Life Expectancy Data Analysis

Ndahimana Raphael, Msc Big data analytics, ID:101228

Abstract

This study analyzes global life expectancy using WHO and UN datasets to understand key drivers and disparities in population health. Using exploratory data analysis and multiple regression, it finds strong associations between life expectancy and factors like education, income, and adult mortality. Results show that regions like Europe and the Americas lead in longevity, while Africa lags. Findings emphasize the role of socio-economic development, access to healthcare, and education in improving life expectancy.

1. Introduction

Life expectancy reflects the average number of years a person is expected to live from birth, based on current mortality rates. It is a core indicator of public health and socio-economic development. Although global life expectancy has improved due to medical and social advancements, disparities persist. High-income countries report averages above 75 years, while many low-income countries, especially in sub-Saharan Africa, remain below 65. This study investigates factors influencing these differences and aims to provide insight into global life expectancy trends using WHO and UN data.

2. Methods

Data was sourced from the WHO Global Health Observatory and UN economic databases, focusing on 193 countries. Selected variables included population, life expectancy, adult mortality, measles incidence, immunization coverage, and education. Analyses were conducted in R, including data cleaning, exploratory data analysis (EDA), and multiple regression modeling. RMSE and adjusted R-squared were used to assess model performance. Only predictors with p-values less than 0.05 were considered significant.

3. Results

The life expectancy dataset had 22 variables and 2,938 observations. No duplicates were found, but missing values occurred in variables like Hepatitis B, GDP, and population over 15. Outliers in population were addressed using the median. China, India, and the US had the largest populations; Japan had the smallest among the top ten. Fertility rates declined more slowly in Nigeria and Pakistan.

Correlation results revealed that life expectancy was positively correlated with schooling ($r = 0.63$) and negatively with fertility rate ($r = -0.64$), both statistically significant ($p < 0.05$). Multiple regression identified several significant predictors:

- Developing country status: $\beta = -1.13$, $p < 0.05$
- Adult mortality: $\beta = -21.35$, $p < 0.05$
- Income composition: $\beta = 22.33$, $p < 0.05$
- Measles cases: $\beta = 1.75$, $p < 0.05$
- Polio immunization: $\beta = 6.23$, $p < 0.05$
- Total expenditure: $\beta = 9.17$, $p < 0.05$
- Schooling: $\beta = 2.18$, $p < 0.05$

Regional analysis in 2015 showed Europe led with an average of 80 years, followed by the Americas, with Africa lowest at 65.2. Within Africa, SADC (65.5 years) had higher life expectancy than EAC (61.4 years).

4. Discussion

This study confirmed that life expectancy is driven by intertwined social, economic, and health-related factors. High adult mortality rates remain a major barrier, especially in low-income regions. Education and income were found to be strong positive predictors, suggesting that investment in human capital directly impacts longevity.

Notably, developing country status was associated with lower life expectancy despite potential medical advancements, indicating persistent systemic disparities. Immunization coverage and healthcare spending also positively impacted longevity, highlighting the importance of preventive care.

Regional disparities remained stark, with Africa lagging significantly. Within Africa, the gap between SADC and EAC emphasizes the role of regional policies and infrastructure in health outcomes. The study supports the call for integrated, multisectoral efforts to address these gaps.

5. Lessons Learned

1. Life expectancy is multifactorial—health, education, and economic conditions all contribute.
2. Regional inequities remain significant, with Africa disproportionately affected.
3. Reliable data quality is crucial for valid analysis.
4. Education and income strongly predict higher life expectancy.
5. High adult mortality has a direct negative impact on longevity.

6. Recommendations

- Invest in healthcare infrastructure in low-income regions.
- Promote access to education and health literacy.
- Target adult mortality with preventive and treatment interventions.
- Strengthen national health data systems.
- Design region-specific health and development strategies.

7. Conclusion

Life expectancy is a critical metric for assessing population well-being. This study highlights that longevity is influenced by adult mortality, economic resources, education, and healthcare investments. Despite global improvements, inequalities persist. Comprehensive policy approaches focusing on education, income equity, healthcare access, and data quality are essential to improving global life expectancy outcomes.