Analyzing the Impact of GDP and Expenditure on Life Expectancy*

Pu Yuan

April 10, 2024

This study investigates the relationship between economic prosperity, measured through Gross Domestic Product (GDP) and expenditure, and its impact on life expectancy across various nations. Utilizing data from international health and economic databases, we employ linear regression models to determine how these economic indicators influence the longevity of populations. The research specifically explores the extent to which increases in GDP and health expenditure as a percentage of GDP are associated with improvements in life expectancy.

1 Introduction

Understanding how long people live and what factors influence life expectancy is crucial for improving public health globally (Roser, Ortiz-Ospina, and Ritchie (2013)). While it's known that richer countries generally have better health outcomes, the exact role of economic strength and healthcare spending needs clearer understanding (Kunze (2014)). This paper investigates how life expectancy correlates with economic indicators like Gross Domestic Product (GDP) and health expenditure across various nations. This approach helps uncover the specific impacts of a nation's wealth and its healthcare investments on the health and longevity of its population.

In this study, we specifically examine the relationship between life expectancy and two main economic factors: GDP and expenditure. The aim is to detail how these factors contribute to health across different global settings. Although existing studies suggest a positive link between a country's economic status and its health outcomes, there remains a significant gap in understanding how health expenditures, particularly in less wealthy nations, influence life

*Code and data are available at: https://github.com/scottyuan6/life_expectancy.git. Bio replication available at: https://osf.io/2d9qn/?view_only=e9821866a4204e3a896d54430d68446f

expectancy. By analyzing international data using regression models, this study fills this gap by detailing how economic output and health investments interact to affect life expectancy.

Our analysis shows that both higher GDP and increased health expenditure are significantly associated with better life expectancy. More importantly, we discover that health spending has a more substantial impact in countries with lower GDP. This finding suggests that targeted health investments in poorer countries can lead to significant improvements in life expectancy, offering a powerful strategy for public health interventions.

Our estimand is to fit linear regression models to explore how changes in these economic indicators directly impact the number of years that residents of a country are expected to live, under the assumption that other influencing factors are held constant.

These results are vital for policymakers and public health officials. They highlight the importance of not just increasing wealth but also strategically investing in health to achieve the best health outcomes. The study advocates for focused health spending as a key tool in public health strategy, especially in less affluent regions.

The rest of the paper is organized as follows: The Section 2 describes the data, variables, limitation of the dataset and the rationale for the selection of this dataset. Plots are also included in this section to better visualize the dataset. The Section 3 describes how to set up the linear model. The Section 4 presents the results of the analysis in a table. The Section 5 provides an in-depth discussion of our findings and reflections on the research process. Finally, the Section A adds details of the models and dataset.

We use the statistical programming language `R` (R Core Team 2023). In the data analysis and visualization process, we also made use of the following `R` packages: `readr` (Wickham et al. 2024), `dplyr` (Wickham et al. 2014), `ggplot2` (Wilkinson 2011), `modelsummary` (Arel-Bundock 2022), `kableExtra` (Abel 2018), `broom` (Robinson 2014),.

2 Data

2.1 Overview

The "Life Expectancy" dataset, hosted on the Global Health Observatory (GHO) data repository under the World Health Organization (WHO), serves as a comprehensive resource for tracking health status and numerous associated factors across all countries. This rich dataset, publicly available for health data analysis, spans data from 2000 to 2015 for 193 countries. It integrates health-related statistics with economic data sourced from the United Nations, reflecting significant strides in global health developments over the last 15 years, particularly in developing nations.

The dataset comprises 22 columns and 2938 rows, encapsulating 20 predictive variables categorized into broad sectors such as Immunization-related factors, Mortality factors, Economic

factors, and Social factors. These variables have been meticulously selected for their strong representational value in health analysis. Among these, indicators such as immunization rates, adult and infant mortality, and economic parameters like GDP offer insights into the health outcomes and infrastructure of a nation.

2.2 Variable

Immunization-related Factors:

- Hepatitis B: Percentage of 1-year-olds immunized against hepatitis B.
- Polio: Immunization coverage among 1-year-olds against poliovirus.
- Diphtheria: Coverage of diphtheria tetanus toxoid and pertussis (DTP3) immunization among 1-year-olds.

Mortality Factors:

- Adult Mortality: Probability of dying between 15 and 60 years per 1000 population.
- Infant Deaths: Number of infant deaths per 1000 live births.
- Under-five Deaths: Number of under-five deaths per 1000 live births.
- HIV/AIDS: Deaths per 1000 live births due to HIV/AIDS (ages 0-4).

Economic Factors:

- GDP: Gross Domestic Product per capita, which reflects the economic state of a country.
- Percentage Expenditure: Expenditure on health as a percentage of GDP.

Social Factors:

- Schooling: Average number of years of schooling of a population.
- Income Composition of Resources: A composite Human Development Index measure that includes knowledge status, income, and longevity.

Other Health-related Factors:

- Alcohol: Recorded per capita (15+) consumption (in liters of pure alcohol).
- BMI: Average Body Mass Index of the adult population.
- Total Expenditure: General government expenditure on health as a percentage of total government expenditure.
- Population: Total population of the country.
- Thinness 1-19 years: Prevalence of thinness among children and adolescents aged 10-19 years.
- Thinness 5-9 years: Prevalence of thinness among children aged 5-9 years.

2.3 Limitation

While the "Life Expectancy" dataset from the WHO's Global Health Observatory provides extensive insights into global health trends and factors affecting life expectancy, it also comes with several limitations that can affect the scope and accuracy of analyses derived from it. The dataset has missing values in several key variables, such as population, GDP, and Hepatitis B vaccination rates, particularly from less-known or smaller countries. This missing data can bias the results and limit the generalizability of the findings.

Despite these limitations, the "Life Expectancy" dataset is a valuable resource for understanding global health trends and the factors affecting life expectancy across countries. Researchers and policymakers should be mindful of these limitations when designing studies, interpreting results, and making decisions based on this data. Adjusting analyses to take these factors into account can help mitigate some of the limitations and lead to more robust and applicable findings.

2.4 Data cleaning and preparation

Cleaning focused on simplifying the dataset to allow for a more streamlined examination. The main cleaning step consisted of handling with missing data. Several columns such as "Alcohol", "Hepatitis B", "Total Expenditure", "GDP", "Population", and various health indices show missing values which required handling for analysis. This process did not construct new variables but rather refined the dataset to the most relevant parts.

2.5 Plots

The scatter plot (Figure 1) illustrates the relationship between GDP per capita and life expectancy, differentiated by country status (Developed vs. Developing). Each point represents a country, with its position determined by its GDP and average life expectancy. A linear model fit (blue line) indicates the trend across the dataset. Typically, this plot shows a positive correlation, suggesting that higher GDP per capita is associated with longer life expectancy. This relationship underscores the impact of economic prosperity on health outcomes, possibly through better access to healthcare, higher living standards, and more effective public health interventions.

The boxplot (Figure 2) compares the distribution of life expectancy between developed and developing countries. This plot highlights median values, quartiles, and potential outliers, providing a clear visual representation of the disparities in health outcomes between different economic statuses. Generally, developed countries tend to have higher life expectancy, demonstrating narrower interquartile ranges and fewer outliers, which implies more uniform health outcomes within these countries compared to developing ones.

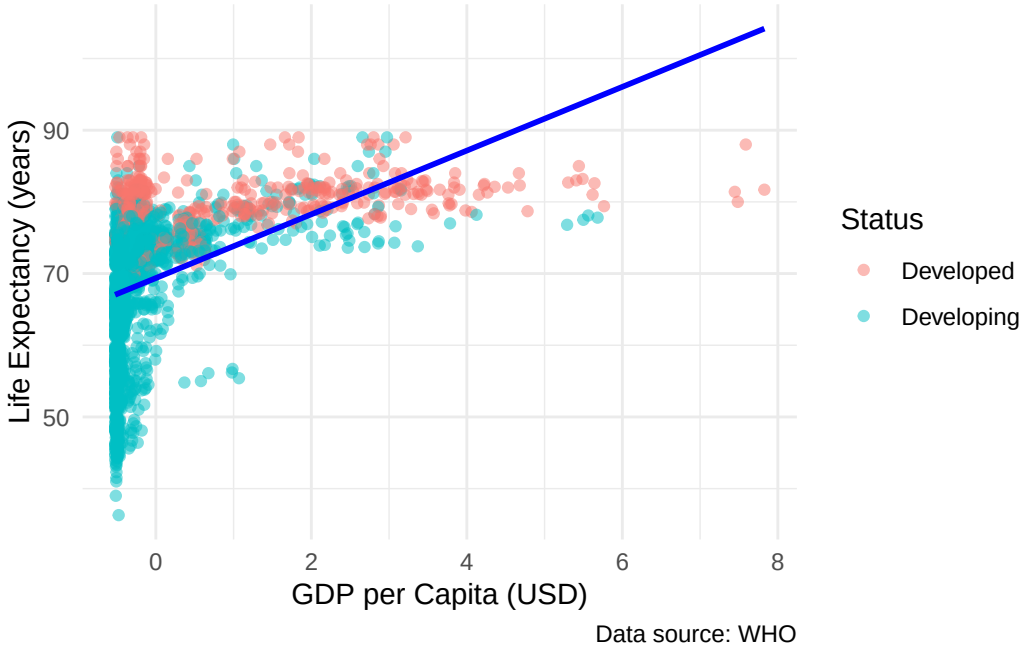

Figure 1: Life Expectancy vs. GDP per Capita

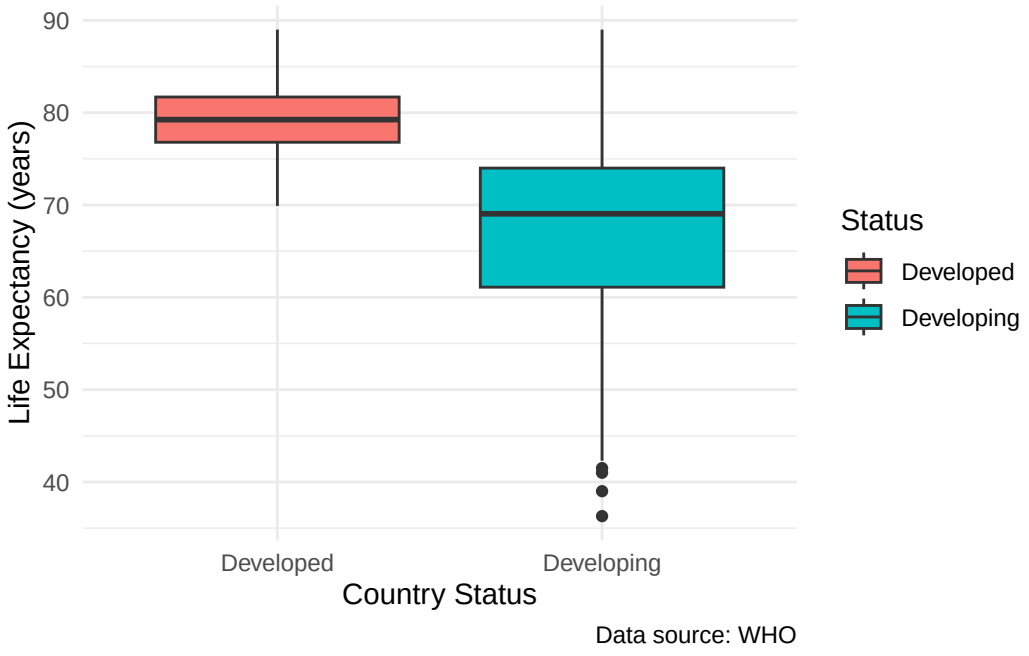

Figure 2: Life Expectancy in Developed vs. Developing Countries

3 Model

In this paper, we explore the influence of economic conditions on life expectancy across various countries by employing regression analysis. We initiate our study by constructing directed acyclic graphs (DAGs) to depict the hypothesized relationships among the variables involved. These DAGs serve to visually demonstrate our assumptions about the interactions between life expectancy and economic factors such as GDP, percentage expenditure , and adult mortality.

The paper aims to find how different economic indicators are significant predictors of life expectancy. We hypothesize that richer nations, indicated by higher GDP, are likely to have greater life expectancy. Furthermore, we predict that higher percentage expenditure relative to GDP positively impacts life expectancy, reflecting better access to healthcare services.

To quantitatively capture these effects, we develop multiple regression models. Our first model, denoted as

$$Y_1 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$$

Here, Y_1 represents the life expectancy in a country. X_1 is the GDP, and X_2 is the percentage expenditure as a percentage of GDP. β_0 is the intercept, representing the expected life expectancy when both GDP and percentage expenditure are zero, which serves as a theoretical baseline. β_1 and β_2 are the coefficients that measure the impact of each unit increase in GDP and percentage expenditure, respectively.

In a similar vein, the second regression model, Y_2, is constructed to explore these relationships further by incorporating additional economic variables such as adult mortality:

$$Y_2 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$

In this model, X_3 represents the adult mortality index, adding depth to our understanding of broader health and economic challenges.

Moreover, we extend our analysis to logistic regression models to estimate the likelihood of countries achieving high life expectancy benchmarks based on their economic conditions:

$$Y_3 = logit^{-1}(\beta_0 + \beta_1 X_1 + \beta_2 X_2)$$
$$Y_4 = logit^{-1}(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3)$$

Here, Y_3 and Y_4 calculate the probability that a country's life expectancy exceeds a certain age, based on its GDP, percentage expenditure, and adult mortality.

These models aim to provide a framework to test our hypotheses about the significant influences of economic conditions on life expectancy. The results from these analyses will not only

Table 1: Summary of Regression Analyses: the Impact of economic Indicators and Adult Mortality on Life Expectancy Across Countries

	Basic Economic Model	Extended Economic Model	Logistic Model for Benchmar
(Intercept)	69.51	76.85	−1.49
	(0.23)	(0.25)	(0.09)
GDP	4.75	2.69	0.83
	(0.39)	(0.30)	(0.15)
percentage.expenditure	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)
Adult.Mortality		−0.05	
		(0.00)	
Num.Obs.	2490	2490	2490
R2	0.213	0.547	
R2 Adj.	0.212	0.546	
AIC	17 758.8	16 387.4	
BIC	17 782.1	16 416.5	
Log.Lik.	−8875.420	−8188.702	
RMSE	8.55	6.49	0.37

contribute to academic knowledge but also offer practical insights for policymakers seeking to improve health outcomes through economic development strategies.

Model check details are included in Appendix A.1.

4 Results

Our results are summarized in Table 1. This table presents the results from three regression models assessing the relationships between life expectancy and economic factors. The Basic Economic Model explores the influence of GDP and percentage of GDP spent on health. The Extended Economic Model includes Adult Mortality to examine additional impacts of health-related demographic factors. Lastly, the Logistic Model for Benchmark assesses the probability of life expectancy surpassing 75 years based on the same economic indicators, using logistic regression. Each model provides estimates, standard errors, and significance levels, facilitating a comprehensive understanding of how economic conditions influence life expectancy.

4.1 Basic Economic Model

The Basic Economic Model explores the relationship between life expectancy and two primary economic indicators: Gross Domestic Product (GDP) and health expenditure as a percent-

age of GDP. The results indicate that both predictors are significantly associated with life expectancy. A higher GDP is generally correlated with increased life expectancy, suggesting that economic prosperity plays a crucial role in improving health outcomes. Similarly, greater health expenditure as a percentage of GDP is positively associated with life expectancy, reinforcing the importance of investment in healthcare services to overall population health. The positive coefficients for both GDP and percentage expenditure confirm that as nations invest more in health relative to their economic output, they tend to achieve better health outcomes.

4.2 Extended Economic Model

Incorporating Adult Mortality into the Basic Economic Model yields the Extended Economic Model, which provides additional insights into the impact of health demographics. The inclusion of Adult Mortality, representing the probability of dying between the ages of 15 and 60, significantly affects the model's interpretation of life expectancy. The results show that higher adult mortality rates are inversely related to life expectancy, highlighting the critical impact of reducing preventable deaths to enhance overall life expectancy. This model underscores the complex interplay between economic factors and more direct health indicators, illustrating how broader socioeconomic conditions and health infrastructure work together to shape public health outcomes.

4.3 Logistic Model

The Logistic Model for Benchmark shifts focus from continuous predictions of life expectancy to the probability of exceeding a benchmark life expectancy of 75 years. This model uses logistic regression to explore how the same economic factors influence the likelihood of achieving higher life expectancy thresholds. The results from this model suggest that both GDP and health expenditure significantly increase the probability of a country achieving a life expectancy above 75 years. This logistic approach provides a useful perspective for policymakers, indicating threshold effects where certain levels of economic health investment significantly boost the chances of surpassing life expectancy benchmarks.

4.4 Overall

Across all models, the data consistently supports the hypothesis that better economic conditions and higher health expenditure are pivotal in improving life expectancy. These findings have profound implications for health policy, emphasizing the need for sustained economic growth and targeted health investments as central strategies for enhancing population health. The clear association between economic indicators and life expectancy also calls for integrated approaches to development and health policy, where economic planning includes explicit considerations of health impacts.

Now we can create a relevant dataset (Appendix A.2).

5 Discussion

5.1 Things we have done in this paper

In this study, we examined the relationship between economic conditions and life expectancy across various countries. Using regression models, we analyzed how GDP and expenditure influence how long people live. The aim was to identify patterns and understand how a country's economic condition might affect the longevity of its citizens.

5.2 Something that we learn about the world

One important finding from our research is the positive link between a nation's economic wealth, indicated by its GDP, and the life expectancy of its people. Countries with higher GDP tend to have populations with longer life spans. This suggests that economic prosperity could lead to better healthcare access, improved nutrition, and overall higher standards of living, all of which contribute to longer lives.

5.3 Another thing that we learn about the world

Our analysis further reveals that not only the size of an economy matters but also how much of the economy is invested in health. Countries that spend a larger proportion of their GDP on healthcare generally achieve better life expectancy figures. This underscores the importance of not just growing an economy but also investing wisely in health infrastructure and services to enhance public health.

5.4 Weaknesses

While our study contributes valuable insights into the relationship between economic metrics and life expectancy, there are several limitations to acknowledge. Primarily, our models did not account for all variables that impact health outcomes, such as environmental quality, public health initiatives, and individual health behaviors, which include diet and exercise. Additionally, using GDP as a proxy for a country's economic health fails to address the complexity of economic disparity within countries. The GDP measure can mask the wealth distribution issues, where a high GDP might not translate into improved health outcomes for all societal segments. This oversight might skew our understanding of how economic prosperity affects public health comprehensively.

Furthermore, the cross-sectional nature of our analysis limits our ability to draw causal inferences. We can identify correlations but cannot definitively say that changes in economic conditions lead to changes in life expectancy without longitudinal data that tracks the same entities over time.

5.5 Next steps

To build on the findings of this study, future research should aim to include additional variables that could influence life expectancy, such as access to clean water, air quality, and rates of non-communicable diseases. Investigating these factors could provide a more nuanced understanding of the interplay between living conditions and public health.

Moreover, incorporating data on income distribution within countries could help researchers understand how wealth is actually used and accessed across different populations. Examining the impact of specific health policies, such as vaccination programs or anti-smoking campaigns, could also shed light on effective strategies for improving health outcomes.

In terms of methodology, conducting longitudinal studies would allow researchers to observe the effects of economic changes over time, providing stronger evidence of causality between economic growth and improvements in life expectancy. Additionally, adopting a multi-country comparative approach could highlight successful policies and practices that could be adapted and applied in different contexts.

Ultimately, advancing this research requires a collaborative approach, engaging economists, public health experts, sociologists, and policymakers. By pooling knowledge and methodologies from various disciplines, we can develop more effective strategies and interventions to enhance life expectancy globally, ensuring that economic gains translate into tangible health benefits for all segments of society.

A Appendix

A.1 Model check

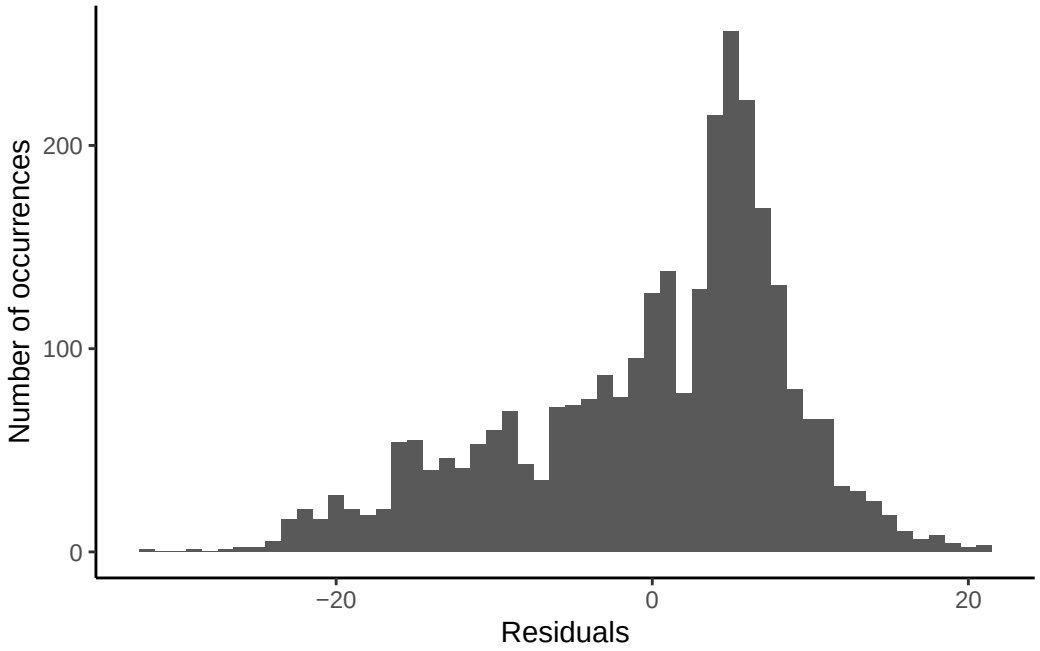

Figure 3: Distribution of residuals

The residual histograms (Figure 3, Figure 4, Figure 5, Figure 6) helps us check if our model's predictions are generally on target. If the model is working well, the residuals—differences between observed and predicted values—should be spread out evenly around the middle of the graph, looking roughly like a bell curve without leaning too much to one side or the other.

The scatterplot with residuals and GDP shows whether our model is consistent for all sizes of economies. We're looking for a random spread of points, indicating that the model works well across the board.

When we compare the actual life expectancy with what the model predicted in the scatterplot, we hope to see the points near a diagonal line. This would mean that the model's predictions are close to reality.

Overall, these plots are quick checks to make sure our model is reasonable and works the same way for countries with different economic conditions.

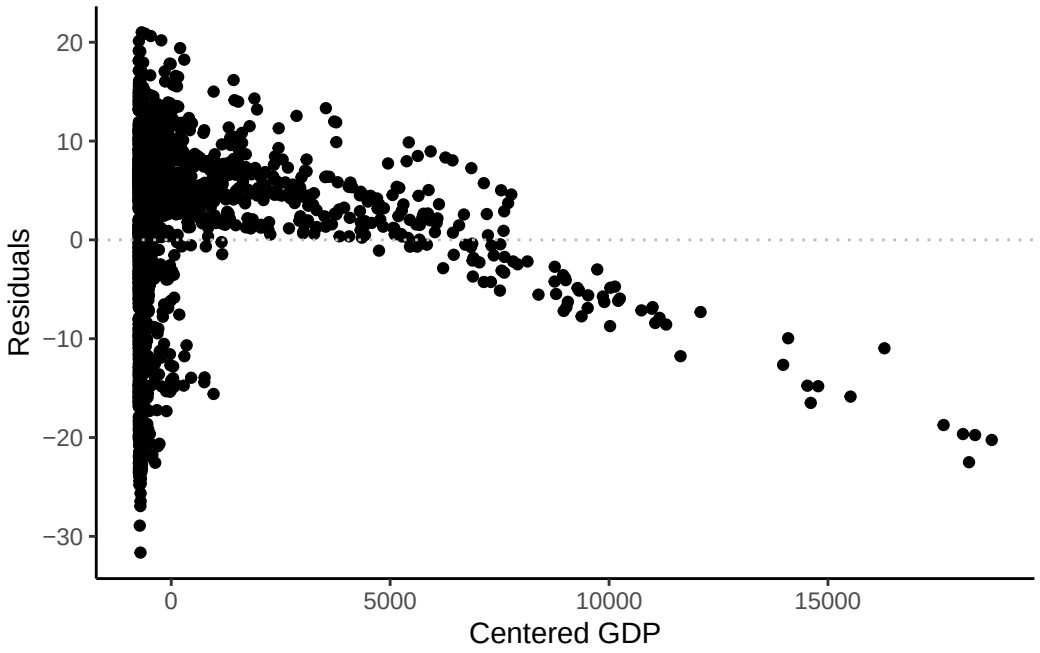

Figure 4: Residuals by GDP

Figure 5: Residuals by life expectancy

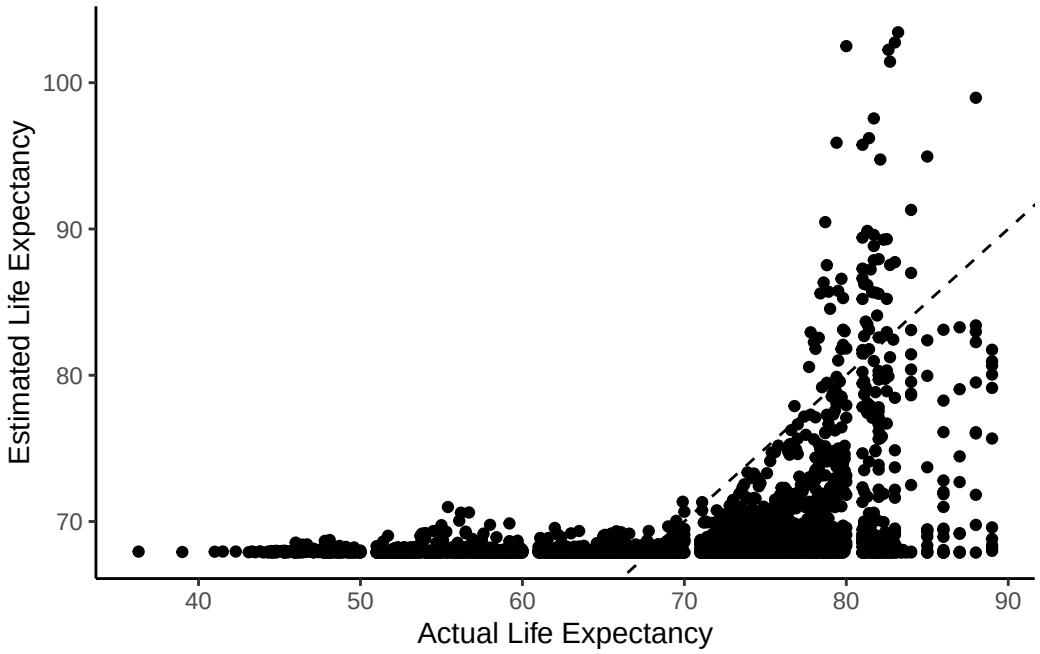

Figure 6: Comparing the fitted value with the actual life expectancy

A.2 Dataset

The newly created dataset provides a focused look at the relationship between life expectancy and key economic indicators.

A.2.1 Dataset Overview

- Country: This column lists the name of each country included in the study, offering a geographical context for the data.
- Year: The year associated with each data entry allows for temporal analysis and helps in understanding trends over time.
- Life Expectancy: The average number of years a person is expected to live based on statistical projections. This is a primary indicator of the health outcomes in a country.
- GDP: This economic measure reflects the economic output per person and provides an indication of the economic prosperity of the population.
- Health Expenditure Percentage: Represents health spending as a percentage of GDP, indicating the priority given to health in economic terms.
- Human Development Index: A composite index that measures average achievements in a country in three basic aspects of human development: health, education, and income. The HDI provides a broader context to the economic capabilities and quality of life.

- Average Schooling Years: The average number of years of schooling received by people ages 25 and older, which correlates with education levels and economic opportunities.

A.2.2 Uses of the Dataset

This dataset is particularly valuable for researchers and policymakers interested in understanding how economic factors influence health outcomes. By analyzing the relationship between GDP, health expenditure, HDI, schooling, and life expectancy, stakeholders can identify patterns, strengths, and areas needing improvement in public health policies.

A.2.3 Analytical Opportunities

The dataset allows for multiple analytical approaches, including: - Trend Analysis: Observing how life expectancy has changed over time in relation to economic growth and investment in health. - Comparative Analysis: Comparing different countries or regions to understand how similar economic conditions might result in different health outcomes. - Regression Analysis: Employing statistical methods to quantify the impact of economic indicators on life expectancy, which can be crucial for effective policy-making.

A.2.4 Potential Insights

From such analyses, insights can be drawn about the effectiveness of health expenditure, the role of education in promoting health, and how overall economic development contributes to improving life expectancy. These insights can further guide targeted interventions and policy adjustments to promote better health outcomes globally.

By focusing on these specific indicators, the dataset not only simplifies the complexity inherent in global health data but also highlights the critical intersections of economics and public health.

References

Abel, Guy J. 2018. "Estimates of Global Bilateral Migration Flows by Gender Between 1960 and 2015." *International Migration Review* 52 (3): 809–52. https://doi.org/10.1111/imre.12327.

Arel-Bundock, Vincent. 2022. "Modelsummary: Data and Model Summaries in r." *Journal of Statistical Software* 103: 1–23.

Kunze, Lars. 2014. "Life Expectancy and Economic Growth." *Journal of Macroeconomics* 39: 54–65.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Robinson, David. 2014. "Broom: An r Package for Converting Statistical Analysis Objects into Tidy Data Frames." *arXiv Preprint arXiv:1412.3565*.

Roser, Max, Esteban Ortiz-Ospina, and Hannah Ritchie. 2013. "Life Expectancy." *Our World in Data*.

Wickham, Hadley, R Francois, L Henry, and K Müller. 2014. "Dplyr." *A Grammar of Data Manipulation 2020 [Last Accessed on 2020 Aug 12] Available from*, Rproject.

Wickham, Hadley, Jim Hester, Romain Francois, Jennifer Bryan, Shelby Bearrows, J Jylänki, and M Jørgensen. 2024. "Package 'Readr'." *Read Rectangular Text Data. Available Online: Https://Cran. R-Project. Org/Web/Packages/Readr/Readr. Pdf (Accessed on 23 August 2023)*.

Wilkinson, Leland. 2011. "Ggplot2: Elegant Graphics for Data Analysis by WICKHAM, h." Oxford University Press.