Economic, Social, and Health Determinants of Life Expectancy: A Cross-Country Analysis of Developing and Developed Nations

Motivation

Life expectancy varies dramatically worldwide, ranging, shaped by complex economic, social, and health factors. Prior research shows that:

- Higher income generally leads to a higher life expectancy (Soares, 2007).
- Education plays a causal role, with each additional year of schooling increasing life expectancy (Llieras-Muney, 2006).

Understanding these relations can guide policy interventions in developing nations.

Research Question

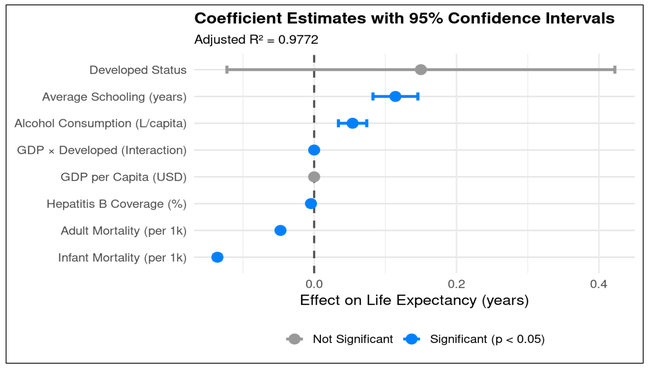
How do economic, social, and health-related factors influence life expectancy across countries, and do these effects differ between developing and developed nations?

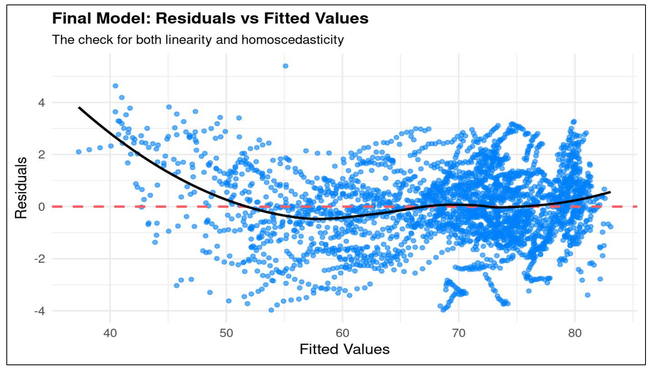
Data Collection

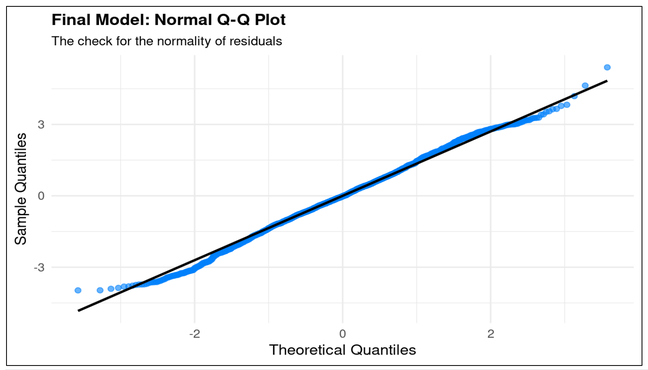
The dataset was obtained from Kaggle, originally compiled from the Global Health Observatory (GHO) under the World Health Organization (WHO). The WHO collects this data to monitor global health trends and track progress towards international health goals, making it a trustworthy source (Rajarshi, 2018). Data gaps were filled using World Bank development data, applying reasonable methods such as averaging from the last three years (Gochiashvili, 2022).

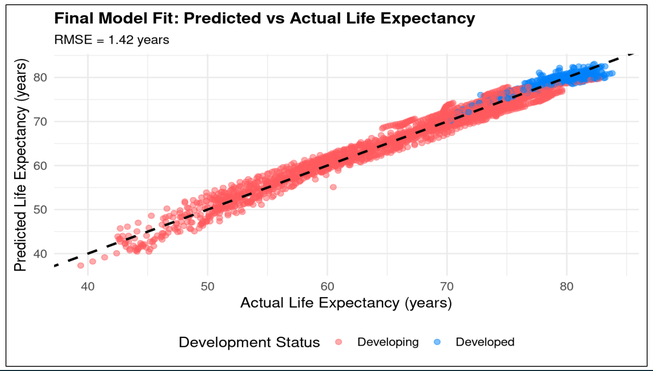
The final dataset contained 2,864 observations, where each is a unique country-year pair. As each row corresponds to a country in a specific year, treating observations as independent is reasonable. The variables (life expectancy, GDP per capital, mortality rates, etc.) were continuous and appropriately scaled, making the dataset well-suited for analysis via multiple linear regression.

Results

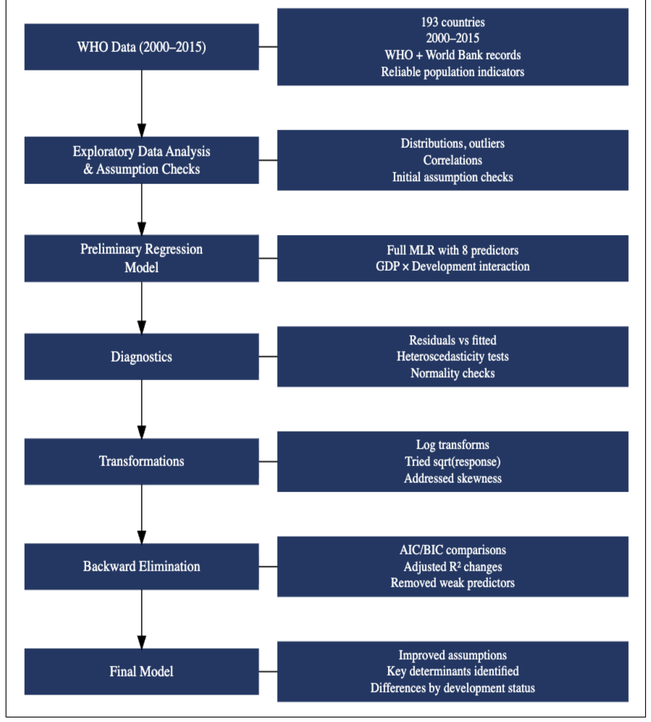








Methods of Analysis



Conclusion

Key Findings: Effects on Life Expectancy

Predictor	Effect Size	Unit	Note
Infant Mortality	−0.14 years	per death/1,000 births	Strongest negative effect
Adult Mortality	−0.05 years	per death/1,000 adults	Significant mortality impact
Average Schooling	+0.11 years	per additional year	Education benefits
GDP (Developing Countries)	+0.007 years	per $1,000 increase	Modest GDP effect
GDP (Developed Countries)	+0.030 years[1]	per $1,000 increase	4× stronger than developing

[1] Developed countries show 4× larger GDP effect than developing countries

Answer to Research Question

Health and education factors have the strongest direct effects on life expectancy. Economic effects (GDP) differ significantly between developing and developed countries, with developed countries showing an almost 400% greater return with GDP increases, suggesting diminishing health returns once basic societal infrastructure has been established.

Findings align with prior research of higher income leading to higher life expectancy (Soares, 2007) as well as education playing a causal role in increasing life expectancy (Llieras-Muney, 2006).

Implications

For developing countries, prioritizing health interventions to reduce infant and adult mortality as well as expanding education will yield greater life expectancy gains rather than solely focusing on economic growth.

References

Gochiashvili, L. (2022). *Life Expectancy (WHO) Fixed.* Kaggle. https://www.kaggle.com/datasets/lashagoch/life-expectancy-who-updated/data

Lleras-Muney, A (2006, July). *The Relationship between Education and Adult Mortality in the United States*. ResearchGate. https://www.researchgate.net/publication/228209433_The_Relationship_between_Education_and_Adult_Mortality_in_the_United_States_Adriana_Lleras-MuneyErratum

Rajarshi, K. (2018, February 10). *Life expectancy (WHO)*. Kaggle. https://www.kaggle.com/datasets/kumarajarshi/life-expectancy-who

Soares, R. R. (2007, May 29). *On the Determinants of Mortality Reductions in the Developing World.* Wiley Online Library. https://onlinelibrary.wiley.com/doi/10.1111/j.1728-4457.2007.00169.x

World Health Organization (2020, December). *WHO methods and data sources for life tables 1990-2019*. Department of Data and Analytics: Division of Data, Analytics and Delivery for Impact. WHO, Geneva. https://cdn.who.int/media/docs/default-source/gho-documents/global-health-estimates/ghe2019_life-table-methods.pdf?sfvrsn=c433c229_5

Limitations

Some heteroscedasticity (i.e., uneven variance) remained even after various transformations, suggesting the model may underestimate uncertainty for countries with either high life expectancy or high GDP. Using more granular data (e.g., multiple measurements per year) as opposed to the current country-to-year aggregation could help build a better model with better predictions, reliability, and interpretability.

Results

- From preliminary model, attempted log transformations on GDP, population, infant mortality, and adult mortality.
  - These transformations made heteroscedasticity slightly worse.
  - Indicates this violation might be caused by the response or a different pattern in variance
- Box-Cox recommended an alpha=0.5, suggesting a square root transformation on the response variable
  - Original predictors with original response: p-value $8.72938e^{-42}$
  - Log predictors with original response: p-value $6.508188e^{-57}$
  - Original predictors with root of response: p-value $1.605268e^{-51}$
  - Log predictors + root of response: p-value $1.220236e^{-57}$
  - Suggests heteroscedasticity is stubborn
- Tried combined approach, but since the p-value for the original predictors with original response gave the largest p-value, we shall stick with it
- Next, backward elimination using adjusted R^2 fdsfds
  - Step 1: removed PopulationMillions predictor as adjusted R^2 improved from 0.97716 to 0.97717
  - Step 2: removed Thinness10to19Percentage as adjusted R^2 stayed the same
  - Thus, our new model contains just 8 predictors
  - Final adjusted R2 is 0.97717 (i.e., 97.717% of variance is explained by the remaining predictors)
- Final model then has just 8 predictors (2 less predictors than the original) with no loss in adjusted R^2
  - $R^2 = 0.97724$
  - Adjusted $R^2 = 0.97717$
  - RMSE = 1.419 years
  - Normality is acceptable: minor tail deviations
  - Heteroscedasticity is still present (BP p-value < 0.05) —tried various transformations with no improvement
  - Moderate multicollinearity: VIF 5-8 on some predictors—acceptable since not severe
  - 180 influential points (similar to the preliminary model

Methods of Analysis

A multistep modelling workflow began with data cleaning and exploratory data analysis, followed by fitting in an initial multiple linear regression model. Diagnostic checks were performed to check for linearity, homoscedasticity, normality, and influential points, then various transformations were applied to address those violations. Model selection was performed using a backward elimination process guided by AIC/BIC as well as adjusted R^2. A final model was determined that satisfies assumptions and provides interpretable effects of the economic, social, and health factors on life expectancy.