International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 10, October 2026)

Socio-Economic Determinants of Life Expectancy Across Indian States: A Panel Data Regression and Cluster Analysis

Dr. J. Subashini

Assistant Professor, Departmaent of Mathematics, Sri Ramakrishna College of Arts & Science for Women, India

***Abstract*— Life expectancy at birth is an important indicator of population health and socio-economic development. Differences in life expectancy across Indian states may reflect variations in income, education, public health expenditure, infant mortality, sanitation and access to healthcare. This study proposes an empirical investigation of the socio-economic determinants of life expectancy across selected Indian states using secondary data from official statistical publications. The research combines panel data regression with cluster analysis to examine associations between development indicators and longevity and to identify groups of states with similar health and socio-economic characteristics. The proposed dataset will integrate life-expectancy estimates from the Sample Registration System (SRS) with relevant state-level indicators from the Reserve Bank of India and other government sources. Regression analysis will examine associations between life expectancy and selected explanatory variables, while cluster analysis will classify states based on health and development profiles. Model selection, diagnostic testing, missing-data assessment and sensitivity analysis will be undertaken to strengthen reliability. Because life-expectancy estimates are generally published for overlapping multi-year periods, the temporal structure and comparability of observations must be evaluated before selecting the final regression model. Actual regression coefficients and cluster memberships are not reported in this draft because the source observations have not yet been assembled and analysed.**

***Keywords*— Life expectancy; socio-economic determinants; Indian states; panel data regression; cluster analysis; health expenditure; infant mortality; regional inequality.**

I. INTRODUCTION

Life expectancy at birth represents the average number of years a newborn would live if the mortality conditions observed during a specified period continued throughout life. It is widely used as an indicator of population health because it reflects mortality risks across different age groups.

Improvements in life expectancy are often associated with better nutrition, higher educational attainment, improved living conditions, access to healthcare, safe drinking water and effective public health programmes.

However, economic development alone may not ensure equivalent improvements in health outcomes across regions.

India provides an important setting for examining these relationships. Its states differ in economic development, demographic conditions, educational attainment, healthcare infrastructure and public service provision. These differences make state-level analysis useful for understanding the relationship between socio-economic development and longevity.

The Sample Registration System (SRS) Abridged Life Tables published by the Office of the Registrar General and Census Commissioner provide life-expectancy estimates for India and larger states and union territories. The 2019–2023 report estimated life expectancy at birth for India at 70.3 years, with 72.5 years for females and 68.5 years for males. These are national estimates and are not results of the proposed state-level regression.

1.1 Statement of the problem

Differences in life expectancy across Indian states cannot be adequately explained by a single socio-economic indicator. Income, education, health expenditure and environmental conditions may operate together, and their relationships may differ across states and time periods. A further challenge concerns data comparability: life expectancy is often reported as a multi-year estimate, whereas several socio-economic variables are available annually. Combining these sources without considering their reference periods may produce misleading regression results.

1.2 Research gap and proposed contribution

The study is designed to address four analytical needs: (i) joint examination of multiple socio-economic and health indicators; (ii) careful matching of multi-year life-expectancy estimates with explanatory variables; (iii) complementary use of regression and cluster analysis; and (iv) translation of state profiles into policy-relevant evidence. These are proposed contributions, not claims that no previous study has addressed similar questions. A final novelty statement requires a systematic literature review.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 10, October 2026)

II. Review Of Literature

2.1 Economic development and population health

Higher income can improve nutrition, housing, sanitation and access to healthcare. Nevertheless, the strength of this relationship may depend on income distribution and the accessibility of public services. Per capita net state domestic product is proposed as an indicator of economic development.

2.2 Education and health outcomes

Education can influence health through improved awareness, the ability to use healthcare information, employment opportunities and household decision-making. Educational attainment may also reflect broader development conditions, so its relationship with life expectancy should be assessed alongside other explanatory variables.

2.3 Healthcare expenditure and longevity

Public healthcare expenditure may support preventive care, primary healthcare, maternal services, immunisation and disease control. Its effectiveness may depend on allocation, health-worker availability, infrastructure and service quality. National health expenditure figures provide context but cannot substitute for state-level expenditure observations.

2.4 Infant mortality and life expectancy

Infant mortality measures deaths among children under one year of age per 1,000 live births. Higher infant mortality is expected to be associated with lower life expectancy at birth because infant deaths contribute directly to the mortality schedule used to calculate life expectancy. The magnitude and statistical significance of this relationship must be estimated from the selected data.

2.5 Sanitation and living conditions

Access to safe drinking water, sanitation and adequate housing can reduce exposure to infectious diseases and environmental hazards. These conditions may contribute to population health alongside economic growth and healthcare provision.

2.6 Contribution of the proposed study

The proposed study combines analysis of socio-economic differences between states, statistical associations between development indicators and life expectancy, and classification of states into groups with comparable health and development profiles. Cluster analysis complements rather than replaces regression analysis.

III. Objectives Of The Study

1. To examine differences in life expectancy across selected Indian states.
2. To analyse the association between life expectancy and per capita state income.
3. To investigate the relationships of education, health expenditure, infant mortality and sanitation with life expectancy.
4. To estimate an appropriate panel regression model if the data structure and coverage support panel estimation.
5. To classify states into socio-economic and health clusters.
6. To suggest policy priorities based on estimated relationships and resulting state profiles.

IV. Research Hypotheses

The following hypotheses are proposed for empirical testing. Expected directions are theoretical expectations, not established findings for the selected sample.

Hypothesis	Null hypothesis	Expected direction
H1	Per capita income is not associated with life expectancy.	Positive
H2	Educational attainment is not associated with life expectancy.	Positive
H3	Public health expenditure is not associated with life expectancy.	Positive
H4	Infant mortality is not associated with life expectancy.	Negative
H5	Sanitation access is not associated with life expectancy.	Positive

V. Research Methodology

5.1 Research design

The study adopts a quantitative, observational design based on secondary data. The unit of analysis is the Indian state or union territory, depending on the final coverage of the official data. The final sample will include states with comparable life-expectancy estimates and sufficient information for the selected explanatory variables.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 10, October 2026)

5.2 Study period and panel structure

The latest official life-table period identified for this draft is 2020–2024. Earlier periods may also be considered where state estimates and explanatory variables are comparable. These life-table periods may overlap and are not independent annual observations. The analysis must avoid treating repeated overlapping estimates as separate calendar-year outcomes.

Two designs may be considered. A panel model may be used if multiple compatible reference periods can be matched with period-level explanatory variables and the overlap issue can be handled appropriately. A cross-sectional model using the latest common period is preferable if reliable temporal alignment is not possible. The final choice must be determined by the assembled data.

5.3 Selection of states

The sample will include states with comparable life-expectancy estimates and sufficient data. State-boundary changes and administrative classification will be documented. For example, some SRS estimates for Jammu and Kashmir refer to the erstwhile state and include the present union territories of Jammu and Kashmir and Ladakh. Such estimates must not be assigned to present-day units without a defensible matching procedure.

5.4 Data sources

Variable/data requirement	Potential official source	Use in study
Life expectancy at birth	Office of the Registrar General and Census Commissioner, SRS Abridged Life Tables	Dependent variable
Per capita state income	RBI Handbook of Statistics on Indian States; state income statistics	Economic development
Literacy/education	RBI state statistics; Census-based official publications	Educational attainment
Infant mortality rate	SRS and official health statistics	Mortality conditions
State health expenditure	State budgets and official health expenditure publications	Public investment in health
Sanitation and drinking water	National Family Health Survey and other official reports	Living conditions

Availability in an official publication does not guarantee that a variable is available for every state and every period. The final dataset must record the source, reference period, unit and transformation for each observation.

Variable	Symbol	Measurement	Expected relationship
Life expectancy at birth	LE	Years	Dependent variable
Per capita income	PCI	Real per capita NSDP, preferably at constant prices	Positive
Educational attainment	EDU	Literacy percentage or comparable education indicator	Positive
Public health expenditure	HE	Real per capita expenditure or consistently defined expenditure ratio	Positive
Infant mortality rate	IMR	Deaths under age one per 1,000 live births	Negative
Sanitation access	SAN	Percentage of households using an improved sanitation facility	Positive

Real per capita income is preferable to nominal income where consistent constant-price data are available. Public health expenditure must be defined consistently; budget allocations, actual expenditure, per capita expenditure and expenditure as a share of state domestic product are different measures and should not be mixed.

5.6 Data preparation

1. Download the relevant official tables and retain publication details.
2. Standardise state names and administrative identifiers.
3. Match life-expectancy periods with corresponding explanatory-variable periods.
4. Convert income and expenditure values to consistent units.
5. Identify missing values and examine their patterns.
6. Check outliers and possible data-entry errors against the original publications.
7. Create a data dictionary documenting definitions and transformations.
8. Prepare an analysis-ready file for SPSS.

Missing values will not automatically be replaced by means or arbitrary estimates. Any imputation must be justified, documented and assessed through sensitivity analysis.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 10, October 2026)

VI. Statistical Analysis

6.1 Descriptive statistics

Means, standard deviations, minimum values and maximum values will be calculated for all variables. State-level tables will describe the distribution of life expectancy and socio-economic conditions. Rankings may be reported only when based on a consistent reference period and comparable state sample.

6.2 Correlation analysis

Pearson correlation coefficients will be used for approximately linear relationships between continuous variables when assumptions are reasonable. Spearman rank correlations may be used as a supplementary analysis where outliers or non-normal distributions make rank-based comparisons suitable. Correlation does not establish causation.

6.3 Panel regression model

If multiple comparable periods are available, the proposed baseline model is:

$$LE_{it} = \alpha_i + \lambda_t + \beta_1 PCI_{it} + \beta_2 EDU_{it} + \beta_3 HE_{it} + \beta_4 IMR_{it} + \beta_5 SAN_{it} + \varepsilon_{it}$$

Here, i denotes the state, t denotes the reference period, α_i represents state-specific effects, λ_t represents period effects, β coefficients represent associations to be estimated, and ε_{it} is the error term. Fixed effects can control for unobserved state characteristics that remain constant over the observed periods. Period effects can control for common changes across states.

A random-effects model may also be estimated where its assumptions are defensible. If only one common reference period is available, panel regression is not appropriate and cross-sectional multiple regression should be used.

6.4 Regression diagnostics

- Multicollinearity using variance inflation factors.
- Heteroskedasticity in regression residuals.
- Serial correlation where a panel structure is used.
- Cross-sectional dependence across states.
- Influential observations and residual patterns.
- Sensitivity to alternative specifications and variable definitions.

A conventional Hausman test can inform comparison of fixed-effects and random-effects estimators under its assumptions. It should not replace substantive model assessment. Since the number of states may be limited, standard errors and diagnostic procedures must be selected with sample size and dependence structure in mind.

6.5 Cluster analysis

Cluster analysis will classify states by life expectancy, real per capita income, education, public health expenditure, infant mortality, and sanitation. Clustering variables will be standardised before distance-based analysis so measurement units do not dominate.

We may initially explore hierarchical clustering using Ward's method and squared Euclidean distance. K-means clustering may then be considered for an interpretable number of groups. The final number of clusters should be chosen after assessing separation, group size, stability and substantive interpretation. Possible labels such as higher-development, intermediate-profile and greater-challenge states are illustrative only; actual memberships must be computed.

VII. Results And Analysis: Evidence And Outstanding Estimates

This section separates published background evidence from the proposed study's uncomputed empirical results.

7.1 Verified national background figures

Indicator	Published value
Life expectancy at birth, total population	70.3 years
Life expectancy at birth, males	68.5 years
Life expectancy at birth, females	72.5 years
Female–male difference	4.0 years

7.2 Regression results table to complete after estimation

Explanatory variable	Coefficient	Standard error	p-value
Per capita income	Not estimated	Not estimated	Not estimated
Education	Not estimated	Not estimated	Not estimated
Public health expenditure	Not estimated	Not estimated	Not estimated
Infant mortality rate	Not estimated	Not estimated	Not estimated
Sanitation access	Not estimated	Not estimated	Not estimated

The final results should also report the number of states and observations, model specification, standard-error method, goodness-of-fit statistics and diagnostic results.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 10, October 2026)

7.3 Cluster results table to complete

Cluster	States included	Mean life expectancy	Development profile
Cluster 1	To be estimated	To be calculated	To be interpreted
Cluster 2	To be estimated	To be calculated	To be interpreted
Cluster 3, if supported	To be estimated	To be calculated	To be interpreted

VIII. Discussion

The study is designed to evaluate whether differences in life expectancy across Indian states are associated with differences in income, education, public healthcare expenditure, infant mortality and living conditions. Regression analysis will assess the direction and magnitude of associations while accounting for state and period effects where the dataset permits. Cluster analysis will identify groups of states with similar combinations of health and development characteristics.

Interpretation must distinguish statistical association from causal impact. For example, a positive association between public health expenditure and life expectancy would not by itself establish that additional expenditure caused the observed difference. Healthier and wealthier states may also differ in governance, infrastructure, social conditions and historical development.

IX. Policy Implications

1. Targeted healthcare investment: identify states where health outcomes are relatively weak despite available economic resources.
2. Education and health awareness: assess whether education indicators are associated with longevity and whether further investigation of health literacy is warranted.
3. Maternal and child health: examine the role of infant mortality in interstate differences in life expectancy.
4. Sanitation and prevention: assess whether differences in household sanitation are associated with health outcomes.
5. State-specific planning: use clusters to distinguish broad development profiles rather than recommending identical interventions for every state.

These are potential policy applications. Specific state-level recommendations should be made only after examining the estimated results and contextual evidence.

X. Conclusion

This research proposes a combined panel regression and cluster analysis framework for examining socio-economic determinants of life expectancy across selected Indian states. It brings together demographic, economic, educational, health and living-condition indicators to investigate interstate differences in longevity.

Official publications provide a starting point: SRS supplies life-expectancy estimates, while RBI and other government sources provide relevant socio-economic and health indicators. A defensible empirical conclusion requires a harmonised dataset, an appropriate model for the temporal structure of life-expectancy estimates and reproducible statistical analysis. Regression coefficients, statistical significance, diagnostic results and cluster memberships should be reported only after calculations have been completed.

REFERENCES

[1] Office of the Registrar General & Census Commissioner, India. (2026). Sample Registration System—Abridged Life Tables 2020–2024. Government of India

[2] Office of the Registrar General & Census Commissioner, India. (2025). Sample Registration System—Abridged Life Tables 2018–2022. Government of India.

[3] Office of the Registrar General & Census Commissioner, India. (2026). Sample Registration System—Abridged Life Tables 2019–2023. Government of India.

[4] Reserve Bank of India. (2023). Handbook of Statistics on Indian States 2022–23.

[5] Ministry of Health and Family Welfare, Government of India. (2026). National Health Accounts Estimates for India 2022–23. Press Information Bureau, 27 May 2026.

[6] Wooldridge, J. M. (2010). Econometric Analysis of Cross Section and Panel Data (2nd ed.). MIT Press.

[7] Cameron, A. C., & Trivedi, P. K. (2005). Microeconometrics: Methods and Applications. Cambridge University Press.

[8] Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). Cluster Analysis (5th ed.). Wiley.

[9] Ward, J. H., Jr. (1963). Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association, 58(301), 236–244.