Factors Explaining Average Life Expectancy:
An Examination Across Nations

By: Akansha Maity, Emelie Rhenman, and Elijah Sanders
Abstract

This paper analyzes the health across several countries worldwide. Data was collected from the World Health Organization as well as the World Bank. The data sets collected contain variables for nation population, GNI per capita (PPP), poverty headcount ratio at $1.00, life expectancy at birth for males and females as well as the averages between the two, the expenditure on health per capita, the completion rate of secondary education, physicians per 1000 individuals as well as the number of hospital beds per 1000, and the adequacy of social protection (Social Security). Regressions on between these variables show whether or not the variables are correlated as well as what the degree of correlation. This regression will then give insight as to how strongly health is affected by the world’s varying societal factors. The motivation for seeking this information is that we are interested in how different elements affect the health of a population.

I. Introduction

As the current era of relative peace presses on, quite a number of people have become increasingly interested in maintaining it for the entire global community. Thus came the inception of the United Nations’ Sustainable Development Goals (SDGs) in an attempt to secure this future for the world (or at least its member nations). One of such goals is that of the third one, good health and well-being, the pursuit of seeking relatively healthy lives and wellbeing for every person at any age. However in order to ensure these things, one must first understand what factors may affect them and to what degree.

Although a great many factors can be said to affect the health and wellbeing of a population, it is only realistic to cover a comparatively small number of such factors for the sake of statistical analysis. Thus the question arises: what exactly are the factors that have such an effect on a population? In order to account for this, variables must be chosen and tested for correlation against other variables that will be used as a relative measure of health and/or wellbeing. The variables we will be testing for the sake of finding this information will be: Gross National Income per capita as Purchasing Power Parity, because there is the belief that the number of goods an individual can buy will undoubtedly affect their health and wellbeing; the poverty headcount ratio at $1.00, due to the understanding that poverty can be detrimental to a person’s well-being; life expectancy, because this value serves as a good measure of how healthy the members of a population are at a given time; the expenditure on health per capita, because it is important to know how a nation’s government is making an attempt to remedy its citizens’ health in this test; completion rate of secondary education, because schools are a place where one can learn about health; the number of physicians per 1000 individuals as well as the number of hospital beds per 1000 individuals, because it takes into account the proportion of the population can access medical services at any given
time; and the adequacy of social protection (Social Security), because it is a good measure of how much the government is putting into supporting its citizens with no or inadequate income. If one tests these variables for correlation with a multiple linear regression one may believe that one will come closer to understanding what factors affect the health and wellbeing of a population.

II. Literature Review

Prior to studying the relationship between average lifespan and poverty, it was crucial to examine other studies that dealt with a similar topic. Several papers acknowledged the relationship between income inequality and population health, though not exactly what this paper will examine. Three papers were specifically chosen for this literature review to show that the project topic is significant. Pickett and Wilkinson (2005) examined other research papers’ finds on the relationship, Lillard, Burkhauser, Hahn, and Wilkins (2014) looked at a self-reported health survey and income inequality, and Wilkinson and Pickett (2014) looked at specific health issues and income inequality.

Pickett and Wilkinson (2005) decided to examine 155 papers that had conducted research on the relationship between income inequality and population health and suggested why the results might be “wholly supportive,” “unsupportive,” and “partially supportive” of the claim that these two variables were related. “Wholly supportive” meant that the relationship between the two variables had only positive statistically significant associations. “Unsupportive” implied that there were no statistically significant positive associations. “Partially supportive” signified that only some of the relationships had statistically significant positive associations. 70% of the studies implied that when there was larger income inequality, the health of the population suffered from poorer health. The paper found that it was important to sample a large area to show the true nature of income inequality. For example, studies that looked at large subnational regions were not as likely to prove the relationship between income inequality in health as international studies or studies examining sub-national regions. Another issue in a few of the studies was identifying the proper control variables. For example, the authors acknowledge that as countries are wealthier per capita, the relationship between life expectancy and GNI per capita becomes less prevalent. Once two issues were identified, Wilkinson and Pickett reviewed all of the papers and found that only 8% of them were unsupportive of the claim that health and income inequality were related. Therefore, the variables of health and income inequality ought to be associated.

Lillard, Burkhauser, Hahn, and Wilkins (2014) investigated the relationship between a US-born adult’s self-reported health and income inequality. The dependent variable was in a range from 1-5 (1 being “poor” and 5 being “excellent”). The independent variable was the share held by the top 1% from the age of 0-4 and also whether or not the child was considered as poor growing up. The main find of this
research paper was that if individuals suffered from income inequality early in their lives, they were more likely to have worse health and this association is statistically significant for both genders. For example, if a male had grown up in a high income inequality society, they would be more likely to have worse health. However, there are some issues with this paper that the researchers acknowledge. Since the income inequality measure only changes over time and does not differ across groups that live in different regions of the US, there may be omitted variable bias. Furthermore, the paper uses a linear model between inequality and health, when the true model may in fact be nonlinear. Though the paper does not suggest why health and income inequality may be associated, it does encourage future studies to examine the mechanism of the relationship.

Wilkinson and Pickett (2014) later on studied and examined new ways of seeing the relationship between health and income inequality. For example their earlier paper looked at the relationship between an index of health and social problems and income inequality in wealthy countries. The index used data on life expectancy, mental illness, obesity, infant mortality, teenage births, etc. This study showed that there was a clear positive correlation between the two variables. Interestingly enough, suicides seemed more prevalent in low income inequality countries, whereas depression was common in high income inequality nations. Generally though, more equal societies were deemed to have a healthier population, possibly because inequality can have a significant physiological effect. Furthermore the authors studied specific health issues’ effect of country-level median household income and state-level income inequality. For example, infant mortality and all cause working age were positively correlated with income inequality and negatively correlated with median income. One major reason for health being strongly associated with income inequality is that income inequality creates more division among social classes. For example people who suffer from income inequality may not have a strong enough voice to receive adequate health services.

All of these studies above point to the fact that there should be a correlation between health and income inequality. Unlike some of the other studies, this paper will use average life span after birth as our dependent variable and use the poverty headcount ratio at $1.00 per day (% of population) as our measure of income inequality for the independent variable. A different combination of variables may produce unique results in comparison to other papers. In comparison to Lillard et al.’s paper, this analysis may produce different results since average life span is an easier variable to measure. Lillard et al.’s paper relied on a Likert scale to measure health, which is an incredibly subjective value. Additionally, some people may have chosen to report themselves as being healthy, when in fact they aren’t reporting their true health state. Furthermore, like Wilkinson and Pickett’s 2004 paper encourages, this paper uses international data to get a true relationship between socioeconomic groups and health. Though this paper
does not dig as deep into specific health issues as Wilkinson and Pickett’s 2014 paper, average lifespan
should be a good enough measure of health since it can capture a general idea of the many illnesses that
can contribute to a shorter lifespan. Additionally, this paper considers income similar to Wilkinson and
Pickett’s 2014 paper to see if there is a relationship between average lifespan and a nation’s income. A
variable that was not discussed in these papers was a nation’s expenditure on health, which should
provide a direct influence on average lifespan.

III. Data

There were two stages in which the project was completed. The first stage of the project had only
a select few independent variables under examination: poverty headcount ratio, GNI per capita,
expenditure on health per capita, and lower secondary completion rates. In the second stage of the project,
one used additional variables to try to predict average life expectancy. Below is a table of all of the
variables used in the stages of the project, including the abbreviations and the sources of the datasets.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Variable Type</th>
<th>Abbreviations</th>
<th>Year</th>
<th>Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. Life Expectancy</td>
<td>Dependent</td>
<td>Life Exp.</td>
<td>2015</td>
<td>WHO</td>
</tr>
<tr>
<td>Poverty Headcount Ratio at $1.00</td>
<td>Independent</td>
<td>Poverty</td>
<td>2007-2011</td>
<td>World Bank</td>
</tr>
<tr>
<td>GNI per capita (PPP)</td>
<td>Independent</td>
<td>Income</td>
<td>2013</td>
<td>WHO</td>
</tr>
<tr>
<td>Expenditure on Health per Capita</td>
<td>Independent</td>
<td>Expenditure</td>
<td>2014</td>
<td>WHO</td>
</tr>
<tr>
<td>Lower Secondary completion rates (%)</td>
<td>Independent</td>
<td>Education</td>
<td>2007-2011</td>
<td>World Bank</td>
</tr>
<tr>
<td>Physicians per 1000</td>
<td>Independent</td>
<td>Doctors</td>
<td>2007-2011</td>
<td>World Bank</td>
</tr>
<tr>
<td>Hospital beds per 1000</td>
<td>Independent</td>
<td>Beds</td>
<td>2007-2011</td>
<td>World Bank</td>
</tr>
<tr>
<td>Adequacy of Social Protection</td>
<td>Independent</td>
<td>SS(social security)</td>
<td>2007-2011</td>
<td>World Bank</td>
</tr>
<tr>
<td>Gini Index</td>
<td>Independent</td>
<td>Ginidum</td>
<td>2007-2011</td>
<td>World Bank</td>
</tr>
</tbody>
</table>
In this paper, one will examine the relationship between average life span and the poverty headcount ratio at $1.00 for the simple regression model. The average life span used data from 2015 WHO on the average life expectancy at birth for both males and females. These datasets were taken and averaged, based on the assumption that 50% of the population is male and 50% of the population is female. This dependent variable will serve as a measure of health of the population. Though it may not be a perfect representative of health, it serves as a good basis. How long a person lives is easy to measure across countries and should be a good indicator of the health of a population. The headcount ratio is the percentage of population living on less than $1.00 a day at 2011 international prices, with data provided by the World Bank from 2007-2011. This variable serves as the independent variable for the simple regression model. This data is also easily accessible and serves as a good indicator of income inequality.

For the multiple regression model, there were more independent variables added to the model. One used the expenditure of health per capita, which would show how much a country spends on its inhabitants. This 2014 data was provided by the WHO. This data exemplifies how much of a country’s resources are used on providing health services to its population. In theory, a country that provides more health services to its population should have a healthier population. It is important to acknowledge, however, that the money may be spent poorly and could be allocated better. This task is difficult to measure though. Furthermore, one looked at GNI per capita (PPP), which was in dollars that was adjusted with 2013 exchange rates. This 2013 data was provided by the WHO. The independent variable was mentioned in Wilson and Pickett’s 2014 paper and was correlated to specific health issues, therefore it could be interesting to examine income as an independent variable and average lifespan as the dependent variable. Additionally, one added the completion rate of lower secondary education (%), with 2007-2011 data provided by World Bank.

The data for the project was compiled manually from the World Health Organization and the World Bank website. Only the poverty headcount ratio was sourced from the World Bank Development Indicators databank, and with a year range from 2007 to 2011 since not every country had the latest data. The observation consists of 192 countries who are the member states of the WHO.

In Table 1.1 are the summary statistics for the different variables used in the analysis. It is important to note that some data was excluded since in one dataset (for example in poverty head count)
there might not have been reported data. Therefore these countries were removed from the analysis, and one can see that the number of observations vary for each variable. Life expectancy showed a medium level of standard deviation, but poverty, income, expenditure, and education had extremely high standard deviations relative to their means. Furthermore, the range of values for poverty, income, expenditure, and education were extremely high, contributing to the high standard deviation. Additional notes have been provided for countries that had the maximum and minimum values.

Table 1.1: Summary Statistics of Variables to be Used

<table>
<thead>
<tr>
<th>Variable</th>
<th>Observations</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Life Exp</strong></td>
<td>183</td>
<td>71.10</td>
<td>7.90</td>
<td>50</td>
<td>83.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Sierra Leone)</td>
<td>(Japan)</td>
</tr>
<tr>
<td><strong>Poverty</strong></td>
<td>129</td>
<td>15.42</td>
<td>21.48</td>
<td>0.1</td>
<td>78.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Belarus, Cyprus, Czechia, Finland, France, Hungary, Iceland, Jordan, Kazakhstan, Luxembourg, Netherlands, Norway, Russia, Serbia, Switzerland, Thailand, Ukraine)</td>
<td>(Madagascar)</td>
</tr>
<tr>
<td><strong>Income</strong></td>
<td>183</td>
<td>2959.86</td>
<td>12005.06</td>
<td>1</td>
<td>88170</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Benin, Burkina Faso, Comoros, Eritrea, Ethiopia, Gambia, Guinea, Haiti, Madagascar, Mali, Mozambique, Rwanda, Sierra Leone, Solomon Islands, Togo, Uganda, Tanzania, Zimbabwe)</td>
<td>(Kuwait)</td>
</tr>
<tr>
<td><strong>Expenditure</strong></td>
<td>192</td>
<td>1303.93</td>
<td>1608.23</td>
<td>25</td>
<td>9403</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Central African Republic)</td>
<td>(United States of America)</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td>159</td>
<td>73.76</td>
<td>27.46</td>
<td>9.3</td>
<td>133.8</td>
</tr>
<tr>
<td><strong>Doctors</strong></td>
<td>157</td>
<td>1.78</td>
<td>1.50</td>
<td>0.1</td>
<td>6.9</td>
</tr>
</tbody>
</table>
The graph below is the two way scatter plot diagram where the y variable is average life expectancy at birth and the x variable is poverty headcount ratio at $1.00 a day. As one can see, there is a negative correlation and the data fits the line of best fit pretty well. However, several points on the far right will skew the regression. Furthermore, it can be noted that several countries have a 0 value from the poverty headcount ratio. In retrospect, the countries with this 0 value should have been removed from the analysis, but for simplification purposes, these values shall stay in our sample. In variables such as income and expenditure, the points create a logarithmic-looking line. In the first stage of the project, the regular, non-logarithmic form will be used. In the extensions portion of the paper, one will examine a different functional form, causing some of the variables to be better linearized.

Figure 1.1: Relationship between Average Life Expectancy and Poverty Headcount Ratio

<table>
<thead>
<tr>
<th></th>
<th>Beds</th>
<th>SS(social security)</th>
<th>Gini_dum</th>
</tr>
</thead>
<tbody>
<tr>
<td>167</td>
<td>3.18</td>
<td>23.26</td>
<td>131</td>
</tr>
<tr>
<td>3.18</td>
<td>2.55</td>
<td>14.67</td>
<td>1.44</td>
</tr>
<tr>
<td>2.55</td>
<td>0.1</td>
<td>0.5</td>
<td>0.54</td>
</tr>
<tr>
<td>0.1</td>
<td>16.5</td>
<td>60.3</td>
<td>1</td>
</tr>
<tr>
<td>16.5</td>
<td></td>
<td></td>
<td>3</td>
</tr>
</tbody>
</table>

One needs to examine whether or not the Gauss Markov assumptions hold in this analysis. First of all, the parameters in the future regression will be linear. Next, the second assumption states that the data was randomly sampled. This assumption does not stand true for this dataset. To begin with, we only
have 194 observations and taking a sample from this set is counterintuitive. Therefore, the second assumption stands void. The third assumption states that there is no perfect linear correlation and the total sum of squares of the explanatory variables are greater than 0. This can be proven since the values for the explanatory variables are not the same. The fourth assumption states that there is zero conditional mean, which is incredibly difficult to prove. The error term should be 0 in theory, but this may not be the case. Therefore this assumption cannot be proven as true. The Gauss Markov assumptions are furthermore justified by the correlation matrix if the variables used are not extremely correlated. This correlation matrix can be seen in Table 1.2. It can be noted that some correlations are rather high, which implies that there is an issue with multicollinearity and the assumptions are violated. However since these values are not perfectly 1, then they do not violate the Gauss Markov assumptions.

Table 1.2:

<table>
<thead>
<tr>
<th></th>
<th>Life Exp</th>
<th>Poverty</th>
<th>Income</th>
<th>Expenditure</th>
<th>Education</th>
<th>Doctors</th>
<th>Beds</th>
<th>SS</th>
<th>Gini</th>
</tr>
</thead>
<tbody>
<tr>
<td>LifeExp</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poverty</td>
<td>-0.809 0</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Income</td>
<td>0.6503</td>
<td>-0.6103</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expenditure</td>
<td>0.6302</td>
<td>-0.5576</td>
<td>0.9186</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>0.7120</td>
<td>-0.7733</td>
<td>0.5969</td>
<td>0.5124</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doctors</td>
<td>0.5184</td>
<td>-0.6147</td>
<td>0.5877</td>
<td>0.5930</td>
<td>0.6767</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beds</td>
<td>0.2550</td>
<td>-0.4004</td>
<td>0.4564</td>
<td>0.3924</td>
<td>0.6375</td>
<td>0.7091</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SS</td>
<td>0.1120</td>
<td>-0.0757</td>
<td>0.2943</td>
<td>0.2822</td>
<td>0.1727</td>
<td>0.3848</td>
<td>0.3817</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Gini_dum</td>
<td>-0.151 1</td>
<td>0.3297</td>
<td>-0.0476</td>
<td>0.0207</td>
<td>-0.2909</td>
<td>-0.3777</td>
<td>-0.5338</td>
<td>-0.1006</td>
<td>1</td>
</tr>
</tbody>
</table>
III. Results

Below is the first stage of the project, in which four independent variables were examined, without making any transformations to the independent variables.

Table 2.1: Summary table

<table>
<thead>
<tr>
<th>LifeExp</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poverty</td>
<td>-0.314***</td>
<td>-0.315***</td>
<td>-0.294***</td>
<td>-0.186***</td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.021)</td>
<td>(0.020)</td>
<td>(0.032)</td>
</tr>
<tr>
<td>Income</td>
<td>-</td>
<td>0.000</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Expenditure</td>
<td>-</td>
<td>-</td>
<td>0.658***</td>
<td>0.001***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.166)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>Education</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.083</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.026)</td>
</tr>
<tr>
<td>Intercept</td>
<td>76.133***</td>
<td>76.141***</td>
<td>71.182***</td>
<td>66.515***</td>
</tr>
<tr>
<td></td>
<td>(0.545)</td>
<td>(0.561)</td>
<td>(1.351)</td>
<td>(2.245)</td>
</tr>
<tr>
<td>Observations</td>
<td>128</td>
<td>127</td>
<td>128</td>
<td>113</td>
</tr>
<tr>
<td>R-Squared</td>
<td>0.6494</td>
<td>0.6485</td>
<td>0.6886</td>
<td>0.7355</td>
</tr>
</tbody>
</table>

The following the simple regression model with the average life expectancy in years as the dependent variable and the poverty headcount ratio under $1.00.

\[
\text{LifeExp} = B_0 + B_1 \text{Poverty} + u
\]

As one can see in Table 2.1, the coefficient $B_1$ is negative and small. This observation implies that as the poverty headcount ratio increases, the average life expectancy should decrease according to the model. More factors may explain this better.

The following model was used by adding income as a variable:
The coefficients have changed now. GNI per capita is very small and negative, showing that it may have little impact on the average life span. However, it also suggests that as GNI per capita increases, the average life expectancy should decrease. A third regression analysis included the expenditure on health per capita. The following model was used:

\[ \text{LifeExp} = B_0 + B_1 \text{Poverty} + B_2 \text{Income} + u \]

In this model, the coefficients have changed. Poverty’s coefficient is slightly less negative and expenditure has a large coefficient in comparison to poverty. Furthermore, the R-squared value has grown as well.

The final regression analysis was performed below:

\[ \text{LifeExp} = B_0 + B_1 \text{Poverty} + B_3 \text{Expenditure} + B_4 \text{Education} + u \]

This model has smaller coefficients, with one of them not being statistically significant at less than 10%. Therefore, the third model is the best model. Thus, this one will be analyzed. The three stars signify that the null hypothesis is rejected at the 1% level. Therefore these variables are statistically significant. Furthermore the coefficients are small, but still economically significant. Poverty’s coefficient signifies that as the poverty headcount ratio at $1.00 increases by 1%, average life expectancy will decrease by 0.294 years. Furthermore if expenditure on health per capita increases by 1, average life expectancy will increase by 0.658 years.

### IV. Pertaining to the Dummy Variable

We decided to use a dummy variable for the regression against the GINI index because we believed that since it measures the degree of income inequality in degrees out of 100, it would be a trivial matter to divide the values into levels. It is also much easier to describe a country’s income inequality through descriptive categories such as high, medium, and low levels of income inequality than to simply state the percentages -- which hold relatively little emphasis on their own. The levels in question that we categorized the values of the GINI index into were:

- 0) Low income inequality: ginidum < 20
- 1) Medium income inequality: 20 < ginidum < 40
- 2) High income inequality: 40 < ginidum < 60
- 3) Very high income inequality: ginidum > 60.
Unfortunately for us, it ended up not holding an incredible amount of significance in our regression models. However, we gained knowledge on how to use dummy variables, which might be more relevant to other models. In the future we will try dividing the GINI index into more concise partitions because, looking back on it, the Very high income inequality group has very few members and does not benefit from being its own group. Instead, it will simply be divided into Low, Medium, and High.

Table 3.1: Summary Table 1

<table>
<thead>
<tr>
<th>LifeExp</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poverty</td>
<td>-0.314***</td>
<td>-0.297***</td>
<td>-0.201***</td>
<td>-0.161***</td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.022)</td>
<td>(0.028)</td>
<td>(0.035)</td>
</tr>
<tr>
<td>ln(Income)</td>
<td>-</td>
<td>0.578**</td>
<td>-0.005</td>
<td>0.0172</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.257)</td>
<td>(0.260)</td>
<td>(0.279)</td>
</tr>
<tr>
<td>ln(Expenditure)</td>
<td>-</td>
<td>-</td>
<td>2.47***</td>
<td>1.497**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.497)</td>
<td>(0.569)</td>
</tr>
<tr>
<td>Education</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.075***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.028)</td>
</tr>
<tr>
<td>Doctors</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Beds</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SS(social security)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Intercept</td>
<td>76.133***</td>
<td>74.449***</td>
<td>59.369***</td>
<td>58.541***</td>
</tr>
<tr>
<td></td>
<td>(0.545)</td>
<td>(0.897)</td>
<td>(3.301)</td>
<td>(3.66)</td>
</tr>
<tr>
<td>Observations</td>
<td>128</td>
<td>127</td>
<td>127</td>
<td>112</td>
</tr>
<tr>
<td>R-Squared</td>
<td>0.6494</td>
<td>0.6633</td>
<td>0.7149</td>
<td>0.7267</td>
</tr>
</tbody>
</table>

The following regression model was used:

\[ \text{LifeExp} = B_0 + B_1 \text{Poverty} + B_2 \ln(\text{Income}) + B_3 \ln(\text{Expenditure}) + B_4 \text{Education} + u \]
The final model shows that lower secondary completion rates have the largest absolute value for a coefficient out of all of the variables in the regression. It also is positive, showing that higher lower secondary completion rates should increase average life expectancy. Though no variable seems to provide a strong correlation between the independent variable and average life expectancy, two of the variables seem to show some relationship to the dependent variable.

After the presentation of the data, more variables were suggested to be examined. However, multiple models found that these variables were not statistically significant. More than these models were used and in the binary models with the poverty dependent variable and some of the new variables, the P-value was above 0.1 for social protection as well as the dummy Gini variable. These models are not included in this report because they are so simple. However, with the doctors or beds variables, the p-values were 0, suggesting to reject the null hypothesis. Some of the models are listed in the table below, but the most interesting one is model 11. Model 11 suggests that all of the variables are statistically significant below 1%. If poverty headcount ratio increased by 1, then life expectancy would decrease by -0.196. For expenditure, this is a level log model. Therefore the coefficient means that as one increases expenditure on health per capita by 1, life expectancy increases by 0.026. Education shows that as one increases lower secondary completion rates by 1, life expectancy increases by 0.097.

Table 3.2: Summary Table 2

<table>
<thead>
<tr>
<th>LifeExp</th>
<th>Model 5</th>
<th>Model 6</th>
<th>Model 7</th>
<th>Model 8</th>
<th>Model 9</th>
<th>Model 10</th>
<th>Model 11</th>
<th>Model 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poverty</td>
<td>-0.194**</td>
<td>-0.275**</td>
<td>-</td>
<td>-</td>
<td>-0.206***</td>
<td>-0.196***</td>
<td>-0.193***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0334)</td>
<td>(0.023)</td>
<td></td>
<td></td>
<td>(0.034)</td>
<td>(0.033)</td>
<td>(0.849)</td>
<td></td>
</tr>
<tr>
<td>ln(Income)</td>
<td>0.340</td>
<td>-0.421</td>
<td>-0.267***</td>
<td>0.353</td>
<td>0.353</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.251)</td>
<td>(2.033)</td>
<td>(0.857)</td>
<td>(0.341)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln(Expenditure)</td>
<td>3.613***</td>
<td>-0.421</td>
<td>-0.267***</td>
<td>0.353</td>
<td>0.353</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.153)</td>
<td>(2.033)</td>
<td>(0.857)</td>
<td>(0.341)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>0.066</td>
<td>0.160**</td>
<td>0.107***</td>
<td>0.066**</td>
<td>0.097***</td>
<td>0.100***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
<td>(0.029)</td>
<td>(0.033)</td>
<td>(0.027)</td>
<td>(0.026)</td>
<td>(0.027)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doctors</td>
<td>0.533</td>
<td>2.908***</td>
<td>-0.151</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.611)</td>
<td>(0.796)</td>
<td>(0.612)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The new correlation table between the variables help us to support our Gauss-Markov Assumption still.

### Unrestricted Model

```
regress averagellifeexpectancy povertyheadcountpoverty1 lnIncome lnExpenditure lowersecondarycompletionrates
>    tsec plsianp per1000 hospitalbeds per1000 adequacyofsocialprotection gini_dum
```

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 49</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1230.83698</td>
<td>8</td>
<td>153.83462</td>
<td>F(8, 40) = 10.71</td>
</tr>
<tr>
<td>Residual</td>
<td>328.979848</td>
<td>40</td>
<td>8.22448369</td>
<td>Prob &gt; F = 0.000</td>
</tr>
<tr>
<td>Total</td>
<td>1559.81622</td>
<td>48</td>
<td>32.4961795</td>
<td>R-squared = 0.789</td>
</tr>
</tbody>
</table>

| average_lifeexpectancy | Coef.     | Std. Err. | t      | P>|t|   | [95% Conf. Interval] |
|-------------------------|-----------|-----------|--------|-------|---------------------|
| povertyheadcountpoverty1| -.1561256 | .0470999  | -3.31  | 0.002 | -.2513172           | -.060936 |
| lnIncome                | 3.048102  | 1.106285  | 2.76   | 0.009 | .8122157            | 5.283988 |
| lnExpenditure           | -1.894259 | .9559926  | -1.96  | 0.058 | -.3.82619           | .226665 |
| lowersecondarycompletionrates | .0625104  | .0370655  | 1.69   | 0.099 | .0124019            | .1374226 |
| physiciansper1000      | .1756521  | .0236262  | 7.49   | 0.000 | .1.041706           | 1.393073 |
| hospitalbedsper1000    | -.8002833 | .3023237  | -2.65  | 0.012 | -.1.411307          | -.189259 |
| adequacyofsocialprotection | .0776991  | .0382593  | 2.02   | 0.043 | .0064375            | .1489808 |
| gini_dum               | .0239548  | 1.121333  | 0.02   | 0.983 | -.2.242344          | 2.290253 |
| _cons                  | 72.71476  | 4.990118  | 14.57  | 0.000 | 62.62936            | 82.80017 |
**F-Test**

**Model1:** The unrestricted model contains all the variables that we used in our project that have an impact on average life expectancy. For our restricted model, we dropped the two variables that were not significant from our unrestricted model (doctors and gini dum) to if the dropped variables had joint significance or not.

Unrestricted Model

\[
\text{Avg. life exp} = B_0 + B_{\text{poverty}} + B_{\ln(\text{income})} + B_{\ln(\text{expenditure})} + B_{\text{education}} + B_{\text{doctors}} + B_{\text{beds}} + B_{\text{SS}} + B_{\text{ginidum}}
\]

Restricted Model

\[
\text{Avg. life exp} = B_0 + B_{\text{poverty}} + B_{\ln(\text{income})} + B_{\ln(\text{expenditure})} + B_{\text{education}} + B_{\text{beds}} + B_{\text{SS}}
\]

\[H_0: B_{\text{doctors}} = 0; B_{\text{ginidum}} = 0\]

\[H_1: H_0 \text{ is not true.}\]

\[F = 3, c = 3.23,\]

Therefore, we failed to reject the null hypothesis at 5% level of significance. Hence, the variables are not jointly significant at 5% level of significance.

**Model2:** Our unrestricted model have the variables that we thought would influence the average life expectancy the most (poverty, income, social security and gini dum). Our restricted model has only income and social security because, once we ran the regression, we found that poverty, education and gini had low low significance, hence we dropped them.

Unrestricted Model

\[
\text{Avg. life exp} = B_0 + B_{\text{poverty}} + B_{\ln(\text{income})} + B_{\text{SS}} + B_{\text{ginidum}}
\]

Restricted Model

\[
\text{Avg. life exp} = B_0 + B_{\ln(\text{income})} + B_{\text{SS}}
\]

\[H_0: B_{\text{poverty}} = 0; B_{\text{education}} = 0\]

\[H_1: H_0 \text{ is not true.}\]

\[F = 51, c = 2.76\]

Therefore, we reject the null hypothesis. Hence, the variables are jointly significant at 5% level of significance.
Conclusion

In summary, even though there is a model that has statistically significant and possibly economically significant independent variables, the variables chosen are probably not the best measures of average life expectancy. As one knows, average life expectancy can be influenced by gender, genetics, lifestyle, etc. and though these variables might be correlated with some of the variables that were studied in this paper, using the true influencers might have been better suited for modelling. Therefore further analysis should be done in order to study good health and well-being and promote the SDG. One recommendation would be to use a dependent variable such as infant mortality, which may be more easily affected by the dependent variables studied in this paper.

Reference

Countries, World Health Organization, 2017


World Development Indicators|Data Bank, World Bank, 2017
List of Country
Afghanistan
Albania
Algeria
Andorra
Angola
Antigua and Barbuda
Argentina
Armenia
Australia
Austria
Azerbaijan
Bahamas
Bahrain
Bangladesh
Barbados
Belarus
Belgium
Belize
Benin
Bhutan
Bolivia
Bosnia and Herzegovina
Botswana
Brazil
Brunei Darussalam
Bulgaria
Burkina Faso
Burundi
Cabo Verde
Cambodia
Cameroon
Canada
Central African Republic
Chad
Chile
China
Colombia
Comoros
Congo
Cook Islands
Costa Rica
Cote d'Ivoire
Croatia
Cuba
Cyprus
Czechia
Democratic Republic of Korea (North)
Democratic Republic of Congo
Denmark
Djibouti
Dominica
Dominican Republican
Ecuador
Egypt
El Salvador
Equatorial Guinea
Eritrea
Estonia
Ethiopia
Fiji
Finland
France
Gabon
Gambia
Georgia
Germany
Ghana
Greece
Grenada
Guatemala
Guinea
Guinea-Bissau
Guyana
Haiti
Honduras
Hungary
Iceland
India
Indonesia
Iran
Iraq
Ireland
Israel
Italy
Jamaica
Japan
Jordan
Kazakhstan
Kenya
Kiribati
Kuwait
Kyrgyzstan
Lao PDR
Latvia
Lebanon
Lesotho
Liberia
Libya
Lithuania
Luxembourg
Madagascar
Malawi
Malawi
Malaysia
Maldives
Mali
Malta
Marshall Islands
Mauritania
Mauritius
Mexico
Micronesia
Monaco
Mongolia
Montenegro
Morocco
Mozambique
Myanmar
Namibia
Nauru
Nepal
Netherlands
New Zealand
Nicaragua
Niger
Nigeria
Niue
Norway
Oman
Pakistan
Palau
Panama
Papua New Guinea
Paraguay
Peru
Philippines
Poland
Portugal
Qatar
Republic of Korea (South)
Republic of Moldova
Romania
Russian Federation
Rwanda
Saint Kitts and Nevis
Saint Lucia
St Vincent and the Grenadines
Samoa
San Marino
Sao Tome and Principe
Saudi Arabia
Senegal
Serbia
Seychelles
Sierra Leone
Singapore
Slovakia
Slovenia
Solomon Islands
Somalia
South Africa
South Sudan
Spain
Sri Lanka
Sudan
Suriname
Swaziland
Sweden
Switzerland
Syrian Arab Republic
Tajikistan
Thailand
The former Yugoslav Republic of Macedonia
Timor-Leste
Togo
Tonga
Trinidad and Tobago
Tunisia
Turkey
Turkmenistan
Tuvalu
Uganda
Ukraine
United Arab Emirates
United Kingdom
United Republic of Tanzania
United States of America
Uruguay
Uzbekistan
Vanuatu
Venezuela
Vietnam
Yemen
Zambia
Zimbabwe