Do Inhabitants of Wealthier Nations Live Longer?

John S. Garvey

Abstract— This study explores the relationship between world development indicators like GDP per capita and health-related aspects such as the prevalence of underweight children and the life expectancy for a nation. Using data from www.worldbank.org, we analyzed these correlations through a handmade predictive linear regression model. GDP per capita demonstrated a strong positive correlation with life expectancy, signifying its role in longer lifespans. Conversely, the percentage of underweight children exhibited a weaker negative correlation. Combining both indicators improved predictive accuracy, highlighting the interplay between economic prosperity and health outcomes. Our findings support the notion that residing in a developed nation leads to a longer life expectancy.

Index Terms— World Development Indicators, Developed Nation, Developing Nation, GDP, Predictive Regression Models

1 Introduction

For this project we will be using the vast amount of data found at www.worldbank.org. The datasets that we will be using are from every country in the world for the years from 1960 to 2022 and the development indicators we will be analyzing are ...
- Life Expectancy from Birth
- Gross Domestic Product per capita
- Higher-Technology Exports Percentage
- Population Ratio below National Poverty Lines
- Prevalence of Underweight Children

Our primary focus is on modern rates of life expectancy, and the factors that influence them. The simple question we are asking is as follows. "Does living in a wealthy nation mean a longer life?" Our research does not consider the quality of life, but rather, we can reframe the hypothesis to be, "Are people living in abundance healthier?"

For this project, we will be using a crude classification for countries with different levels of prosperity. We will be using the terminology used by The World Bank which is "developed" and "developing" nations or countries. Our analysis recognizes this classification as a spectrum, and that a country can fall on one end or somewhere in between. For the purposes of this project we are defining “developed” countries as countries with abundance, wealth, and access to modern technology. For "developing" we will be considering nations with scarcity, limited access to resources including technology and that are under distressed or impoverished economic circumstances. For our research objectives, we will concentrate on three indicators that reflect a country's development level: its citizens' economic prosperity, their food supplies and technological advancement.

The following are the speculations and objectives that spawned this project, all of which are believed to have intellectual merit for determining factors that affect life expectancy in the twenty-first century. The knowledge we hope to gain is finding evidence that supports one of two following contradictory notions. On one hand, the citizens of developed countries are likely to be healthy because they have no shortage of food or resources to keep them healthy, as well as access to modern medicines and technologies that help keep them alive when they get sick or hurt. On the other hand, people that live in societies of over-abundance tend to suffer from health issues such as addiction and obesity. Which is reflected in the fact that the leading cause of death in the USA is and has been heart disease for quite some time [1].

The flip side of this coin is that people in developing countries obviously have much more limited access to resources such as food and medicine which affect their populations' health considerably. Additionally, another sad fact is that even though tuberculosis has been curable since the 1940s[2] it has also been, "1 of the top 10 leading causes of death worldwide, and is the second leading cause of death from a single infectious agent behind only COVID-19 and ranking above HIV/AIDS."[3] This lack of access to tuberculosis medication means that, "99% of deaths occur in developing countries, with the greatest burden in Sub-Saharan Africa and Southeast Asia"[4]. Meaning if you get sick in a developing country, there is a much greater chance that you will not be able to receive the help you need. However, again, on the other hand, people not living in a "developed" country are not necessarily less healthy and in fact for many societies that are still largely agricultural, "the major chronic diseases have low prevalence among the rural peasantry and tribal groups, with the exception of gastro-esophageal cancer, osteoarthritis and cataract."[5] With all of this in mind, for this project we are seeking evidence to support the hypothesis that living in a developed country actually means having a longer life.

1.1 Related Work

In exploring previous work related to predictors influencing life expectancy, several studies have contributed significantly to this domain. One relevant study is the "Glob-

al Burden of Disease Study 2019"[6] by GBD 2019 Diseases and Injuries Collaborators. This research provides comprehensive global health estimates, highlighting trends in life expectancy and various factors affecting health outcomes across different regions. Another significant work is Angus Deaton's research on "Income, Health, and Well-Being around the World: Evidence from the Gallup World Poll." [7] Deaton's study delves into the relationship between income, health, and overall well-being on a global scale, shedding light on how economic factors impact health outcomes. Moreover, Raj Chetty et al.'s study on "Life Expectancy and Income"[8] focuses on exploring the correlation between income levels and life expectancy. This research particularly investigates how income inequality affects life expectancy within different geographical areas in the United States. Additionally, the Institute for Health Metrics and Evaluation (IHME) conducts vital research on "Global Health Metrics and Evaluation." [9] Their work provides extensive data on life expectancy trends and factors influencing health outcomes worldwide. Furthermore, "The State of Global Air Report"[10] by Health Effects Institute provides insights into air quality and its impact on life expectancy globally. This report highlights the association between air pollution and health, offering crucial data on its effects on life expectancy. Finally, the "OECD Health Statistics"[11] by the Organization for Economic Co-operation and Development (OECD) furnishes comprehensive health statistics across OECD countries, emphasizing health determinants influencing life expectancy and health outcomes. These studies collectively contribute valuable insights into the factors influencing life expectancy, the impact of economic indicators on health outcomes, and trends in global health.

2 Exploratory Data Analysis

The dataset under analysis is a subset of the World Development Indicators (WDI) available at worldbank.org's DataBank. This specific dataset provides the most up-to-date and accurate global development data, offering estimates at national, regional, and global scales. Comprising a total of 57 WDIs, our focus narrows down to five key indicators, notably the total life expectancy for a nation, along with four others: GDP per capita, poverty ratio, percentage of higher-technology exports, and prevalence of underweight children. These selected indicators serve to depict a nation's economic status, technological prowess, and the adequacy of its food supply.

For instance, our initial analysis reveals that the life expectancy of the United States of America starts at 69.8 in the 1960s, steadily climbing to a peak of 78.8 in 2019. Then there is a dip in the subsequent two years due to the impact of COVID-19. Two more notable findings from our initial analysis are available in our summary table within the processed data folder. This table highlights two significant statistics: firstly, the highest GDP (per capita) recorded in the past 63 years stands at $234,317.08. Upon investigating the dataset, we find this record belongs to Monaco for the year 2022, consistently recognized as one of the wealthiest nations in the world. Secondly, the lowest life expectancy within the past 63 years was 11.99 years old, which is attributed to Cambodia in 1975. Upon investigation, this tragic figure coincides with the Cambodian genocides perpetrated by the Khmer Rouge under Pol Pot's regime.

The dataset from the World Bank comprises of data collected from a total of 217 countries. Consequently, the primary limitation for our datasets resides in the sample sizes pertaining to our development indicators. The largest sample size among our indicators, such as GDP, encompasses data from 214 countries. Conversely, the national poverty ratio exhibits the smallest sample size, containing records from only 155 countries. Below in Figure 1, the sample sizes for each indicator data category are illustrated.

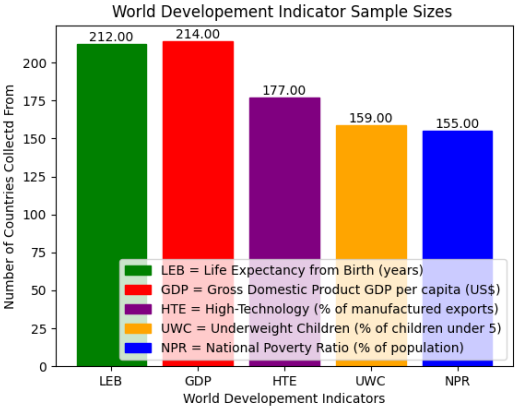

Fig. 1. Sample size or number of countries with records for the five world development indicator data categories

3 Methodology

For this research project, we have chosen to use the most fundamental machine learning approach of linear regression to construct a predictive model for a nation's mean life expectancy. This model utilizes development indicators, like GDP per capita, as parameters to create a model that predicts the mean life expectancy of the citizen living in a country based on the indicators given.

The initial model is constructed by utilizing the indicator that exhibited the strongest correlation with life expectancy, as identified during our data analysis. The correlations between not just the life expectancy but between all the indicators can be found in the correlation table in the data processed folder. Among these indicators, gross domestic product per capita emerges with the highest correlation rate of 0.68 concerning life expectancy. Therefore, the first predictive machine learning model is trained using the correlation data between life expectancy and the GDP of nations worldwide. Leveraging this data, we developed the initial model capable of predicting the average life expectancy of a nation's citizens based on its current GDP per capita.

The second model is based on the indicator with the second strongest correlation with a nation's life expectancy, namely the prevalence of underweight children or the percentage of children under five years old who are un-

derweight. This particular indicator displays a negative correlation, albeit not as robust as the GDP correlation, with a correlation coefficient of -0.293. Despite its relatively weaker strength compared to GDP, it stands as the second strongest correlation with life expectancy among the chosen indicators. Therefore, the second predictive machine learning model is built using the correlation data between life expectancy and the percentage of underweight children across nations worldwide. Utilizing this data, we established a model capable of predicting the average life expectancy of a nation's citizens based on its current percentage of underweight children.

Our final model uses both these indicators to create a predictive linear regression model. This model leverages both GDP per capita and the percentage of underweight children as parameters to predict the current life expectancy of a nation's citizens.

4 Results and Discussion

For our initial predictive model, we employed the correlation between a nation's citizens' life expectancy and its GDP per capita. Below in Figure 2 is a scatterplot depicting life expectancy by GDP per capita for all nations worldwide, followed by a graph illustrating our first predictive linear regression model.

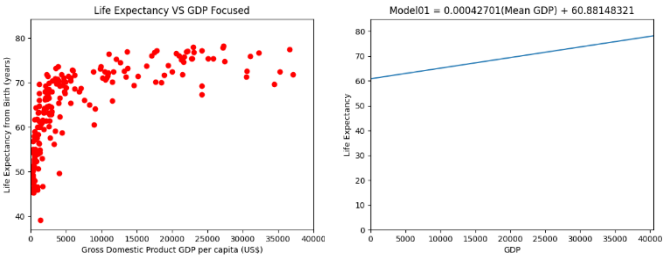

Fig. 2. (Left)Scatter plot of all the world's nations life expectancy vs GDP per capita. (Right) The first predictive linear regression model for a nation's life expectancy for its GDP per capita.

After running our first predictive model 20 times and averaging its Mean Standard Error (MSE) score and Accuracy in years, we observed that our initial model achieves a predictive rate with an MSE score of 56.74 and an average accuracy of 6.4 years. This indicates that, on average, the model can predict a nation's life expectancy accurately within 6.4 years.

In our second predictive model, we utilized the correlation between a nation's citizens' life expectancy and its percentage of underweight children. Below in Figure 3, you'll find a scatterplot illustrating life expectancy plotted against the percentage of underweight children for all nations globally. This is followed by a graph displaying our second predictive linear regression model.

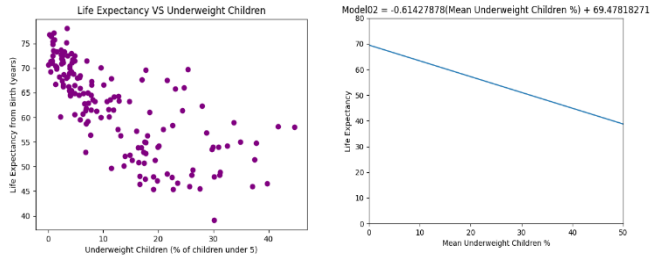

Fig. 3. (Left)Scatter plot of all the world's nations life expectancy vs their percentage of underweight children. (Right) The second predictive linear regression model for a nation's life expectancy for its percentage of underweight children.

After running our second predictive model a total of 20 times and averaging its Mean Standard Error (MSE) score and Accuracy in years, we discovered that our second model can achieve a predictive rate with a 41.08 MSE score and an average accuracy of 5.9 years. Despite the lower correlation between the two indicators, on average, the second model exhibits slightly higher accuracy in its predictions.

Finally after running our final predictive model a total of 20 times and averaging its Mean Standard Error (MSE) score and Accuracy in years, we determined that our last model achieves a predictive rate with a 31.34 MSE score and an average accuracy of 5.5 years. Integrating both indicators produces the most accurate version among our predictive linear regression models.

On average, all our models demonstrate an accuracy rate of approximately 5-7 years difference between our predictions and a nation's actual life expectancy. Meaning for a given indicator our models can accurately predict a nation's life expectancy within 5-7 years. Further discussion of these results to follow.

5 Conclusion

Our analysis aimed to unravel the relationships between specific world development indicators and their impact on a nation's life expectancy. The primary objective of our research was to discern whether residing in a wealthier nation correlates with a longer lifespan.

Utilizing linear regression models, we sought to predict life expectancy based on two primary indicators: GDP per capita and the prevalence of underweight children. Our models aimed to utilize the predictive power of these indicators on a nation's average life expectancy.

Throughout our analysis, it became evident that GDP per capita exhibits a relatively strong positive correlation with life expectancy. Nations with higher GDPs per capita tended to demonstrate increased life expectancy rates, suggesting a plausible association between economic prosperity and longer lifespans. Conversely, the percentage of underweight children exhibited a weaker negative correlation with life expectancy. While not as influential as GDP per capita, this indicator revealed a trend wherein nations with lower percentages of underweight children tended to have higher life expectancy rates.

Our study supports the idea that residing in a developed country tends to result in a longer life expectancy. The relationship between a nation's GDP per capita and life expectancy highlights the crucial role of economic prosperity in influencing longevity. Furthermore, disparities in accessing essential resources significantly impact variations in life expectancy among different regions. In summary, our research sheds light on the intricate relationship between economic well-being, health, and life expectancy.

REFERENCES

[1] (CDC, n.d.). Centers for Disease Control and Prevention. "FastStats - Leading Causes of Death." Retrieved from https://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm

[2] (ERS Journals, n.d.). European Respiratory Journal. "Major historical landmarks of tuberculosis control and the lessons they teach." Retrieved from https://erj.ersjournals.com/content/20/36_suppl/87S#:~:text=The%20major%20historical%20landmarks%20of,in%20the%201970s%20that%20isoniazid

[3] (Government of Canada, n.d.). Government of Canada. "Tuberculosis (TB)." Retrieved from https://www.international.gc.ca/world-monde/issues_development-enjeux_developpement/global_health-sante_mondiale/tuberculosis-tuberculose.aspx?lang=eng

[4] (MPRA, n.d.). Munich Personal RePEc Archive. "Global Health and Economic Development." Retrieved from https://mpra.ub.uni-muenchen.de/82851/

[5] (PubMed, n.d.). PubMed. "Globalization and health: A framework for analysis and action." Retrieved from https://pubmed.ncbi.nlm.nih.gov/24398278/

[6] GBD 2019 Diseases and Injuries Collaborators. "Global Burden of Disease Study 2019." [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/33069326/

[7] A. Deaton, "Title of the article: Subtitle." *Journal Title*, vol. volume, no. issue, pp. page range, year. [Online]. Available: https://www.aeaweb.org/articles?id=10.1257/jep.22.2.53

[8] R. Chetty et al., "Title of the article: Subtitle." *Journal Title*, vol. volume, no. issue, pp. page range, year. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4866586/

[9] Institute for Health Metrics and Evaluation. "Global Health Metrics and Evaluation."[Online]. https://www.healthdata.org/news-events/newsroom/videos/global-health-metrics-and-evaluation-2013

[10] Health Effects Institute. "The State of Global Air Report." [Online].Available: https://www.stateofglobalair.org/sites/default/files/soga_2017_report.pdf

[11] Organisation for Economic Co-operation and Development (OECD). "OECD Health Statistics." [Online]. Available: https://www.oecd.org/health/health-data.htm