QUBAHAN ECOLEAD JOURNAL
VOL. 5, NO. 4, November 2026
https://doi.org/10.48161/qelj.v5n4a61

Regression-Based Identification of Key Determinants in Multivariate Data: Evidence from Global Life Expectancy Indicators

Yogeesh N[1*]

[1] Mathematics and Natural Sciences, Gulf University for Science & Technology, Hawalli, Kuwait;

*** Corresponding author:** yogeesh.r@gmail.com.

ABSTRACT: Multivariate datasets are increasingly used to support evidence-based decision-making, yet their practical value depends on whether researchers can identify which variables most strongly explain an outcome after competing predictors are considered simultaneously. This study develops and applies a regression-based statistical analysis framework for identifying key factors in multivariate data. The empirical case uses an updated global life-expectancy dataset covering 2,864 country-year observations, 179 countries, 9 regions, and 21 variables from 2000 to 2015. Life expectancy was modeled as the dependent variable, while mortality, immunization, disease-burden, economic, demographic, schooling, body-mass, regional, and development-status indicators were used as explanatory variables. The analytical procedure combined descriptive statistics, correlation analysis, least absolute shrinkage and selection operator (LASSO)-assisted variable selection, heteroscedasticity-robust multiple regression, multicollinearity assessment, residual diagnostics, and train-test plus 10-fold cross-validation. The strongest bivariate associations with life expectancy were adult mortality, under-five deaths, infant deaths, schooling, polio immunization, diphtheria immunization, body mass index (BMI), gross domestic product (GDP) per capita, and human immunodeficiency virus (HIV) incidence. The final robust regression model retained mortality, development status, schooling, GDP per capita, BMI, HIV incidence, and regional indicators as key explanatory factors. The validation results showed strong explanatory and predictive performance, with a test coefficient of determination (R^2) of 0.982, root mean squared error (RMSE) of 1.232, mean absolute error (MAE) of 0.981, and a 10-fold cross-validated R^2 of 0.983. The findings demonstrate that regression-based analysis remains a transparent and reliable method for identifying interpretable determinants in multivariate datasets when combined with variable selection, diagnostics, and validation. The study contributes a reproducible empirical framework for key-factor detection in applied statistical research.

Keywords: multiple regression; multivariate analysis; variable selection; diagnostic testing; predictive validity.

I. INTRODUCTION

Multivariate data analysis is central to modern empirical research because social, health, economic, educational, and business outcomes are rarely shaped by a single factor. Researchers are frequently required to evaluate how several predictors operate together and how the influence of one factor changes once other variables are controlled. In this setting, regression analysis remains one of the most transparent and interpretable statistical tools. Unlike purely black-box prediction, regression models can estimate effect direction, statistical significance, uncertainty, and model fit in a form that is understandable to researchers, policy analysts, and decision-makers.

The problem addressed in this study is not only prediction but key-factor identification. Many empirical datasets contain numerous candidate predictors, and simple correlation may exaggerate the importance of variables that are actually proxies for broader social or economic conditions. A predictor may appear strongly related to an outcome in isolation but lose importance after controlling for mortality, economic development, schooling, disease burden, or regional differences. Therefore, regression-based modeling requires an integrated workflow that includes exploratory analysis, model estimation, variable selection, diagnostic testing, and validation.

This study applies that workflow to global life expectancy indicators. Life expectancy is a useful empirical outcome because it reflects cumulative population-level conditions, including mortality, infectious-disease burden, health-service access, educational attainment, and economic capacity. The selected dataset contains country-year observations from 2000 to 2015 and includes variables that are appropriate for multivariate modeling. The empirical aim is to identify which factors remain most influential after considering several related predictors together.

II. LITERATURE REVIEW

Cross-country evidence from the MENA region shows that healthcare expenditure-outcome relationships vary by regional context and efficiency, supporting the inclusion of regional indicators in life-expectancy modeling [1]. Similar evidence from OECD Asia/Pacific countries indicates that health expenditure and socioeconomic variables help explain differences in longevity across countries [2]. Transparent-reporting guidance for multivariable prediction models also emphasizes clear predictor definition, model-development procedures, validation, and performance reporting [3].

Global health-access analysis shows that publicly funded healthcare and social determinants jointly contribute to differences in life expectancy, which supports the use of health and socioeconomic indicators together [4]. Applied regression methodology further stresses that interpretation, uncertainty, and model checking should be integrated rather than presented as isolated numerical results [5]. Time-series evidence from Turkey demonstrates that economic development and life expectancy are connected over time, justifying the inclusion of income-related variables [6].

Regression modeling strategy literature frames model building as a complete process of specification, validation, and interpretation rather than a mechanical coefficient-estimation task [7]. Sparsity-oriented statistical learning supports LASSO-type selection when candidate predictors are numerous or conceptually overlapping [8]. Variable-selection research cautions that automated selection should be paired with substantive judgment and validation because correlated predictors may produce unstable models [9].

Cross-country and time-series evidence on health expenditure and life expectancy confirms the usefulness of regression-based analysis for macro-health determinants [10]. Statistical learning guidance explains how cross-validation, shrinkage, and model assessment help balance interpretability and predictive reliability [11]. Methodological work on multicollinearity shows that correlated predictors can inflate standard errors and weaken confidence in adjusted effects when diagnostic checks are ignored [12].

Predictive modeling guidance positions preprocessing, resampling, tuning, and performance assessment as necessary steps before substantive interpretation [13]. Panel evidence from OECD countries shows that public and private health expenditure may affect life expectancy differently across national spending contexts [14]. Evidence from EU accession candidate countries shows that socioeconomic development, GDP, and infant mortality are closely associated with life expectancy [15].

Evidence from new EU member states indicates that educational attainment and economic freedom are associated with life expectancy, emphasizing the importance of human-capital indicators [16]. The TRIPOD reporting statement supports transparent documentation of multivariable prediction-model development, validation, and performance evaluation [17]. Policy-oriented well-being analytics further shows that life-expectancy modeling can support cross-national decision-making when indicators are interpreted carefully [18].

Regional panel evidence from SAARC-ASEAN countries shows that health expenditure, sanitation, and GDP should be modeled jointly when explaining health outcomes [19]. Multi-country analysis of universal

health coverage links health-service access with life expectancy and healthy life expectancy [20]. Dynamic global panel analysis shows that health expenditure, longevity, and child mortality are interdependent, reinforcing the need for multivariate regression designs [21].

Sample-size guidance for prediction-model development emphasizes that model complexity should match the available data and validation strategy [22]. Longitudinal OECD evidence on life expectancy at birth demonstrates the usefulness of multivariable regression and selection procedures for identifying durable determinants [23]. Validation methodology shows that internal, internal-external, and external validation are necessary when regression models are intended to support applied decisions [24].

Long-run OECD evidence demonstrates the close relationship between health and GDP over extended periods, strengthening the justification for economic predictors [25]. Diagnostic guidance on variance inflation factors supports systematic multicollinearity screening before interpreting adjusted coefficients [26]. Calibration research cautions that predictive models should not be judged only by fit or discrimination because model reliability depends on broader validation evidence [27].

Epidemiologic regression methodology confirms that multicollinearity can distort coefficient interpretation, which justifies diagnostic testing within the present framework [28]. Systematic evidence on sociodemographic, macroeconomic, and health-resource determinants shows that life expectancy is repeatedly associated with multiple predictor groups rather than a single domain [29]. Robust covariance-estimation methodology supports using heteroscedasticity-consistent standard errors when regression assumptions are violated [30].

Taken together, the reviewed evidence supports the present framework by connecting substantive determinants of life expectancy with regression-based modeling, variable selection, diagnostics, robust inference, and validation.

Table 1. Literature review synthesis of regression-based and determinant studies.

Ref.	Study	Focus / method	Relevance to current study
[1]	Balkhi et al. (2021)	Cross-country analysis of healthcare expenditures and health outcomes in the MENA region and comparison countries.	Shows that spending-outcome associations differ by regional context and efficiency.
[2]	Blazquez-Fernandez et al. (2017)	Analysis of health expenditure and socioeconomic determinants of life expectancy in OECD Asia/Pacific countries.	Supports use of regional and socioeconomic control variables.
[3]	Collins et al. (2015)	Detailed reporting guidance for multivariable prediction-model studies.	Strengthens reporting standards for empirical regression studies.
[4]	Galvani-Townsend et al. (2022)	Global analysis of publicly funded healthcare, access, and social determinants of health.	Shows that healthcare access and social determinants jointly predict longevity.
[5]	Gelman et al. (2020)	Applied regression, uncertainty, model checking, and interpretation for real data problems.	Reinforces interpretation, uncertainty, and model checking as core regression practices.
[6]	Halicioglu (2011)	Time-series regression evidence on life expectancy and economic development.	Shows that economic development can be modeled as a determinant of life expectancy.
[7]	Harrell (2015)	Regression modeling strategies for linear, logistic, ordinal, and survival models.	Supports treating regression as a full modeling strategy with specification, validation, and interpretation.
[8]	Hastie et al. (2015)	Sparse statistical learning and LASSO-based regularization.	Justifies penalized selection for reducing complexity in multivariate datasets.
[9]	Heinze et al. (2018)	Review and recommendations for variable selection in applied statistics.	Guides cautious use of automated and penalized selection.
[10]	Jaba et al. (2014)	Cross-country and time-series analysis of health expenditure and life expectancy.	Demonstrates the value of regression for macro-health determinants.

Ref.	Study	Focus / method	Relevance to current study
[11]	James et al. (2021)	Statistical learning methods, cross-validation, shrinkage, and model assessment.	Supports balancing interpretability and predictive reliability.
[12]	Kim (2019)	Practical explanation of multicollinearity and misleading regression results.	Supports diagnostic interpretation of coefficient instability.
[13]	Kuhn and Johnson (2013)	Applied predictive modeling with preprocessing, resampling, and validation.	Supports workflow-based modeling and validation before interpretation.
[14]	Linden and Ray (2017)	Public and private health expenditure and life expectancy in OECD countries.	Shows that expenditure effects can vary by national spending context.
[15]	Miladinov (2020)	Socioeconomic development and life expectancy in EU accession candidate countries.	Highlights GDP and infant mortality as core explanatory indicators.
[16]	Moga Rogoz et al. (2022)	Economic freedom, education, and life expectancy in new EU member states.	Confirms education as a durable longevity-related factor.
[17]	Moons et al. (2015)	Transparent reporting statement for multivariable prediction models.	Supports transparent reporting of predictors, model building, and validation.
[18]	Murtin et al. (2022)	Policy-oriented well-being analytics and life-expectancy modeling.	Positions regression-based analytics as a tool for policy interpretation.
[19]	Rahman et al. (2018)	Panel evidence on health expenditure, sanitation, GDP, and health outcomes.	Shows that spending, sanitation, and GDP require joint modeling.
[20]	Ranabhat et al. (2018)	Multi-country study of universal health coverage, life expectancy, and healthy life expectancy.	Links health-service coverage to longevity outcomes.
[21]	Ray and Linden (2020)	Dynamic panel analysis of health expenditure, longevity, and child mortality.	Supports examining mortality and expenditure jointly.
[22]	Riley et al. (2020)	Sample-size considerations for developing multivariable prediction models.	Supports adequacy checks for model complexity relative to sample size.
[23]	Roffia et al. (2022)	Longitudinal OECD study of determinants of life expectancy at birth.	Uses multiple regression and selection procedures close to this study.
[24]	Steyerberg and Harrell (2016)	Internal, internal-external, and external validation of prediction models.	Strengthens the validation logic used in the current regression framework.
[25]	Swift (2011)	Long-run relationship between health and GDP across OECD countries.	Supports inclusion of income-related indicators in longevity modeling.
[26]	Thompson et al. (2017)	Extraction and interpretation of VIF and multicollinearity diagnostics.	Provides technical support for multicollinearity screening.
[27]	Van Calster et al. (2019)	Calibration and predictive analytics.	Emphasizes that model performance must be validated and interpreted cautiously.
[28]	Vatcheva et al. (2016)	Multicollinearity in epidemiologic regression analyses.	Justifies VIF assessment before interpreting adjusted coefficients.
[29]	Wirayuda et al. (2021)	Systematic review of sociodemographic, macroeconomic, and health-resource factors.	Provides synthesized evidence for including multiple determinant groups.
[30]	Zeileis et al. (2020)	Robust and clustered covariance estimation.	Supports robust standard errors when heteroscedasticity is present.

III. RESEARCH GAP

The reviewed literature confirms that life expectancy has been examined through health, economic, demographic, educational, and institutional determinants, but many applied studies emphasize either substantive determinants or statistical modeling rather than integrating both in one reproducible workflow. A remaining gap is the need for a compact framework that moves from correlation screening to variable selection, robust regression, multicollinearity checks, residual diagnostics, and validation while preserving clear factor interpretation. This study addresses that gap by using a multivariate public-health dataset to demonstrate how regression can identify key determinants without reducing the analysis to either simple bivariate association or prediction-only modeling.

IV. CONCEPTUAL FRAMEWORK AND HYPOTHESES DEVELOPMENT

The conceptual framework assumes that life expectancy is shaped by a combination of mortality conditions, public-health protection, economic resources, education, disease burden, nutritional status, demographic scale, and regional-development context. Mortality and disease-burden indicators are expected to show negative associations with life expectancy, whereas schooling, immunization, and GDP per capita are expected to show positive associations. Regional and development-status variables are expected to capture structural contextual differences not fully reflected by numeric predictors.

- H1: Mortality and disease-burden indicators are significantly associated with lower life expectancy after controlling for other predictors.
- H2: Schooling and GDP per capita are significantly associated with higher life expectancy after controlling for mortality and regional conditions.
- H3: Variable selection improves model interpretability by retaining a smaller set of statistically and substantively meaningful predictors.
- H4: The final regression model demonstrates stable validation performance across holdout and cross-validation procedures.

V. PROPOSED METHOD

The proposed method is a six-stage regression-based framework for identifying key factors in multivariate data. The stages are: data audit, descriptive and correlation screening, regression estimation, variable selection, diagnostic testing, and validation. The method is designed for empirical studies where the main objective is not only to predict an outcome but to identify interpretable predictors that remain important after adjustment.

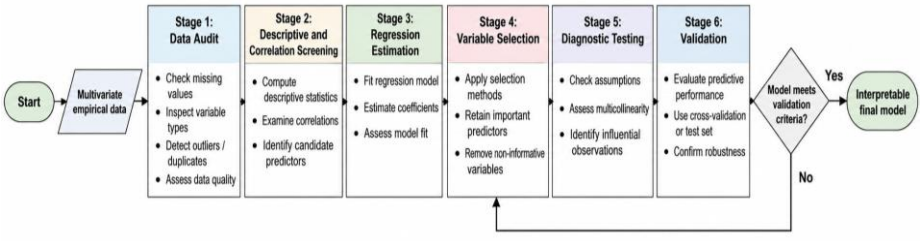


FIGURE 1. Proposed regression-based key-factor identification framework.

FIGURE 1 summarizes the complete analytical workflow used to move from raw multivariate data to interpretable key-factor identification and validation.

VI. STUDY CONTRIBUTION

The study makes three main contributions.

First, it presents a practical regression-based framework for identifying key factors in multivariate datasets. Second, it demonstrates the framework using a real public-health dataset with sufficient sample size and predictor diversity.

Third, it combines interpretability-oriented regression with model validation, which helps avoid overreliance on statistical significance alone. The approach is therefore suitable for applied researchers who need both explanatory clarity and predictive reliability.

VII. METHODOLOGY

1. RESEARCH DESIGN

The study used a quantitative explanatory-predictive design. The explanatory component identifies adjusted relationships between predictors and life expectancy, while the predictive component evaluates whether the selected regression model performs consistently on unseen data.

2. DATA SOURCE AND SAMPLE

The empirical dataset was an updated global life-expectancy dataset covering 2,864 country-year observations, 179 countries, 9 world regions, and yearly observations from 2000 to 2015. The dataset contains 21 variables representing mortality, immunization, disease burden, economic conditions, demographic scale, schooling, body-mass indicators, regional classification, and development status. No missing values were detected in the selected analytical file, making it appropriate for direct regression modeling without imputation.

Table 2. Dataset profile.

Criterion	Value
Observations	2,864
Countries	179
Regions	9
Years	2000-2015
Variables	21
Missing values	0
Dependent variable	Life_expectancy

Table 2 shows that the selected dataset has adequate scale, coverage, and completeness for multivariate regression analysis.

3. VARIABLES AND MEASURES

The dependent variable was Life_expectancy. Candidate predictors included adult mortality, infant deaths, under-five deaths, alcohol consumption, hepatitis B, measles, BMI, polio, diphtheria, HIV incidence, GDP per capita, population, thinness indicators, schooling, region, and economy-status indicators. Numeric predictors were standardized for coefficient comparability, and region was represented using dummy variables.

4. ANALYTICAL PROCEDURE

The analysis followed seven steps: descriptive statistics, correlation analysis, preprocessing and standardization, LASSO-assisted selection, heteroscedasticity-robust OLS regression, VIF and residual diagnostics, and model validation through an 80/20 train-test split plus 10-fold cross-validation. Robust standard errors were used because the Breusch-Pagan test indicated heteroscedasticity.

5. REGRESSION SPECIFICATION

The regression specification and validation metrics were formalized using Word professional equation objects as follows. In Equation (1), the response is life expectancy for each observation, the X terms represent standardized predictors, beta coefficients represent adjusted effects, and the residual term captures unexplained variation.

$$y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i \quad (1)$$

$$KeyFactor_j = \{ X_j : p_j < 0.05, VIF_j < \tau, validation\ stable \} \quad (2)$$

$$RMSE = \sqrt{\frac{\Sigma_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \quad (3)$$

$$MAE = \frac{\Sigma_{i=1}^{n} abs(y_i - \hat{y}_i)}{n} \quad (4)$$

$$R^2 = 1 - \frac{\Sigma_{i=1}^{n}(y_i - \hat{y}_i)^2}{\Sigma_{i=1}^{n}(y_i - \bar{y})^2} \quad (5)$$

VIII. RESULTS

1. DESCRIPTIVE STATISTICS

Table 3. Descriptive statistics for selected variables.

Variable	Mean	SD	Min	Median	Max
Life_expectancy	68.86	9.41	39.4	71.4	83.8
Adult_mortality	192.25	114.91	49.38	163.84	719.36
Under_five_deaths	42.94	44.57	2.3	23.1	224.9
Schooling	7.63	3.17	1.1	7.8	14.1
GDP_per_capita	11540.92	16934.79	148.0	4217.0	112418.0
BMI	25.03	2.19	19.8	25.5	32.1
Incidents_HIV	0.89	2.38	0.01	0.15	21.68

Table 3 reports core distributional information for the dependent variable and main explanatory variables retained for interpretation.

2. CORRELATION ANALYSIS

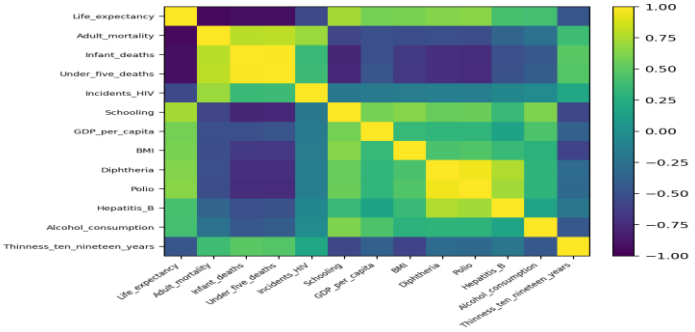


FIGURE 1. Correlation structure among selected variables.

FIGURE 2 shows the direction and strength of bivariate relationships before adjusted regression estimation.

Table 4. Strongest correlations with life expectancy.

Variable	Correlation with Life_expectancy
Adult_mortality	-0.945
Under_five_deaths	-0.92
Infant_deaths	-0.92
Schooling	0.733
Polio	0.641
Diphtheria	0.627
BMI	0.598
GDP_per_capita	0.583
Incidents_HIV	-0.553
Measles	0.49

Table 4 indicates that mortality-related indicators had the strongest negative bivariate relationships, while schooling, immunization, BMI, and GDP per capita had strong positive relationships.

3. FINAL REGRESSION MODEL

Table 5. Final robust regression coefficients.

Variable	Coef.	Robust SE	t	p	95% CI
Adult_mortality	-5.440	0.077	-70.95	<0.001	[-5.590, -5.290]
Under_five_deaths	-3.592	0.083	-43.50	<0.001	[-3.754, -3.430]
Region_Central America and Caribbean	1.778	0.125	14.26	<0.001	[1.533, 2.022]
Region_South America	1.550	0.113	13.69	<0.001	[1.329, 1.772]
Economy_status_Developed	1.026	0.059	17.49	<0.001	[0.911, 1.141]
Region_Oceania	-1.013	0.126	-8.06	<0.001	[-1.260, -0.767]
Region_European Union	-0.761	0.152	-5.01	<0.001	[-1.059, -0.463]
Region_North America	0.551	0.230	2.39	0.017	[0.099, 1.002]
Region_Asia	0.041	0.093	0.44	0.660	[-0.141, 0.223]
GDP_per_capita	0.372	0.039	9.58	<0.001	[0.296, 0.448]
Schooling	0.420	0.055	7.62	<0.001	[0.312, 0.528]
BMI	-0.273	0.039	-6.98	<0.001	[-0.349, -0.196]
Region_Rest of Europe	0.184	0.106	1.73	0.084	[-0.024, 0.393]
Region_Middle East	0.105	0.110	0.95	0.341	[-0.111, 0.321]
Incidents_HIV	0.232	0.057	4.07	<0.001	[0.120, 0.343]

Table 5. presents heteroscedasticity-robust estimates. Adult mortality and under-five deaths were the strongest negative predictors, while development status, schooling, GDP per capita, and several regional indicators remained meaningful explanatory factors.

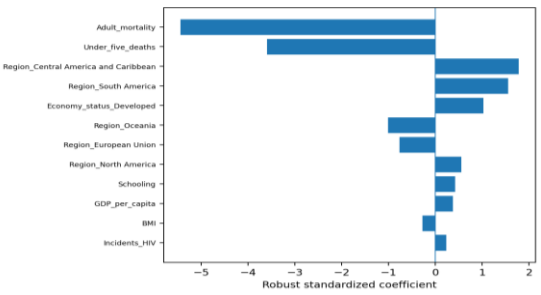


FIGURE 3. Estimated effect directions in the final regression model.

FIGURE 3. Visualizes the direction and relative magnitude of standardized coefficient estimates in the final model.

4. DIAGNOSTICS AND VALIDATION

Table 6. Diagnostic and validation summary.

Metric	Value
Test R-squared	0.982
Test RMSE	1.232
Test MAE	0.981
10-fold CV R-squared	0.983 ± 0.002
10-fold CV RMSE	1.230 ± 0.058
Durbin-Watson	1.988
Breusch-Pagan p-value	<0.001
Jarque-Bera p-value	0.010

Table 6 shows that the model had high predictive stability; the heteroscedasticity result justified robust standard errors, while the Durbin-Watson value indicated no major residual autocorrelation pattern.

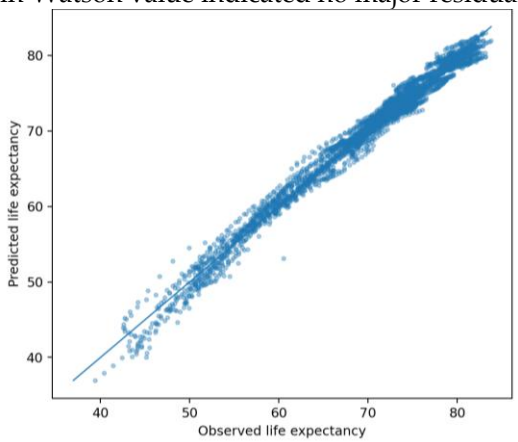


FIGURE 4. Observed and predicted life expectancy.

FIGURE 4 confirms that predicted values closely followed observed life expectancy values, supporting the model's validation results.

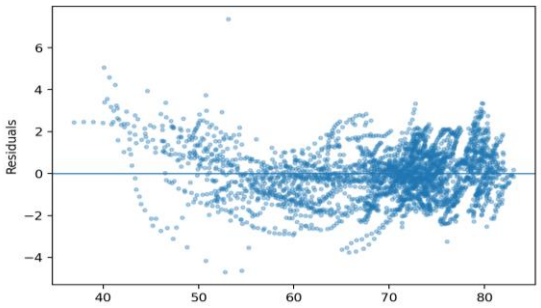


FIGURE 5. Residual dispersion in the final model.

FIGURE 5. displays residual variation around the fitted values and supports the use of diagnostic checks alongside coefficient interpretation.

IX. DISCUSSION

The findings confirm that regression-based analysis can identify interpretable key factors in multivariate data when it is supported by diagnostic and validation procedures. The strongest negative adjusted effects were associated with adult mortality and under-five deaths, which is substantively consistent with life expectancy as a survival-related population outcome. The results show that mortality indicators remain important even when socioeconomic and regional variables are included.

Schooling and GDP per capita remained positive predictors, suggesting that social and economic development contribute to life expectancy beyond direct mortality indicators. Development status also remained significant, indicating that structural differences between developed and developing economies capture additional variation not fully explained by GDP alone. Some regional indicators were significant, showing that geography and regional context may reflect unobserved differences in institutions, infrastructure, epidemiological profile, and health-system organization.

The correlation results also illustrate why multivariate regression is necessary. Several variables were strongly correlated with life expectancy in bivariate analysis, but adjusted regression helps distinguish direct explanatory contribution from shared variation among predictors. The final model preserved interpretability while achieving strong validation performance. This supports the central claim that regression remains useful not only for estimation but also for identifying key factors for evidence-based decision-making.

At the same time, the results should be interpreted cautiously. The dataset is observational and country-year based, so the coefficients should not be treated as causal effects. Some variables may proxy deeper structural conditions, and country-level aggregation may hide within-country inequalities. The heteroscedasticity test also indicates that conventional standard errors would be inappropriate, which is why robust inference was used.

X. PRACTICAL IMPLICATIONS

The results imply that policy-oriented statistical analysis should prioritize variables that remain important after adjustment and validation. Public-health planning should not rely only on simple associations; it should use multivariate models to separate mortality, socioeconomic, and regional effects. For researchers, the framework provides a reproducible template for analyzing multivariate datasets in healthcare, education, economics, business, and social sciences.

XI. LIMITATIONS AND FUTURE RESEARCH

The study is limited by its observational design, country-level aggregation, and reliance on variables available in the dataset. Future research can extend the framework using panel fixed effects, mixed-effects models, nonlinear specifications, interaction terms, country-clustered standard errors, and additional validation across independent datasets. Further work may also compare regression-based key-factor identification with nonlinear ensemble approaches to assess whether interpretability and predictive accuracy can be improved together.

XII. CONCLUSION

This study developed and empirically evaluated a regression-based statistical framework for identifying key factors in multivariate data. Using a global life-expectancy dataset with 2,864 country-year observations, the analysis showed that transparent regression remains a strong analytical strategy when combined with variable selection, robust inference, diagnostic testing, and validation. The final model showed that adult mortality and under-five mortality were the most influential negative determinants of life expectancy, while schooling, GDP per capita, development status, and selected regional indicators retained meaningful explanatory value after adjustment. These findings confirm that identifying key factors requires more than simple correlation because bivariate relationships can be altered by shared variation among mortality, socioeconomic, and regional variables. The study's methodological contribution lies in presenting regression as a staged evidence framework rather than a single estimation procedure. The use of LASSO-assisted

screening, heteroscedasticity-robust regression, multicollinearity assessment, residual diagnostics, train-test validation, and cross-validation provides a balanced structure for researchers who need both explanatory clarity and predictive reliability. The model achieved strong validation performance, with a test R-squared of 0.982, RMSE of 1.232, MAE of 0.981, and cross-validated R-squared of 0.983, indicating that the selected predictors explained variation in life expectancy with high stability. At the same time, the study avoids overstating the evidence. The data are observational and country-year based; therefore, the results should be interpreted as adjusted statistical associations rather than causal effects. Future research should extend this framework through panel fixed effects, mixed-effects specifications, interaction analysis, nonlinear terms, and country-clustered uncertainty estimates. Overall, the study shows that regression-based analysis can remain highly publishable and policy-relevant when it is executed as a complete, validated, and diagnostically transparent workflow for multivariate evidence generation.

Author Contributions

The author conducted the conceptualization, methodology, data analysis, investigation, writing, review, editing, and final approval of the manuscript.

Funding

This research received no external funding.

Data Availability

The dataset will be available from the author upon reasonable request.

Conflicts of Interest

The author declares no conflict of interest.

REFERENCES

[1]. Balkhi, B., Alshayban, D., & Alotaibi, N. M. (2021). Impact of healthcare expenditures on healthcare outcomes in the Middle East and North Africa (MENA) region: A cross-country comparison, 1995-2015. Frontiers in Public Health, 8, Article 624962. https://doi.org/10.3389/fpubh.2020.624962

[2]. Blazquez-Fernandez, C., Cantarero-Prieto, D., & Pascual-Saez, M. (2017). Health expenditure and socio-economic determinants of life expectancy in the OECD Asia/Pacific area countries. Applied Economics Letters, 24(3), 167-169. https://doi.org/10.1080/13504851.2016.1173174

[3]. Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. M. (2015). Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): Explanation and elaboration. Annals of Internal Medicine, 162(1), W1-W73. https://doi.org/10.7326/M14-0698

[4]. Galvani-Townsend, S., Martinez, I., & Pandey, A. (2022). Is life expectancy higher in countries and territories with publicly funded healthcare? Global analysis of healthcare access and the social determinants of health. Journal of Global Health, 12, Article 04091. https://doi.org/10.7189/jogh.12.04091

[5]. Gelman, A., Hill, J., & Vehtari, A. (2020). Regression and other stories. Cambridge University Press.

[6]. Halicioglu, F. (2011). Life expectancy and economic development: Evidence from Turkey. Applied Economics Letters, 18(6), 559-562. https://doi.org/10.1080/13504851003761770

[7]. Harrell, F. E., Jr. (2015). Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis (2nd ed.). Springer. https://doi.org/10.1007/978-3-319-19425-7

[8]. Hastie, T., Tibshirani, R., & Wainwright, M. (2015). Statistical learning with sparsity: The lasso and generalizations. CRC Press. https://doi.org/10.1201/b18401

[9]. Heinze, G., Wallisch, C., & Dunkler, D. (2018). Variable selection: A review and recommendations for the practicing statistician. Biometrical Journal, 60(3), 431-449. https://doi.org/10.1002/bimj.201700067

[10]. Jaba, E., Balan, C. B., & Robu, I. B. (2014). The relationship between life expectancy at birth and health expenditures estimated by a cross-country and time-series analysis. Procedia Economics and Finance, 15, 108-114. https://doi.org/10.1016/S2212-5671(14)00454-7

[11]. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An introduction to statistical learning: With applications in R (2nd ed.). Springer. https://doi.org/10.1007/978-1-0716-1418-1

[12]. Kim, J. H. (2019). Multicollinearity and misleading statistical results. Korean Journal of Anesthesiology, 72(6), 558-569. https://doi.org/10.4097/kja.19087

[13]. Kuhn, M., & Johnson, K. (2013). Applied predictive modeling. Springer. https://doi.org/10.1007/978-1-4614-6849-3

[14]. Linden, M., & Ray, D. (2017). Life expectancy effects of public and private health expenditures in OECD countries 1970-2012: Panel time series approach. Economic Analysis and Policy, 56, 101-113. https://doi.org/10.1016/j.eap.2017.06.005

[15]. Miladinov, G. (2020). Socioeconomic development and life expectancy relationship: Evidence from the EU accession candidate countries. Genus, 76, Article 2. https://doi.org/10.1186/s41118-019-0071-0

[16]. Moga Rogoz, A. T., Moga, T. L., Teodorescu, C., & Nica, E. (2022). Impact of economic freedom and educational attainment on life expectancy: Evidence from the new EU member states. Frontiers in Public Health, 10, Article 907138. https://doi.org/10.3389/fpubh.2022.907138

[17]. Moons, K. G. M., Altman, D. G., Reitsma, J. B., Ioannidis, J. P. A., Macaskill, P., Steyerberg, E. W., Vickers, A. J., Ransohoff, D. F., & Collins, G. S. (2015). Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): The TRIPOD statement. Annals of Internal Medicine, 162(1), 55-63. https://doi.org/10.7326/M14-0697

[18]. Murtin, F., Mackenbach, J., Jasilionis, D., & d'Ercole, M. M. (2022). Well-being analytics for policy use: The OECD how's life framework and life expectancy modeling. OECD Publishing. https://doi.org/10.1787/d6e2d305-en

[19]. Rahman, M. M., Khanam, R., & Rahman, M. (2018). Health care expenditure and health outcome nexus: New evidence from the SAARC-ASEAN region. Globalization and Health, 14, Article 113. https://doi.org/10.1186/s12992-018-0430-1

[20]. Ranabhat, C. L., Atkinson, J., Park, M.-B., Kim, C.-B., & Jakovljevic, M. (2018). The influence of universal health coverage on life expectancy at birth and healthy life expectancy: A multi-country cross-sectional study. Frontiers in Pharmacology, 9, Article 960. https://doi.org/10.3389/fphar.2018.00960

[21]. Ray, D., & Linden, M. (2020). Health expenditure, longevity, and child mortality: Dynamic panel data approach with global data. International Journal of Health Economics and Management, 20(1), 99-119. https://doi.org/10.1007/s10754-019-09272-z

[22]. Riley, R. D., Ensor, J., Snell, K. I. E., Harrell, F. E., Jr., Martin, G. P., Reitsma, J. B., Moons, K. G. M., Collins, G., & van Smeden, M. (2020). Calculating the sample size required for developing a clinical prediction model. BMJ, 368, m441. https://doi.org/10.1136/bmj.m441

[23]. Roffia, P., Bucciol, A., & Hashlamon, M. (2022). Determinants of life expectancy at birth: A longitudinal study on OECD countries. International Journal of Health Economics and Management, 23, 189-212. https://doi.org/10.1007/s10754-022-09338-5

[24]. Steyerberg, E. W., & Harrell, F. E., Jr. (2016). Prediction models need appropriate internal, internal-external, and external validation. Journal of Clinical Epidemiology, 69, 245-247. https://doi.org/10.1016/j.jclinepi.2015.04.005

[25]. Swift, R. (2011). The relationship between health and GDP in OECD countries in the very long run. Health Economics, 20(3), 306-322. https://doi.org/10.1002/hec.1590

[26]. Thompson, C. G., Kim, R. S., Aloe, A. M., & Becker, B. J. (2017). Extracting the variance inflation factor and other multicollinearity diagnostics from typical regression results. Basic and Applied Social Psychology, 39(2), 81-90. https://doi.org/10.1080/01973533.2016.1277529

[27]. Van Calster, B., McLernon, D. J., van Smeden, M., Wynants, L., Steyerberg, E. W., & Topic Group Evaluating Diagnostic Tests and Prediction Models of the STRATOS Initiative. (2019). Calibration: The Achilles heel of predictive analytics. BMC Medicine, 17, Article 230. https://doi.org/10.1186/s12916-019-1466-7

[28]. Vatcheva, K. P., Lee, M., McCormick, J. B., & Rahbar, M. H. (2016). Multicollinearity in regression analyses conducted in epidemiologic studies. Epidemiology, 6(2), Article 227. https://doi.org/10.4172/2161-1165.1000227

[29]. Wirayuda, A. A. B., Sari, T. W. A., & Putri, R. G. S. (2021). A systematic review of sociodemographic, macroeconomic, and health resources factors on life expectancy. Asia Pacific Journal of Public Health, 33(4), 335-356. https://doi.org/10.1177/1010539520983671

[30]. Zeileis, A., Köll, S., & Graham, N. (2020). Various versatile variances: An object-oriented implementation of clustered covariances in R. Journal of Statistical Software, 95(1), 1-36. https://doi.org/10.18637/jss.v095.i01