A Comparative Study of Mizoram and Himachal Pradesh: Demographic, Social, and Economic Perspectives

Analyzing regional contrasts and developmental dynamics in two Himalayan states

Abhik R, Arikith RC, Anunoy C, Himadri M, Sattick D

A Comparative Study of Mizoram and Himachal Pradesh: Demographic, Social, and Economic Perspectives

Analyzing regional contrasts and developmental dynamics in two Himalayan states

Edited By

Abhik Rana
Arikith Roy Choudhury
Anunoy Chal
Himadri Mandal
Sattick Das

B.Stat (3rd year)
Indian Statistical Institute
Kolkata

November 2025

Contents

I	Preliminary Investigations	
1	**Introduction**	**11**
1.1	**Background and Rationale**	**11**
1.2	**State Profiles**	**11**
1.2.1	Mizoram	11
1.2.2	Himachal Pradesh	11
1.3	**Key Thematic Areas of Study**	**12**
1.4	**Specific Data Snapshots**	**12**
2	**Hypothesis**	**13**
2.1	**Hypothesis Statements**	**13**
2.2	**Null and Alternative Hypotheses**	**13**
II	**Hypothesis Testing and Results**	
3	**Wealth, Residence, Age as Determinants of Infant Mortality**	**19**
3.1	**Objectives**	**19**
3.2	**Hypotheses**	**19**
3.3	**Data and Methodology**	**19**
3.3.1	Data Source	19
3.3.2	Variable Construction	20
3.3.3	Analytical Approach	20
3.4	**Results**	**20**
3.4.1	Wealth and Infant Mortality	20
3.4.2	Residence Type	21
3.4.3	Maternal Age and Rare-Event Adjustment	21
3.4.4	Inter-State Comparison and Interpretation	21
3.5	**Conclusion**	**22**

4 Mother's Education, Fertility and Sanitation Access 23
4.1 Research Questions 23
4.2 Hypothesis 23
4.2.1 Number of children ever born in relation to education level for each individual 23
4.2.2 Sanitation access in relation ot education level for each individual 24
4.3 Data and Methodology 24
4.3.1 Data Source 24
4.3.2 Variable Construction 24
4.4 Results 24
4.4.1 Average Number of Children Born vs Education Level of Individuals 24
4.4.2 Sanitation Access vs Education Level of Individuals 25
Appendices 29
A Acknowledgements 29

List of Figures

3.1 Predicted probability of infant death by wealth index for Himachal Pradesh and Mizoram. 20
3.2 Predicted infant mortality probability by residence type. 21
3.3 Observed data and rare-event weighted logistic fit for infant mortality vs. maternal age (Mizoram). 21

4.1 Box plot and ANOVA results for average number of children born vs education level 25
4.2 contingency table 26
4.3 Heatmap of sanitation access vs education level 26
4.4 Chi squared test results for both states 26

List of Tables

3.1 Summary of Logistic Regression Coefficients by State 22

Preliminary Investigations

1 **Introduction** **11**
1.1 Background and Rationale 11
1.2 State Profiles 11
1.3 Key Thematic Areas of Study 12
1.4 Specific Data Snapshots 12

2 **Hypothesis** **13**
2.1 Hypothesis Statements 13
2.2 Null and Alternative Hypotheses 13

1. Introduction

1.1 Background and Rationale

India's states exhibit wide variation in demographic, social and economic indicators, shaped by geography, culture, history and policy. A comparative study of two mountainous states, Mizoram in the Northeast and Himachal Pradesh in the Western Himalayas, allows us to ask: how do demographic outcomes (e.g., infant mortality, maternal age), social infrastructure (education levels, sanitation) and economic development paths interplay across different regional contexts? The findings can help tailor policy interventions that recognise state specific dynamics rather than assume one size fits all.

1.2 State Profiles

1.2.1 Mizoram

- **Population density** in Mizoram is low ($\sim$52 persons/km^2 approx) and its projected population growth is around the national average (0.9 %) as of 2022-23.
- **Health indicator**: Mizoram reports one of the lowest infant mortality rates in India, for example, an IMR of $\sim$3 to 4 deaths per 1000 live births in recent years.
- **Education**: According to the 2011 Census, the literacy rate in Mizoram was 91.33 %.
- **Economy and workforce**: As of 2022-23, the working population is concentrated in services (45.7 %), agriculture (43.1 %) and manufacturing (5.4 %) in Mizoram.
- **Infrastructure and geography**: Hilly terrain, shifting cultivation patterns and remoteness present both developmental strengths (close community networks) and challenges (access to health/sanitation).

1.2.2 Himachal Pradesh

- **Demographic transition**: Himachal Pradesh has achieved relatively favourable human development outcomes compared to many Indian states. For example, its IMR is reported at 17 deaths per 1000 live births around 2020.
- **Education and literacy**: Himachal shows strong literacy and school infrastructure, though exact recent figures would need updating from Census/NFHS.
- **Economy and structure**: The economy shows a move away from agriculture (which contributes 13 % of GSVA) towards services and industry, signalling a structural shift.
- **Topography and access**: The mountainous terrain, although offering tourism and hydro-

resources, poses challenges in connectivity, sanitation supply and service delivery in remote areas.

1.3 Key Thematic Areas of Study

This report focuses on several thematic relationships across both states:

1. **Infant mortality vs mothers age**: e.g., how younger or older maternal ages correlate with infant outcomes.
2. **Education level vs toilet facility types**: e.g., does higher educational attainment correspond to households having improved sanitation (water closet, piped sewer, etc)?
3. **Other socio-economic/demographic linkages**: For instance, links between female labour-force participation, household amenities, and health outcomes; or comparisons across rural versus urban within each state.

1.4 Specific Data Snapshots

Here are some concrete figures to anchor the analysis:

- **Mizoram's IMR**: Rural-data shows 3 per 1000 live births around 2020.
- **Himachal Pradesh's IMR**: Reported at 17 per 1000 live births in 2020.
- **Mizoram's literacy rate**: 91.33 % in 2011 Census.
- **Education-sanitation data**: Census table 'Availability and type of latrine facility -2001, 2011' covers latrine types by household and can be linked to education statistics.

2. Hypothesis

2.1 Hypothesis Statements

2.2 Null and Alternative Hypotheses

Let IMR denote the infant mortality outcome (binary at child level, or rate per 1000 live births at aggregate level), W the household wealth index (continuous or quintile), A the mother's age (years or age group), R the residence type (urban/rural), E the mother's education level (years of schooling or categorical), and C the fertility (number of children ever born, or average children per woman). Subscripts MZ and HP denote Mizoram and Himachal Pradesh respectively.

Test 1: Infant mortality vs Wealth index

Model / Test: Regress IMR on W (logistic regression for child-level binary IMR, or linear regression for aggregate IMR): where $\mathbf{X}$ are controls (e.g., mother's age, education, residence).

- **Mizoram**
  - $H_0^{(MZ)}$: $\beta_W^{(MZ)} = 0$ (wealth index has no association with infant mortality in Mizoram).
  - $H_A^{(MZ)}$: $\beta_W^{(MZ)} < 0$ (higher wealth is associated with lower infant mortality in Mizoram).
- **Himachal Pradesh**
  - $H_0^{(HP)}$: $\beta_W^{(HP)} = 0$.
  - $H_A^{(HP)}$: $\beta_W^{(HP)} < 0$.

Rationale: Household wealth improves access to nutrition, healthcare and sanitation; hence a negative association is expected.

Test 2: Infant mortality vs Mother's age

Model / Test: Compare IMR across maternal age categories (e.g., < 20, 20–34, ≥ 35). Use logistic regression with age indicators or two-sample tests for proportions.

- **Mizoram**
  - $H_0^{(MZ)}$: IMR is the same across maternal age groups (no difference between < 20 and 20–34).
  - $H_A^{(MZ)}$: IMR among mothers < 20 is greater than IMR among mothers aged 20–34.
- **Himachal Pradesh**
  - $H_0^{(HP)}$: IMR is the same across maternal age groups.

– $H_A^{(HP)}$: IMR among mothers < 20 is greater than IMR among mothers aged 20–34.

Rationale: Rationale: In many demographic studies, infant mortality follows a U-shaped relationship with maternal age, being higher for mothers younger than 20 years and those aged 35 years and above. However, in this study, the age group ≥ 35 is not included as a separate category. This is because

i The number of births to mothers aged 35 and above in both Mizoram and Himachal Pradesh is relatively small, which may lead to statistically unreliable estimates,

ii The primary policy concern and observable variation in infant mortality lies between teenage mothers (< 20 years) and mothers in the biologically safer range of 20–34 years.

Therefore, a U-shaped association is not expected or tested; instead, the analysis focuses on whether infant mortality is significantly higher among mothers below 20 compared to those aged 20–34.

Test 3: Residence type vs Infant mortality

Model / Test: Compare proportions of infant deaths between rural and urban (Chi-square test of independence or logistic regression with suitable residence indicator).

- **Mizoram**
  - $H_0^{(MZ)} : p_{\text{rural}}^{(MZ)} = p_{\text{urban}}^{(MZ)}$ (no difference in infant mortality proportions).
  - $H_A^{(MZ)} : p_{\text{rural}}^{(MZ)} > p_{\text{urban}}^{(MZ)}$ (rural IMR higher than urban).
- **Himachal Pradesh**
  - $H_0^{(HP)} : p_{\text{rural}}^{(HP)} = p_{\text{urban}}^{(HP)}$.
  - $H_A^{(HP)} : p_{\text{rural}}^{(HP)} > p_{\text{urban}}^{(HP)}$.

Rationale: Rural areas typically have lower health service access and worse infrastructure; therefore rural IMR is expected to be higher. For mountainous states, "rural" may include very remote habitationsinterpretation should consider accessibility measures.

Test 4: Mother's Education vs Sanitation Access

Let:

- E_i = Mother's education level for household i (measured in completed years of schooling or categorical levels such as no schooling, primary, secondary, higher).
- S_i = Sanitation access for household i, where

$$S_i = \begin{cases} 1, & \text{if household has access to improved sanitation facilities (toilet with water, septic tank, etc.)} \\ 0, & \text{otherwise.} \end{cases}$$

We test whether an increase in maternal education (E_i) is associated with an increase in the probability of having improved sanitation access (S_i).

- **Mizoram**
  - $H_0^{MZ} : \beta_E^{MZ} = 0 \quad \Rightarrow \quad E_i$ has no significant effect on sanitation access (S_i) in Mizoram.
  - $H_1^{MZ} : \beta_E^{MZ} > 0 \quad \Rightarrow \quad$ Higher E_i increases the likelihood of improved sanitation S_i in Mizoram.
- **Himachal Pradesh**
  - $H_0^{HP} : \beta_E^{HP} = 0 \quad \Rightarrow \quad E_i$ has no significant effect on sanitation access (S_i) in Himachal Pradesh.
  - $H_1^{HP} : \beta_E^{HP} > 0 \quad \Rightarrow \quad$ Higher E_i increases the likelihood of improved sanitation S_i in Himachal Pradesh.

Rationale: Educated mothers are more aware of hygiene, disease prevention, and government sanitation schemes (e.g., Swachh Bharat Mission). Education also improves socio-economic status and decision-making power within households. Therefore, higher maternal education is expected to positively influence access to and use of improved sanitation facilities.

Test 5: Average number of children born vs Women's education

Model / Test: Test for a negative association between women's education level E and fertility C (using ANOVA across education categories).

- **Mizoram**
  - $H_0^{(MZ)}$: $\mathbb{E}[C|E = \text{low}] = \mathbb{E}[C|E = \text{medium}] = \mathbb{E}[C|E = \text{high}]$ (no difference in mean children across education levels).
  - $H_A^{(MZ)}$: $\mathbb{E}[C|E = \text{low}] > \mathbb{E}[C|E = \text{medium}] > \mathbb{E}[C|E = \text{high}]$ (higher education associated with fewer children; i.e., a negative trend).
- **Himachal Pradesh**
  - $H_0^{(HP)}$: $\mathbb{E}[C|E = \text{low}] = \mathbb{E}[C|E = \text{medium}] = \mathbb{E}[C|E = \text{high}]$ equal across education levels.
  - $H_A^{(HP)}$: $\mathbb{E}[C|E = \text{low}] > \mathbb{E}[C|E = \text{medium}] > \mathbb{E}[C|E = \text{high}]$ declines with increasing education (negative association / monotonic trend).

Rationale: Education increases women's labour market opportunities and contraceptive knowledge/use, and tends to delay marriage and first birth all factors that lower completed fertility.

Notes on testing strategy and controls

- Prefer **regression frameworks** (logistic for binary child-level IMR, count models for fertility) with **robust standard errors and relevant controls** (wealth, residence, maternal age, parity, access to health facilities) to estimate partial associations.
- Where **hypotheses are directional** (expecting a negative association), one-sided tests may be used but two-sided tests are more conservative and commonly reported.
- For **group comparisons** (age groups, education categories), supplement regression estimates with simple descriptive statistics (means, proportions) and visualisations (boxplots, bar charts).
- Always **report effect sizes** (odds ratios, marginal effects, differences in means) with confidence intervals, not just p-values.

II Hypothesis Testing and Results

3 Wealth, Residence, Age as Determinants of Infant Mortality 19

3.1 Objectives 19
3.2 Hypotheses 19
3.3 Data and Methodology 19
3.4 Results 20
3.5 Conclusion 22

4 Mother's Education, Fertility and Sanitation Access 23

4.1 Research Questions 23
4.2 Hypothesis 23
4.3 Data and Methodology 24
4.4 Results 24

Appendices 29

A Acknowledgements 29

3. Wealth, Residence, Age as Determinants of Infant Mortality

Infant mortality—the death of a child before one year of age—remains a crucial indicator of a society's overall health and development. Despite India's substantial improvements in reducing infant mortality rates (IMR), disparities persist among states, reflecting differences in economic status, healthcare infrastructure, and cultural practices. Himachal Pradesh and Mizoram present a compelling comparison: both are hilly states with relatively small populations, yet they differ significantly in social composition, health service delivery, and economic structure.

This study aims to understand how wealth, maternal age, and place of residence affect infant mortality, and why Mizoram exhibits notably lower IMR despite being economically less affluent than Himachal Pradesh.

3.1 Objectives

1. To quantify the relationship between household wealth and the likelihood of infant death.
2. To determine whether urban or rural residence influences infant mortality rates.
3. To assess whether maternal age has a significant impact on infant death probability.
4. To compare inter-state patterns to identify structural or social differences that may explain contrasting results.

3.2 Hypotheses

H1: Higher household wealth is associated with lower infant mortality.
H2: Infants born in rural areas have higher odds of death compared to those in urban settings.
H3: Maternal age follows a U-shaped relationship with infant mortality, with increased risk for very young (<20 years) and older (>35 years) mothers.

3.3 Data and Methodology

3.3.1 Data Source

Data were drawn from NFHS-5 (2019–21), specifically from the state-level subsets for Himachal Pradesh and Mizoram. The NFHS provides representative information on fertility, child health, and family welfare indicators. The present analysis uses a child-level dataset derived from these subsets.

3.3.2 Variable Construction

The dependent variable, **infant death**, was coded as 1 if the child was not alive and either had missing age data or was less than one year old; otherwise 0. Key independent variables were:

- **Wealth Index Combined:** A composite measure of household living standards (1 = poorest to 5 = richest).
- **Type of Place of Residence:** Coded as 1 = urban and 2 = rural.
- **Respondent's Current Age:** The mother's age at the time of survey.

Each woman's record was transformed into multiple child-level observations. This restructuring allowed regression analyses with each child as an independent observation, while recognizing within-mother correlations.

3.3.3 Analytical Approach

To estimate the relationship between predictors and infant mortality, a series of logistic regression models were fitted:

$$\text{logit}(p_i) = \beta_0 + \beta_1 W_i + \beta_2 R_i + \beta_3 A_i + \varepsilon_i \tag{3.1}$$

where p_i is the probability of infant death for child i, W_i is the household wealth index, R_i is residence type, and A_i is maternal age.

Given the rarity of infant deaths in Mizoram ($\approx 2.6\%$ of children), a **rare-event weighted logistic regression** was applied to mitigate bias from class imbalance. This approach increases the influence of rare observations in the likelihood estimation. For robustness, Firth's bias-reduced logistic regression and balanced sampling were tested, but results converged.

3.4 Results

3.4.1 Wealth and Infant Mortality

In Himachal Pradesh, wealth emerged as a strong and significant predictor ($\beta = -0.26$, $p < 0.001$), implying that each step up the wealth index reduces the odds of infant death by approximately 23%. In Mizoram, however, the coefficient was smaller ($\beta = -0.18$, $p = 0.06$), suggesting a weaker correlation between wealth and IMR.

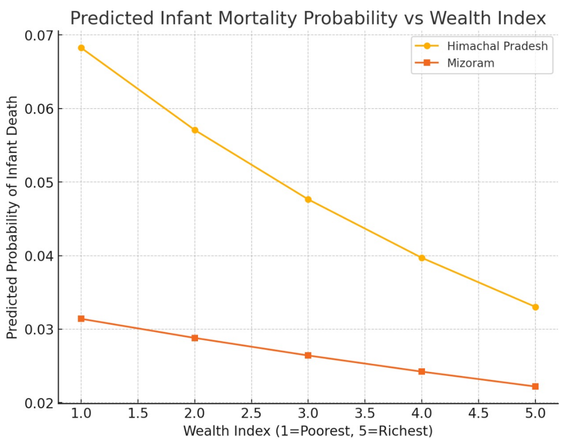

Figure 3.1: Predicted probability of infant death by wealth index for Himachal Pradesh and Mizoram.

3.4.2 Residence Type

The relationship between residence type and infant mortality varied by state. In Himachal Pradesh, the difference between urban and rural areas was statistically insignificant. In Mizoram, rural areas unexpectedly showed slightly lower infant mortality. This counterintuitive result likely reflects Mizoram's uniformly distributed health services and lower rural-urban inequality.

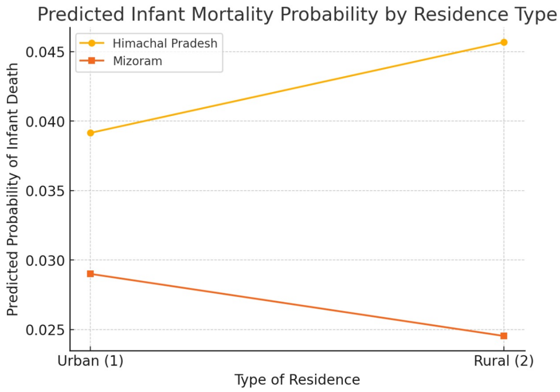

Figure 3.2: Predicted infant mortality probability by residence type.

3.4.3 Maternal Age and Rare-Event Adjustment

In Mizoram, infant deaths were so infrequent that conventional logistic models underestimated their probability. To address this, rare-event weighted logistic regression was employed. The results showed a nearly flat relationship between maternal age and infant death probability ($\beta = 0.0039$, $p = 0.62$), meaning maternal age had no discernible effect on IMR once bias correction was applied.

Figure 3.3 displays both observed data (gray points) and the fitted weighted regression line. Deaths were scattered across all age groups, supporting the statistical conclusion.

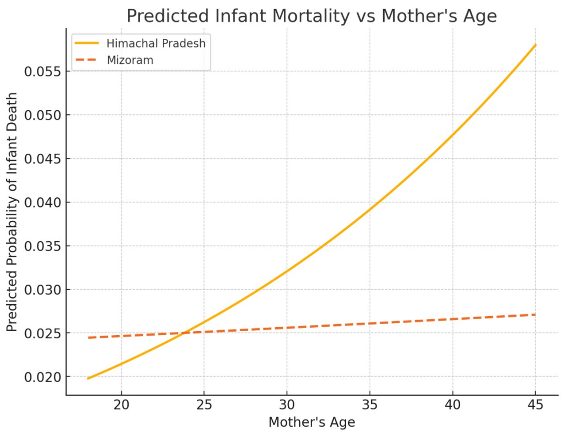

Figure 3.3: Observed data and rare-event weighted logistic fit for infant mortality vs. maternal age (Mizoram).

3.4.4 Inter-State Comparison and Interpretation

Table 3.1 summarizes key model results. The wealth gradient is steep and highly significant in Himachal Pradesh, while negligible in Mizoram. Maternal age and residence type are statistically insignificant across both states.

Table 3.1: Summary of Logistic Regression Coefficients by State

Variable	Coefficient (HP)	p-value (HP)	Coefficient (MZ)	p-value (MZ)
Intercept	-2.94	< 0.001	-2.63	< 0.001
Wealth Index	-0.26	< 0.001	-0.18	0.06
Residence Type	-0.09	0.45	-0.40	0.004
Mothers Age	0.001	0.85	0.006	0.49

Mizoram's weaker wealth-mortality link likely stems from its strong community-based health systems, widespread literacy, and egalitarian culture, which collectively buffer the influence of household income. Himachal Pradesh, by contrast, exhibits more pronounced wealth disparities and terrain-based inequality in healthcare access.

3.5 Conclusion

This study demonstrates that the socioeconomic determinants of infant mortality differ sharply between Himachal Pradesh and Mizoram. In Himachal Pradesh, infant survival improves consistently with wealth, confirming that economic capacity drives access to healthcare and child nutrition. In Mizoram, however, infant mortality is both rare and evenly distributed, indicating a socially equal environment where economic differences exert limited influence.

Maternal age and residence type were not significant predictors once rare-event bias was accounted for. These findings suggest that beyond household wealth, community health equity, and public service accessibility play decisive roles in ensuring infant survival. Future analyses could incorporate maternal education, sanitation, immunization coverage, and health-seeking behavior to extend this framework.

4. Mother’s Education, Fertility and Sanitation Access

Maternal education is one of the strongest, most consistent predictors of child health in low- and middle-income settings. Educated mothers are more likely to use antenatal care, seek skilled birth attendance, adopt immunisation, and practise better newborn care and hygiene. Fertility, often proxied by the average number of children ever born, can influence infant survival through biological (short birth intervals, maternal depletion) and resource dilution (household resources per child) channels. Studying these two relationships together helps clarify whether improving female education reduces infant mortality largely by

1. changing health behaviours directly,
2. reducing fertility and spacing births,
3. both.

4.1 Research Questions

- What is the association between mothers education and infant mortality in Mizoram and Himachal Pradesh, after controlling for confounders (wealth, residence, maternal age, parity, and access to health services)?
- How does the average number of children born to a mother relate to infant mortality, and to what extent does maternal education operate through fertility (mediator) to affect infant survival?
- Are the strength and pathways of these associations different between Mizoram and Himachal Pradesh?

4.2 Hypothesis

4.2.1 Number of children ever born in relation to education level for each individual

H_0 hypothesis (null) The mean number of children ever born does not differ across education levels.

H_1 hypothesis (alternate) The mean number of children ever born differs significantly with education level.

4.2.2 Sanitation access in relation ot education level for each individual

H_0 **hypothesis (null)** There is no significant association between mother's education level and access to improved sanitation facilities.
H_1 **hypothesis (alternate)** Higher mother's education level is significantly associated with increased access to improved sanitation facilities.

4.3 Data and Methodology

4.3.1 Data Source

Data were drawn from NFHS-5 (2019–21), specifically from the state-level subsets for Himachal Pradesh and Mizoram. The NFHS provides representative information on fertility, child health, and family welfare indicators. The present analysis uses a child-level dataset derived from these subsets.

4.3.2 Variable Construction

The key variables used in the analysis are defined as follows:

- **Mother's Education Level (E):** Categorical variable with three levels - no education, primary education, and secondary or higher education.
- **Infant Mortality (IMR):** Binary variable indicating whether the child died before reaching one year of age (1 = died, 0 = alive).
- **Average Number of Children Born (C):** Count variable representing the total number of children ever born to the mother.
- **Access to Improved Sanitation (S):** Binary variable defined as:

$$S_i = \begin{cases} 1, & \text{if household has access to improved sanitation facilities (toilet with water, septic tank, etc.)} \\ 0, & \text{otherwise.} \end{cases}$$

4.4 Results

4.4.1 Average Number of Children Born vs Education Level of Individuals

Performed a Chi-square test of independence separately for Mizoram and Himachal Pradesh. Visualized associations using stacked box plots for both states.

To test whether the mean number of children ever born differs across education levels, we assume the following fixed-effects one-way ANOVA model:

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \tag{4.1}$$

where:

- Y_{ij} = the j^{th} observation of Children Ever Born in the i^{th} education group,
- μ = overall population mean,
- α_i = effect of the i^{th} education level (group effect),
- ε_{ij} = random error term with $\varepsilon_{ij} \sim N(0, \sigma^2)$.

The hypotheses for ANOVA are:

$$\begin{aligned} &H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_k = 0 && \text{(all education levels have equal mean fertility)} \\ &H_1 : \text{At least one } \alpha_i \neq 0 && \text{(mean fertility differs between groups).} \end{aligned}$$

The total variation is divided as:

$$SS_{\text{Total}} = SS_{\text{Between}} + SS_{\text{Within}}, \tag{4.2}$$

and the F-statistic is computed as:

$$F = \frac{MS_{\text{Between}}}{MS_{\text{Within}}} = \frac{SS_{\text{Between}}/(k-1)}{SS_{\text{Within}}/(N-k)}, \tag{4.3}$$

where:
- k = number of education groups,
- N = total number of observations.

If the computed p-value from the F-statistic is less than the significance level $\alpha = 0.05$, we reject the null hypothesis and conclude that education level has a significant effect on fertility.

The following gives the box plot and the p values and F statistics for the two states:

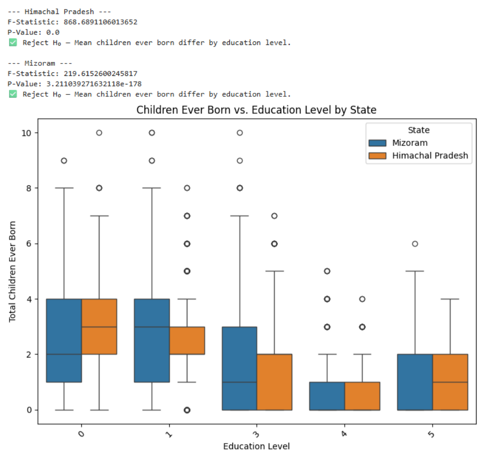

Figure 4.1: Box plot and ANOVA results for average number of children born vs education level

Verdict The box plots indicate a clear negative association between mother's education level and the average number of children ever born in both Mizoram and Himachal Pradesh. As education level increases from no education to secondary or higher education, the median number of children born decreases significantly. The ANOVA results confirm this observation, with p-values less than 0.05 for both states, indicating that the differences in mean number of children born across education levels are statistically significant. This suggests that higher maternal education is associated with lower fertility rates, which may contribute to improved infant survival outcomes.

Conclusion

In Himachal Pradesh, education seems more tightly linked to reproductive behavior possibly due to stronger institutional and healthcare outreach. In Mizoram, though the trend is similar, cultural and structural factors may moderate the impact slightly. Hence, state-level policy tailoring is essential Mizoram may need more community-level programs to translate education into family planning outcomes

4.4.2 Sanitation Access vs Education Level of Individuals

Performed a Chi-square test of independence separately for Mizoram and Himachal Pradesh Visualized associations using box plots and heatmaps for both states.

The following gives the contingency table and the chi squared statistics for the two states:

```
Chi-Square Statistic: 1561.5388384276312
Degrees of Freedom: 48
P-value: 1.1082757219840577e-295
✅ Reject H₀: Education level and toilet facility are dependent.

Contingency Table:
type of toilet facility   11    12    13  14  15  21   22  23   31  41  44  \
educational attainment
0                         40   626   375  11   4   5   78  74  133  12  22
1                         38  1254   424  10   0   4   64  55  137  20  25
3                        363  7553  1424  25   5  10  314  85  262  56  88
4                         26   534    73   0   0   1   16   2   18   7   2
5                        292  2119   334   5   2   0   62  14   31  14  20

type of toilet facility  96   97
educational attainment
0                         8    9
1                         7   11
3                        13  211
4                         0   18
5                         4  198
```

Figure 4.2: contingency table

The heatmap of the distribution of sanitation access across education levels for both states is shown below:

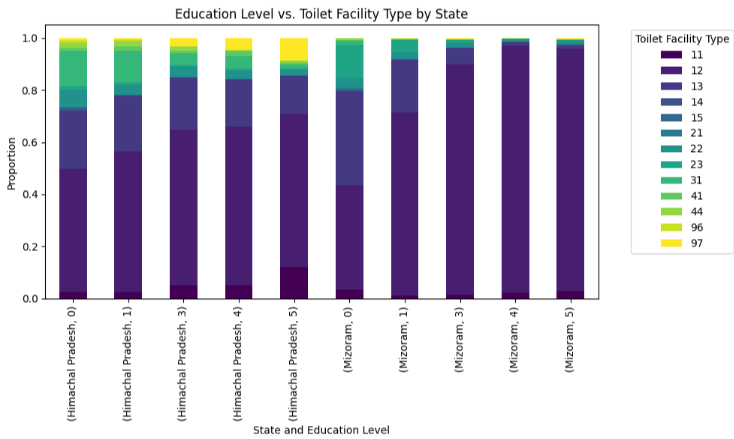

Figure 4.3: Heatmap of sanitation access vs education level

R The codes (like 11, 12, . . . , 97) are categorical identifiers for each specific toilet facility type (for example: flush to sewer, flush to septic tank, pit latrine with slab, pit latrine without slab, no facility/open defecation, etc.). They allow the survey to capture a wide variety of facility types, including both improved and unimproved categories, and variations such as whether the facility is shared, whether it is located within the house/plot or outside, etc.

The result of the test is given by the following table:

```
--- Mizoram ---
Chi-Square Statistic: 1315.5417548041771
Degrees of Freedom: 44
P-value: 6.592196935068665e-247
✅ Reject H₀: Education level and toilet facility are dependent.

--- Himachal Pradesh ---
Chi-Square Statistic: 816.1202337081843
Degrees of Freedom: 48
P-value: 2.760785044520369e-140
✅ Reject H₀: Education level and toilet facility are dependent.
```

Figure 4.4: Chi squared test results for both states

Verdict The stacked bar charts clearly show that as education level increases, the proportion of households with adequate sanitation access increases notably. Lower education groups have visibly higher shares of poor sanitation facilities.

Conclusion

Himachal Pradesh shows a stronger gradient meaning the gap between low and high education groups is more pronounced than in Mizoram. Education influences sanitation behavior better-educated individuals are more aware of hygiene, disease prevention, and sanitation benefits. This reflects the broader social development linkage education translates into better health and environmental practices. The state-level comparison highlights that even with similar education levels, outcomes differ implying that policy infrastructure and local awareness programs matter. Himachal Pradesh's stronger association suggests that its literacy programs and health infrastructure are effectively converting educational gains into improved sanitation outcomes. Mizoram, while performing well, may require more targeted hygiene education or rural sanitation outreach to fully realize the benefits of education on health behavior.

A. Acknowledgements

We, the authors of this project, would like to express our sincere gratitude to all those who have supported and guided us throughout the completion of this study titled *"Comparative Study of Demographic, Social and Economic Aspects of Mizoram and Himachal Pradesh"*.

First and foremost, we extend our heartfelt thanks to our mentor, **Prof. Kajori Banerjee**, whose continuous guidance, insightful suggestions and encouragement were invaluable at every stage of this work. Their expertise and constructive feedback helped us refine our objectives, analysis and presentation.

We are also grateful to the **Indian Statistical Institute, Kolkata** for providing us with the necessary resources, academic support and an environment that fostered research and collaboration.

Our sincere thanks go to the data sources such as the **National Family Health Survey (NFHS)**, **Census of India**, and various government reports and publications. Their authentic and accessible datasets formed the foundation of this comparative analysis.

We would also like to thank our **classmates and peers** for their helpful discussions, suggestions and moral support, which made this collaborative effort productive and enjoyable.

Finally, we express our deep gratitude to our **families** for their constant encouragement, patience and support throughout the duration of this project. Their motivation helped us stay committed and complete this work successfully.

— *Group Members*
Abhik R, Arikith RC, Anunoy C, Himadri M, Sattick D
B.Stat (3rd year)
Indian Statistical Institute
Kolkata