen variables are **GDP per Capita of India**, which would include several other indicators like Consumer Expenditure, Investment Spending and Net Exports, the **Homicide rate** and the **Total Expenditure on Healthcare per Capita**.

There were many sites providing data regarding the three chosen variables, however, in order to predict a function of life expectancy as accurately as possible, I have used the data from authentic sites such as World Bank and Knoema and have and collated the data of **Homicide Rate**, **Total Expenditure on Healthcare per Capita** and **GDP per Capita** from **2009 to 2017** in a tabular form in the table below.

[1] "Life expectancy." *Authoritative information and statistics to promote better health and wellbeing*, Australian Institute of Health and Welfare, aihw.gov.au/deaths/life-expectancy/.

Year	Homicide rate (per 100,000) - in %[2]	Total Expenditure on Healthcare per Capita - in US$[3]	GDP per Capita - in US$[4]	Life Expectancy - in years[5]
2009	3.8	38	1101.961	66.244
2010	3.8	45	1,357.564	66.693
2011	3.8	49	1,458.104	67.13
2012	3.8	49	1,443.88	67.545
2013	3.6	56	1,449.606	67.931
2014	3.7	57	1,573.881	68.286
2015	3.4	59	1,605.605	68.607
2016	3.2	61	1,732.564	68.897
2017	3.1	69	1,981.651	69.165

Table 1 – Data collected from 2009 to 2017

Relationship between the Dependent and Independent Variables

To make a linear equation with several variables, it is essential to understand the distinct relationships between the dependent variable and the respective independent variables. Thus, by using the data in Table 1 and Microsoft Excel, 3 scatterplots have been constructed.

From the three scatterplot charts, we see an important measure that is different for each of the graphs. This is the **R^2** value. This measure is called the coefficient of determination and is also known as the coefficient of multiple determination for multiple regression. With a value between 0

---

2 "India Homicide Rate, 1990-2019 - Knoema.Com". *Knoema*, 2020, https://knoema.com/atlas/India/Homicide-rate.

3 "India Health Expenditure Per Capita, 1960-2019 - Knoema.Com". *Knoema*, 2020, https://knoema.com/atlas/India/topics/Health/Health-Expenditure/Health-expenditure-per-capita#:~:text=India%20%2D%20Current%20expenditure%20on%20health%20per%20capita&text=In%202017%2C%20health%20expenditure%20per,average%20annual%20rate%20of%208.71%25.

4 "GDP Per Capita (Current US$) - India | Data". *Data.Worldbank.Org*, 2020, https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?end=2017&locations=IN&name_desc=false&start=2009&view=chart.

5 "Life Expectancy At Birth, Total (Years) - India | Data". *Data.Worldbank.Org*, 2020, https://data.worldbank.org/indicator/SP.DYN.LE00.IN?end=2017&locations=IN&start=2009.

to 1, the coefficient of determination is calculated as the square of the correlation coefficient (R) between the sample and predicted data.[6] **R^2** has been used for this investigation because it indicates the degree to which the change in the dependent variable is caused by a change in the independent variable.[7]

Sample calculation for R^2 value

The equation[8] to finding the value is as follows:

$$R^2 = \frac{[N(\sum xy) - (\sum x)(\sum y)]^2}{[N\sum x^2 - (\sum x)^2][N\sum y^2 - (\sum y)^2]}$$

Where, N is the total number of trials, y is the **value of the dependent variable** and x is the value of the respective **independent variables**.

Taking the first case of life expectancy versus homicide rate, the different values were calculated for the above equation and have been tabulated below, In this case, x represents the **Homicide Rate** and y is the **Life Expectancy**.[9]

---

[6] "Coefficient Of Determination Lesson." *Chegg Study*, Chegg Inc., www.chegg.com/homework-help/definitions/coefficient-of-determination-31.

[7] Fernando, Jason. "R-Squared." *Investopedia*, Investopedia, 18 Nov. 2020, www.investopedia.com/terms/r/r-squared.asp.

[8] Andale. "Coefficient of Determination (R Squared): Definition, Calculation." *Statistics How To*, Statistics How To, 19 Apr. 2012, www.statisticshowto.com/what-is-a-coefficient-of-determination.

[9] Similar tables have been created to calculate Coefficient of Determination values of Total Expenditure on Healthcare per Capita and GDP per Capita are made in Appendix 1.

	x	y	xy	x^2	y^2
	3.8	66.244	251.7272	14.44	4388.267536
	3.8	66.693	253.4334	14.44	4447.956249
	3.8	67.13	255.094	14.44	4506.4369
	3.8	67.545	256.671	14.44	4562.327025
	3.6	67.931	244.5516	12.96	4614.620761
	3.7	68.286	252.6582	13.69	4662.977796
	3.4	68.607	233.2638	11.56	4706.920449
	3.2	68.897	220.4704	10.24	4746.796609
	3.1	69.165	214.4115	9.61	4783.797225
$\sum(var)$	32.2	610.498	2182.2811	115.82	41420.10055

Table 2 – Data for R^2 value Calculation

Hence, substituting the values from the above table we get:

$$R^2 = \frac{[(9\times 2182.2811) - (32.2\times 610.498)]^2}{\left[(9\times 115.82) - (32.2)^2\right]\left[(9\times 41420.10055) - (610.498)^2\right]}$$
$$= \frac{306.4495325}{[5.54\times 73.097]}$$
$$= 0.756745$$

Hence, by using the above method, the $\mathbf{R^2}$ values were found to be **0.75675**, **0.95273** and **0.86067** for Figures 1,2 and 3 respectively.

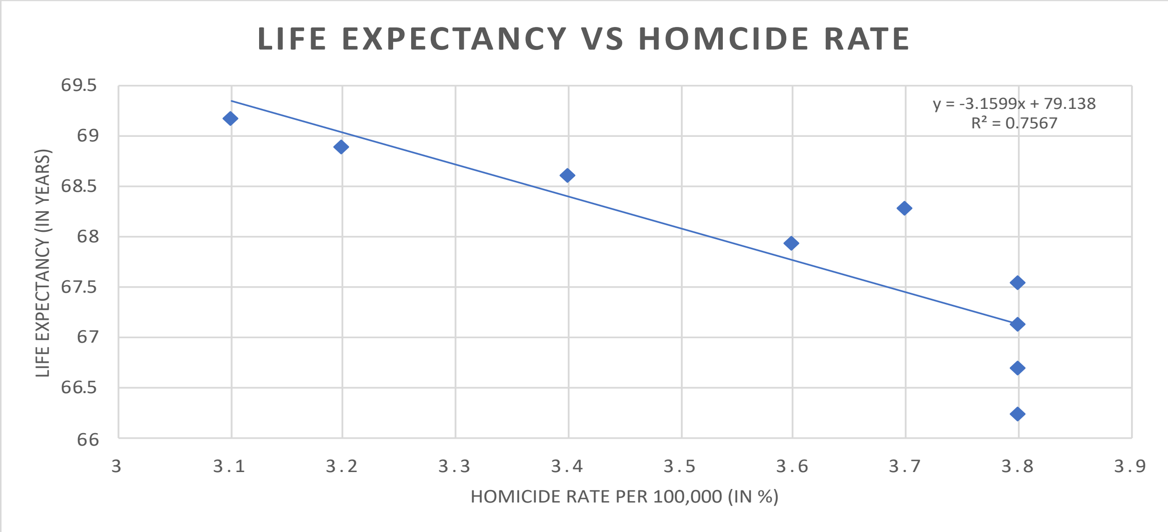

Figure 1 – Scatterplot of Life Expectancy and Homicide Rate

From the first scatterplot chart above, it is evident that there is a linear relation between life expectancy and homicide rate. Furthermore, both these variables have moderate negative correlation, telling us that as the explanatory / independent variable (in this case Homicide Rate) increases, the response / dependent variable (Life Expectancy) reduces. Looking at the **R^2** value of this graph, being **0.75675**, it can be inferred that homicide rate is not an important variable that has a powerful influence on life expectancy.

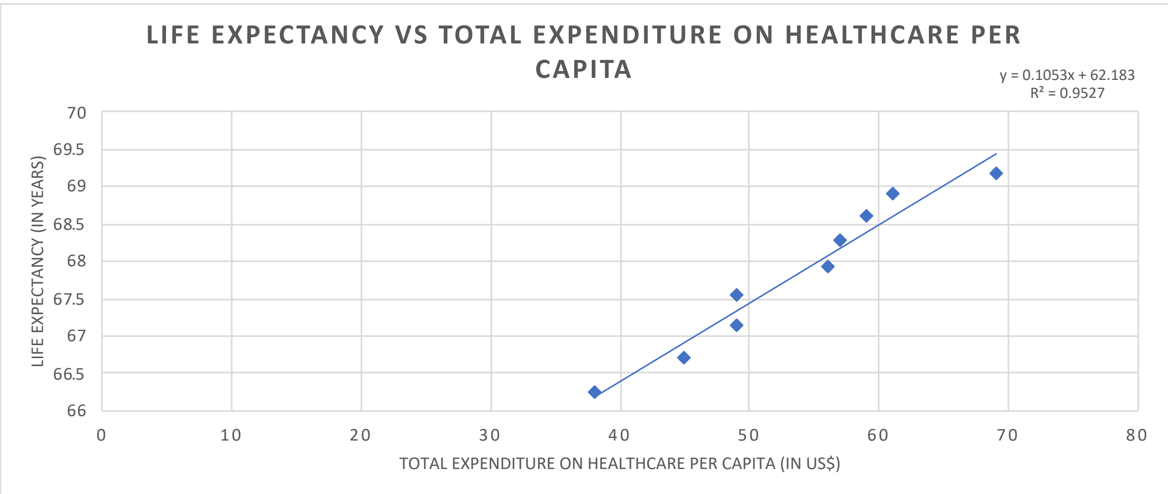

Figure 2 – Scatterplot of Life Expectancy and Expenditure on Healthcare per Capita

From the scatterplot above, just as in Figure 1, it can be seen that life expectancy and total expenditure on healthcare per capita share a strong linear correlation. But in this case they have a strong positive correlation, informing us that as the independent variable (in this case Total Expenditure on Healthcare per Capita) increases, the dependent variable (Life Expectancy) also increases. The $\mathbf{R^2}$ value for this scatter plot is **0.95273**, revealing that the independent variable (Total Expenditure on Healthcare per Capita) is crucial in the calculation of life expectancy.

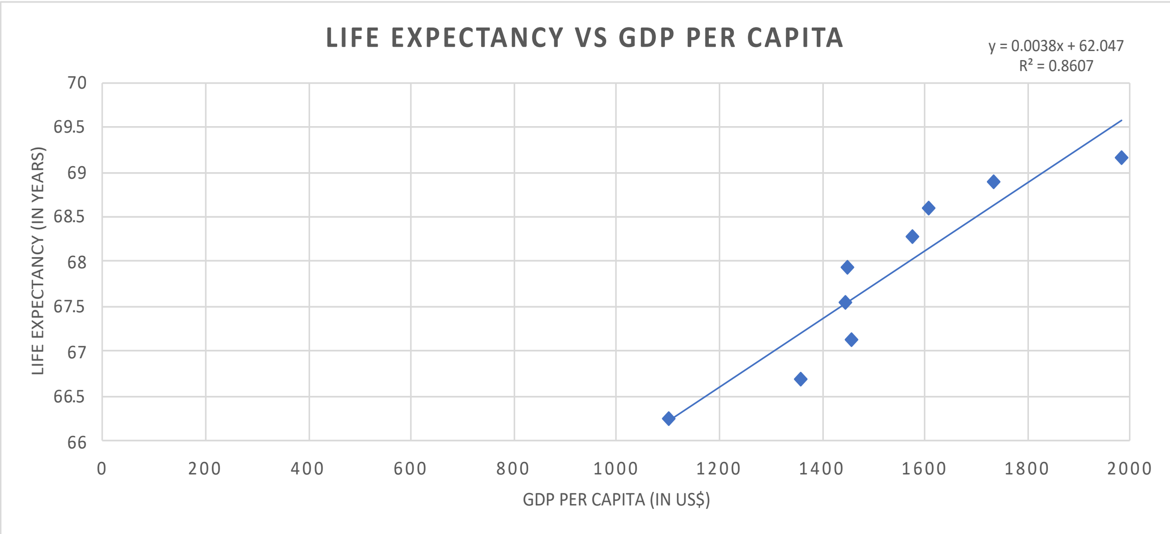

Figure 3 – Scatterplot of Life Expectancy and GDP per Capita

Similarly, the graph in Figure 3 shows a strong linear association between life expectancy and GDP per Capita. Just as in Figure 2, both of these factors have a strong positive correlation telling us that as the GDP per Capita rises, so does the life expectancy. Reflecting on the $\mathbf{R^2}$ value of **0.86067**, we see that there is a strong correlation between the two variables, indicating that the GDP per Capita has a significant impact on the calculation of life expectancy.

Looking at the $\mathbf{R^2}$ values, it can be understood that the Total Expenditure on Healthcare per Capita is one of the most important measures among many for predicting a near perfect model for Life Expectancy. As a result, among the three, homicide rate is the least representative factor that can influence the model.

Multivariable Regression is a tool that helps devise a model to anticipate a value with the help of two or more firmly related quantities. This tool originates from an old concept of ordinary least squares and is otherwise called Multiple Linear Regression. Accordingly, I will be using this statistical process in predicting a relevant model for life expectancy by using the previously mentioned three independent variables. Likewise, in predicting the model, the correlation between the factors is significant which can be viewed as the following equation:

$$Y = b_1 + b_2(H) + b_3(E) + b_4(G)$$

Where Y is the life expectancy and the $b\,i$ is a constant where i = 1,2,3,4. **H** represents the **Homicide rate**, **E** represents the **Total Expenditure on Healthcare per Capita** and **G** represents the **GDP per capita.**

Using the action of Summation in the above equation, we get:

$$\sum Y = \sum b_1 + b_2 \sum H + b_3 \sum E + b_4 \sum G \quad \text{(Equation 1)}$$

By using MS Excel and the data from Table 1, the respective values are substituted in equation 1 forming the second equation:

$$610.498 = b_1 \times 9 + b_2 \times 32.2 + b_3 \times 483 + b_4 \times 13704.816 \quad \text{(Equation 2)}$$

Now in the skeletal equation, **H (Homicide Rate)** is multiplied with each term. Again, using summation on both sides, the following equation is formed:

$$\sum YH = b_1 \times 9 \sum H + b_2 \sum H^2 + b_3 \sum EH + b_4 \sum GH \quad \text{(Equation 3)}$$

$$\rightarrow 2182.2811 = b_1 \times 289.8 + b_2 \times 115.82 + b_3 \times 1710 + b_4 \times 48562.0554$$

Now in the skeletal equation, **E (Total Expenditure on Healthcare per Capita)** is multiplied with each term. Using summation on both sides, the following equation is formed:

$$\sum EH = b_1 \times 9 \sum E + b_2 \sum EH + b_3 \sum E^2 + b_4 \sum GE \qquad \text{(Equation 4)}$$

$$\rightarrow 32836.885 = b_1 \times 4347 + b_2 \times 1710 + b_3 \times 26619 + b_4 \times 753202.285$$

Lastly, **G (GDP per Capita)** is multiplied with each term from the skeletal equation. Using summation on both sides, the following equation is formed:

$$\sum GH = b_1 \times 9 \sum G + b_2 \sum GH + b_3 \sum EG + b_4 \sum G^2 \qquad \text{(Equation 5)}$$

$$\rightarrow 931840.035 = b_1 \times 123343.344 + b_2 \times 48562.0554 + b_3 \times 753202.285 + b_4 \times 21353299.9$$

Thus, we have got 4 equations that can be solved via the concept of *simultaneous linear equations*. From this we can find the respective bi constants where i = 1,2,3,4 with the help of a Casio FX CG-50 GDC (Graphic Display Calculator). The values of the constants are (adjusted to 4 decimal places) shown in Figure 4 below:

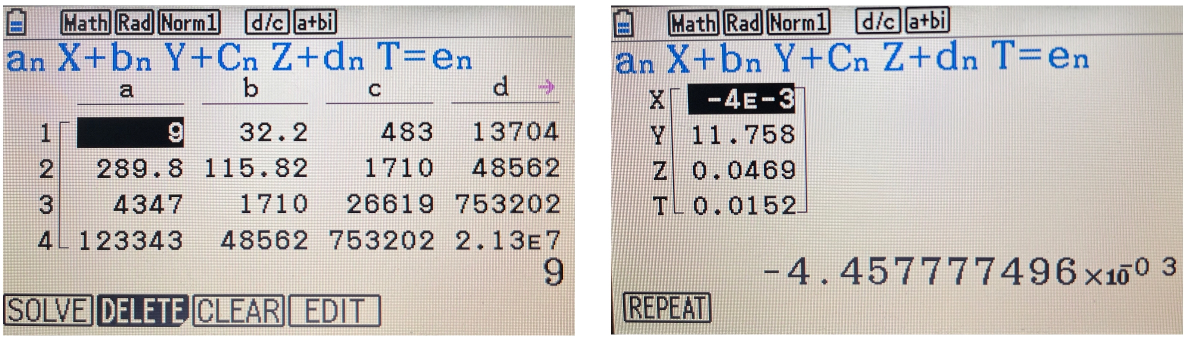

Figure 4 – Screenshots from GDC

Thus, the values of the constants are found to be:

$b_1 = -0.0045$

$b_2 = 11.758$

$b_3 = 0.0469$

$b_4 = 0.0152$

Therefore, by incorporating the above constants and the respective independent variables, which are **Homicide Rate(H)**, **Total Expenditure on Healthcare per Capita(E)** and **GDP per Capita(G)**, the model equation that I have formed is:

$$\text{Life Expectancy} = -0.0045 + 11.758(\text{H}) + 0.0469(\text{E}) + 0.0152(\text{G})$$

Verification

To prove the above equation, I will take the values of all the independent variables from the year **2011** and substitute them into the equation. In this year, the **Homicide Rate** is **3.8%**, the **Expenditure on Healthcare per Capita** is **$49** and **GDP per Capita** is **$1458.104**.

Doing this we get:

$$\text{Life Expectancy} = -0.0045 + 11.758(\text{H}) + 0.0469(\text{E}) + 0.0152(\text{G})$$
$$= -0.0045 + 11.758(3.8) + 0.0469(49) + 0.0152(1458.104)$$
$$= 69.137$$

It is quite evident that there is a difference of +2.007 from the original value of 67.13. This is because life expectancy is dependant on several other factors which have not been taken into consideration. These factors include the level of education acquired, proper nutrition and exercise, effect of diseases, access to technology and many more. Hence, a more precise prediction can be inferred if more variables are taken into consideration. The next section explores a more detailed version of the equation mentioned above.

Error Analysis

To understand the importance of the standard error, the initial equation has to be modified to obtain a more detailed regression equation as follows:

$$Y = b_1 + b_2(H) + b_3(E) + b_4(G) \pm S_{res}$$

In the above equation, S_{res} is the Residual Standard Deviation, which is also known as the Error Term of the Regression Equation. This term makes up for the difference in the values. It is calculated using the following formula[10]:

$$S_{res} = \sqrt{\frac{\sum(Y - Y_{est})^2}{n - 2}}$$

Where, S_{res} is the error, Y are the actual values of Life Expectancy, Y_{est} are the expected values of Life Expectancy and n is the number of observations. Using MS Excel, the Expected values of Life Expectancy are tabulated below.

	Actual (Y)	**Expected** (Y_{est})	$Y - Y_{est}$	$(Y - Y_{est})^2$
	66.244	63.208	3.036	9.217296
	66.693	67.421	-0.728	0.529984
	67.13	69.137	-2.007	4.028049
	67.545	68.921	-1.376	1.893376
	67.931	66.985	0.946	0.894916
	68.286	70.096	-1.81	3.2761
	68.607	67.145	1.462	2.137444
	68.897	66.817	2.08	4.3264
	69.165	69.802	-0.637	0.405769
$\sum(var)$				26.709334

Table 3 – Data for Residual Error Calculation

10 "Regression." *MedCalc*, MedCalc Software bvba, www.medcalc.org/manual/regression.php.

Substituting all the above values in the formula, we get:

$$S_{res} = \sqrt{\frac{26.709334}{9-2}}$$

$$S_{res} = \sqrt{\frac{26.709334}{7}}$$

$$= 1.953$$

The residual error value is calculated to be 1.953, which brings the calculated value of Life Expectancy to be 67.184 years. But the predicted value for Life Expectancy is still larger than the actual by 0.054 years. Hence, it can be said that the independent variables which were not taken into consideration will count for the difference in the values.

Descriptive Statistics of Data

With the help of the software MS Excel, the quantitative dataset is summarised and analysed below. The data is described in the following table:

Homicide rate (per 100,000) - in %	
Mean	3.577777778
Standard Error	0.092462872
Median	3.7
Mode	3.8
Standard Deviation	0.277388616
Sample Variance	0.076944444
Kurtosis	-0.795802667
Skewness	-0.925153782
Range	0.7
Minimum	3.1
Maximum	3.8
Sum	32.2
Count	9

Total Expenditure on Healthcare per Capita - in US$	
Mean	53.66666667
Standard Error	3.113590282
Median	56
Mode	49
Standard Deviation	9.340770846
Sample Variance	87.25
Kurtosis	-0.13921068
Skewness	-0.098336935
Range	31
Minimum	38
Maximum	69
Sum	483
Count	9

GDP per Capita - in US$	
Mean	1522.757333
Standard Error	82.00531793
Median	1458.104
Mode	#N/A
Standard Deviation	246.0159538
Sample Variance	60523.84952
Kurtosis	1.170145077
Skewness	0.28497695
Range	879.69
Minimum	1101.961
Maximum	1981.651
Sum	13704.816
Count	9

Life Expectancy - in years	
Mean	67.83311111
Standard Error	0.335862962
Median	67.931
Mode	#N/A
Standard Deviation	1.007588885
Sample Variance	1.015235361
Kurtosis	-1.154891308
Skewness	-0.273727657
Range	2.921
Minimum	66.244
Maximum	69.165
Sum	610.498
Count	9

Table 4 – Descriptive Statistics of the Data

This tool aids us in better understanding the essential aspects of the datasets by calculating and condensing the different findings in a tabular form. From the years 2009 to 2017 and from the above table, we see that the mean **Homicide Rate** is supposed to be **3.58%** having **variance** of **0.0769**. As for the **GDP per Capita of India**, **$1522.76** is the mean output of each individual and its **variance** is **60523.85** and the mean **Total Expenditure on Healthcare per Capita** is **$53.67** with a **variance** of **87.25**. Lastly, the mean **Life Expectancy** is shown to be **67.83** years and having a **variance** of **1.015**.

Regression Mathematical Model

Multivariable Regression, also known as Multiple Regression, is where linear regression is performed but with more than two variables. As this concept is not part of my syllabus, I had searched for several research projects and online reports that could add to my knowledge for this theory of mathematics.

Microsoft Excel was used to predict the most accurate form of an equation for Life Expectancy. Thus, the regression statistics collected can be seen in table 5 below:

Regression Statistics	
Multiple R	0.978385041
R Square	0.957237288
Adjusted R Square	0.931579661
Standard Error	0.263557864
Observations	9

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	64.10702318	3.543677049	18.09053768	9.48271E-06	54.99771132	73.21633503	54.99771132	73.216335
Total Expenditure on Healthcare per Capita - in US$3	0.117426574	0.039014601	3.009810981	0.029762064	0.01713635	0.217716797	0.01713635	0.2177168
Homicide rate (per 100,000) - in %2	-0.360041108	0.695582322	-0.517611067	0.626807414	-2.14809239	1.428010174	-2.14809239	1.42801017
GDP per Capita - in US$4	-0.000845609	0.001433723	-0.589799774	0.580967979	-0.004531111	0.002839892	-0.00453111	0.00283989

Table 5 – Regression Statistics

The above table tells us very important information about the predicted model. It can be seen that the I**ntercept value** of the **coefficient** is **64.107**, which represents the minimum value of **Life Expectancy of India** using the data values from **2009-2017**. It can be seen that the **coefficient of determination** is very close to 1, being at **0.957**; it tells us that Life Expectancy can be accurately predicted. For each of the 3 independent variables the values of the coefficient have been calculated. From the 3 variables, the **Total Expenditure on Healthcare per Capita** has a positive correlation telling us that a **1% increase in Total Expenditure on Healthcare per Capita** would result in an **increase in Life Expectancy by 0.117%**.

However, the **Homicide Rate** and **GDP per Capita** have a negative sign in the value of their coefficients, which means that an **increase** in **Homicide Rate** and **GDP per Capita by 1%** would result in a **decrease in Life Expectancy by 0.3604% and 0.0008% respectively**. The P-value and T-stat are tools that are used in statistical hypothesis testing. A P-value measures the strength of evidence in support of a null hypothesis.[11] Or it is simply a test that helps us in understanding the importance of the results. A P-value below the significance level testifies to the acceptance of the hypothesis and vice versa. The significance level is considered to be arbitrary but is conventionally chosen to be 5% which I have done. Hence, the P-value of the Expenditures on Healthcare is statistically significant as it is less than 0.05 but the other two independent variables have a significantly larger value, thereby having weak statistical significance. A T-test is used to decide whether or not to reject the null hypothesis.[12] The T-stat value works alongside the P-value. The larger absolute value of the T-stat tells us that there is proof of the variable being statistically

[11] "Statistics and Probability Dictionary: P-Value." *Stat Trek - Teach yourself statistics*, StatTrek.com, stattrek.com/statistics/dictionary.aspx?definition=P-value.

[12] "T Statistic: Definition, Types and Comparison to Z Score." *Statistics How To*, Statistics How To, www.statisticshowto.com/t-statistic/.

significant and vice versa. From Table 5, we see that from the 3 independent variables, **Expenditure on Healthcare** has the **largest T-stat value** and hence is significant to the model.

ANOVA					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	7.774569151	2.59152305	37.30809887	0.000758494
Residual	5	0.347313737	0.069462747		
Total	8	8.121882889			

Table 6 – ANOVA Statistics

The Analysis of Variance (ANOVA) is a statistical tool used to check the acceptability of a hypothesis. It is an extension of the multivariable regression statistics and helps in understanding the statistical significance. From all the calculated information in the table, there are 2 outputs that are most important for my regression equation. These are F and P-value. Both these values tell us the significance level of the regression. The F Test is very similar to the T-test explained above. The only difference between both these tests is that the F-test reveals the significance of a group of variables whereas the T-test tells us the significance of each variable. In my case, the regression model has a F-value of 37.3081 and a P-value of 0.000758, which is highly statistically significant as the value of F is quite large and the P-value is <0.001. Hence, this tells us that my predicted regression model is accurate for the given data.

From Table 5, a precise version of the equation can be formed, which is as follows:

$$\text{Life Expectancy} = 64.107 - 0.36004(H) - 0.0008(G) + 0.117(E)$$

Verification

Just like before, I will take the respected values of the variables from **2011** and substitute them into the equation mentioned above to test the new model. In this year, the **Homicide Rate** is **3.8%**, the **Total Expenditure on Healthcare per Capita** is **$49** and **GDP per Capita** is **$1458.104**.

Thus, we get:

$$\text{Life Expectancy} = 64.107 - 0.36004(3.8) - 0.0008(1458.104) + 0.117(49)$$
$$\rightarrow \text{Life Expectancy} = 67.305$$

The expected value of Life Expectancy, through the use of software, has produced a closer estimate value that from the hand produced equation. This prediction has a much smaller difference of 0.175 from the actual value of 67.13. This model bears values with extremely low inaccuracies, thus it can be said that this is an ideal model that fits with the given data.

Acceptance of this model for other Countries

Since, this model has been predicted to calculate the Life Expectancy for India, it is ideal to check whether the same model can be used for other countries to decide the same. For this test, I will take another developing nation and use the information and data from a specific year from 2009 to 2017 to check whether it is a best fit. Taking **Indonesia** into consideration, the data is tabulated below.

Year	Homicide rate (per 100,000) - in %[12]	Total Expenditure on Healthcare per Capita - in US$[13]	GDP per Capita - in US$[14]	Life Expectancy - in years[15]
2013	0.6	107.20	3623.912	70.179

Table 7 – Data for Indonesia for the year 2013

Substituting the above values in the respective model, we get:

$$\text{Life Expectancy} = 64.107 - 0.36004(0.6) - 0.0008(3623.912) + 0.117(107.2)$$
$$\rightarrow \text{Life Expectancy} = 73.53$$

We see that there we have obtained an estimation of 73.53. There is a difference of +3.351 from the actual value of 70.179 years. From this, it can be inferred that each nation doesn't rely upon mutual factors rather region specific factors also need to be taken into consideration to get an accurate estimate of the value of Life Expectancy. In this case, a nation like Indonesia would have factors like several diseases, sustenance, number of doctors and specialists and the level of water pollution/ contamination would enormously affect Life Expectancy. Therefore, it can be said that this model

[13] "Indonesia Homicide Rate, 1990-2019." *Knoema*, knoema.com/atlas/Indonesia/Homicide-rate.

[14] "Indonesia Health Expenditure per Capita, 1960-2020." *Knoema*, knoema.com/atlas/Indonesia/topics/Health/Health-Expenditure/Health-expenditure-per-capita.

[15] "GDP per Capita (Current US$) - Indonesia." *Data*, data.worldbank.org/indicator/NY.GDP.PCAP.CD?end=2019&locations=ID&start=2013.

[16] "Life Expectancy at Birth, Total (Years) - Indonesia." *Data*, data.worldbank.org/indicator/SP.DYN.LE00.IN?end=2018&locations=ID&start=2011.

can only be used till a certain extent. But some modifications would have to be made to fit the given information and factors. This will be the case for every nation and each one of them will have their own distinctive model.

Conclusion and Limitations

To conclude, with the help of Multivariable Regression I have been able to achieve my aim of predicting an appropriate model that can ascertain the Life Expectancy for any country. However, one of the noteworthy drawbacks of my exploration is that I have just taken three factors into consideration, from which, one of them doesn't have a high $\mathbf{R^2}$ value. Most relevant variables like transferable infections, doctors and other independent variables should have been incorporated to get a more precise model. I didn't have access to a wide range of data, as it wasn't accessible on the internet which would have made my model more extensive. There was difficulty in choosing the underlying three variables due to this fact. The predicted model can't be used to predict Life Expectancy for developed nations like USA, Europe and Australia. This was on the grounds of enormous discrepancies between actual and predicted values, as GDP per Capita and Total Expenditure on Healthcare were pretty high. This could be possible because of the fact that I had predicted my model based on the statistical information of India and consequently, was not entirely a representative for a developed nation. Additionally, my prediction was based on the average values of the three variables. In a specific region of the country, the Homicide Rate could be very high whereas another part could have a very small value. Hence, my model can't be used to calculate region specific life expectancy. These drawbacks could potentially affect my prediction.

I believe that the predicted model is fairly accurate, however not perfect. There is an opportunity in the improvement of the model by incorporating few more independent variables and replacing the ineffective variables with suitable ones. The model equation could likewise be modified in a manner where it could be acceptable in calculating Life Expectancy for specific regions. Lastly, a separate model could likewise be constructed that would work with more developed nations. But, every model would have a few uncertainties and errors that would make it acceptable to a certain degree.

Bibliography

"Life expectancy." Authoritative information and statistics to promote better health and wellbeing, Australian Institute of Health and Welfare, aihw.gov.au/deaths/life-expectancy/.

"India Homicide Rate, 1990-2019 - Knoema.Com". Knoema, 2020, https://knoema.com/atlas/India/Homicide-rate.

"India Health Expenditure Per Capita, 1960-2019 - Knoema.Com". Knoema, 2020, https://knoema.com/atlas/India/topics/Health/Health-Expenditure/Health-expenditure-per-capita#:~:text=India%20%2D%20Current%20expenditure%20on%20health%20per%20capita&text=In%202017%2C%20health%20expenditure%20per,average%20annual%20rate%20of%208.71%25.

"GDP Per Capita (Current US$) - India | Data". Data.Worldbank.Org, 2020, https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?end=2017&locations=IN&name_desc=false&start=2009&view=chart.

"Life Expectancy At Birth, Total (Years) - India | Data". Data.Worldbank.Org, 2020, https://data.worldbank.org/indicator/SP.DYN.LE00.IN?end=2017&locations=IN&start=2009.

"Coefficient Of Determination Lesson." Chegg Study, Chegg Inc., www.chegg.com/homework-help/definitions/coefficient-of-determination-31.

Fernando, Jason. "R-Squared." Investopedia, Investopedia, 18 Nov. 2020, www.investopedia.com/terms/r/r-squared.asp.

Andale. "Coefficient of Determination (R Squared): Definition, Calculation." Statistics How To, Statistics How To, 19 Apr. 2012, www.statisticshowto.com/what-is-a-coefficient-of-determination.

"Regression." MedCalc, MedCalc Software bvba, www.medcalc.org/manual/regression.php.

"Statistics and Probability Dictionary: P-Value." Stat Trek - Teach yourself statistics, StatTrek.com, stattrek.com/statistics/dictionary.aspx?definition=P-value.

"T Statistic: Definition, Types and Comparison to Z Score." Statistics How To, Statistics How To, www.statisticshowto.com/t-statistic/.

"Indonesia Homicide Rate, 1990-2019." Knoema, knoema.com/atlas/Indonesia/Homicide-rate.

"Indonesia Health Expenditure per Capita, 1960-2020." Knoema, knoema.com/atlas/Indonesia/topics/Health/Health-Expenditure/Health-expenditure-per-capita.

"GDP per Capita (Current US$) - Indonesia." Data, data.worldbank.org/indicator/NY.GDP.PCAP.CD?end=2019&locations=ID&start=2013.

"Life Expectancy at Birth, Total (Years) - Indonesia." Data, data.worldbank.org/indicator/SP.DYN.LE00.IN?end=2018&locations=ID&start=2011.

Appendices

Appendix 1 - Data for R^2 calculation

Taking the case of **Life Expectancy** versus **Total Expenditure on Healthcare per Capita**, the different values were calculated for the above equation and have been tabulated below, In this case, *x* represents the **Total Expenditure on Healthcare per Capita** and *y* is the **Life Expectancy**.

	x	y	xy	x^2	y^2
	38	66.244	2517.272	1444	4388.267536
	45	66.693	3001.185	2025	4447.956249
	49	67.13	3289.37	2401	4506.4369
	49	67.545	3309.705	2401	4562.327025
	56	67.931	3804.136	3136	4614.620761
	57	68.286	3892.302	3249	4662.977796
	59	68.607	4047.813	3481	4706.920449
	61	68.897	4202.717	3721	4746.796609
	69	69.165	4772.385	4761	4783.797225
$\sum(var)$	483	610.498	32836.885	26619	41420.10055

Taking the case of **Life Expectancy** versus **GDP per Capita**, the different values were calculated for the above equation and have been tabulated below, In this case, x represents the **GDP per Capita** and y is the **Life Expectancy**.

	x	y	xy	x^2	y^2
	1101.961	66.244	72998.304484	1214318.045521	4388.267536
	1,357.56	66.693	90539.74908	18,42,969.1536	4447.956249
	1,458.10	67.13	97882.253	21,26,055.61	4506.4369
	1,443.88	67.545	97526.8746	20,84,789.4544	4562.327025
	1,449.61	67.931	98473.45691	21,01,369.1521	4614.620761
	1,573.88	68.286	107473.96968	24,77,098.2544	4662.977796
	1,605.61	68.607	110156.08527	25,77,983.4721	4706.920449
	1,732.56	68.897	119368.18632	30,01,764.1536	4746.796609
	1,981.65	69.165	137060.82225	39,26,936.7225	4783.797225
$\sum(var)$	13704.816	610.498	931479.701594	21353284.018221	41420.10055