How Long Will I Live? A Linear Mixed Effects Analysis of Factors Influencing Life Expectancy in Developing Countries

Amos O. Okutse Monica Colon-Vargas

Introduction

Life expectancy is a key indicator of human health and development and a major determinant of the quality and quantity of life. However, life expectancy varies widely across and within countries, reflecting the complex interplay of biological, environmental, social, and behavioral factors influencing mortality and morbidity. Understanding the factors that affect life expectancy and how they differ across countries is crucial for designing and evaluating effective and equitable public health policies and interventions that can improve health outcomes and reduce health disparities.

Studying factors associated with life expectancy in different countries is not a new topic. This is an active and evolving research area as new data, methods, and challenges emerge. For example, the COVID-19 pandemic highlighted the importance and limitations of life expectancy as a measure of health and the need to account for the direct and indirect effects of pandemics on mortality and morbidity. Moreover, the availability and quality of data on life expectancy and its determinants vary across countries and require careful harmonization and adjustment to ensure comparability and validity. The methods and models used to analyze life expectancy and its factors need to be robust and flexible and able to handle the heterogeneity, nonlinearity, and multicollinearity of the data.

This study aims to contribute to the existing literature on life expectancy and its associated determinants by using novel and rigorous methods and data to address some of the current gaps and challenges in the field. Specifically, we employ linear mixed-effects methods (LMM) to model determinants of life expectancy in the developing world. We conduct sensitivity analyses and compare the performance of the fitted LMM to a fixed-effect linear model based on ordinary least squares estimation.

Methods

Data and variables

- Source of data and basic characteristics
- Dependent and explanatory variables (Table listing)
- What informed choice of explanatory variables?

Statistical modeling

Linear and linear mixed effects model (LMM)

Denote $\mathbf{Y_i} = (y_{m_i,i}, \cdots, y_{m_i,i})^T$ be an m_i-dimensional vector of repeated responses for the i^{th} (independent) individual, $\mathbf{X_i}$ and $\mathbf{Z_i}$ be $m_i \times p$ and $m_i \times q$-dimensional vector of covariates, respectively, and be a p-dimensional vector of fixed effects. The standard LMM is then (Ariyo et al. 2022)

$$
\begin{aligned}
&\mathbf{Y_i} = \beta \mathbf{X_i} + \mathbf{b}_i \mathbf{Z_i} + \epsilon_i,\ (i = 1, \cdots, n) \\
&\epsilon_i \sim N_{m_i}(0, \Sigma_i)
\end{aligned}
\tag{1}
$$

where ϵ_i is the residual vector and Σ_i is $m_i \times m_i$ positive definite (p.d) covariance matrix.

The $q \times 1$ vector of random effects, $\mathbf{b_i} \sim N_q(0, \mathbf{D})$ where $\mathbf{D}$ denotes the $q \times q$ p.d covariance matrix. The classical LMM represented by Equation (1) combines fixed effects β with the subject specific random effects $\mathbf{b_i}, \cdots, \mathbf{b_n}$ and inference might be focused on the regression coefficients, β, the subject-specific coefficients, $\mathbf{b_i}$, or the variance-covariance structure, Σ_i and $\mathbf{D}$. This is in contrast to a linear model defined as in Equation (2) that assumes the subject-specific coefficients $\mathbf{b_i}$.

$$
\begin{aligned}
&\mathbf{Y_i} = \beta \mathbf{X_i} + \epsilon_i,\ (i = 1, \cdots, n) \\
&\epsilon \sim N_{m_i}(0, \Sigma_i)
\end{aligned}
\tag{2}
$$

Model specifications

- See Roffia et al for specification of the OLS model and justification for its use!
- See Ariyo 2019 for LME version and Lin

What are the specific forms of the fitted models? Why?

Sensitivity analysis

Model selection

Results and discussion

References

Ariyo, Oludare, Emmanuel Lesaffre, Geert Verbeke, and Adrian Quintero. 2022. "Model Selection for Bayesian Linear Mixed Models with Longitudinal Data: Sensitivity to the Choice of Priors." *Communications in Statistics-Simulation and Computation* 51 (4): 1591–1615.