

The Value of Life: A Macro Investigation of GDP and Life Expectancy Correlation

Abigail Tata* Kenneth Moreno† Joshua Greene‡ Fabiola Cruz§

2024-01-12

Abstract

This paper seeks to investigate the relationship between GDP per capita and life expectancy rates worldwide. We do so by utilizing analytical techniques within R using 2017 data from the World Bank to help establish a statistically significant relationship between the dependent variable (life expectancy) and the independent variables (GDP per capita and healthcare expenditures) - we also include additional variables for control. The results show that GDP per capita and healthcare expenditures are statistically significant predictors of life expectancy.

1 Introduction

One of the best indicators to measure socio-economic stability in a country is life expectancy, it not only helps us understand how countries are effectively providing basic services to citizens, but serves as a clue to the country's overall governance infrastructure which can determine access to health care, jobs, education, crime rates, and much more. In addition, to fund the social services that a society needs to thrive, the country also needs to ensure a robust and thriving economy. Yet, even with a strong economy, disparities in quality of life and access to services can increase. In the United States for example National Academies of Sciences (2015) illustrates how the gap between income and life expectancy has widened over recent decades. Thus, we seek to ask: Does a higher GDP per capita have an impact on life expectancy?

We hypothesize that higher GDP per capita will have a direct effect on life expectancy. We believe that Countries with higher GDPs per capita have a higher life expectancy average compared to countries that have a lower GDP. For example, the quality of healthcare within the United States often relies on an individual's income National Academies of Sciences (2015). Using data and statistical software, we aim to substantiate our claims that GDP per capita and life expectancy are directly correlated. We examined this relationship by pulling data on life expectancy and GDP per capita from the World Development Indicators (2015). When we measured for GDP per capita and life expectancy alone, we found a strong correlation between the two. While this initially proved our hypothesis, after expanding our research, we accounted for other independent factors like Birth rate per 1000 people, current health expenditures (%GDP), inequality index, physicians per 1000 people, safe sanitation usage, current education expenditure, literacy rates, child immunization

*American University

†American University

‡American University

§American University

rates, incidence of HIV, incidence of TB, access to electricity, CO2 emissions, urban population, population density, internet usage, fertility rates, gender equality, poverty gap, poverty headcount ratio, income share, government expenditure on education, population - we then ran a multiple linear regression. The regression showed that both healthcare expenditure and GDP per capita were incredibly significant within the model. A more detailed analysis of the theoretical evidence, the data, our methods, and the results are outlined below.

2 Framing the Conversation

Scholars have sought to understand how the economic status of nations around the world correlates with the health of their citizens since the late 1900s. Low-income nations often experience the highest mortality rates among their population, a phenomenon our research seeks to attribute to economic performance. Ultimately, studies done on the health outcomes of nations are often bogged down by the governance and statistical capacity deficits. Health issues are multidimensional, often incorporating social and economic factors that are hard to measure. However, the research is incredibly important to fully understand wealth and life expectancy disparities in society, in hopes that it leads us towards sustainable solutions. Through a deeper understanding of the ways in which these factors influence each other, we can better understand why some countries remain economically stagnant and why others prosper.

While medicinal innovation has been booming since the late 1940s with the creation of organizations like the World Health Organization (WHO), many countries still lack the means to access lifesaving care. Acemoglu and Johnson (2007) argue that this inaccessibility affects life expectancy rates, leading to an unstable workforce and slow economic growth. Ultimately, the economic strain puts low-income countries out of the fighting ring for health resources. In order to address these issues, these nations routinely spend more time and resources on fertility, growing their population to make up for high mortality rates. However, Simona (2014) argues that this practice has the opposite effect and instead produces a population that the nation's economy routinely fails to support, leading to further economic and health strains. Additionally, the international community has attempted several interventions to address the declining population in low-income countries. On occasion, global organizations like the UN supply temporary resources to developing nations. However, this fails to produce sustainable economic growth. Such unsuccessful efforts towards improving mortality rates and life expectancy is exactly why research into GDP per capita and life expectancy correlation is so vital. A clear, concise route to health improvement and economic prosperity starts with a base understanding of how these two factors continually rely on each other.

Beyond domestic effects, the endurance of this discrepancy can also lead to geopolitical consequences. As argued by National Academies of Sciences (2015), a gap between income and life expectancy globally has major implications for economic sustainability, eliciting fears of a quickly decreasing rate of global production due to a concentration of resources exclusive to the highest-earning nations. In times of global or regional conflict, the material differences between developing and developed countries are exacerbated as well. Whether or not the country has the resources to mitigate the suffering of their citizens in times of political strife is directly related to their economic capability.

Inarguably, the implications of this research are vast. An informed understanding of the economic effects of mortality rates and overall health can lead to a more substantive response, helping ensure future economic stability and the upwards mobility of developing nations. Furthermore, it is important to note that this research does not only seek to draw focus on low-income nations but nations from middle to high-income as well. Measuring the relationship between GDP per capita, healthcare expenditures, and life expectancy has implications that go across the macro-economic

spectrum. What follows is the collection and analysis of hard data from the World Bank, which supplements the conversation of the importance health outcomes have on economic prosperity – fostering what we hope to be a step towards a more sustainable solution to global health inequity.

3 Data

The World Bank aggregates data from various sources on development indicators each year. This data accounts for 266 recognized countries and territories and encompasses many relevant fields of study, including economic development, healthcare, education, and more.

Our dataset was extracted from the World Bank’s World Development Indicators database. We used data from 2017 for all 266 countries and territories, due to more recent years having more sizable gaps in data. The key outcome we accounted for is life expectancy at birth, measured in years. For our research, the key predictor of life expectancy was GDP per capita measured in US dollars in the year of 2015. To enhance our model, we examined several other independent variables.

As we reviewed preliminary figures and statistical tests of our indicators, we found that scatterplots of GDP were significantly skewed due to the wide range of the GDP data points. The way we overcame this issue was by logging two of the original columns Healthcare expenditure and GDP. The logging process created a more normal distribution and allowed us to conduct statistical tests on the relationship between the independent and dependent variables. These two logged variables were then added to the data set to then be displayed in future visualizations.

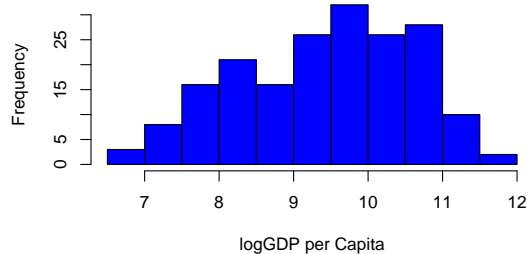
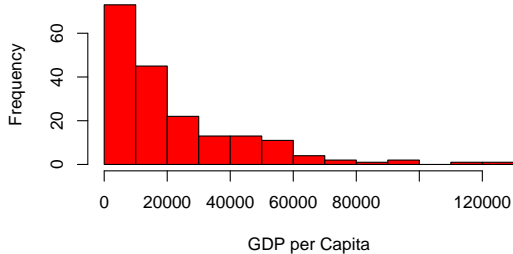


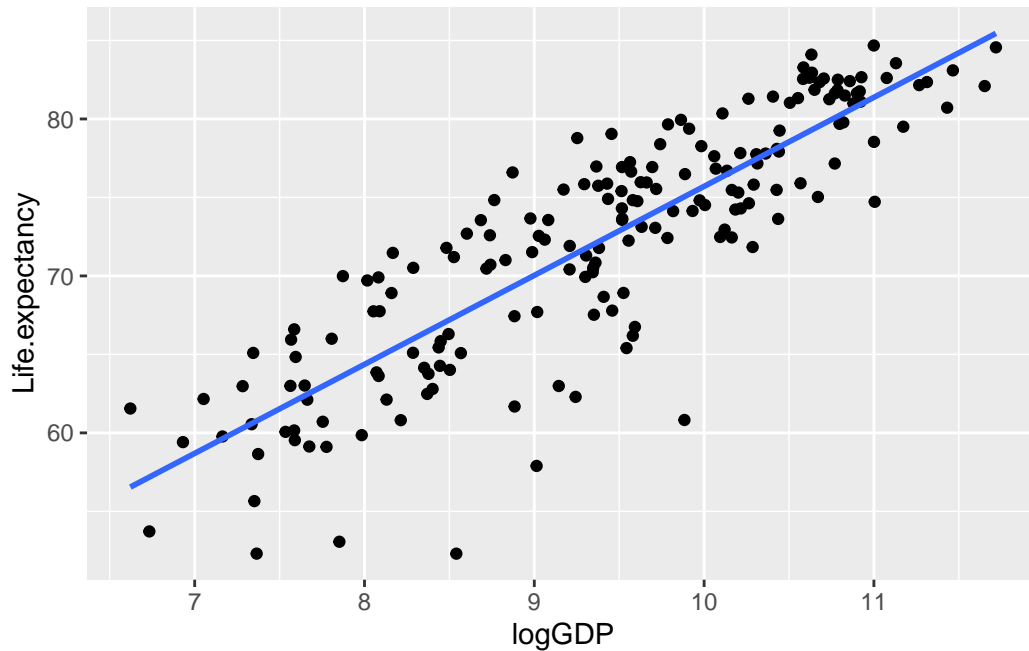
Figure 1: GDP per Capita Distribution Before Log Figure 2: LogGDP per Capita Distribution

The analysis was conducted in 4.3.2 RStudio (2023).

We utilized the following R packages: stargazer, Sleuth3, tidyverse, car, mosaic, ggthemes, gmodels, DescTools, dplyr, ggplot2, tidyr, tidymodels, tidyr, modelr, rpart.plot, and sjPlot. Citations for each can be found at the end of this paper.

4 Results

Our analysis sought to establish a significant correlation between the log of GDP per capita and life expectancy. A scatterplot of the data visually reveals what appears to be a positive correlation between the independent and dependent variables, as well as a clear division between the countries in each World Bank classification.



Below is a table showcasing linear regression.

Table 1: Linear Regression of GDP and Life Expectancy

Predictors	Life expectancy		
	Estimates	CI	p
(Intercept)	19.00	14.41 – 23.59	<0.001
logGDP	5.67	5.18 – 6.16	<0.001
Observations	185		
R ² / R ² adjusted	0.744		
	/		
	0.743		

Linear model testing further reveals a highly significant positive correlation between logGDP and life expectancy ($p \approx 0$). We find that each 1% increase in GDP per capita is correlated with a .0567-year difference in life expectancy. These findings support our hypothesis that GDP per capita has a positive correlation with life expectancy.

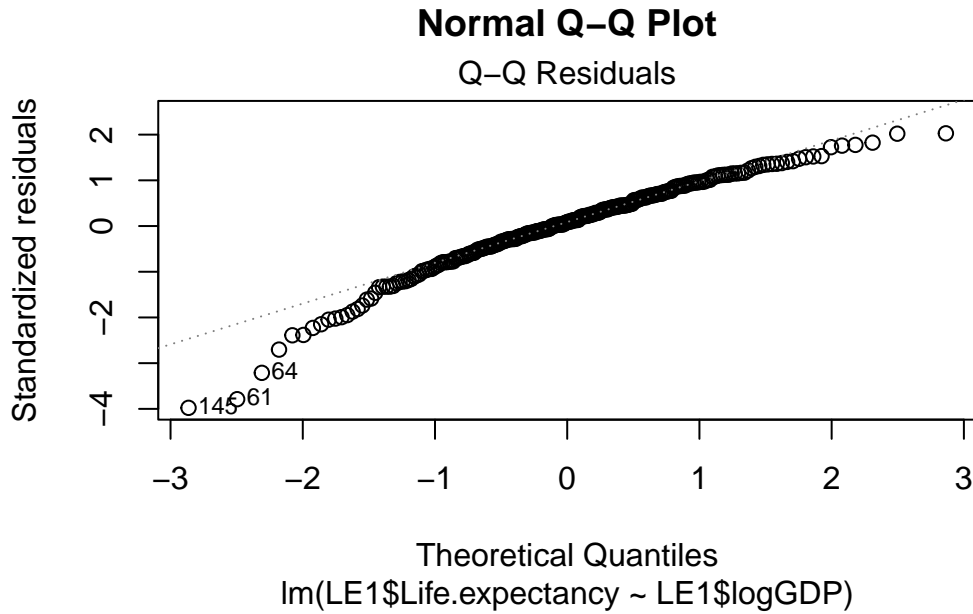
```

      Df Sum Sq Mean Sq F value Pr(>F)
class    3  222.7   74.24   394.8 <2e-16 ***
Residuals 184   34.6    0.19
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
21 observations deleted due to missingness

```

Examining the QQ plot of our data revealed a mostly normal distribution of residuals with a departure from normality in both extremes of the dataset. In a QQ plot, these extraneous points denote a separation between the theoretical distribution of normality and the observed residuals

from the linear model. To address this disparity, we chose to test the data for heteroskedasticity using the estimatr package in R. A comparison between our standard linear model and the robust model showed a slight decrease in standard error while maintaining the same coefficient, which both models labelled as highly significant. Thus, accounting for heteroskedasticity did not have a significant effect on our model, as our data showed robustness to the adjustment.



Below is a table of our Robust Model.

```
lm_robusted <- lm_robust(Life.expectancy ~ logGDP, data = merged_df)
tab_model(lm_robusted, title = "Robust Model")
```

Table 2: Robust Model

	Life.expectancy		
Predictors	Estimates	CI	p
(Intercept)	19.00	14.73 – 23.28	<0.001
logGDP	5.67	5.24 – 6.10	<0.001
Observations	185		
R ² / R ² adjusted	0.744		
	/		
	0.743		

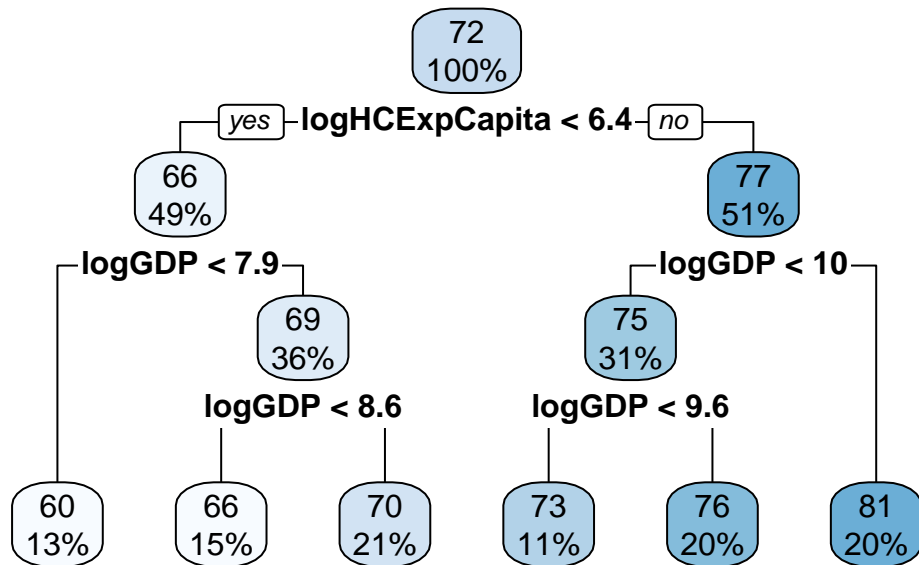
To expand the framework of our model, we applied a multiple linear regression model to our dataset to test the significance of GDP per capita against life expectancy when exposed to the influence of other independent variables. Our added predictors included the log of healthcare expenditure per capita, population, and compulsory education duration (years). These results align with Simpson’s paradox, and therefore suggest that GDP per capita may have been confounded in the multiple linear regression model, or the model was influenced by multicollinearity. We found that GDP per capita and healthcare expenditures were highly significant in predicting life expectancy.

Table 3: Multi-factor Linear Model

Predictors	Estimates	Life.expectancy CI	p
(Intercept)	24.95	17.62 – 32.27	<0.001
logGDP	3.21	1.59 – 4.82	<0.001
logHCExpCapita	2.40	0.98 – 3.83	0.001
Population	0.00	-0.00 – 0.00	0.158
Compulsory education duration	0.12	-0.14 – 0.38	0.358
Observations	163		
R ² / R ² adjusted	0.767 / 0.761		

Our final model of analysis was a decision tree predictive model made with the rpart package. The goal of this model was to predict life expectancy given the logarithm of any GDP per capita and healthcare expenditures. The dataset was split via a training:testing ratio of 70:30, and the model was tasked with determining predictions in the form of linear regression. This model exemplifies the statistical significance of healthcare expenditures and GDP per capita in predicting life expectancy.

For example, if we began at the top of the model with France, whose logHealthcare expenditures is 8.3, translating to just over \$5,071 per capita, we determine that this country does not fall under the 6.4 limit. Thus, we continue to the next branch where we compare logGDP per capita. France has a logGDP of 10.7 or \$44,577 per capita. This puts them over 10 and thus predicts the life expectancy to be 81. Following the same method for other countries produces the same outcome: a prediction on life expectancy based GDP and healthcare expenditures per capita.



5 Future Implications

In undertaking this topic, we aimed to find and measure a correlation between GDP per capita and life expectancy around the world. Understanding how these two factors influence each other is the keystone to tackling significant global issues, like poverty and high mortality rates. We found that measured on their own, GDP per capita and life expectancy had a strong correlation: the higher the GDP per capita, the higher the life expectancy. We had predicted this outcome early on, attributing the relationship to a country's access or lack of access to life-sustaining resources and quality of life.

Throughout this research, we developed a working theory as to what all this research means. Countries with low GDP per capita lacked the resources to sustain their population, thus leading to high mortality rates. As Simona (2014) argued, high mortality rates incentivized citizens to focus on fertility to “make up” for the deaths. This, however, adds strain to an economy that is already unable to support its current population. Thus, more of the population is left economically stratified in poor living conditions, heightening the risk of death to preventable illness or disease. An unsustainable population means an unsustainable workforce, leading to the limited production, innovation, and a lack of participation in the national economy, culminating in a low GDP per capita. This cycle is doomed to continue to repeat itself.

This research was limited in some ways. We did not account for a country's socio-cultural issues, for example levels of inequality among a country's population. As stated before, many factors that contribute to a country's life expectancy are notoriously multidimensional and are difficult to measure. Ultimately, our topic focused on the factors we could accurately attest for – GDP per capita and healthcare expenditures – and we discovered that GDP per capita and healthcare expenditures were both statistically significant.

Undoubtedly, this topic is broad and nuanced. In future replications of this research, one should be cognizant of the historical and geopolitical factors that contribute to a country's life expectancy or GDP per capita. While we found relationships that are significant and important, a holistic approach to this topic has the possibility of producing a meaningful report of the issues plaguing low-income countries. It also has the potential to be incredibly beneficial in informing future international economic policy and the mitigation of risk for high-income countries. While our research was limited in scope, we hope it contributes to a broader conversation on health and overall socio-economic equity.

References

- Acemoglu, Daron, and Simon Johnson. 2007. "Disease and Development: The Effect of Life Expectancy on Economic Growth." *Journal of Political Economy* 115 (6): 925–85. <https://doi.org/10.1086/529000>.
- Arnold, Jeffrey B. 2023. *Ggthemes: Extra Themes, Scales and Geoms for 'Ggplot2'*. <https://CRAN.R-project.org/package=ggthemes>.
- F. L. Ramsey, Original by, D. W. Schafer; modifications by Daniel W. Schafer, Jeannie Sifneos, Berwin A. Turlach; vignettes contributed by Nicholas Horton, Linda Loi, Kate Aloisio, Ruobing Zhang, and with corrections by Randall Pruim. 2019. *Sleuth3: Data Sets from Ramsey and Schafer's "Statistical Sleuth (3rd Ed)"*. <https://CRAN.R-project.org/package=Sleuth3>.
- Fox, John, and Sanford Weisberg. 2019. *An R Companion to Applied Regression*. Third. Thousand Oaks CA: Sage. <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>.
- Hlavac, Marek. 2022. *Stargazer: Well-Formatted Regression and Summary Statistics Tables*. Bratislava, Slovakia: Social Policy Institute. <https://CRAN.R-project.org/package=stargazer>.
- Kuhn, Max, and Hadley Wickham. 2020. *Tidymodels: A Collection of Packages for Modeling and Machine Learning Using Tidyverse Principles*. <https://www.tidymodels.org>.
- Lüdecke, Daniel. 2023. *sjPlot: Data Visualization for Statistics in Social Science*. <https://CRAN.R-project.org/package=sjPlot>.
- Milborrow, Stephen. 2022. *Rpart.plot: Plot 'Rpart' Models: An Enhanced Version of 'Plot.rpart'*. <https://CRAN.R-project.org/package=rpart.plot>.
- National Academies of Sciences, and Medicine (U.S.), Engineering, ed. 2015. *The Growing Gap in Life Expectancy by Income: Implications for Federal Programs and Policy Responses*. Washington, D.C: National Academies Press.
- Posit team. 2023. *RStudio*. Boston, MA: Posit Software, PBC. <http://www.posit.co/>.
- Pruim, Randall, Daniel T Kaplan, and Nicholas J Horton. 2017. "The Mosaic Package: Helping Students to 'Think with Data' Using r." *The R Journal* 9 (1): 77–102. <https://journal.r-project.org/archive/2017/RJ-2017-024/index.html>.
- Signorell, Andri. 2023. *DescTools: Tools for Descriptive Statistics*. <https://CRAN.R-project.org/package=DescTools>.
- Simona, Dogan Mihaela. 2014. "The Relationship Between Life Expectancy and GDP." *Anale. Seria Stiinta Economice. Timisoara* 20: 93–97. <https://www.proquest.com/conference-papers-proceedings/relationship-between-life-expectancy-gdp/docview/1626834093/session2?accountid=8285>.
- Warnes, Gregory R., Ben Bolker, Thomas Lumley, Randall C Johnson. Contributions from Randall C. Johnson are Copyright SAIC-Frederick, Inc. Funded by the Intramural Research Program, of the NIH, National Cancer Institute, and Center for Cancer Research under NCI Contract NO1-CO-12400. 2022. *Gmodels: Various r Programming Tools for Model Fitting*. <https://CRAN.R-project.org/package=gmodels>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- . 2023. *Modelr: Modelling Functions That Work with the Pipe*. <https://CRAN.R-project.org/package=modelr>.
- Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.
- Wickham, Hadley, Davis Vaughan, and Maximilian Girlich. 2023. *Tidyr: Tidy Messy Data*. <https://CRAN.R-project.org/package=tidyr>.
- World Bank. 2015. "World Development Indicators Data Bank." <https://databank.worldbank.org/source/world-development-indicators>.