# **“Does Wealth Matter for Health? An Investigation of the Link Between National Income and Life Expectancy”**

# *by Ian Musumba*

Dec, 2022

**Introduction**

Every nation, no matter how wealth, strives to enhance the health of its population. The majority of attempts are made to decrease mortality and improve health, though not always with the same rates of success. In order to allocate resources to the population groups most in need, mortality assessments are of great interest to academics, policymakers, medical researchers, and other stakeholders. Mortality forecasts are of great importance in providing policy-relevant information, and therefore, governments making institutional arrangements for retirement and health care should be aware of the actual prospects of cohorts survival (Shkolnikov, Jdanov, Andreev, & Vaupel, [2011](https://genus.springeropen.com/articles/10.1186/s41118-019-0071-0#ref-CR45)). One of the most significant and dominating aspects of any nation or region's society and economy will continue to be population dynamics. To achieve the desired socioeconomic effects, it is helpful to evaluate and analyze existing demographic patterns and the anticipated outcomes and repercussions. At the macro level, preserving, increasing, and enhancing human population health is seen as one of the most important policies for sustainable development and overall life expectancy. A common summary measure for describing population health and longevity is life expectancy.

Life expectancy is a convenient and important summary measure of mortality and more intuitive than mortality rates (Klenk, Rapp, Büchele, Keil, & Weiland, [2007](https://genus.springeropen.com/articles/10.1186/s41118-019-0071-0#ref-CR28)). Therefore, life expectancy at birth is a reliable and significant predictor of the community's overall well-being. Although health is a multifaceted concept, life expectancy is one of the most frequently used indices of population health. Life expectancy at birth is used as a proxy for population health. A crucial synthetic indicator for evaluating the social and economic progress of a nation or region is life expectancy. With that in mind, it can be claimed that defining good health is based on a number of socioeconomic preconditions, including a decrease in low levels of education, a decrease in unemployment, better diet, access to immunization, and improvement of living conditions.

This paper aims to investigate the effect of the socioeconomic development on life expectancy. These goals will be met through attempts to assert or deny previously made claims regarding factors that heavily influence life expectancy. With that being said, the research question that will be addressed and investigated through this paper is as follows: Whether a nation’s economic standing has an impact on life expectancy? And what variables most accurately predict life expectancy in our dataset.

**Methods**

We will utilize R as the programming language because of its capability to provide a favorable environment for statistical computation and design, which will help you grasp the stages involved in EDA (Exploratory Data Analysis). We begin by importing each of the necessary libraries, including ggplot for graphic functions, tidyverse for data structures, and mosaic for statistics and calculus instructions.

We will be using dataset for life expectancy that includes 2938 obervations for 193 countries, which can be classified as either developed or developing. This dataset was obtained from the World Health Organization’s data repository website and the United Nations website for corresponding economic information. Our focus will be on examining the relationship between life expectancy and subset of 22 variables, including Status, Adult Mortality, Infant Mortality, Hepatitis B, BMI, GDP, Population, Income Composition of Resources, and Schooling. To get a quick overview of the data, we can use the glimpse() method to retrieve the first few variables of each column in the data frame.

Starting our EDA, we checked if our data contains any null values. If we spotted any potential outliers in our data, handling the missing values by replacing them with the median value is advised. In order to investigate the Life expectancy variable on its own, we first used R to compute the five summary statistics. According to Table 1, the average global life expectancy in this data is 69 years old, with a minimum age of 36 and a maximum age of 89. Figure 1 shows the histogram distribution for the same variable to be left-skewed when explored graphically. In this case, the model’s skewness is a good indication that the average longevity is high.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Min** | **Q1** | **Median** | **Q3** | **Max** | **Mean** | **SD** | **N** | **Missing** |
| 36.3 | 63.1 | 72.1 | 75.7 | 89 | 69.22 | 9.52 | 2928 | 10 |

(5 number Summary and Quartiles of *Life expectancy*)

{Table 1}

Chart, histogram

Description automatically generated

Next, we explored the relationship between age and global life expectancy, which is found under the Appendix as *Figure 2*. Over the past decade and a half, from year 2000-2015, we notice that the global life span had an upwards trend. In the early 2000, the minimum years. This could be attributed to many countries taking measurements to combat low life expectancy by promoting better healthier lifestyles, diet, and even medical care, which resulted to an increase in longevity. The Status variable is evaluated on its own in *Figure 3* located in the appendix. It goes to show that there are way more countries that have the status of developing than developed.

**Results**

Next, we would like to examine the research questions to whether a country’s status has an influence on life expectancy. *Figure 4* demonstrates these two variables side to side graphically. There is an undeniable distinction between the population density of developed and developing nations. Therefore, it is not surprising that governments' perceptions of the nation's death rate vary depending on development level. Life expectancy in industrialized countries started to increase while it stayed low in the rest of the world. This led to a very high inequality in how health was distributed across the world. Good health in the rich countries and persistently bad health in those countries that remained poor. As a result, life expectancy is generally higher in industrialized nations as opposed to developing or less developed nations. T Freeman says, “While in general a country’s life expectancy increases with national income, some countries ‘punch above their weight’, while some ‘punch below their weight’ – achieving higher or lower life expectancy than would be predicted by their per capita income.” (Freeman, 2020)

Chart, line chart

Description automatically generated

Discovering which variables heavily impact the *Life expectancy* variable is going to be a key contributing point in this research paper. More precisely, we will perform linear regressions models to help determine the characteristics and strengths of these contributors and whether it is critical to improving population health overall.

Before fitting linear models, we must integrate useful packages. In this scenario, we want to call on **jtools** package which help understand and report the results of regressions models. The first step we took towards creating the final model is fitting all the desirable variables mentioned above in paragraph 2 and seeing the outcome. All the variables with a high p-value were dropped in order to increase the effectiveness of our final model. Thus, we had to delete *Population*, which had a 0.9 p-value, and *Year*, which had a significant0.09 p-value.

The overall results section of our study, after running the appropriate statistical tests, we found that p-value was extremely low, indicating a significant difference between the means. Based on this results, we can reject the null hypothesis and concluded that a significant difference does exist between national income and life expectancy. This supports the idea that wealth and income have an impact on health outcomes, as previously discussed in out investigation.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Coefficients:** | **Estimate** | **Std.Error** | **t-value** | **P-value** | **95% CI**  **2.5% 97.5%** | |
| **(Intercept)** | 5.699e+01 | 6.258e-01 | 91.953 | <2e-16 | 55.758 | 58.212 |
| **StatusDeveloping** | -1.859e+00 | 2.618e-01 | -7.099 | 1.57e-12 | -2.372 | -1.345 |
| **Adult.Mortality** | -2.127e-02 | 9.579e-04 | -24.792 | <2e-16 | 0.023 | -0.020 |
| **Hepatitis.B** | 1.213e-02 | 3.618e-03 | 3.352 | 0.000813 | 0.005 | 0.019 |
| **BMI** | 6.233e-02 | 4.936e-03 | 12.628 | <2e-16 | 0.053 | 0.072 |
| **GDP** | 4.303e-05 | 7.221e-06 | 5.960 | 2.83e-09 | 0.000 | 0.000 |
| **Income composition of resources** | 6.837e+00 | 6.848e-01 | 9.985 | <2e-16 | 5.495 | 8.180 |
| **HIV.AIDS** | -4.840e-01 | 1.896e-02 | -25.528 | < 2e-16 | -0.521 | -0.447 |
| **Schooling** | 8.436e-01 | 4.464e-02 | 18.898 | < 2e-16 | 0.756 | 0.931 |

(Summary Statistics and Confident Interval)

{Table 2}

Chart, histogram

Description automatically generated

(Histogram of Final Model)

{Figure 5}

A screenshot of a computer

Description automatically generated with medium confidence

(Detailed Summary Statistics Table of Final Model)

{Table 3}

**Discussion**

Based on our investigation of the link between national income and life expectancy, it appears that wealth does play a role in health outcomes. Countries with higher levels of GDP and other measures of wealth tend to have a higher life expectancy, indicating that access to resources and quality healthcare can impact a population’s overall health and longevity. However, it is important to note that wealth is not only the factor influencing health outcomes, and other variables such adult mortality, infant mortality, and access to education also play a role. Further research is needed to fully understand other complex relationships between wealth and health to identify the most effective strategies for improving life expectancy outcomes in both developed and developing.

One potential area for improvement in future studies on this topic would be to more thoroughly investigate the other variables that may influence the relationship between national income and life expectancy. Additionally, it may be useful to focus on specific health outcomes in different context. Future research could also examine the effectiveness of various strategies for improving life expectancy, such as investing in healthcare infrastructures or promoting healthy lifestyles, to identify the most impactful interventions for improving overall population health.

There were no substantial limitations, however having more years of data would have helped us study big changes and critical time periods. As for my areas of weakness, I'd say that I wasn't able to delve deeply into some variables since I couldn't figure out how to seamlessly transition to other variables or how to provide enough explanation without diverging from the research topic. Strengths include the fact that I was able to locate all the required codes and packages to assist with the visualizations and related tables. I would advise future researchers to figure out how to simultaneously graph multiple qualitative and quantitative variables.

**Annotated Appendix**

**Year:** Year

**Status:** Developed or Developing status

**Life expectancy:** Life Expectancy in age

**Hepatitis B:** Hepatitis B (HepB) immunization coverage among 1-year-olds (%)

**BMI:** Average Body Mass Index of entire population

**HIV/AIDS:** Deaths per 1 000 live births HIV/AIDS (0-4 years)

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

**Population:** Population of the country

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

**Schooling:** Number of years of Schooling(years)

**A picture containing chart

Description automatically generated**

*Figure 2* explored the relationship of Years vs Life expectancy. The upward trend is a good predictor of growing life expectancy. The chart has blatant outliers, but we decided to disregard them.

**Chart, bar chart

Description automatically generated**

*Figure 3* is a side-by-side bar chart comparing the count of the binary variable Status. It looks like ¼ of the countries are considered developed.

Visualizing this model graphically *(Figure 5)*, we plotted a histogram that showed the distribution of our residuals was able to achieve the desired bell-shaped curvature. Since the normality condition was meet, there was no need to apply any type of transformation.

**Cited Sources**

Freeman, Toby, et al. “Why Do Some Countries Do Better or Worse in Life Expectancy Relative to Income? an Analysis of Brazil, Ethiopia, and the United States of America - International Journal for Equity in Health.” *BioMed Central*, BioMed Central, 10 Nov. 2020, <https://equityhealthj.biomedcentral.com/articles/10.1186/s12939-020-01315-z>.

Miladinov, Goran. “Socioeconomic Development and Life Expectancy Relationship: Evidence from the EU Accession Candidate Countries - Genus.” *SpringerOpen*, Springer International Publishing, 10 Jan. 2020, <https://genus.springeropen.com/articles/10.1186/s41118-019-0071-0>.

“Baby Vaccines at Birth.” *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 25 Feb. 2020, <https://www.cdc.gov/vaccines/parents/by-age/newborn-birth.html>.

Disabled World. “Longevity: Extending Life Span Expectancy.” *Disabled World*, Disabled World, 14 Apr. 2022, <https://www.disabled-world.com/fitness/longevity/>.